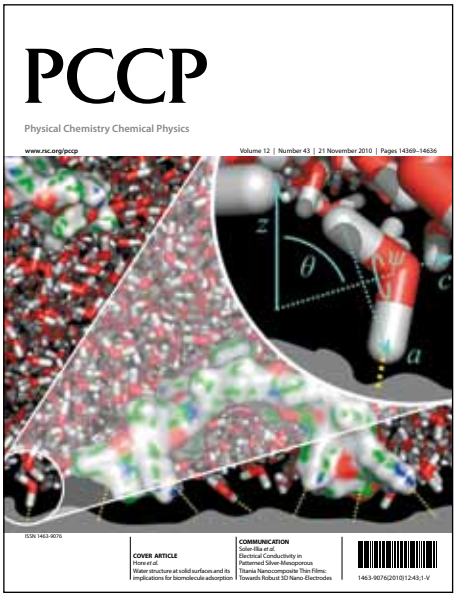
PCCP

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the RSC Publishing peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, which is prior to technical editing, formatting and proof reading. This free service from RSC Publishing allows authors to make their results available to the community, in citable form, before publication of the edited article. This *Accepted Manuscript* will be replaced by the edited and formatted *Advance Article* as soon as this is available.

To cite this manuscript please use its permanent Digital Object Identifier (DOI®), which is identical for all formats of publication.

More information about *Accepted Manuscripts* can be found in the **Information for Authors**.

Please note that technical editing may introduce minor changes to the text and/or graphics contained in the manuscript submitted by the author(s) which may alter content, and that the standard **Terms & Conditions** and the **ethical guidelines** that apply to the journal are still applicable. In no event shall the RSC be held responsible for any errors or omissions in these *Accepted Manuscript* manuscripts or any consequences arising from the use of any information contained in them.

The Intrinsic Conformational Features of Amino Acids from Protein Coil Library and Their Applications in Force Field Development

Fan Jiang[1], Wei Han[2], and Yun-Dong Wu[1,2,3]*

1. *Lab of Computational Chemistry and Drug Design, Laboratory of Chemical Genomics, Peking University Shenzhen Graduate School, Shenzhen 518055, China*

2. *Department of Chemistry*, *The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China*

3. *College of Chemistry, Peking University, Beijing, 100871, China*

*To whom correspondence should be addressed.

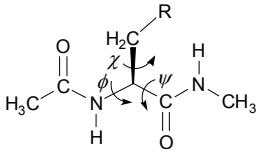
E-mail: chydwu@ust.hk

Abstract

The local conformational (ϕ, ψ, χ) preferences of amino acid residues remain an active research area, which are important for the development of protein force fields. In this perspective article, we first summarize spectroscopic studies of alanine-based short peptides in aqueous solution. While most studies indicate a preference for P_{II} conformation in unfolded state over α and β conformations, significant variations are also observed. A statistical analysis from various coil libraries of high-resolution protein structures is then summarized, which gives a more coherent view of the local conformational features. The ϕ, ψ, χ distributions of the 20 amino acids have been obtained from a protein coil library, considering both backbone and side-chain conformational preferences. The intrinsic side-chain χ_1 rotamer preference and χ_1-dependent Ramachandran plot can be generally understood by combining the interaction of side-chain Cγ/Oγ atom with two neighboring backbone peptide groups. Current all-atom force fields such as AMBER ff99sb-ILDN, ff03 and OPLS-AA/L do not reproduce these distributions well. A method has been developed by combining the ϕ, ψ plot of alanine with the influence of side-chain χ_1 rotamers to derive the local conformational features of various amino acids. It has been further applied to improve the OPLS-AA force field. The modified force field (OPLS-AA/C) reproduces experimental ^3J coupling constants for various short peptides quite well. It also better reproduces the temperature-dependence of the helix-coil transition for alanine-based peptides. The new force field can fold a series of peptides and proteins with various secondary structures to their experimental structures. MD simulations of several globular proteins using the improved force field give significantly less deviation (RMSD) to experimental structures. The results indicate that the local conformational features from coil libraries are valuable for the development of balanced protein force fields.

1. Introduction

The marginal stability of proteins, the function-related protein dynamics,[1-5] and the recent discovery of intrinsically disordered proteins[6-12] all imply a delicate balance between different protein conformations. One important factor for both folded and unfolded states of proteins is the intrinsic preference of the backbone to adopt certain ϕ, ψ dihedral angles (Scheme 1) without the influence of other residues. Although it has been long realized based on both statistical analysis of protein crystal structures[13,14] and guest-host studies of peptides in aqueous solution[15 - 20] that the common 20 amino acids can have different secondary structure propensities, the study of intrinsic local conformational preferences is mostly a recent effort.



Scheme 1. The dipeptide model and the definition of ϕ, ψ, χ_1 dihedral angles.

Recently, with an increasing interest in the protein unfolded state,[21,22] a number of spectroscopic studies (including CD, NMR, IR) of short peptides in aqueous solution tried to resolve the intrinsic ϕ, ψ preference of amino acids.[23-44] Most studies agree that the unfolded short peptides of Ala and some other amino acids mainly exist in the polyproline II (P_{II}) conformation. This is different from the distributions observed in folded proteins, in which the conformations related to secondary structures (α-helix, β-sheet, β-turn) can be stabilized by inter-residue backbone hydrogen bonds (H-bonds).

In 1995, Swindells et al. derived the ϕ, ψ preferences of various amino acids based on statistical analysis of coil residues (outside regular secondary structures) in protein crystal structures.[45] Since then, a number of statistical analyses of protein coil libraries have been published.[46 - 60] The statistical results from the PDB coil library not only correlate with secondary structure propensities of different amino acids,[45,56] but also agree well with the NMR experiments of unfolded peptides/proteins in solution.[49,61] Especially, the *J*-coupling constants of amino acid residues calculated based on coil library statistics agree very well with those measured from dipeptides in water.[57] The statistical potentials of the ϕ, ψ distributions from coil libraries have been used in knowledge-based simulations of the denatured or intrinsically disordered proteins.[62-67]

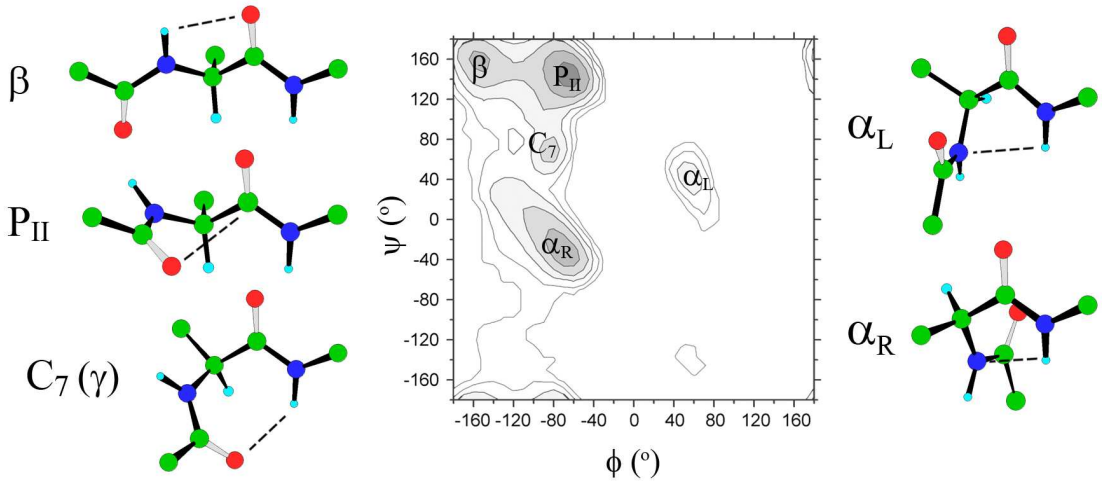


Figure 1. Five major backbone conformational basins in the Ramachandran (ϕ, ψ) plot of Ala residue from protein coil library. Contours are drawn every *RT* free energy difference (i.e. natural logarithmic scale for probability densities). The same scale is used throughout this work. The representative structures are from QM calculations at MP2/6-31+G** level[68] with CPCM[69,70] implicit solvent model of water. The key attractive interaction for each conformer is labeled and the methyl hydrogen atoms are omitted for clarity. Green, dark blue, and red are for carbon, nitrogen, and oxygen atoms, respectively.

From the coil library ϕ, ψ plot of alanine (Ala) residue,[60] five major backbone conformers can be identified (Fig. 1). They can also be located from high-level quantum mechanics (QM) calculations[71] using Ala dipeptide with the solvent effect of water. The most extended β conformer is also referred to as the C_5 conformer, due to its five-member-ring H…O hydrogen bond (H-bond). The P_{II} (polyproline II) conformer is also extended, but without intra-molecular H-bonding. C_7 conformer, also referred to as γ-turn, has seven-member-ring H-bond. The α_R and α_L conformers are related to helical structures with opposite rotation directions. They are stabilized by N … H electrostatic attraction between two neighboring peptide groups,[72] and also involved in various types of β-turns.[52] Noticeably, the α_R basin has a diagonal shape and asymmetric (skewed) density probability distribution. The average α-helix conformation ($\phi \sim$ -62°, $\psi \sim$ -42°) is close to the steep edge of α_R basin. From Fig. 1, the relative abundances of the five conformers are in the order $P_{II} > \beta \sim \alpha_R > C_7 > \alpha_L$. This preference has recently been supported by special QM/MM simulations which can treat the alanine dipeptide at the MP2 level.[73] The differences between their probabilities are quite large, indicating backbone

conformational behaviors may differ from a pure random coil model.

To study the dynamics of peptides and proteins in various states, molecular dynamics (MD) simulations using physics-based force fields are more generally applicable than knowledge-based scoring functions.[74-76] In most force fields, the backbone torsional ϕ, ψ parameters are optimized for small model systems (often Ala and Gly dipeptides) by reproducing their ϕ, ψ preferences from *ab initio* quantum mechanics (QM) calculations in gas phase.[77-84] However, recent simulations of various systems in aqueous solution identified secondary structure biases in various force fields.[85-100] Recent QM methodology development made it possible to carry out highly accurate calculations on biomolecules including short peptides, which agree well with gas-phase experiments.[89,101,102,103] However, protein force fields are usually used with aqueous environment. The force fields optimized for the gas-phase energetics may not capture well the strong solvent effect of water. Since modeling the solvent effect in QM calculations with high accuracy is still challenging, force field improvement using condensed-phase experimental data have becomes a practical approach.

Recently, fine-tunings of the backbone ϕ, ψ parameters have been reported to achieve a better balance among various secondary structures[104-111], and a better agreement with NMR data.[112-114] The recent efforts to improve protein force fields have not directly taken into account the available experimental data on ϕ, ψ preferences. In addition, the difference in the intrinsic conformational features of alanine and other amino acids were not considered. Recently, we have develop a new force field named PACE (Protein in Atomistic details coupled with Coarse-grained Environment)[115,116], which is used with the coarse-grained water model of Marrink et al.[117] The torsion potentials in the PACE force field were mainly parameterized by comparing dipeptide simulations with the coil library statistics. This force field can achieve a good balance between α-helix and β-sheet secondary structures in folding simulations of various peptides. We therefore try to address whether it is possible to apply the accumulated knowledge about intrinsic local conformational features from protein coil library to improve all-atom biomolecular force fields.

Here, we first analyze the local conformational preference of Ala from two different sources: the solution spectroscopy of unfolded peptides, and the statistical analysis of protein crystal

structures. Then, other amino acids with side-chain rotamers are considered. We discuss some important conformational features of amino acids with χ_1 side-chain rotamers, focusing on the interactions between side-chain γ-atom and backbone peptide groups. The χ_1 rotamers (*g*+/*t*/*g*-) distributions from the coil library were then compared with results from current force fields. To support the statistical rotamer-dependent ϕ, ψ plots using *ab intio* QM calculations, we introduce the ϕ, ψ decomposition scheme to reduce the dimensionality of the conformational space. We then apply a similar strategy to the all-atom force field OPLS-AA. We demonstrate that excellent agreement with the coil library results can be achieved for the ϕ, ψ distributions of Ala and other amino acids. Finally, we show that the new force field performs better than the original one in several different aspects.

