PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

A comprehensive library of blocked dipeptides reveals intrinsic backbone conformational propensities of unfolded proteins

Kwang-Im Oh,[1] Kyung-Koo Lee,[1] Eun-Kyung Park,[1] Youngae Jung,[2] Geum-Sook Hwang,[2] and Minhaeng Cho[1,2]*

[1] Department of Chemistry, Korea University, Seoul 136-701, Korea
[2] Division of Analytical Research, Korea Basic Science Institute, Seoul 136-713, Korea

ABSTRACT

Despite prolonged scientific efforts to elucidate the intrinsic peptide backbone preferences of amino-acids based on understanding of intermolecular forces, many open questions remain, particularly concerning neighboring peptide interaction effects on the backbone conformational distribution of short peptides and unfolded proteins. Here, we show that spectroscopic studies of a complete library of 400 dipeptides reveal that, irrespective of side-chain properties, the backbone conformation distribution is narrow and they adopt polyproline II and *β*-strand, indicating the importance of backbone peptide solvation and electronic effects. By directly comparing the dipeptide circular dichroism and NMR results with those of unfolded proteins, the comprehensive dipeptides form a complete set of structural motifs of unfolded proteins. We thus anticipate that the present dipeptide library with spectroscopic data can serve as a useful database for understanding the nature of unfolded protein structures and for further refinements of molecular mechanical parameters.

Proteins 2012; 80:977–990.

Key words: peptide conformation; blocked dipeptide; polyproline II; unfolded protein structure.

INTRODUCTION

The chemical diversity of amino-acid side-chains expands nature's inventory of protein structures and functions. It is, however, not clear whether the same side-chain diversity plays any role in determining the local structures of unfolded or denatured proteins. Definition of the unfolded state of proteins is particularly essential for understanding their stability and folding on biological timescales. An intriguing hypothesis on the unfolded protein structures was suggested by Tiffany and Krimm, that is, "the backbone conformations of unfolded proteins are not random coils but include short stretches of polyproline II (P_{II}) structural motifs interspersed with turns and bends".[1–5] They discovered similarities between the circular dichroism (CD) spectra of denatured proteins and those of homopolymers of proline, which provided a clue about the backbone conformations of unfolded proteins. Subsequently, numerous spectroscopic and theoretical investigations for short alanine- and proline-based peptides verified this P_{II} hypothesis,[2,4–12] even though there also exist compelling evidences for presence of residual local structures in unfolded proteins and a variety of turn conformations in a series of blocked oligopeptides.[7,13–19] Particularly, Sosnick and coworkers obtained Ramachandran distributions of all amino-acid residues and emphasized the importance of nearest neighbor interactions, which can be asymmetric. Furthermore, Schweitzer-Stenner and coworkers showed that various β-turns are populated in several GXG peptides. Regarding to the stability of P_{II} conformation, Bartlett *et al.*[20] argued that an electronic effect of the $n \rightarrow \pi^*$ interaction between the non-bonding *n*-orbital of the *i*th peptide carbonyl oxygen atom and the anti-bonding π^*-orbital of the (*i*-1)th peptide carbonyl carbon atom provides substantial stabilization to the left-handed P_{II} and other types of helices. Nonetheless, whether such a high P_{II} propensity found in a limited number of short peptides is sufficiently generic for a much wider range of peptides is still an open question.[2] Are there distinct structural propensities of amino-acids in unfolded proteins? Similar to native protein structure determined by the

Additional Supporting Information may be found in the online version of this article.
Grant sponsor: The Ministry of Education, Science and Technology, National Research Foundation (NRF), Korea; Grant number: 20090078897; Grant sponsor: Korea Basic Science Institute (KBSI); Grant number: T31401; Grant sponsor: Korea University.
*Correspondence to: Minhaeng Cho, Department of Chemistry, Korea University, Seoul 136-701, Korea.
E-mail: mcho@korea.ac.kr.
Received 2 October 2011; Revised 7 November 2011; Accepted 10 November 2011
Published online 22 November 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24000

protein's amino-acid sequence not by the conformational preferences of the constituent amino-acids,[21] are conformational propensities[22–25] of amino-acids important information to understand the unfolded protein structures? Furthermore, is the same chemical diversity of amino-acid side-chains an important asset to the definition of unfolded state of proteins?

In this article, we address these questions and discuss about intricate details of the intrinsic backbone conformational propensities and their side-chain dependences, carrying out NMR and CD studies in combination with molecular dynamics (MD) simulations of all the 400 blocked dipeptides (X_1X_2; Ac-X_1aa-X_2aa-NH_2) in water. In comparison to blocked oligopeptides, the dipeptides are ideal and minimal size model systems for investigating possible neighbor interaction effects on intrinsic conformational preferences of amino-acids. Such spectroscopic studies reveal that both P_{II} and β-strand conformers are preponderant, indicating a significant deviation from coil structures, and that their aqueous solution structures do not strongly depend on side-chains. Consistent with these findings, we shall consider only P_{II} and β-strand conformers of all dipeptides, a close approximation to the actual distribution of conformers. Anchoring support for this approximation is presented in detail in Results and Discussion. Any excursions from this may well exist, but they would have only a minor impact on our results and conclusions. We shall additionally show that the present dipeptide library can be a complete collection of structural motifs for elucidating local conformations of unfolded proteins, for example, acylphosphatase[26] and lysozyme.[27] We anticipate our assay to be a supporting evidence on the notion that, in contrast to native protein structures dictated by Anfinsen dogma,[21] unfolded protein structures are determined by conformational propensities of individual amino-acids with weak dependences on side-chains, indicating that the chemical diversity of amino-acid side-chains may not be important in defining the unfolded state of proteins.

EXPERIMENTAL METHODS

Materials

Dipeptides were synthesized on the X-CTRa(Bead Tech, Seoul) with a LibrakitTM (Tokyo Rikakikai) solid-state synthesizer. Fmoc-amino acids were purchased from Bead Tech (Seoul, Korea). Fmoc-XX-OH was taken HBTU coupling for 1.5–2 h. Peptides were acetylated at the N terminus and amidated at the C terminus. All couplings were monitored using a ninhydrin test. Solvent washes with DMF, MeOH, CH_2Cl_2, and then DMF were performed right after deprotection, coupling, and capping. Peptide resin cleavage was performed with trifluoroacetic acid. Purity of the peptide lyophilized was determined using HPLC T1260 Infinity Micro-scale Purification/Spotting System(Agilent). The syntheses and characterizations of all the dipeptides were performed by Bead Tech (Seoul, Korea).

NMR spectroscopy

All NMR experiments were carried out on a Varian VnmrS600 NMR spectrometer equipped with a 5-mm $^1H\{^{13}C/^{15}N\}$ salt tolerant triple resonance cold probe. The WET and NOESY pulse sequences were employed for water suppression in one-dimensional (1D) and two-dimensional (2D) spectroscopy, respectively. Chemical shifts were referenced to an internal standard: DSS (2,2-dimethyl-2-silapentane-5-sulfonate, sodium salt) in D_2O/H_2O (1:9, pH 2–5). 1D ^1H NMR spectra were recorded on 10–15 m*M* dipeptide in D_2O/H_2O (1:9, pH 2) at 25°C. Here, temperature was controlled with a L99 temperature controller (Varian) and a TC-84 nitrogen air cooler (FTS Systems). A set of 38,462 complex data points was collected and 64 scans were averaged for aqueous solution. The original free induction decay (FID) data were zero-filled to 524,288 points before the Fourier transformation. The NMR data were analyzed with ACD/SpecManager (ACD/Labs). The peak-to-peak frequency corresponding to the $^3J(H^N,H^\alpha)$ coupling constant was measured using the resultant spectra obtained from peak-fitting with Gaussian+Lorentzian functions. In this work, the coupling constant $^3J(H^N,H^\alpha)$ is denoted as $^3J_{HN\alpha}$ for the sake of notational simplicity.

Circular dichroism

The CD spectra of all the dipeptides except for Ac-GG-NH_2 were measured on a JASCO J-815 spectrometer equipped with a 450-W Xe arc lamp, where dipeptide was dissolved in H_2O at about 1.0 m*M*. The CD spectra were measured with a scan speed of 100 nm/min using a 110-QS quartz cell (Hellma) with a path length of 1 mm and averaged over 10 scans. Temperature was controlled with a PTC-423S/15 Peltier system (JASCO). CD spectra were measured at five different temperatures (10, 20, 30, 50, and 70°C).

MD simulation

MD simulations of all the four hundred blocked dipeptides in water were carried out with the Sander module of the AMBER9 program package and the *ff*03 force fields.[28] A cubic periodic boundary condition was employed and the integration of the equations of motion was carried out using the Verlet leapfrog algorithm. SHAKE constraints were applied to all bonds between hydrogen and heavy atoms. Long range electrostatic interactions were fully taken into account, by using the particle mesh Ewald (PME) method.[29] The distance criterion to switch from direct sum to the PME calculation was assumed to be 10 Å. The number of TIP4P water

molecules in the box depends on the size of dipeptide, but typically it varies from 1100 to 1700. A fully extended dipeptide conformation was used as the initial structure for the MD simulation. The initial system was minimized by 500 steps of steepest-descent minimization and 500 steps of conjugate-gradient method with the solute molecule fixed. Then, the system was heated and equilibrated at 298 K for 500 ps in the NVT ensemble, by using the Berendsen temperature coupling method with a coupling constant of 1.0 ps. Subsequently, a 30 ns production run for each dipeptide solution was implemented under the constant temperature condition (298 K) in NPT ensemble. Since typical free energy barriers between different conformations appear to be similar to thermal energy at 298 K, multiple conformational transitions in each MD trajectory were observed, which guarantee that the entire Ramachandran space were explored by 30-ns simulation. For all the 400 dipeptide systems, the overall MD simulation time is 12 μs in total. All the constant temperature and pressure conditions were implemented using the weak coupling algorithm of Berendsen *et al.* The time step for the equilibration and production runs is set to be 1 fs.