2. Spectroscopic Experiments of Unfolded Peptides

Solution spectroscopic studies of the intrinsic backbone ϕ, ψ preference are mostly focused on alanine-based short peptides. Some early studies tried to fit the experimental observations by assuming only one single conformation for each residue. Most of these studies yielded ϕ values between -60° ~ -105° and ψ values between 125° ~ 170°, which coincide with the P_{II} conformation.[23-26] As previously reviewed by Kallenbach et al,[118] growing evidence has supported the hypothesis, first proposed more than 40 years ago,[119] that P_{II} is the dominant conformation for unfolded peptides and proteins. Of course, this information is not sufficient for force field parameterization.

Table 1. Conformational preferences of the Ala residue in short peptides from reported spectroscopic experiments in aqueous solution.

System	source	year	P_{II}%	β%	α_R%	$C_7(\gamma)$%	ref.
Ala dipeptide	amide III	2008	60	29	11		27
Ala dipeptide	Rama skeletal	2008	76	6	18		27
Ala_3 (tri-alanine)	CD	2003	~50	~50			28
Ala_3	2D amide I	2002	~80		~20		29
Ala_3	*J*-coupling	2007	92	8			30
Ala_3	amide I, *J*-coupling	2009	84	8	4	4	31
Ala_3	CD, *J*-coupling	2010	~55	~45			32
Ala_3	UVRR amide III	2010	~92[a]				33
Ala_4 (tetra-alanine)	UVRR amide III	2010	~91[a]				33

Ala_4	amide I′, CD	2004	>78				[34]
Ala_4	amide I′, *J*-coupling	2007	70	15	15		[35]
AAKA	amide I′, *J*-coupling	2007	60	19	20		[35]
GAG	amide I′, *J*-coupling	2010	79	6	5	5	[36]
Ac-GGAGG-NH_2	*J*-coupling	2005	82	18			[37]
AAAAAW	amide I′, *J*-coupling	2010	73	10	10	5	[38]
see note[b]	NMR SRE[c]	2007	79	18	3		[39]
XAO peptide[d]	*J*, NOE, CD	2002	~90	~10			[40]
XAO peptide[d]	amide I′	2007	~50	~23			[41]

[a] including 2.5_1 helix.
[b] $O_2TP_3AP_3AO_2$ and $O_2TP_3A_4O_2$, O is ornithine.
[c] spin relaxation enhancement.
[d] Ac-XXA_7OO-NH_2, X is diaminobutyric acid.

Besides P_{II}, there are other backbone (ϕ, ψ) conformations relevant to various secondary structures (α-helix, β-sheet, β-turn, γ-turn). It is also important to obtain information on the population of the less preferred conformations. Table 1 gives a summary of recent experiments reporting the local conformational preferences of Ala residues. Most studies in Table 1 did not report the percentages of all four major conformations (P_{II}, β, α_R, C_7), so some spaces in Table 1 are leaved empty. Considering all of these results together, it is clear that P_{II} is dominant in Ala residue but that other conformations are also populated. The population of the P_{II} conformation is between 50% ~ 92%, the population of β is below 50% with large variations, and the percentage of α is close to or below 20%. Most studies indicate that the β conformation is favored over the α conformation. However, some previous ^{13}C NMR[43] and Raman optical activity (ROA)[44] spectra studies of Ala dipeptide preferred the α conformation over the β conformation, which is supported by recent QM/MM simulations of Ala dipeptide.[73]

As shown in Table 1, there are large uncertainties for the reported percentages of major conformations. For NMR measurements such as *J*-coupling, NOE, and chemical shift, the results are averaged over accessible conformational states. Resolving the experimental spectra crucially relies on (1) the assumption of the ϕ, ψ distributions of certain conformations (such as P_{II}, β, α) (2) the relationship between a given set of ϕ, ψ and its corresponding spectroscopic measurement. For example, the NMR $^3J(H_N,H_\alpha)$ coupling constant is related to backbone ϕ torsion by the Karplus relationship,[120-123] a cosine function where ϕ = -160° and ϕ = -80° give the same value. Thus, some highly extended β conformations might be mistakenly resolved as P_{II}.[124] Especially, a number of studies assumed ϕ around -120° for the β conformation based on average β-sheet structures, which differs from the bottom of β basin (ϕ ~ -150°) in Fig. 1. Also,

α_R and P_{II} have similar ϕ values (around -70°) and $^3J(H_N,H_\alpha)$ coupling constants. Thus, α_R conformations might also be mistakenly resolved as P_{II}.

Vibrational spectra can probe structures in a much shorter time period, but still on a large number of molecules in solution. Due to significant overlap, information about the position, relative intensity, and shape of the active bands for certain conformer is needed to decompose the observed spectra into contributions from major conformers. The interpretation may require extensive computational studies.[125 - 130] Gaigeot recently reported DFT-based molecular dynamics simulations of the Ala dipeptide in the gas-phase and in aqueous solution.[131-133] The comparison with experimental IR spectra favors P_{II}/β conformations towards α_R. However, it is still difficult to precisely resolve experimental vibrational spectra due to the structural heterogeneity and interaction with solvent. Due to large uncertainties, these reported conformational preferences still can not be used reliably for calibrating a molecular force field.

Besides these experimental studies, *ab initio* QM calculations of the Ala dipeptide and its analogue were carried out in both the gas-phase[134] and with implicit solvent models.[135-137] In the gas-phase, the global energy minimum for Ala dipeptide is the C_7 conformation (ϕ, $\psi \sim$ -80°, +80°), which was confirmed by rotational spectra.[138] QM calculations also indicate that the P_{II} conformation is unstable in the gas-phase and less polar solvents, but is an energy minimum in water. Experimental studies also indicate the P_{II} preference is promoted by the polar/aqueous environment.[139,140] Pappu et al. used molecular simulations to show that the P_{II} conformation has minimal intra-peptide steric repulsion while the water screens the local electrostatic interactions that favor other conformations.[141,142] The preference for the P_{II} conformation might also benefit from the n -> π* orbital interaction between the backbone oxygen and the carbonyl carbon on the adjacent residue.[143,144] Recent *ab initio* QM/MM indicate that PII conformation has favorable interactions with solvent water molecules[73]

3. Coil Library from Protein Crystal Structures

3.1 Alanine

Current spectroscopic experiments give only very limited information on the local conformational preferences. On the other hand, statistical analysis of PDB coil libraries can give much more detailed ϕ, ψ distributions for various amino acids. Although all coil libraries excluded the α-helix and β-sheet structures to avoid the bias due to backbone H-bonding, some of them have further restrictions. As shown in Table 2, six different coil libraries were

constructed with increasingly stringent criteria from the same protein crystal structure database (6178 PDB structures with resolution <2.0 Å and R factor < 0.2 and 50% sequence identity cutoff). Here, we use the same definitions of P_{II}, β, α_R, α_L and C_7 conformational regions in the Ramachandran plot as our previous work (see Fig. 1).[60] Briefly, all regions except α_L have values of $\phi < -30°$. Both P_{II} and β conformations have values of $\psi > 100°$ or $\psi < -160°$, with $\phi < -100°$ for β region. The C_7 region has $40° < \psi < 100°$ and the α_R region has $-70° < \psi < 40°$. α_L region is within $30° < \phi < 100°$ and $-20° < \psi < 80°$. Besides, α_H (-60°, -40°) and α_T (-90°, 0°) conformations within the α_R region are related to different secondary structures: α-helix and β-turn, so their probabilities are sampled separately here with a Gaussian kernel of $\sigma = 10°$. The similarity coefficient S between two distributions $\{x_i\}$ and $\{y_i\}$ is defined as in our previous work[60]:

$$S_{XY} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2} \cdot \sqrt{\sum y_i^2}} \qquad (1)$$

Table 2. Conformational preferences of Ala residues from coil libraries of various restrictions.

Entry	Description	$N_{res.}$	P_{II}%	β%	α_R%	α_L%	C_7%	α_H%	α_T%
Coil-1	outside helices/sheets	34617	40.6	12.1	37.0	3.7	5.3	8.8	6.6
Coil-2	Coil-1 without pre-Pro residues	31215	38.2	11.3	40.7	3.9	4.4	9.7	7.3
Coil-3	Coil-2 without turn residues	20761	50.7	17.0	21.8	2.5	6.3	5.4	2.8
Coil-4	Coil-3 without secondary structure neighbors	7638	42.4	20.1	27.2	2.1	6.2	7.5	3.2
Coil-5	Coil-4 without 15% most exposed residues	6650	43.0	22.2	24.2	2.1	6.3	6.1	3.0
Coil-6	Coil-5 without neighboring to SDNVITFYHW	1668	45.9	23.1	21.0	2.0	6.1	5.5	1.7

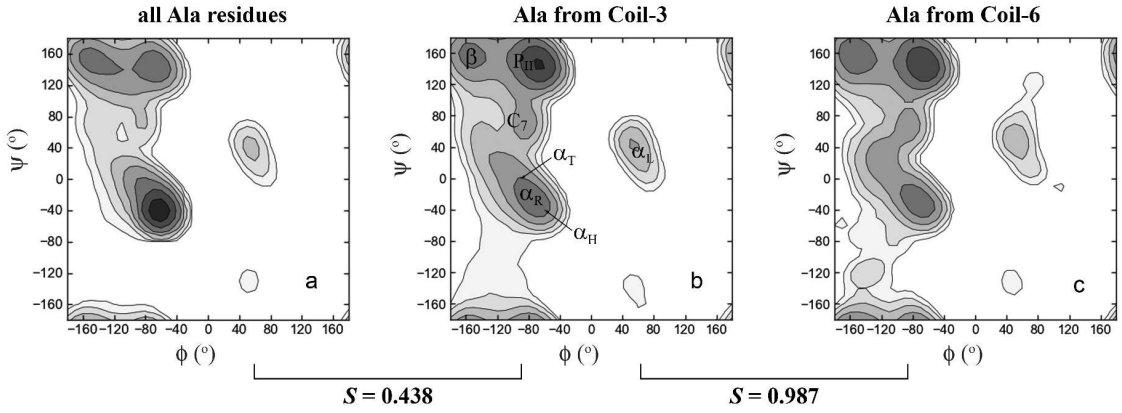


Figure 2. Ramachandran plots of Ala obtained from whole protein structures (a), the protein coil library (b, Coil-3) and the highly restricted coil library (c, Coil-6).

It is well known that the residues preceding proline (pre-Pro) show a distinct ϕ, ψ distribution.[53,145] Therefore, they should be excluded from the coil library (from Coil-1 to Coil-2). The residues in β-turns mostly occupy both the right-handed α_R and left-handed α_L helical region. In a recent statistical analysis of ϕ, ψ distributions by Dunbrack et al., the α_R content from turn residues (42.4%) is significantly higher than from coil residues (17.1%).[146] Removing turn residues will significantly decrease α_R% as shown in Table 2 (Coil-2 to Coil-3), which leads to a better agreement with recent spectroscopic experiments of short peptides. To remove β-turn residues is reasonable because they are stabilized by i ← i+3 backbone H-bonding. Noticeably, although β-turn residues were excluded in a very early coil library by Serrano,[46] they were not excluded in some later studies.[51,52,55]

In a recent coil library by Jha et al, removing residues adjacent to secondary structure elements (α-helix, β-sheet, various H-bonded turns) slightly decreased the population of the P_{II} and increased the population of α_R and β.[56] They also found that most exposed residues especially favor α_R against β, and excluded them from their restricted coil library. Besides, residues with short polar side-chains (Asp, Asn, Ser, Thr) can form H-bonds with the neighboring backbone amide groups,[147,148] and β-branched or aromatic residues might increase the β-propensity of neighboring residues.[49,56,149] For the most stringent coil library, Coil-6 in Table 2, residues adjacent to these amino acids were all removed to minimize the nearest neighbor effect. The percentages of P_{II}, β and α_R conformations obtained from Coil-6 are 46%, 23% and 21%, respectively. These are very similar to the populations of P_{II}, β and α_R reported from previous restricted coil library by Jha et al.: 48%, 25% and 24%.[56]

As shown in Fig. 2, the ϕ, ψ distribution of Ala from the whole sequence (i.e. all secondary structures) is significantly different (S = 0.44) from coil library results. From the studies by Jha et. al., only 17.1% of all Ala residues in protein crystal structures are in the P_{II} conformation,[56] which is very different from spectroscopic experiments of unfolded peptide. On the other hand, the ϕ, ψ distributions from different coil libraries Coil-3 and Coil-6 are very similar (S = 0.99), even though the size of the coil library decreases to less than 1/10 from Coil-3 to Coil-6. The invariance indicates the effects of neighboring residues either may not be very strong or can be statistically averaged out. Also, the coil library ϕ, ψ plot from database containing only mainly-α proteins is also very similar (S = 0.986) to that from only mainly-β proteins. The comparison supports that coil libraries do reflect the intrinsic conformational preferences of amino acid residues.