RESULTS AND DISCUSSION

NMR coupling constants and chemical shifts

To obtain information on the dipeptide backbone conformations, we first used ^{1}H NMR to measure the $^3J_{HN\alpha}$ coupling constants as well as the peptide N−H^N and C$^\alpha$−H$^\alpha$ proton chemical shifts, denoted as δ_{NH} and $\delta_{HC\alpha}$, respectively [see Fig.1(A) and Tables S1–S4 in Supporting Information]. The spin-spin coupling constant $^3J_{HN\alpha}$ is a function of the dihedral angle ϕ_i (C'_{i-1} − N_i − C_i^α − C'_i)) through the Karplus equation.[30] Despite that the chemical shifts, δ_{NH} and $\delta_{HC\alpha}$, are complicatedly dependent on local backbone conformation, electronic effects of side-chain, and solvent environment, Avbelj *et al.* showed that $\delta_{HC\alpha}$ is partly related to the extent of solvent exposure[31] of each peptide. Thereby, these NMR quantities sensitively reflect peptide backbone conformation and local solvation environment. Dipeptides considered here have two N—H groups (two protons at the C-terminal amide group excluded) yielding two $^3J_{HN\alpha}$ values for each dipeptide. To make resonant peak assignments, one should carry out appropriate ^{15}N-isotope-labeling and 2D HSQC experiments, which are prohibitively difficult for the entire 400 dipeptide samples. Therefore, we rather consider average $^3J_{HN\alpha}$ value (=($^3J_{HN\alpha}$ (N-terminal) + $^3J_{HN\alpha}$ (C-terminal))/2) for each dipeptide [see the color matrix plot in Fig. 1(B)]. Similarly, the average δ_{NH} and $\delta_{HC\alpha}$ values for each dipeptide are analyzed here. Despite these averaging approximations, as will be shown later, the peptide backbone conformation analysis results and the estimated population distributions are little affected.

In Figure 1(A), to qualitatively examine the dipeptide conformational distribution, the $^3J_{HN\alpha}$, $\delta_{HC\alpha}$, and $\delta_{HC\alpha}$ data of the blocked dipeptides (red spheres) are plotted and directly compared with those of coils defined by the absence of α-helix and β-sheet, β-sheets, and α-helices.[23,32–34] The projected data (closed circles) onto the ($^3J_{HN\alpha}$, δ_{NH}), ($^3J_{HN\alpha}$, $\delta_{HC\alpha}$) and (δ_{NH}, $\delta_{HC\alpha}$) planes are also shown in this figure. Quite evidently, the three-dimensional distribution of the dipeptide $^3J_{HN\alpha}$, δ_{NH}, and $\delta_{HC\alpha}$ values significantly overlaps with that (blue spheres) of coil residues. This immediately suggests that the ϕ angle and backbone conformation distributions appear to be similar to those of coils, which supports the P_{II} hypothesis mentioned above. However, this observation does not necessarily confirm that the dipeptide conformations in aqueous solutions are coils, which are traditionally defined as non-α-helical and non-β-sheet. More exacting tests are consequently needed, particularly those that involve peptide backbone structure determination.

Circular dichroism

CD spectroscopy has been used for quantitative assays of protein secondary structures. To extract complementary information on the peptide backbone conformations, we thus measured the temperature-dependent CD spectra of 399 dipeptides except for the glycine dipeptide (Supporting Information Fig. S1). In Figure 1(C), all the CD spectra at 20°C are depicted. Despite variations in CD lineshapes, intensities, and negative and/or positive peak wavelengths, most of them exhibit a strong negative peak at ∼198 nm and often an additional weak positive peak at ∼210 nm. The thick solid line in Figure 1(C) is the average over all the 399 CD spectra, which clearly resembles typical P_{II} CD spectra of alanine-based oligopeptides and homopolymers of proline.[1,3,9,12,35] To our surprise, more than 300 dipeptide CD spectra exhibit at least one isodichroic point in the wavelength range from 200 to 210 nm (Supporting Information Table S5). Even the average CD spectra at different temperatures ranging from 10 to 70°C [see the inset of Fig. 1(C)] also show a single isodichroic point at 202.5 nm. This is strong evidence in support of the two-state approximation to overall conformational distributions of dipeptides.

The presence of an isodichroic point alone does not provide information about what the two preponderant conformers are. Therefore, we carried out detailed component analyses of the CD spectra using a modified singular value decomposition (SVD) method recently developed to determine the relative populations of all the possible trialanine conformers.[12] The outcomes of such analyses are (i) the molar CD spectra of the two dominant conformers, which should exhibit a single isodichroic

point and (ii) the *T*-dependent changes of relative populations of the two dominant conformers present in each dipeptide solution. Comparing the resultant molar CD spectra with those of trialanine, we found that the two conformers are essentially P_{II} and β. This is fully consistent with the vibrational spectroscopic studies of 19 *N*-acetylated and C-amidated amino-acids by Grdadolnik *et al.*,[36] where they showed that, with a possible exception of the Gly residue, the two conformers are P_{II} and β. The small amount of α-helical conformer (<10%) is neglected in the present analysis of P_{II} and β because they do not affect the interpretation, as shown below. In this case, Figure 1(A) clearly shows that the distribution of the dipeptide $^3J_{HN\alpha}$, δ_{NH}, and $\delta_{HC\alpha}$ values is distinctively different from that of α-helical residues. Therefore, the two dominant conformers should be P_{II} and β and they are discriminated by solvent water.