Compared with most spectroscopic studies of short peptides, coil library indicates a slightly lower population of the P_{II} conformation. A few possible reasons may account for this: (1) The terminal effects in peptides can promote P_{II} conformation and reduce α content,[100] such that Ala_3 shows more P_{II} and less α contents than Ala dipeptide. Gnanakaran and Garcia reported higher P_{II} propensity for Ala_3 (~80%) than for Ala dipeptide (~59%) based on their modified AMBER force field simulations.[150] (2) Some studies assuming a value for ϕ around -120° for the β conformation but the actual conformation in the β region can be much more extended (ϕ ~ -160°). This can mistakenly assign β structures as P_{II}, as discussed before. (3) The P_{II} conformation is stabilized by solvation in water, but residues in protein coil region may not be fully solvated, which can lead to the decrease of P_{II} preference.

3.2 Consideration of Side-chain rotamers

Most amino acids can be regarded as derivatives of Ala with different substitutions at the β-carbon (C_β) site. As we discussed earlier,[60] they can be divided into five types according to different local conformational features: (1) ordinary amino acids, (2) Ser, (3) Asp and Asn, (4) Val and Ile, (5) Thr. The type-1 (ordinary, Ord for short) amino acids have single non-polar (aliphatic or aromatic) carbon attached to C_β, which have very similar local conformational features from recent studies.[151] As shown in Fig. 3, there is a strong coupling between the side-chain χ_1 distributions and backbone conformations,[60,152] which is different for different types of amino acids. From Fig. 3, the χ_1 distribution can be sensitive to small change in backbone conformation. Significant difference between α_A and α' are observed, both of which are belong to the α_R conformational basin (Fig. 1). On the other hand, all these different χ_1

distributions are related to the rotation around Csp3-Csp3 bond, resulting discrete conformers called rotamers (Scheme 2).

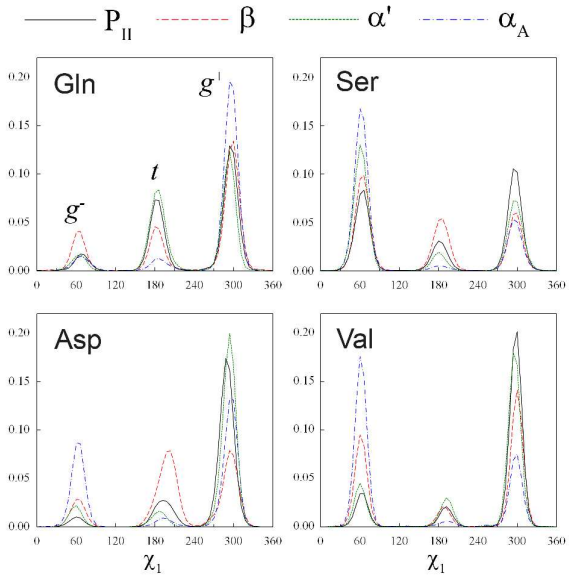
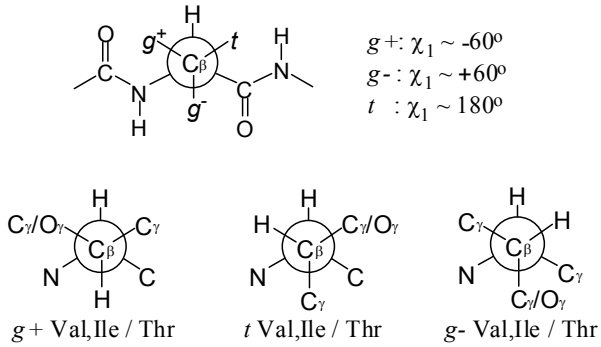


Figure 3. The χ_1 probability distributions for four representative amino acids in four different ϕ, ψ regions. The α_A region defined in our previous work is related to α-helix conformation (similar to the α_H), and α' is the remaining region of the α_R basin.



Scheme 2. Definition of side-chain χ_1 rotamers based on the relative orientation of γ atom(s), illustrated by Newman projections down the C_β–C_α bond for single-β-substituted (upper left) and β-branched (bottom) amino acids. The χ_1 distribution of lysine is also given as example.

Therefore, the rotamer-dependent ϕ, ψ plots $p(\phi,\psi|\chi_1)$ can be a good representation of the conformational features of amino acid residues. Indeed, $p(\phi,\psi|\chi_1)$ distributions have been used to develop a new backbone-dependent rotamer library very recently.[153] As shown in Fig. 4, the Ramachandran plots of the three rotamers of an amino acid can have significantly greater differences than those between different amino acids under the same rotamer. Our previous work

also showed that the α-helix, β-sheet, and type-I β-turn preferences of different rotamers of various amino acid types can be captured by their intrinsic ϕ, ψ, χ preferences from our coil library. Therefore, it is important to incorporate the intrinsic side chain rotamer preferences in future development of protein force fields.

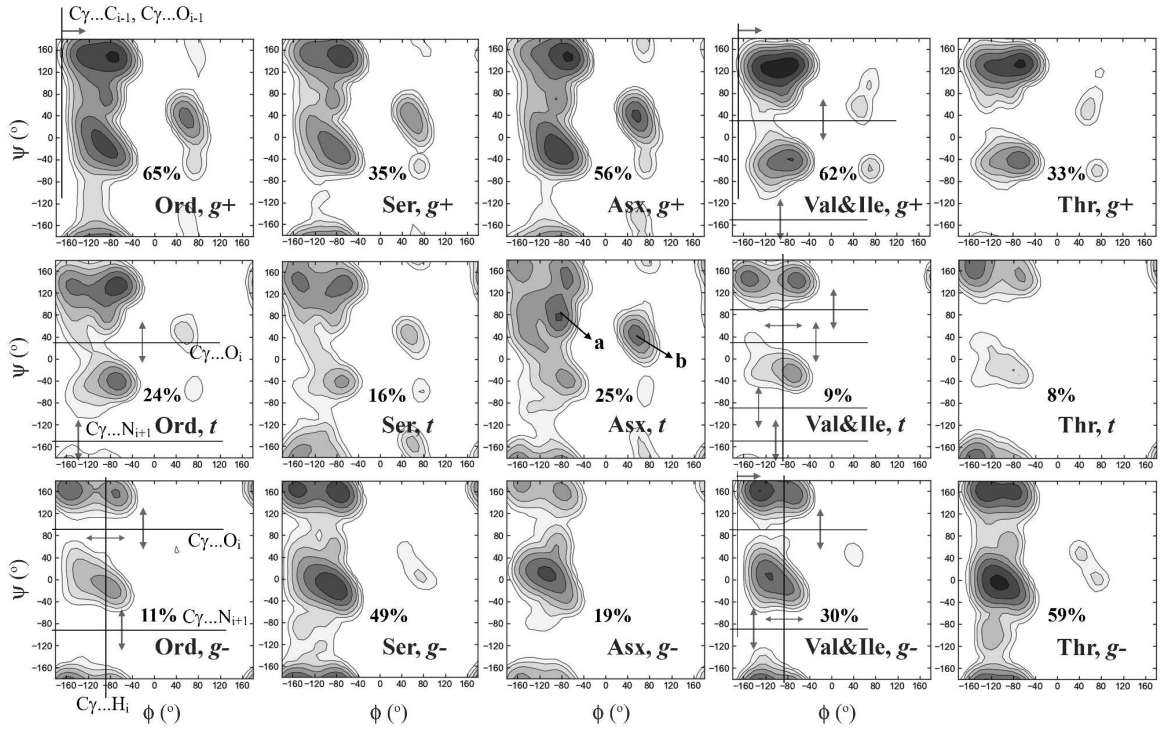


Figure 4. ϕ, ψ plots of χ_1 rotamers ($g^+/t/g^-$) of various residue types from protein coil library (Coil-3). The leftmost column is for the ordinary amino acids, which have single non-polar (aliphatic or aromatic) γ carbon atom attached to C_β atom. The unfavorable regions due to Cγ/backbone repulsions are marked with vertical (ϕ regions) or horizontal (ψ regions) lines. The especially preferred conformations of *t* rotamer of Asx are indicated.

As show in Scheme 2, for the single-β-substituted amino acids, the γ atom in g^+ rotamer only directly interacts with the N-terminal peptide group (ϕ-pep for short), and the γ atom in *t* rotamers only directly interacts with the C-terminal peptide group (ψ-pep for short). In g^- rotamer, the γ-atom interacts with both ϕ- and ψ- peptide groups. Comparing $p(\phi,\psi|\chi_1)$ plots of ordinary amino acids in Fig. 4 with the ϕ, ψ plot of Ala in Fig. 2, regions around some ϕ or ψ values are disfavored due to the repulsions from Cγ. These repulsions mainly involve atoms separated by four bonds (1-5 interactions), as suggested by Dunbrack and Karplus.[154] From Fig. 4, Asp and Asn have special ϕ, ψ preferences of their *t* rotamers, with C_7-like and α_L conformers

significantly favored. The two structures optimized by *ab initio* QM method have short O_{δ}…C and O…Cγ distances (~3.1Å), respectively (Fig. 5). In addition to electrostatic attraction between opposite charges, there may be stabilization from n → π* interaction between the long pair of oxygen and anti-bond π orbital of C=O group.[155]

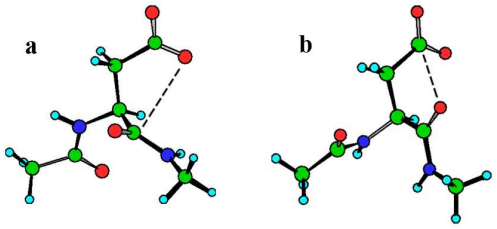
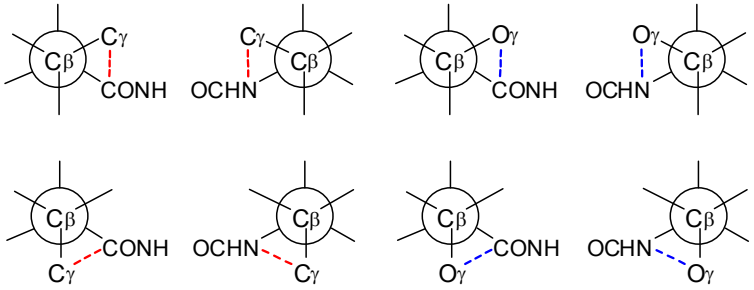


Figure 5. Structures of special conforms of *t* rotamer of Asp indicated in Fig. 4.



$E(C\gamma...\psi\text{-pep}) > E(C\gamma...\phi\text{-pep}) > 0 > E(O\gamma...\psi\text{-pep}) > E(O\gamma...\phi\text{-pep})$

Scheme 3. The total effects of γ-atom/backbone interactions for understanding side-chain rotamer preferences. $E > 0$ indicates repulsive interactions and $E < 0$ indicates attractive interactions.