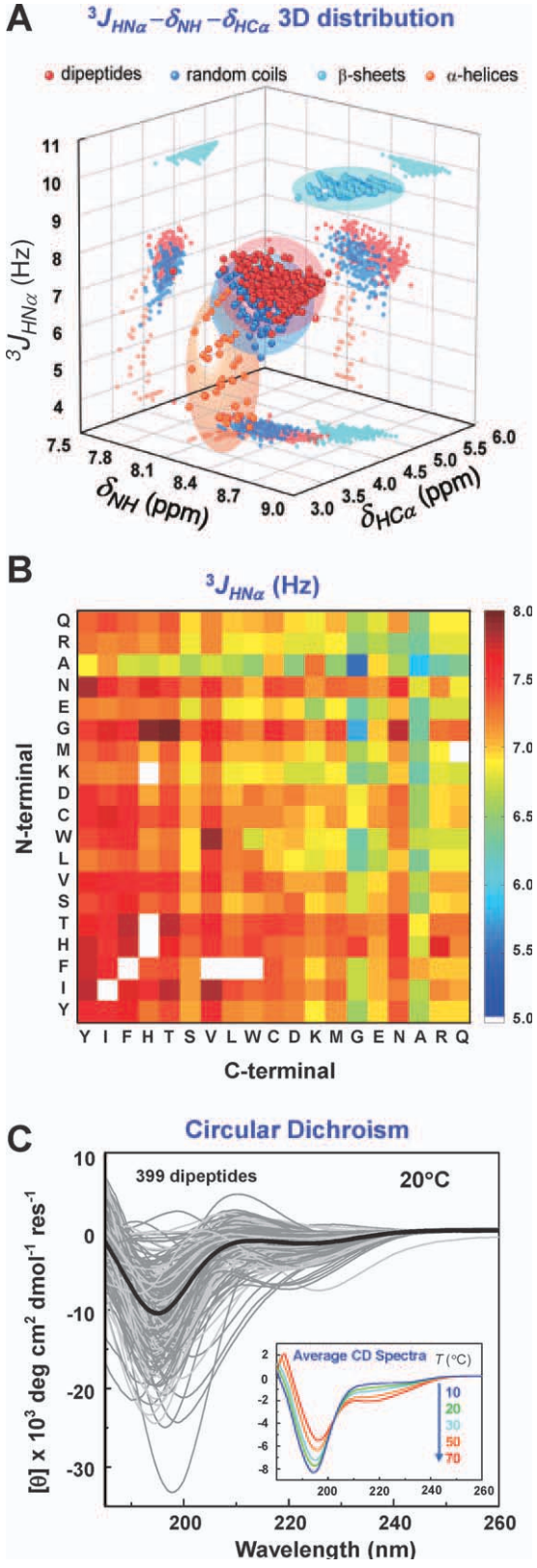

Figure 1

NMR coupling constants $^3J_{HN\alpha}$ (in Hz) and chemical shifts δ_{NH} and $\delta_{HC\alpha}$ (in ppm) of the N—H and C_α-H protons, respectively, of the dipeptides (at 20°C) are compared with those of random coils (blue), β-sheets (cyan), and α-helices (orange) obtained from Refs. 23,32–34 (**A**). The average coupling constants, $^3J_{HN\alpha}$, are shown in **B**. Here, Pro-containing dipeptides are not included. For each dipeptide, there are two different $^3J_{HN\alpha}$, δ_{NH}, and $\delta_{HC\alpha}$ values. Here, we consider their averages for a given dipeptide. Experimentally measured CD spectra (at 20°C) of all the dipeptides except for Gly dipeptide are shown in **C**. The thick solid line is the average CD spectrum. The temperature-dependent average CD spectra are also shown in the inset of **C**.

P_{II}/β equilibrium constant

Once the major conformers of the dipeptides are determined to be P_{II} and β, the relative P_{II} populations ($P_{P_{II}}$) of most of the dipeptides, except for those containing Pro or having very low solubility, can be calculated by using the reference coupling constants of individual amino-acids, which were reported by Shi *et al.* in Ref. 33, and the following linear equation,

$$J_{\exp} = P_{P_{II}} J_{P_{II}} + P_\beta J_\beta = P_{P_{II}} J_{P_{II}} + (1 - P_{P_{II}}) J_\beta, \qquad (1)$$

where $J_{\exp}$ is the experimentally measured average $^3J_{HN\alpha}$ value of a given dipeptide. $J_{P_{II}}$ and J_β are the reference coupling constants of P_{II} and β conformers of a given amino-acid, respectively. Shi *et al.* used the coil library given in Ref. 37. More recently, Sosnick and coworkers have constructed improved coil libraries and reported the conformational distributions of amino-acids in coil structures.[19] However, we couldn't retrieve the reference coupling constants $J_{P_{II}}$ and J_β from their coil libraries so that those in Ref. 33 are used to estimate the relative P_{II} population in the present work. From the estimated P_{II} and β populations of the dipeptides, we could further determine the P_{II} propensity of each individual amino-acid. Let us denote the P_{II} population of the X_1X_2 dipeptide as $P_{P_{II}}$

(X_1, X_2), which is essentially an $n \times n$ matrix with $n = 19$ (Supporting Information Table S6). Then, one can calculate the partially averaged values defined as

$$\bar{P}^N_{P_{II}}(Xaa) = \frac{1}{n}\sum_{X_2=A}^{Y} P_{P_{II}}(Xaa, X_2)$$
$$\bar{P}^C_{P_{II}}(Xaa) = \frac{1}{n}\sum_{X_1=A}^{Y} P_{P_{II}}(X_1, Xaa). \quad (2)$$

Here, $\bar{P}^N_{P_{II}}(Xaa)$ and $\bar{P}^C_{P_{II}}(Xaa)$ are the average P_{II} population of the *Xaa* amino-acid when it is the N- and C-terminal residue in the dipeptides, respectively. Then, the P_{II} propensity of each individual amino-acid *Xaa* can be calculated by

$$\bar{P}_{P_{II}}(Xaa) = \{\bar{P}^N_{P_{II}}(Xaa) + \bar{P}^C_{P_{II}}(Xaa)\}/2. \quad (3)$$

The quantity $\bar{P}_{P_{II}}(Xaa)$ is in fact a doubly averaged P_{II} propensity of Xaa. One can sort out the amino-acids in the ascending order of the relative P_{II} populations ($\bar{P}_{P_{II}}(Xaa)$). This order of amino acids is used to draw color-matrix plots in this article. In Figure 2(A), the calculated $P_{P_{II}}(X_1, X_2)$ values are shown. The P_{II} population averaged over all the dipeptides is found to be 63.0% with surprisingly small standard deviation of 7.6%. This shows that, despite the variations in polarity, size, and hydrophobicity of amino-acid side-chains, the dipeptide backbone conformational distribution is strikingly narrow and peaks at two extended conformations, that is, P_{II} and β. In addition, careful inspections of the results in Figure 2(A) and of the estimated $\bar{P}_{P_{II}}(Xaa)$ values bear out that the dipeptides containing bulky and nonpolar side-chains such as Tyr, Ile, and Phe have comparatively small P_{II}/β ratio, whereas those with polar and small side-chains, for example, Gln, Arg, Ala, Asn, and Glu, have slightly higher P_{II} propensity.

To further investigate neighbor interaction effects on dipeptide conformational distributions, we compared the P_{II} populations of the X_1X_2 dipeptides with those of the X_2X_1 dipeptides [Fig. 2(B)], i.e., $P_{P_{II}}(X_1, X_2)$ versus $P_{P_{II}}(X_2, X_1)$. The standard deviation (0.169) of the diagonal distribution is about 5.6 times larger than that (0.030) of the antidiagonal distribution. Such small aspect ratio of the distribution in Figure 2(B) indicates that the P_{II}/β propensity of a given dipeptide does depend on the

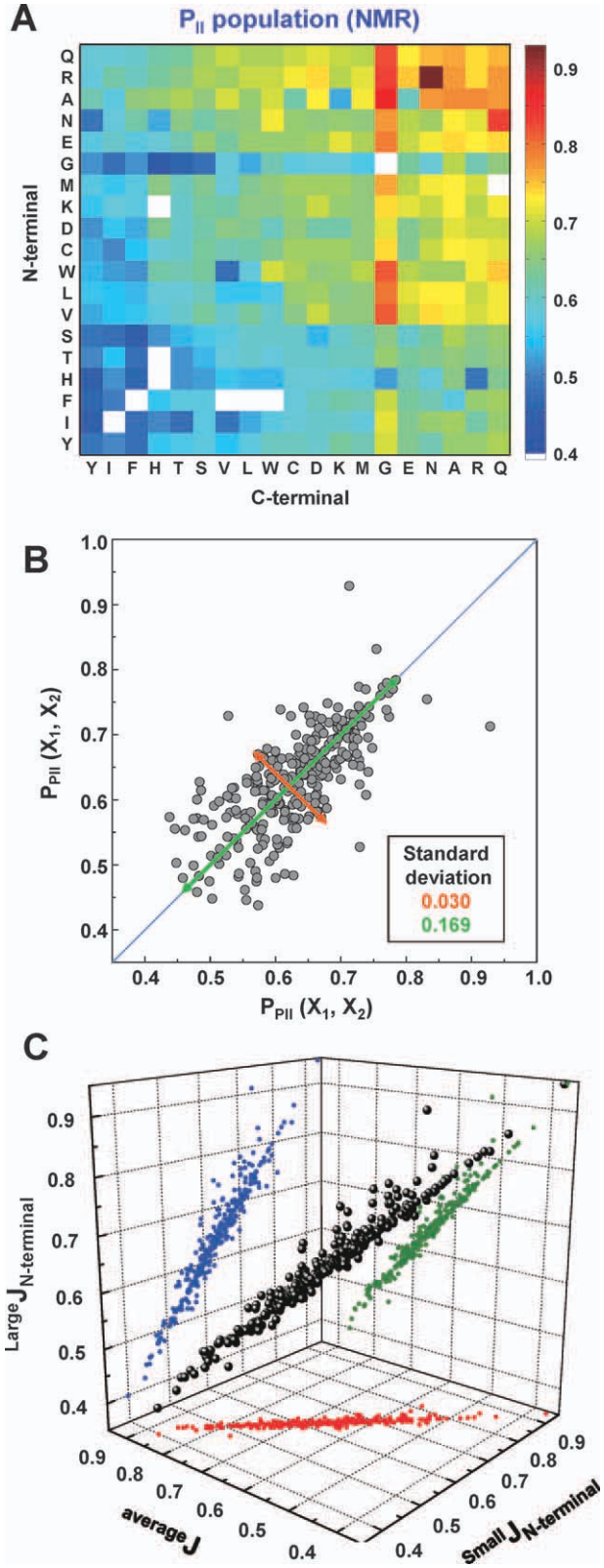

Figure 2

The relative P_{II} populations are shown in **A**, where Pro-containing dipeptides are not included. Data not available are represented by white blanks in the color matrix plots and the X- and Y-labels of amino-acids are in the ascending order of $\bar{P}_{P_{II}}(Xaa)$. To examine the neighbor interaction effect on P_{II} population, $P_{P_{II}}(X_2,X_1)$ is plotted with respect to $P_{P_{II}}(X_1,X_2)$ in **B**. The standard deviations along the diagonal and antidiagonal distributions are found to be 0.169 and 0.030, respectively. In **C**, $P_{P_{II}}(X_1,X_2)$ (at 25 °C) estimated by using (i) average $^3J_{HN\alpha}$ value for each dipeptide (the axis labeled with "$^{\text{average}}J$"), (ii) $^3J_{HN\alpha}$(N-terminal) = smaller $^3J_{HN\alpha}$ and $^3J_{HN\alpha}$(N-terminal) = larger $^3J_{HN\alpha}$ (the axis labeled with "$^{\text{Small}}J_{\text{N-terminal}}$"), and (iii) $^3J_{HN\alpha}$(N-terminal) = larger $^3J_{HN\alpha}$ and $^3J_{HN\alpha}$(N-terminal) = smaller $^3J_{HN\alpha}$ (the axis labeled with "$^{\text{Large}}J_{\text{N-terminal}}$").

constituent amino-acid but not on its amino-acid sequence. This reminds us of Flory's random coil model even though our dipeptides are not sufficiently long polymers at all.[38] Moreover, this relatively small neighbor interaction effect suggests that dipeptide conformation is largely determined by backbone peptide solvation as well as electronic effects, such as steric and $n \rightarrow \pi^*$ interactions, not by the properties and inter- and intramolecular interactions of side-chains.