The different χ_1 preferences can also be understand by the interactions between the γ atom(s) and backbone peptide groups, based on the assumed average energetics shown in Scheme 3. In general, non-polar C_{γ} atom is more likely to have repulsive interactions with polar backbone peptide group than with polar O_{γ} atom. Besides the steric repulsion of C_{γ} atom, the screening of strong interactions between backbone peptide group and water can also contribute to the unfavorable effective interaction energy.[149] On the other hand, O_{γ} atom can form favorable interactions with both backbone carbonyl C atom and amide H atom. The interaction of C_{γ} with ψ-pep is more repulsive than with ϕ-pep, which can be understood by the 1-4 interaction with larger backbone C atom for ψ-pep compared with N atom for ϕ-pep. Beside, 1-5 interactions are also contributed to the preference. From Fig. 4, Cγ interactions with ψ-pep atoms resulted in

significantly unfavorable or even disallowed ψ regions, while interactions with ϕ-pep atoms only decrease the probabilities of related regions. The more attractive interaction of O_γ with ϕ-pep can be understood by 1-5 interactions with positively-charged carbonyl C and amide H atoms attached to backbone N atom. For amino acids with single γ atom, the observed g^+% > t% >g^-% for Ord and g^-% > g^+% > t% for Ser can easily can obtained from the assumed energetics.

Among β-branched amino acids, Val and Ile have two C_γ atoms, and Thr has one O_γ atom and one C_γ atom. Therefore, the features of their each rotamer can be understood by combining those of two rotamers of Ord or Ser. For example, g^+ rotamer of Val (g^+-Val for short) has the features of g^+-Ord and t-Ord. Similarly, the conformational features of g^--Thr are the mix of g^--Ser and g^+-Ord. Based on the interactions in Scheme 3, g^+% > g^-% > t% for Val&Ile and g^-% > g^+% > t% for Thr can also be understood.

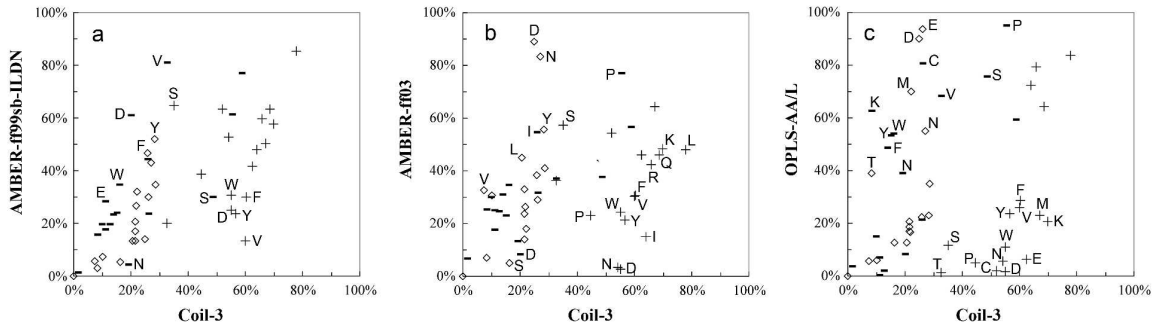


Figure 6. Side-chain χ_1 rotamer distributions of the 17 amino acids obtained by dipeptide simulations against coil library statistics. The force fields include (a)AMBER-ff99sb-ildn (b) AMBER-ff03, and (c) OPLS-AA/L. Plus sign (+) is for g^+ rotamers; horizontal bar (-) is for g^- rotamers; and open diamond is for t rotamers. Amino acids with large deviations are labeled.

Previous NMR studies of the side-chain rotamer populations in unfolded peptides and proteins showed close agreements with the statistical results from protein coil residues.[156-158] The χ_1 distributions are also consistent among various coil libraries. We have shown that excluding the residues adjacent to Ser, Asx, and β-branched and aromatic amino acids from the restricted coil library did not change (S = 0.997) the obtained $g^+/t/g^-$ percentages. The results also correlate very well with those from an extremely restricted coil library in which the residues

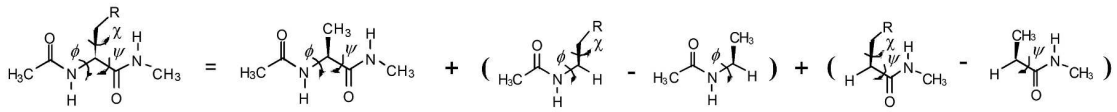
can only be adjacent to Gly or Ala.[60] Therefore, the coil library can give reliable intrinsic conformational preferences without significant influence of the NN effect.

On the other hand, commonly used force fields such as OPLS-AA/L and AMBER-ff03 can not reproduce the observed side-chain χ_1 rotamer distributions of some amino acids.[60] Lindorff-Larsen et. al. also noticed that AMBER-99sb simulations give side-chain rotamer distributions for Ile, Leu, Asp, and Asn significantly differ from PDB distributions, and they refitted the side-chain torsional parameters according to the gas phase QM calculations.[159] Interestingly, as shown in Fig. 6, this improved AMBER-ff99sb-ildn force field[160] still can not fully reproduce the rotamer distributions from the PDB coil library. The problem is most serious for residues with short polar side chains (D, N, S), for aromatic residues (Y, F, W), and for β-branched residues (V, I, T). This suggests the need for further improvement of these force fields.

4. Force Fields Developments

4.1 ϕ, ψ Decomposition and QM Calculations

Unlike Ala dipeptide, it has been difficult for dipeptides with C_β-substitution to carry out systematic conformational studies because the coupling of ϕ, ψ potentials with χ_1 and sometimes χ_2 results in a higher dimensionality of the conformational space and significantly increase the number of grid points. Although there are previous QM studies trying to locate minima on the ϕ, ψ, χ energy hyper-surface of dipeptides or their analogues,[161-165] full scans of the potential energy surface have only been reported for Ala and Gly dipeptides. One way to overcome the problem is to decompose the effect of the substituent group(s) on the backbone ϕ and ψ torsions, as shown in Scheme 4. The idea of separating ϕ and ψ was previously used by Dunbrack and Cohen to obtain the distributions for the Bayesian statistics.[166].



$$G_X(\chi,\phi,\psi) = G_{Ala}(\phi,\psi) + \Delta G_{X\text{-}Ala}(\chi,\phi) + \Delta G_{X\text{-}Ala}(\chi,\psi)$$

Scheme 4. Separation of the effects of side-chain Cβ substituent on the backbone conformation into ϕ part and ψ part, illustrated using the dipeptide models and monopeptide models. Here χ

belongs to the three χ_1 rotamers (g^+/t/ g^-).

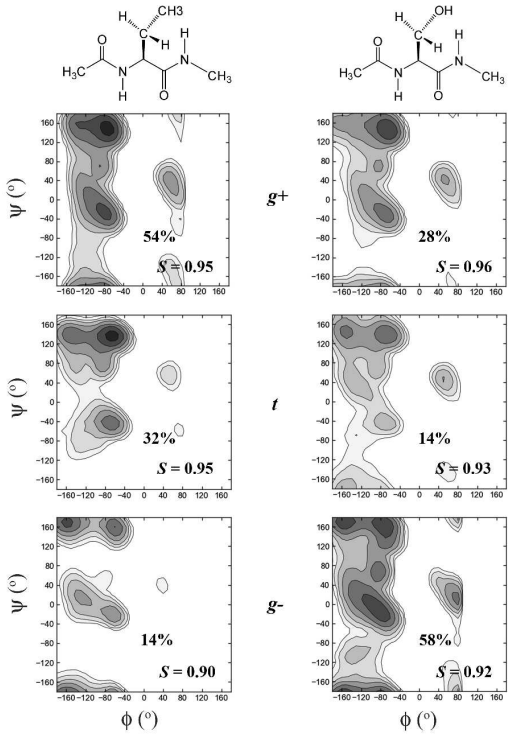


Figure 7. Ramachandran plots of g+, t, g- rotamers of model dipeptides with R = -CH3 (left) and R = -OH (right) calculated at the MP2/6-31+G** level with CPCM solvent model for water.

Based on this strategy, we carried out *ab initio* QM calculations in water with the CPCM solvent model on the ϕ and ψ mono-peptide models. The obtained effective energies $\Delta E_{\text{X-Ala}}(\chi, \phi)$ and $\Delta E_{\text{X-Ala}}(\chi, \psi)$ in water, which approximate the correspond hypothetic free energies $\Delta G_{\text{X-Ala}}(\chi, \phi)$ and $\Delta G_{\text{X-Ala}}(\chi, \psi)$ describing the side-chain substitution effect on the ϕ and ψ peptide groups, were used to reconstruct the χ_1-dependent Ramachandran plots based on Boltzmann relationship.

$$p_{\text{X}}(\chi,\phi,\psi) = p_{\text{Ala}}(\phi,\psi)\cdot\exp[-\frac{\Delta E_{\text{X-Ala}}(\chi,\phi)+\Delta E_{\text{X-Ala}}(\chi,\psi)}{RT}] \quad (2)$$

where $p_{\text{Ala}}(\phi,\psi)$ is the statistical ϕ, ψ distribution of Ala.

Comparing Fig. 7 with Fig. 4, the calculations reproduce the statistical χ_1-dependent ϕ, ψ plots very well (S = 0.90 ~ 0.96). For ordinary amino acids, their t rotamers prefer P_{II} conformation and g- rotamers prefer β conformation. On the other hand, similar preferences for P_{II} and β

conformations are observed for both *t* and *g*- rotamers of Ser. Our MP2/6-31+G** calculations well reproduce these features. The calculations also reproduce the different side-chain rotamer preferences between ordinary amino acids (*g*+% > *t*% > *g*-%) and Ser (*g*-% > *g*+% > *t*%). The very similar local conformational behaviors of Gln, Glu, Lys, Arg and Met are mainly contributed from the γ-carbon of their side-chains. These studies support the rotamer-dependent ϕ, ψ plots from protein coil libraries for evaluating and improving protein force fields. These studies indicate that the effects of the side-chain on backbone ϕ torsion and ψ torsion can be roughly independent. This can be very useful for theoretical studies on the local effect of different side-chains.

Besides, the idea of separating the difference between two Ramachandran plots into ϕ-component and ψ-component can be very useful for force field parameterization. Under the forms of current standard force fields, we can not use different backbone potentials under different side-chain rotamers. However, we can still use the free energy decomposition scheme introduced here to separate the difference between current and target 2D ϕ, ψ free energy surfaces into corrections on individual ϕ or ψ potentials.

For a given type of amino acid, to achieve a good match between its ϕ, ψ plot from MD simulations (p_{MD}) with its ϕ, ψ plot from a coil library (p_{Coil}), we decompose the difference between the two free energy surfaces into ϕ-component and ψ-component:

$$\Delta V_{\phi}(\phi) + \Delta V_{\psi}(\psi) = -RT \ln[\frac{p_{\mathrm{Coil}}(\phi,\psi)}{p_{\mathrm{MD}}(\phi,\psi)}] \qquad (3)$$

Here, ΔV_{ϕ} and ΔV_{ψ} are the correction needed on the current ϕ and ψ torsional potentials. The fitting of parameters is done iteratively until two free energy surfaces match very well. This procedure did not change the fact that the fitting is to match between free energies, but can significantly speed up the parameterization for all amino acids. This can significantly speed up the parameterization, and the detailed procedure will be described elsewhere.

4.2 Alanine

REMD[167] simulations with temperature range between 273K ~ 418K were carried out for the Ala dipeptide solvated with ~320 waters, using common biomolecular force fields and our

recently developed ones. The results shown in Table 3 are from the 297 K replica. All simulations presented in this work were carried out using the *Gromacs* molecular simulation software.[168] All force fields in Table 3 except PACE and OPLS-AA/C were parameterized based on QM calculations. They can give out quite different conformational distributions. In general, the order of similarities to the coil library distributions is: AMBER-99sb > OPLS-AA/L ~ AMBER-96 > AMBER-03 ~ CHARMM-27 > GROMOS-53A6 > AMBER-94 ~ AMBER-99.

Table 3. Conformational preferences of Ala dipeptide from molecular dynamics simulations with various force fields, together with similarity coefficients (*S*) with our coil library results.