Before we close this subsection, the approximate nature of the above procedure that originates from incomplete resonance assignments of the two N—H proton NMR peaks should however be briefly discussed. For a given dipeptide, only the average $^3J_{HN\alpha}$ value was used to estimate the P_{II} and β populations. To examine how large the error in estimating P_{II} population of each dipeptide is caused by such averaging approximation, we deliberately considered two additional cases: (i) the small (large) coupling constant among the two is associated with the N-terminal (C-terminal) peptide and (ii) the small (large) coupling constant among the two is associated with the C-terminal (N-terminal) peptide. With these different (and presumably incorrect) assignments, we could calculate the relative P_{II} populations and compare them with those obtained by using the average values [see Fig. 2(C)]. They all are in excellent correlations with one another, and the standard deviations are as small as about 0.02, which corresponds to 2% error in estimated population. This strongly indicates that the averaging approximation is in fact quite acceptable and that any inadequate assignments of the two peptide N—H peaks may not cause any significant deviations of the estimated P_{II} and β populations of the dipeptides.

ΔG between P_{II} and β

Once the relative populations of P_{II} and β conformers for each dipeptide are determined [Fig. 2(A)], one can obtain the corresponding equilibrium constant $K_{eq}(X_1, X_2)$ ($\equiv P_\beta(X_1, X_2)/P_{P_{II}}(X_1, X_2)$) for the X_1X_2 dipeptide. To extract information on the Gibbs free energy difference between the two conformers, we first calculate the partially averaged values that are defined as

$$\bar{K}^N_{eq}(Xaa) = \frac{1}{n}\sum_{X_2} K_{eq}(Xaa, X_2)$$
$$\bar{K}^C_{eq}(Xaa) = \frac{1}{n}\sum_{X_1} K_{eq}(X_1, Xaa), \quad (4)$$

where $\bar{K}^N_{eq}(Xaa)$ and $\bar{K}^C_{eq}(Xaa)$ are the average equilibrium constants of the *Xaa* amino-acid when it is the N- and C-terminal residue in the dipeptides, respectively. The doubly averaged equilibrium constant of each individual amino-acid Xaa can then be calculated by

$$\bar{K}_{eq}(Xaa) = \{\bar{K}^N_{eq}(Xaa) + \bar{K}^C_{eq}(Xaa)\}/2. \quad (5)$$

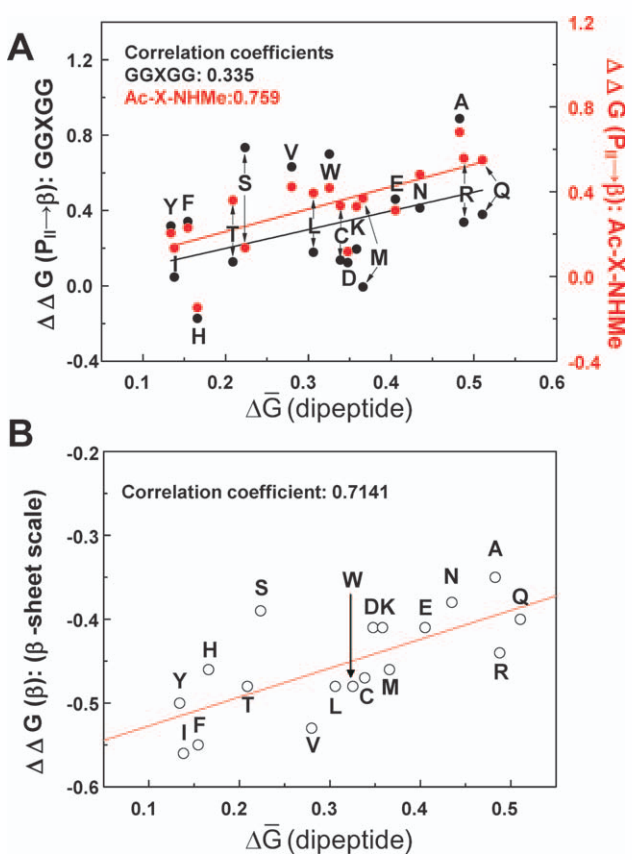

Figure 3

Apparent free energy difference $\Delta\bar{G}(Xaa)$ in kcal/mol of a given amino-acid, which was determined from the average equilibrium constant from Figure 2(**B**) is compared with those from electrostatic screening model results[5] for 18 blocked amino-acids (Ac-X-NHMe) and from amino-acid P_{II} propensities[33] in GGXGG pentapeptides (**A**). Also, the apparent free energy differences extracted from the amino-acid β-sheet scales[25] are compared with our results (B). Note that there is a good anticorrelation between $\Delta\bar{G}(Xaa)$ and β-sheet scale. Pearson correlation coefficients are given in this figure.