Force field	water model	P_{II}%	β%	α_R%	α_L%	C_7%	α_H%	α_T%	*S*
AMBER-ff94	TIP3P	8	3	87	0.4	1	12.4	9.4	0.32
AMBER-ff96	TIP3P	39	46	7	0.0	4	1.3	0.1	0.83
AMBER-ff99	TIP3P	1.4	4	87	0.6	2	4.2	4.6	0.20
AMBER-ff99sb	TIP3P	37	29	27	3	2	2.4	3.8	0.88
AMBER-ff03	TIP3P	33	21	42	0.1	1	4.8	5.6	0.81
CHARMM-27[a]	TIP3P	19	14	34	6	2	6.2	3.9	0.80
OPLSAA/L	TIP3P	40	19	34	0.5	5	2.1	5.0	0.82
OPLSAA/L	TIP4P/Ew	43	18	31	0.6	6	1.9	4.8	0.83
GROMOS 53A6	SPC	26	38	11	0.7	15	1.3	0.2	0.28
PACE	Marrink	40	26	27	1.6	4	2.9	4.8	0.87
OPLS-AA/C	TIP4P/Ew	48	23	20	1.2	6	3.9	2.4	0.98

[a]with C-MAP correction

It is well known that the early AMBER force fields ff94 and ff99 significantly favor the α_R conformation.[88] On the other hand, both ff96 and ff99sb force fields give much higher similarities ($S > 0.8$) with the coil library results. The ff96 and ff99sb force fields are improved versions of ff94 and ff99 force fields for their backbone ϕ and ψ torsional parameters. Although the ff96 parameter set has a known bias towards β-sheet conformations when used with explicit water,[85,87,97,169] it can correctly fold both α- and β- peptides with certain generalized-Born implicit solvent model, achieving a balanced secondary structure preference.[170] Consistent with the high similarity (0.88) to the coil library results, ff99sb force field and its improved version ff99sb-ILDN[160] are recently reported to correctly fold both α- and β- peptides and miniproteins including all-α villin headpiece subdomain and all-β WW domain.[75,171] The AMBER ff99sb is also reported to better reproduce experimental NMR data than several other force fields.[172,173]

The AMBER-03 and CHARMM-27 force field with lower *S* to coil library ϕ, ψ plot over-stabilizes the α-helix structure. The AMBER ff03 force field predicted melting temperature (T_m) of Ac-Ala_{21}-Nme as high as ~ 380 K.[174] The CHARMM-27 force field without CMAP

gives significantly higher α% and less β% and also a quite different ϕ, ψ plot.[94] Even with the C-MAP correction, CHARMM still gives out too little P_{II} population and low β% value compared with high α_R% and α_H%, which relate to its significant bias towards α-helical structure.[92] The OPLS-AA/L gives an *S* value (0.82) comparable to CHARMM-27 and AMBER-ff03, but slightly lower than ff99sb and other improved AMBER variants. It can be seen from Table 3 that changing water model from TIP3P to TIP4P/Ew[175] does not change the distribution significantly.

The united-atom GROMOS 53A6 force field is known to destabilize the α-helical conformations, in agreement with the low value of α_H%. It also gives inadequate P_{II}% and an unrealistic global minimum around $\phi \sim -145°$ and $\psi \sim 120°$. All these factors account for the obtained low similarity coefficients of $S = 0.28$. As shown in Table 3, our recently developed coarse-grained PACE force field gives similarity coefficient comparable to current all-atom force fields. This is because PACE was parameterized against the ϕ, ψ plot from the coil library statistics[60] through dipeptide simulations.[115] The PACE force field can achieve a good balance between α-helix and β-sheet secondary structures in folding simulations of various peptides.[116]

Using a strategy similar to the one used in the development of the PACE force field, we reparameterized the ϕ, ψ, χ potentials in the original OPLS-AA force field to maximize the similarity with the coil library statistics. The new force field is named OPLS-AA/C, where C means both Coil library statistics and Condensed phase simulations. The manual adjustment of local L-J parameters was partly inspired by a previous study on the Ramachandran plot of the Ala residue.[176] The new OPLS-AA/C parameter set gives much better agreement ($S = 0.98$) with the coil library statistics than the parent OPLS-AA/L force field. Among various force fields, the new force field gives out highest P_{II} content (48%) for the Ala dipeptide, in good agreement with the solution spectroscopic experiments. The OPLS-AA/L force field gives lowest α_H%:α_T% ratio, while the results from OPLS-AA/C simulations give significantly improved α_H%:α_T% ratio.

Table 4. *J*-coupling constants of Ala_3 cation from experiment and REMD simulations.

	$^3J(H_N,H_\alpha)$	$^3J(H_N,C')$	$^3J(H_\alpha,C')$	$^1J(N,C_\alpha)$	$^2J(N,C_\alpha)$
Experiment[30]	5.68	1.13	1.84	11.34	8.45
AMBER-99sb	7.1	1.1	2.1	11.4	8.2
AMBER-03	6.7	1.2	1.9	11.3	8.2
CHARMM-27[a]	6.4	1.6	2.0	10.9	7.8
OPLS-AA/L	7.3	0.8	2.0	11.1	8.1

OPLS-AA/C	6.2	1.3	1.7	11.3	8.3

[a]with C-MAP correction.

To allow a direct comparison between results from simulations and from experiments on the same system, we used the trialanine cation (Ala_3) with known experimental NMR *J*-coupling constants[30]. The *J*-coupling constants were calculated from simulated ϕ, ψ distributions based on empirical Karplus relationship[121-123] and compared with the experimental data. In Table 4, the first three columns of *J*-coupling constants are related to ϕ torsion and the last two are related to ψ torsion. Among them, $^3J(H_N,H_\alpha)$ is the most widely used in the studies of unfolded peptides and proteins. Some force fields, such as AMBER-99sb and OPLS-AA/L, give much higher $^3J(H_N,H_\alpha)$ values than the experimental results. Also, OPLS-AA/C gives a $^2J(N,C_\alpha)$ value closer to the experimental value than other force fields (Table 4), in agreement with its lower α_R content. The new OPLS-AA/C parameter set gives better agreement with the experimental *J*-coupling values, suggesting that the approach of using ϕ, ψ distributions from PDB coil libraries as reference data may improve current force fields.

4.3 Other amino acids

For different types of amino acids, common force fields can not consistently give high similarities with coil library ϕ, ψ distributions (Table 5). Among common force fields, the AMBER-99sb with recently improvement[159] gives generally best similarity coefficients. This agrees with its good performance. Still AMBER-99sb gives lower similarity for Val. Although OPLS-AA/L seems better than AMBER and CHARMM force field in Gln and Val, it can not well reproduce the coil library results for some special amino acids (Ser, Asp, Gly). The ϕ, ψ plot of Gly generated by the original OPLS-AA/L force field is very different from the coil library results ($S = 0.35$), agreeing with recent studies.[177]

Table 5. Similarity coefficients (*S*) between coil library ϕ, ψ distributions and simulated results using various force fields, for different amino acid types.

	Gln	Tyr	Val	Ser	Asp	Gly
AMBER-99sb	0.92	0.91	0.77	0.93	0.82	0.91
AMBER-03	0.84	0.62	0.75	0.71	0.82	0.93
CHARMM-27	0.80	0.80	0.75	0.82	0.86	0.62
OPLS-AA/L	0.94	0.74	0.83	0.62	0.61	0.35

OPLS-AA/C	0.98	0.98	0.98	0.97	0.95	0.94

In OPLS-AA/C, all ϕ, ψ, χ torsional parameters were optimized to achieve as good as possible the agreements between dipeptide simulations and the coil library results of all other amino acids. Results for some amino acids are given in Table 5. Compared with OPLS-AA/L, much better agreements with coil library are achieved by the new force field parameters. In addition to the refitted ϕ, ψ potentials, to better account for the coupling with side-chain conformations, a few local (1-5/1-6) Lennard-Jones (L-J) interactions were modified in the new OPLS-AA/C force field. For Ala-derived amino acids, it is important to consider the intrinsic χ_1 rotamer preferences. In our new OPLS-AA/C force field, an excellent agreement with coil library data can be easily achieved (Fig. 8) by adjusting a few parameters of the related torsion potentials.

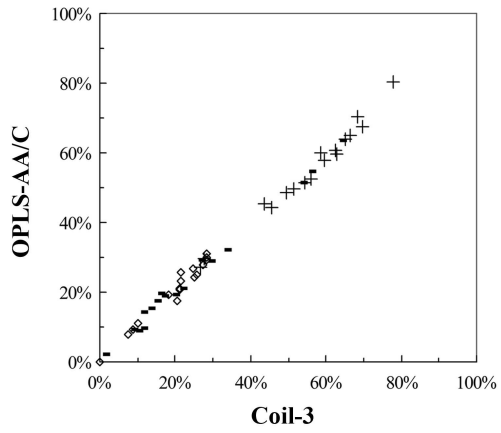


Figure 8. Side-chain χ_1 rotamer distributions from OPLS-AA/C simulations of the 17 amino acids obtained by dipeptide simulations plotted against the coil library statistics. The symbols have the same meaning as in Fig. 6.

Table 6. ^3J(HN,Hα) coupling constants of some amino acids in dipeptide model, comparison between experiments and MD simulations.

	Ala	Gln	Lys	Tyr	Val	RMSD
Experiment[57]	6.06	7.14	6.83	7.13	7.30	
AMBER-ff99sb-ildn	7.4	7.9	7.7	7.8	8.3	0.93
AMBER-ff03	6.8	7.6	6.6	6.7	7.8	0.49
CHARMM-27/CMAP	7.0	7.3	7.3	7.8	7.2	0.56
OPLS-AA/L	7.4	7.8	7.8	8.0	7.9	0.95
OPLS-AA/C	6.5	7.4	7.3	7.5	7.9	0.47

To further validate the new force field, the ^3J(HN,Hα) coupling constants were calculated and compared with the results from other force fields. As shown in Table 5, experimental *J*-coupling constants are in the order Ala < Lys < Gln ~ Tyr < Val, roughly agree with their β-sheet propensities. Experimental *J*-coupling constants for Gln and Tyr are more than 1.0 Hz higher than for Ala, which is correctly reproduced by OPLS-AA/C. Other force fields all give smaller differences between the coupling constants for Ala and Gln/Lys/Tyr. They can not well reproduce the relative differences between amino acids from experiments. Although AMBER-03 and CHARMM-27 give out mean deviations (RMSD) as small as that of OPLS-AA/C, this results from their over-stabilization of α conformation, which gives relatively lower ^3J(HN,Hα) values similar to P_{II}. We also simulated the cationic tripeptides using the new OPLS-AA/C force field. Going from Ala_3 to Val_3, the calculated ^3J(HN,Hα) coupling constant increases from 6.1 Hz to 8.1 Hz, in good agreement with the experimentally observed change from 5.7 Hz to 7.9 Hz.

5. Applications of the new force field

Current standard force fields gave out very different α-helix content of polyAla-based peptides, except for those tuned for better helix-coil balance.[108] Indeed, the cooperatively in α-helix formation can significantly multiply small inaccuracy (<1kJ/mol) in ϕ, ψ energetics of a single residue. Also, current force fields usually give insufficient temperature-dependence of folding/unfolding equilibrium of α-helical peptides.[178,179] We evaluated the performance of the new OPLS-AA/C force field by carrying out REMD simulations of α-helical peptides: Ac-Ala_{14}-Nme (A14) and Ac-$(AAKAA)_5$GY-Nme (AK17). TIP4P/Ew water model was used for both OPLS-AA/L and OPLS-AA/C force fields, while TIP3P water was used for other force fields. To increase the sampling efficiency, the mass of oxygen atom in the TIP4P/Ew model was reduced from 16 amu to 2 amu without changing the thermodynamics.[180] As shown in Fig. 9, AMBER-03 and CHARMM-27 gave too high helical propensity, while AMBER-99sb gave very low helical contents. The coil library-based OPLS-AA/C force field gave much better helicities. The new OPLS-AA/C force field predicts stronger *T*-dependency than the original OPLS-AA/L force field, closer to the experimental results. Nevertheless, the OPLS-AA/C force field slightly overestimates the α-helicity of these peptides. Further improvement of the force field may be achieved by adding very small ϕ/ψ correction, as proposed by Best et al.[108]

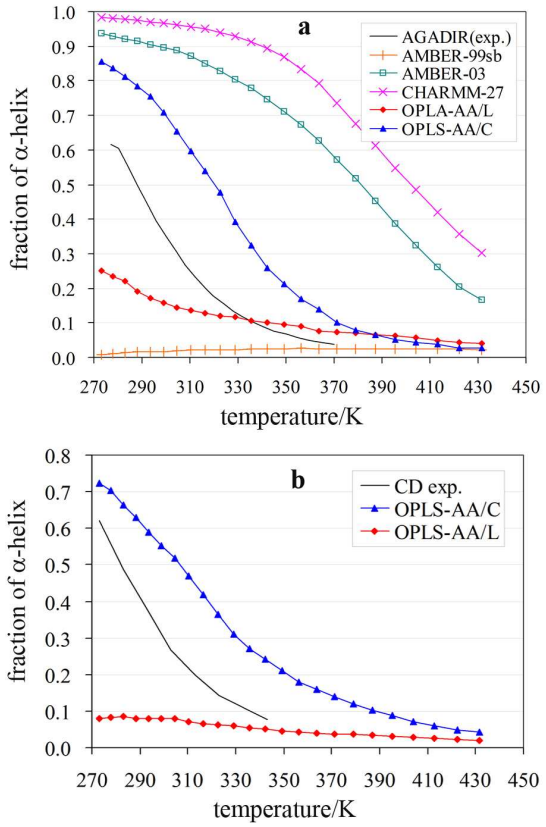


Figure 9. The melting curves of α-helical peptides from experiments[181,182] and various force fields. (a) Ac-Ala_{14}-Nme (A14), (b) Ac-$(AAKAA)_5$GY-Nme (AK17).

A stringent test of the overall performance of a protein force field is through the folding simulations of peptides/proteins, starting from their unfolded structures. We carried out folding simulations of CLN025 (a mutant of chignolin)[183] and Trpzip2[184] peptides, both of which are β-hairpins with protein-like cooperative folding behaviors, and also the Trp-cage [185] mini-protein, which contains an α-helix, a 3_{10}-helix, and a P_{II} segment. We also chose two fastest folding natively-occurring domains: the Nle/Nle double mutant of the villin headpiece subdomain[186,187] and the GTT variant of the Fip35 WW domain[188]. The native structures of former and later consist of three α-helices and three β-strands, respectively. All of them have attracted significant attention in protein folding studies.[189-192]

As shown in Fig. 10, REMD simulations using the OPLS-AA/L force field can only fold the two smallest β-hairpin peptides to their native structures. On the other hand, the representative structures from folding simulations using improved OPLS-AA/C force field can be aligned very well with the experimental structures (Cα RMSD ~ 1 Å) for all the five peptides and proteins. Interestingly, the structures from OPLS-AA/L simulation of villin-HP indeed mainly contain

α-helix structures while simulation of WW domain mainly gives β-sheet structures. Therefore, the problem of OPLS-AA/L may not be the biased secondary structure preference, as people noticed for many force fields. Instead, OPLS-AA/L can not well reproduce the intrinsic conformational features of some special amino acids, and also for the side-chain rotamer preferences. This indicates that only achieving balance between different secondary structures may not be enough for accurate force fields.

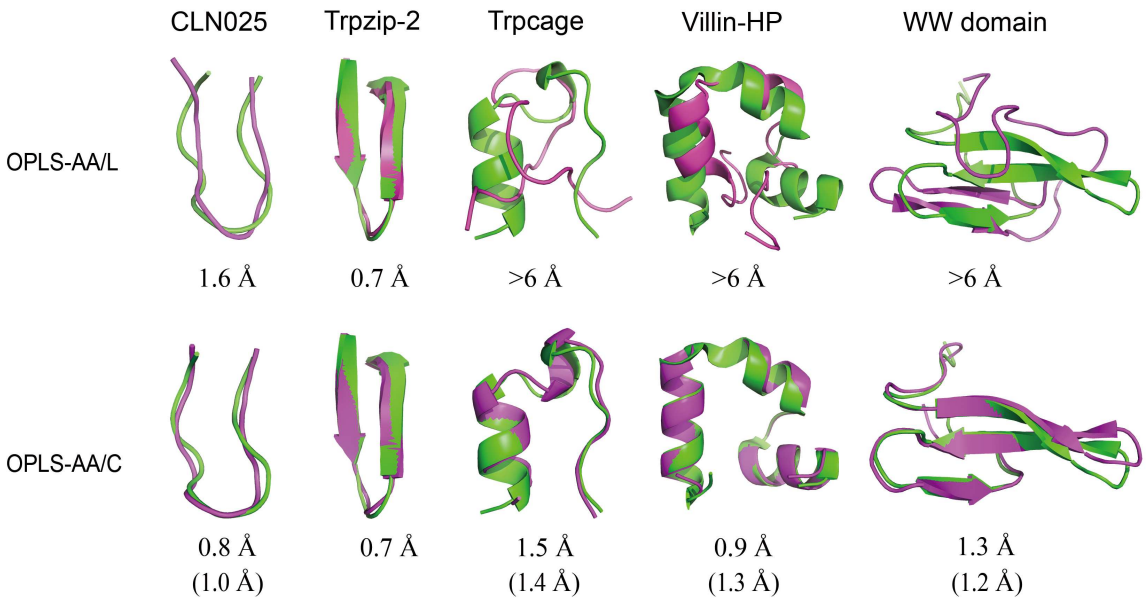


Figure 10. Predicted structure (magenta) from REMD simulation superposed with experimental structures (green). Predicted structure is the representative structure of largest cluster from replica of lowest temperature. Cα RMSD is also given below each structure. The values in parentheses are from recent work[192] of Lindorff-Larsen et al. for comparison.

Especially, the original OPLS-AA/L can not stabilize the native structure of Trpcage mini-protein, from REMD simulation initiated from PDB structure (1L2Y). This may partly due to its inaccurate description of ϕ, ψ preferences of Gly and Ser residues in the loop connecting packed α-helix and P_{II}-helix. The backbone ϕ distribution of Gly from OPLS-AA/L force field is mainly around ±120°, which may lead to a significant destabilization of β-turn conformations. Interestingly, AMBER-99sb force field, which gives quite low α-helicity for Ala-based peptides, can fold Trpcage quite well in a recent study.[193] This agrees with the fact that AMBER-99sb reproduces the coil library ϕ, ψ distributions of Ser and Gly much better than OPLS-AA/L

(Table 5).

Table 7. Comparison between the averaged backbone RMSD (Å) from the experimental structures for the three proteins simulated with OPLS-AA/L and OPLS-AA/C force fields.

	OPLS-AA/L	OPLS-AA/C
1P7E (56[a], 1-55[b])	2.12	0.96
5PTI (58[a], 1-55[b])	1.97	0.67
1UBQ (76[a], 1-71[b])	1.20	0.75

[a]The total number of residues.

[b]The residues involved in RMSD calculation.

Another way to validate force field improvement is through the long time normal MD simulations of globular proteins initiated from their native structures. Here the simulations of B3 domain of Protein G, bovine pancreatic trypsin inhibitor, and the ubiquitin were carried out from PDB entries 1P7E, 5PTI, and 1UBQ solvated in octahedron boxes containing ~2500 TIP4P/Ew water molecules. A 0.8 μs trajectory was generated for each protein at temperature of 300K and 1 atm pressure. The last 0.6 μs of the trajectory was used to calculate the deviation (backbone RMSD) from native structures, as shown in Table 7. Compared with original OPLS-AA/L, the simulated structures from the re-parameterized OPLS-AA/C force field were closer to the native structures for all three proteins, with backbone RMSD values < 1 Å.

6. Summary

Considering all solution experimental results from different sources, these studies gave large uncertainty in the percentages of different conformations (P_{II}, β and α). On the other hand, ϕ, ψ distributions from a statistical analysis of the restricted PDB coil library are much more invariant, regardless of different restrictions, and hence can provide more detail information. From the statistical side-chain-dependent ϕ, ψ plots of various amino acid types, strong coupling between side-chain and backbone conformational preferences can be found. These intrinsic conformational features can be understood by the interaction between γ atom(s) and adjacent backbone peptide groups. AMBER ff99sb, ff03 and OPLS-AA/L force fields are unable to reproduce the intrinsic side-chain rotamer preferences obtained from statistical analysis of protein coil library well, indicating further improvements might be needed.

We also found that the effect of the side-chain in certain rotameric states ($g+/t/g-$) on

backbone ϕ torsion and ψ torsion are additive. Based on this, the reported χ_1-dependent ϕ, ψ plots from protein coil library can be well reproduced by MP2/6-31+G** calculations with solvent effect of water. This ϕ, ψ decomposition scheme can also be very useful for force field parameterization. Most of current all-atom force fields reproduce the ϕ, ψ distribution of Ala residue from coil library reasonably well (similarity $S \sim 0.80$ - 0.91). Among the standard force fields, the results from AMBER-99sb agree best with the coil library statistics, in line with their good performances as reported recently.

The OPLS-AA force field was optimized to achieve an excellent match ($S = 0.98$) with the coil library ϕ, ψ distributions of Ala residue, and good agreement with the χ-dependent ϕ, ψ plots for other amino acids. The improved performances achieved by the new OPLS-AA/C force field were demonstrated on (1) *J*-coupling constants of Ala_3 and dipeptides of Ala and other amino acids, (2) temperature dependence of the α-helicities of Ala-based peptides, (3) predicting the native structures of five sequences with various secondary structure contents, and (4) stabilization of the native structures of globular proteins from long time MD simulations. The results indicate that a statistical analysis of the PDB coil library may be able to provide a good reference data set for force field parameterization.

Acknowledgement

We are grateful for the financial supports from the Research Grants Council of Hong Kong (663509), the Shenzhen Peacock Program (KQTD201103), and the National Science Foundation of China (21133002, 21203004). We also thank Prof. Olaf Wiest for improving the manuscript.

References

[1] P. G. Debrunner and H. Frauenfelder, *Ann. Rev. Phys. Chem.*, 1982, **33**, 283–299.

[2] S. Kumar, B. Ma, C.-J. Tsai, N. Sinha and R. Nussinov, *Protein Sci.*, 2000, **9**, 10–19.

[3] R. G. Smock and L. M. Gierasch, *Science*, 2009, **324**, 198–203.

[4] N. Tokuriki and D. S. Tawfik, *Science*, 2009, **324**, 203–207.

[5] S. R. Tzeng and C. G. Kalodimos, *Curr. Opin. Struct. Biol.*, 2011, **21**, 62–67.

[6] P. Tompa, *Trends Biochem. Sci.*, 2002, **27**, 527–533.

[7] H. J. Dyson and P. E. Wright, *Nat. Rev. Mol. Cell Biol.*, 2005, **6**, 197–208.

[8] T. Mittag, L. E. Kay and J. D. Forman-Kay, *J. Mol. Recognit.*, 2010, **23**, 105–116.

[9] Y. Huang and Z. Liu, *Proteins*, 2010, **78**, 3251–3259.

[10] T. H. Click, D. Ganguly and J. Chen, *Int. J. Mol. Sci.*, 2010, **11**, 5292–5309.

[11] V. N. Uversky, *Chem. Soc. Rev.*, 2011, **40**, 1623–1634.

[12] T. Chouard, *Nature*, 2011, **471**, 151–153.

[13] P. Y. Chou and G. D. Fasman, *Biochemistry*, 1974, **13**, 211–222.

[14] M. Levitt, *Biochemistry*, 1978, **17**, 4277–4285.

[15] V. S. Ananthanarayanan, R. H. Andreatta, D. Poland and H. A. Scheraga, *Macromolecules*, 1971, **4**, 417–424.

[16] K. T. O'Neil and W. F. Degrado, *Science*, 1990, **250**, 646–651.

[17] M. Blaber, X. J. Zhang and B. W. Matthews, *Science*, 1993, **260**, 1637–1640.