This $\bar{K}_{eq}(Xaa)$ value represents the average equilibrium constant of *Xaa* when it is included in dipeptides. Except for Gly and Pro, we have determined 18 doubly averaged free energy differences, which are related to the above equilibrium constants as

$$\Delta\bar{G}(Xaa) = -RT \ln \bar{K}_{eq}(Xaa). \quad (6)$$

In Figure 3, thus calculated $\Delta\bar{G}(Xaa)$'s are directly compared with Avbelj and Baldwin's electrostatic screening model results[39] for 18 blocked amino-acids (Ac-Xaa-NHMe), Shi *et al.*'s amino-acid P_{II} propensities in GGXGG pentapeptides, and Kim and Berg's amino-acid β-sheet scale.[25] Our results quantitatively correlate with those of blocked oligopeptides and with the amino-acid β-sheet scales. Noting that the relative population of β conformer is anticorrelated with that of P_{II} conformer,

such correlation between our $\Delta\bar{G}(Xaa)$ and Kim and Berg's $\Delta\Delta G(\beta; Xaa)$ is important evidence supporting the two-state approximation, where the two conformers are again P_{II} and β.

CD eigenspectra of P_{II} and β conformers

Carrying out the modified SVD analyses of the *T*-dependent CD spectra of more than 200 dipeptides, which exhibit a single isodichroic point, we determined both CD eigenspectra and *T*-dependent populations of P_{II} and β conformers. Although the CD eigenspectra of dipeptide P_{II} conformers display small variations in line shapes and negative peak wavelengths, those of β conformers appear to be strongly dependent on the constituent residues (see Supporting Information Fig. S2 for the P_{II} and β CD eigenspectra of a few representative dipeptides). Furthermore, some of the eigenspectra of β conformers appear to be similar to the CD spectra of typical α-helices, which exhibit double negative peaks, that is, "W"-shape, in the wavelength range from 200 to 240 nm. However, this does not indicate that the dominant component other than P_{II} is α-helical. For instance, we found that the eigenspectrum of the trialanine β conformer in pH = 2 aqueous solution showed two negative peaks at around 200 and 220 nm, even though the trialanine has essentially zero population of α-helical conformer. Second, the eigenspectra of the β conformers of the alanine-containing dipeptides studied here in this work exhibit a similar W-shape pattern also. Thus, in the cases of short peptides, such a shallow maximum in the CD spectrum does not necessarily indicate that the backbone conformation is α-helical. Third, the measured $^3J_{HN\alpha}$ values of the dipeptides are larger than those of P_{II} and α-helical residues [see Fig. 1(A)], which suggests that one of the two conformers should be non-α-helical. Since one of the two eigenspectra (Supporting Information Fig. S2) of each dipeptide can be assigned to that of P_{II}, we could assign the other eigenspectrum to that of β conformer. In addition, it should be mentioned that, unlike sufficiently long polypeptides that can form $i-(I+4)$ intramolecular hydrogen bonds, there is no driving force stabilizing such α-helical conformation for these short dipeptides.

Now, the validity of two-state approximation is further discussed by comparing our results with previous works by other groups. Recently, Schweitzer-Stenner and coworkers studied a series of GXG tripeptides and discussed the conformational propensities of amino-acids. Such GXG, GGXGG, and blocked amino-acid systems have been studied extensively before and the electronic CD and NMR methods were mainly used to reach the conclusion that these amino-acids prefer two conformations that are P_{II} and β. In Ref. 7, Hagarman *et al.* specifically considered eight GXG systems (X = A, V, F, L, S, E, K, and M) and reconfirmed Kallenbach and coworkers' conclusion that these eight tripeptides adopt mainly P_{II} and β-strand with minor populations of γ-turns and helical structures. Later, in Ref. 16, they studied the remaining GXG peptides for X = T, C, D, and N, where these four amino-acids are known to have a high propensity to be located in the i, $i+1$, $i+2$, and $i+3$ positions in β-turns. They showed that these four amino-acids have non-negligible propensities for Types I/II/I′/II′ β-turn populations, even though the sum of P_{II} and β-strand populations is more than 70% except for D. Such conclusion was based on multicomponent analyses of IR and Raman spectra and fitting analyses of various *J* coupling constants extracted from 1D and 2D NMRs. Their approaches and results are very interesting and provide a new insight into the nature of conformational propensities of blocked amino-acids, but there are some issues that are remained to be addressed in the future. First of all, the spectral band analyses of amide I vibrational spectra using multiple Voigtian (or other line shape functions) profiles are often very complicated due to spectral congestion problem. Furthermore, the uncoupled amide I local mode frequencies that are key ingredients in vibrational exciton theory for amide I spectrum were found to be highly dependent on the amino-acid side-chains, local backbone conformation, neighboring residues, and intramolecular electrostatic or H-bonding interactions. For instance, even without any conformational change, the amide I band can be shifted or split into two bands if there is a mixture of species with and without such intramolecular H-bonding interaction. In Ref. 16, it was particularly emphasized that the GDG has fairly large populations of type I/II′ β-turn and upper-right quadrant (in the Ramachandran plot) structures and suggested that such turn conformations are stable due to an intramolecular H-bonding interaction between Asp's COOH group and the backbone N—H group. However, the latter interaction might cause a shift of the amide I local mode frequency so that the exciton model calculation based on amide I coupling constants is not easy.