[18] A. Chakrabartty, T. Kortemme and R. L. Baldwin, *Protein Sci.*, 1994, **3**, 843–852.

[19] C. A. Kim and J. M. Berg, *Nature*, 1993, **362**, 267–270.

[20] C. K. Smith, J. M. Withka and L. Regan, *Biochemistry*, 1994, **33**, 5510–5517.

[21] L. J. Smith, K. M. Fiebig, H. Schwalbe and C. M. Dobson, *Folding and Design*, 1996, **1**, R95–R106.

[22] Z. Shi, K. Chen, Z. Liu and N. R. Kallenbach, *Chem. Rev.*, 2006, **106**, 1877–1897.

[23] W.-G. Han, K. J. Jalkanen, M. Elstner and S. Suhai, *J. Phys. Chem. B*, 1998, **102**, 2587–2602.

[24] C. Poon, E. T. Samulski, C. F. Weise and J. C. Weisshaar, *J. Am. Chem. Soc.*, 2000, **122**, 5642–5643.

[25] R. Schweitzer-Stenner, F. Eker, Q. Huang and K. Griebenow, *J. Am. Chem. Soc.*, 2001, **123**, 9628–9633.

[26] C. F. Weise and J. C. Weisshaar, *J. Phys. Chem. B*, 2003, **107**, 3265–3277.

[27] J. Grdadolnik, S. G. Grdadolnik and F. Avbelj, *J. Phys. Chem. B*, 2008, **112**, 2712–2718.

[28] F. Eker, K. Griebenow and R. Schweitzer-Stenner, *J. Am. Chem. Soc.*, 2003, **125**, 8178–8185.

[29] S. Woutersen, R. Pfister, P. Hamm, Y. G. Mu, D. S. Kosov and G. Stock, *J. Chem. Phys.*, 2002, **117**, 6833–6840.

[30] J. Graf, P. H. Nguyen, G. Stock and H. Schwalbe, *J. Am. Chem. Soc.*, 2007, **129**, 1179–1189.

[31] R. Schweitzer-Stenner, *J. Phys. Chem. B*, 2009, **113**, 2922–2932.

[32] K.-I. Oh, K.-K. Lee, E.-K. Park, D.-G. Yoo, G.-S. Hwang and M. Cho, *Chirality*, 2010, **22**, E186-E201.

[33] B. Sharma and S. A. Asher, *J. Phys. Chem. B*, 2010, **114**, 6661–6668.

[34] R. Schweitzer-Stenner, F. Eker, K. Griebenow, X. Gao and L. A. Nafie, *J. Am. Chem. Soc.*, 2004, **126**, 2768–2776.

[35] R. Schweitzer-Stenner, T. Measey, L. Kakalis, F. Jordan, S. Pizzanelli, C. Forte and K. Griebenow, *Biochemistry*, 2007, **46**, 1587–1596.

[36] A. Hagarman, T. J. Measey, D. Mathieu, H. Schwalbe and R. Schweitzer-Stenner, *J. Am. Chem. Soc.*, 2010,

132, 540–551.

[37] Z. Shi, K. Chen, Z. Liu, A. Ng, W. C. Bracken and N. R. Kallenbach, *Proc. Natl. Acad. Sci. U.S.A.*, 2005, **102**, 17964–17968.

[38] D. Verbaro, I. Ghosh, W. M. Nau and R. Schweitzer-Stenner, *J. Phys. Chem. B*, 2010, **114**, 17201–17208.

[39] K. Chen, Z. Liu, C. Zhou, W. C. Bracken and N. R. Kallenbach, *Angew. Chem. Int. Ed.*, 2007, **46**, 9036–9039.

[40] Z. Shi, C. A. Olson, G. D. Rose, R. L. Baldwin and N. R. Kallenbach, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 9190–9195.

[41] R. Schweitzer-Stenner and T. J. Measey, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 6649–6654.

[42] J. Grdadolnik, V. Mohacek-Grosev, R. L. Baldwin and F. Avbelj, *Proc. Natl. Acad. Sci. U.S.A.*, 2011, **108**, 1794–1798.

[43] M. A. Mehta, E. A. Fry, M. T. Eddy, M. T. Dedeo, A. E. Anagnost and J. R. Long, *J. Phys. Chem. B*, 2004, **108**, 2777–2780.

[44] P. Mukhopadhyay, G. Zuber and D. N. Beratan, *Biophys. J.*, 2008, **95**, 5574–5586.

[45] M. B. Swindells, M. W. Macarthur and J. M. Thornton, *Nat. Struct. Biol.*, 1995, **2**, 596–603.

[46] L. Serrano, *J. Mol. Biol.*, 1995, **254**, 322–333.

[47] K. M. Fiebig, H. Schwalbe, M. Buck, L. J. Smith and C. M. Dobson, *J. Phys. Chem.*, 1996, **100**, 2661–2666.

[48] L. J. Smith, K. A. Bolin, H. Schwalbe, M. W. MacArthur, J. M. Thornton and C. M. Dobson, *J. Mol. Biol.*, 1996, **255**, 494–506.

[49] C. J. Penkett, C. Redfield, I. Dodd, J. Hubbard, D. L. Mcbay, D. E. Mossakowska, R. Smith, C. M. Dobson and L. J. Smith, *J. Mol. Biol.*, 1997, **274**, 152–159.

[50] S. R. Griffiths-Jones, G. J. Sharman, A. J. Maynard and M. S. Searle, *J. Mol. Biol.*, 1998, **284**, 1597-1609.

[51] T. M. O'Connell, L. Wang, A. Tropsha and J. Hermans, *Proteins*, 1999, **36**, 407–418.

[52] S. Hovmöller, T. Zhou and T. Ohlson, *Acta Cryst. D*, 2002, **58**, 768–776.

[53] S. C. Lovell, I. W. Davis, W. Arendall III, P. I. W. de Bakker, J. M. Word, M. G. Prisant, J. S. Richardson and D. C. Richardson, *Proteins*, 2003, **50**, 437–450.

[54] F. Avbelj and R. L. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.*, 2003, **100**, 5742–5747.

[55] N. C. Fitzkee, P. J. Fleming and G. D. Rose, *Proteins*, 2005, **58**, 852–854.

[56] A. K. Jha, A. Colubri, M. H. Zaman, S. Koide, T. R. Sosnick and K. F. Freed, *Biochemistry*, 2005, **44**, 9691–9702.

[57] F. Avbelj, S. G. Grdadolnik, J. Grdadolnik and R. L. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 1272–1277.

[58] L. Ormeci, A. Gursoy, G. Tunca and B. Erman, *Proteins*, 2007, **66**, 29–40.

[59] L. L. Perskie, T. O. Street and G. D. Rose, *Protein Sci.*, 2008, **17**, 1151–1161.

[60] F. Jiang, W. Han and Y. D. Wu, *J. Phys. Chem. B*, 2010, **114**, 5840–5850.

[61] W. Peti, M. Hennig, L. J. Smith and H. Schwalbe, *J. Am. Chem. Soc.*, 2000, **122**, 12017–12018.

[62] A. K. Jha, A. Colubri, K. F. Freed and T. R. Sosnick, *Proc. Natl. Acad. Sci. U.S.A.*, 2005, **102**, 13099–13104.

[63] P. Bernado, L. Blanchard, P. Timmins, D. Marion, R. Ruigrok and M. Blackledge, *P. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 17002–17007.

[64] Y. Fujitsuka, G. Chikenji and S. Takada, *Proteins*, 2006, **62**, 381–398.

[65] M. R. Betancourt, *J. Phys. Chem. B*, 2008, **112**, 5058–5069.

[66] I. A. Rata, Y. Li and E. Jakobsson, *J. Phys. Chem. B*, 2010, **114**, 1859–1869.

[67] A. Soranno, R. Longhi, T. Bellini and M. Buscaglia, *Biophys. J.*, 2009, **96**, 1515–1528.

[68] C. Moller and M. S. Plesset, *Phys. Rev.* 1934, **46**, 618–622.

[69] V. Barone and M. Cossi, *J. Phys. Chem. A* 1998, **102**, 1995–2001.

[70] Y. Takano and K. N. Houk, *J. Chem. Theory. Comput.* 2005, **1**, 70–77.

[71] Z. Wang and Y. Duan, *J. Comput. Chem.*, 2004, **25**, 1699–1716.

[72] B. K. Ho, A. Thomas and R. Brasseur, *Protein Sci.*, 2003, **12**, 2508–2522.

[73] F. F. Garcia-Prieto, G. I. Fdez, M. A. Aguilar and M. E. Martin, *J. Chem. Phys.*, 2011, **135**, 194502.

[74] M. Karplus and J. Kuriyan, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 6679–6685.

[75] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. B. Shan and W. Wriggers, *Science*, 2010, **330**, 341–346.

[76] Y. Q. Zhou, Y. Duan, Y. D. Yang, E. Faraggi and H. X. Lei, *Theor. Chem. Acc.*, 2011, **128**, 3–16.

[77] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.

[78] W. L. Jorgensen, D. S. Maxwell and J. Tiradorives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.

[79] M. D. Beachy, D. Chasman, R. B. Murphy, T. A. Halgren and R. A. Friesner, *J. Am. Chem. Soc.*, 1997, **119**, 5908–5920.

[80] A. D. Mackerell, D. Bashford, Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-Mccarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.

[81] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. B*, 2001, **105**, 6474–6487.

[82] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. M. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee,

J. Caldwell, J. M. Wang and P. A. Kollman, *J. Comput. Chem.*, 2003, **24**, 1999–2012.

[83] A. D. Mackerell, M. Feig and C. L. Brooks, *J. Am. Chem. Soc.*, 2004, **126**, 698–699.

[84] A. D. Mackerell, M. Feig and C. L. Brooks, *J. Comput. Chem.*, 2004, **25**, 1400–1415.

[85] H. Hu, M. Elstner and J. Hermans, *Proteins*, 2003, **50**, 451–463.

[86] T. Yoda, Y. Sugita and Y. Okamoto, *Chem. Phys.*, 2004, **307**, 269–283.

[87] T. Yoda, Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.*, 2004, **386**, 460–467.

[88] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins*, 2006, **65**, 712–725.

[89] H. Fujitani, A. Matsuura, S. Sakai, H. Sato and Y. Tanida, *J. Chem. Theory Comput.*, 2009, **5**, 1155–1165.

[90] T. Z. Lwin and R. Luo, *Protein Sci.*, 2006, **15**, 2642–2655.

[91] P. L. Freddolino, F. Liu, M. Gruebele and K. Schulten, *Biophys. J.*, 2008, **94**, L75–L77.

[92] P. L. Freddolino, S. Park, B. Roux and K. Schulten, *Biophys. J.*, 2009, **96**, 3772–3780.

[93] N. Todorova, F. S. Legge, H. Treutlein and I. Yarovsky, *J. Phys. Chem. B*, 2008, **112**, 11137–11146.

[94] J. Vymětal and J. Vondrášek, *J. Phys. Chem. B*, 2010, **114**, 5632–5642.

[95] L. Wickstrom, A. Okur and C. Simmerling, *Biophys. J.*, 2009, **97**, 853–856.

[96] E. J. Thompson, A. J. Depaul, S. S. Patel and E. J. Sorin, *PLoS ONE*, 2010, **5**, e10056.

[97] T. Wang and R. C. Wade, *J. Chem. Theory Comput.*, 2006, **2**, 140–148.

[98] E. Project, E. Nachliel and M. Gutman, *J. Comput. Chem.*, 2010, **31**, 1864–1872.

[99] D. Matthes and B. L. de Groot, *Biophys. J.*, 2009, **97**, 599–608.

[100] R. B. Best, N. V. Buchete and G. Hummer, *Biophys. J.*, 2008, **95**, L07–L09.

[101] P. Hobza and J. Šponer, *J. Am. Chem. Soc.*, 2002, **124**, 11802–11808.

[102] P. Jurecka, J. Sponer, J. Cerny and P. Hobza, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1985–1993.

[103] S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

[104] N. Kamiya, Y. S. Watanabe, S. Ono and J. Higo, *Chem. Phys. Lett.*, 2005, **401**, 312–317.

[105] J. Chen, W. Im and C. L. Brooks, *J. Am. Chem. Soc.*, 2006, **128**, 3728–3736.

[106] E. Kim, S. Jang and Y. Pak, *J. Chem. Phys.*, 2007, **127**, 145104.

[107] D. Katagiri, H. Fuji, S. Neya and T. Hoshino, *J. Comput. Chem.*, 2008, **29**, 1930–1944.

[108] R. B. Best and G. Hummer, *J. Phys. Chem. B*, 2009, **113**, 9004–9015.

[109] R. B. Best and J. Mittal, *J. Phys. Chem. B*, 2010, **114**, 8790–8798.

[110] Y. Sakae and Y. Okamoto, *Mol. Simul.*, 2010, **36**, 1148–1156.

[111] H. Lei, Z.-X. Wang, C. Wu and Y. Duan, *J. Chem. Phys.*, 2009, **131**, 165105.

[112] P. S. Nerenberg and T. Head-Gordon, *J. Chem. Theory Comput.*, 2011, **7**, 1220–1230.

[113] D.-W. Li and R. Brüschweiler, *Angew. Chem. Int. Ed.*, 2010, **49**, 6778–6780.

[114] D.-W. Li and R. Brüschweiler, *J. Chem. Theory Comput.*, 2011, **7**, 1773–1782.

[115] W. Han, C.-K. Wan, F. Jiang and Y.-D. Wu, *J. Chem. Theory Comput.*, 2010, **6**, 3373–3389.

[116] W. Han, C.-K. Wan and Y.-D. Wu, *J. Chem. Theory Comput.*, 2010, **6**, 3390–3402.

[117] S. J. Marrink, A. H. de Vries and A. E. Mark, *J. Phys. Chem. B*, 2004, **108**, 750–760.

[118] Z. Shi, K. Chen, Z. Liu and N. R. Kallenbach, *Chem. Rev.*, 2006, **106**, 1877–1897.

[119] M. L. Tiffany and S. Krimm, *Biopolymers* 1968, **6**, 1379–1382.

[120] M. Karplus, *J. Chem. Phys.*, 1959, **30**, 11–15.

[121] J. Hu and A. Bax, *J. Am. Chem. Soc.*, 1997, **119**, 6360–6368.

[122] K. Ding and A. M. Gronenborn, *J. Am. Chem. Soc.*, 2004, **126**, 6232–6233.

[123] J. Wirmer and H. Schwalbe, *J. Biomol. NMR*, 2002, **23**, 47–55.

[124] J. Makowska, S. Rodziewicz-Motowidlo, K. Bagińska, M. Makowski, J. A. Vila, A. Liwo, L. Chmurzyński and H. A. Scheraga, *Biophys. J.*, 2007, **92**, 2904–2917.

[125] R. D. Gorbunov, P. H. Nguyen, M. Kobus and G. Stock, *J. Chem. Phys.*, 2007, **126**, 054509.

[126] J. Jeon, S. Yang, J. Choi and M. Cho, *Accounts Chem. Res.*, 2009, **42**, 1280–1289.

[127] J. Choi and M. Cho, *Chem. Phys.*, 2009, **361**, 168–175.

[128] A. Amadei, I. Daidone, A. Di Nola and M. Aschi, *Curr. Opin. Struc. Biol.*, 2010, **20**, 155–161.

[129] H. Torii, *J. Phys. Chem. B*, 2007, **111**, 5434–5444.

[130] K. Kwac, K. K. Lee, J. B. Han, K. I. Oh and M. Cho, *J. Chem. Phys.*, 2008, **128**, 105106.

[131] M. P. Gaigeot, *J. Phys. Chem. B*, 2009, **113**, 10059–10062.

[132] M. P. Gaigeot, *Phys. Chem. Chem. Phys.*, 2010, **12**, 3336–3359.

[133] M. P. Gaigeot, *Phys. Chem. Chem. Phys.*, 2010, **12**, 10198–10209.

[134] T. Head-Gordon, M. Head-Gordon, M. J. Frisch, C. L. Brooks and J. A. Pople, *J. Am. Chem. Soc.*, 1991, **113**, 5989-5997.

[135] M. Iwaoka, M. Okada and S. Tomoda, *J. Mol. Struct.*, 2002, **586**, 111–124.

[136] Z. Wang and Y. Duan, *J. Comput. Chem.*, 2004, **25**, 1699–1716.

[137] Y. K. Kang, *J. Phys. Chem. B*, 2006, **110**, 21338–21348.

[138] R. J. Lavrich, D. F. Plusquellic, R. D. Suenram, G. T. Fraser, A. Walker and M. J. Tubergen, *J. Chem. Phys.*, 2003, **118**, 1253–1265.

[139] V. Madison and K. D. Kopple, *J. Am. Chem. Soc.*, 1980, **102**, 4855–4863.

[140] F. Eker, X. Cao, L. Nafie, Q. Huang and R. Schweitzer-Stenner, *J. Phys. Chem. B*, 2003, **107**, 358–365.

[141] R. V. Pappu and G. D. Rose, *Protein Sci.*, 2002, **11**, 2437–2455.

[142] A. N. Drozdov, A. Grossfield and R. V. Pappu, *J. Am. Chem. Soc.*, 2004, **126**, 2574–2581.

[143] M. P. Hinderaker and R. T. Raines, *Protein Science*, 2003, **12**, 1188–1194.

[144] G. J. Bartlett, A. Choudhary, R. T. Raines and D. N. Woolfson, *Nat. Chem. Biol.*, 2010, **6**, 615–620.

[145] B. K. Ho and R. Brasseur, *BMC Struct. Biol.*, 2005, **5**, 14.

[146] D. Ting, G. Wang, M. Shapovalov, R. Mitra, M. I. Jordan and R. L. Dunbrack, Jr., *PLoS Comput. Biol.*, 2010, **6**, e1000763.

[147] M. Vijayakumar, H. Qian and H. X. Zhou, *Proteins*, 1999, **34**, 497–507.

[148] R. Srinivasan and G. D. Rose, *P. Natl. Acad. Sci. USA*, 1999, **96**, 14258–14263.

[149] F. Avbelj, *J. Mol. Biol.*, 2000, **300**, 1335–1359.

[150] S. Gnanakaran and A. E. Garcia, *J. Phys. Chem. B*, 2003, **107**, 12555–12557.

[151] D. B. Dahl, Z. Bohannan, Q. Mo, M. Vannucci and J. Tsai, *J. Mol. Biol.*, 2008, **378**, 749–758.

[152] M. J. Mcgregor, S. A. Islam and M. Sternberg, *J. Mol. Biol.*, 1987, **198**, 295–310.

[153] M. V. Shapovalov and R. J. Dunbrack, *Structure*, 2011, **19**, 844–858.

[154] R. J. Dunbrack and M. Karplus, *Nat Struct Biol*, 1994, **1**, 334–340.

[155] Pal, T. K.; Sankararamakrishnan, R. *J. Phys. Chem. B* **2010**, *114*, 1038–1049.

[156] N. J. West and L. J. Smith, *J. Mol. Biol.*, 1998, **280**, 867–877.

[157] M. Hennig, W. Bermel, A. Spencer, C. M. Dobson, L. J. Smith and H. Schwalbe, *J. Mol. Biol.*, 1999, **288**, 705–723.

[158] N. Vajpai, M. Gentner, J. Huang, M. Blackledge and S. Grzesiek, *J. Am. Chem. Soc.*, 2010, **132**, 3196–3203.

[159] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins*, 2010, **78**, 1950–1958.

[160] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins*, 2010, **78**, 1950–1958.

[161] I. Jákli, A. Perczel, Ö. Farkas, A. G. Császár, C. Sosa and I. G. Csizmadia, *J. Comput. Chem.*, 2000, **21**, 626–655.

[162] M. W. Klipfel, M. A. Zamora, A. M. Rodriguez, N. G. Fidanza, R. D. Enriz and I. G. Csizmadia, *J. Phys. Chem. A*, 2003, **107**, 5079–5091.

[163] A. Lang, K. Gyorgy, I. G. Csizmadia and A. Perczel, *J. Mol. Struct.*, 2003, **666**, 219–241.

[164] A. Láng, I. G. Csizmadia and A. Perczel, *Proteins*, 2005, **58**, 571–588.

[165] E. Yurtsever, D. Yuret and B. Erman, *J. Phys. Chem. A*, 2006, **110**, 13933–13938.

[166] R. J. Dunbrack and F. E. Cohen, *Protein Sci.*, 1997, **6**, 1661–1681.

[167] Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.*, 1999, **314**, 141–151.

[168] D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701–1718.

[169] E. J. Sorin and V. S. Pande, *Biophys. J.*, 2005, **88**, 2472–2493.

[170] M. S. Shell, R. Ritterson and K. A. Dill, *J. Phys. Chem. B*, 2008, **112**, 6878–6886.

[171] S. Piana, K. Sarkar, K. Lindorff-Larsen, M. H. Guo, M. Gruebele and D. E. Shaw, *J. Mol. Biol.*, 2011, **405**, 43–48.

[172] O. F. Lange, D. van der Spoel and B. L. de Groot, *Biophys. J.*, 2010, **99**, 647–655.

[173] S. A. Showalter and R. Bruschweiler, *J. Chem. Theory Comput.*, 2007, **3**, 961–975.

[174] S. Yang and M. Cho, *J. Phys. Chem. B*, 2007, **111**, 605–617.

[175] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura and T. Head-Gordon, *J. Chem. Phys.*, 2004, **120**, 9665–9678.

[176] B. K. Ho, A. Thomas and R. Brasseur, *Protein Sci.*, 2003, **12**, 2508–2522.

[177] A. E. Aliev and D. Courtier-Murias, *J. Phys. Chem. B*, 2010, **114**, 12358–12375.

[178] R. B. Best and J. Mittal, *J. Phys. Chem. B*, 2010, **114**, 14916–14923.

[179] K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror and D. E. Shaw, *PLoS One*, 2012, **7**, e32131.

[180] I. C. Lin and M. E. Tuckerman, *J. Phys. Chem. B*, 2010, **114**, 15935–15940.

[181] W. Shalongo, L. Dugad and E. Stellwagen, *J. Am. Chem. Soc.*, 1994, **116**, 8288–8293.

[182] P. Luo and R. L. Baldwin, *Biochemistry*, 1997, **36**, 8413–8421.

[183] S. Honda, T. Akiba, Y. S. Kato, Y. Sawada, M. Sekijima, M. Ishimura, A. Ooishi, H. Watanabe, T. Odahara and K. Harata, *J. Am. Chem. Soc.*, 2008, **130**, 15327–15331.

[184] A. G. Cochran, N. J. Skelton and M. A. Starovasnik, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 5578–5583.

[185] J. W. Neidigh, R. M. Fesinmeyer and N. H. Andersen, *Nat Struct Biol*, 2002, **9**, 425–430.

[186] C. J. McKnight, P. T. Matsudaira and P. S. Kim, *Nat Struct Biol*, 1997, **4**, 180–184.

[187] J. Kubelka, T. K. Chiu, D. R. Davies, W. A. Eaton and J. Hofrichter, *J. Mol. Biol.*, 2006, **359**, 546–553.

[188] S. Piana, K. Sarkar, K. Lindorff-Larsen, M. H. Guo, M. Gruebele and D. E. Shaw, *J. Mol. Biol.*, 2011, **405**, 43–48.

[189] C. D. Snow, L. Qiu, Du D, F. Gai, S. J. Hagen and V. S. Pande, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 4077–4082.

[190] J. Zhang, M. Qin and W. Wang, *Proteins*, 2006, **62**, 672–685.

[191] H. Nymeyer, *J. Phys. Chem. B*, 2009, **113**, 8288–8295.

[192] K. Lindorff-Larsen, S. Piana, R. O. Dror and D. E. Shaw, *Science*, 2011, **334**, 517–520.

[193] R. Day, D. Paschek and A. E. Garcia, *Proteins*, 2010, **78**, 1889–1899.