In this work, one of the important observations supporting the two-state approximation is the existence of an isodichroic point in the *T*-dependent CD spectra of dipeptides and in those obtained by averaging over all the 399 dipeptides [see the inset of Fig. 1(C)]. Kallenbach and coworkers also observed a single isodichroic point in their *T*-dependent CD spectra of various GGXGG peptides. In particular, both our *T*-dependent CD spectra of Ac-DD-NH_2 and Kallenbach and coworkers' GGDGG exhibit an isodichroic point in the *T*-dependent CD spectra. These strongly indicate that there are two major conformers. However, Schweitzer-Stenner *et al.* in Ref. 16 suggested that the populations of $(i+2)$ Types I/II′ β-turn and upper-right quadrant structures in GDG peptide are as significantly large as 28 and 20%, respectively, indicating that there are more than two major conformers populated. A possible resolution to this inconsistency between independent observations by three different groups is that the CD spectra of the β-turn,

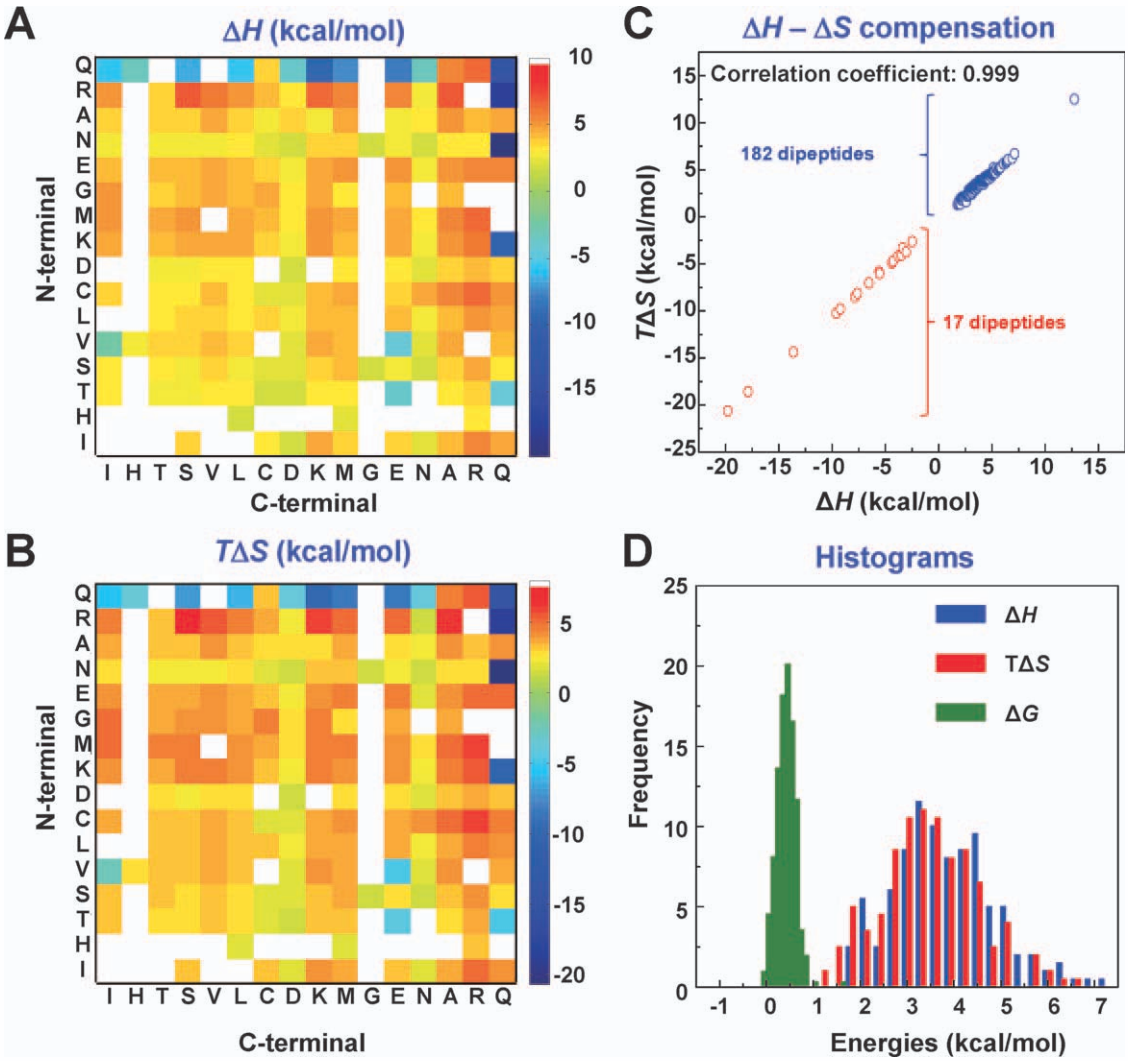

Figure 4

Using the singular value decomposition analyses of the *T*-dependent CD spectra, the ΔH and $T\Delta S$ (at 300 K) values associated with $P_{II} \rightarrow \beta$ transition were obtained and plotted in **A** and **B**, respectively. The distributions of ΔG (green), ΔH (blue), and $T\Delta S$ (red) are plotted in **C**. The distribution of ΔG is very narrow and its standard deviation is 0.2 kcal/mol. The enthalpy–entropy compensation plot is given in **D**.

upper-right quadrant structure, and β-strand should have the same CD intensity at a single wavelength accidentally. An alternative explanation would be that the CD intensity of the β-turn or upper-right quadrant structure are comparatively weak. To further prove whether such turn conformations are significantly populated even in the present dipeptide systems, it would be necessary to perform FTIR, isotropic, and anisotropic Raman, vibrational CD, and 1D and 2D NMR studies for all the 400 dipeptides. We are currently carrying out such investigations for the dipeptides.

ΔH, ΔS, and enthalpy–entropy compensation

As mentioned above, the SVD analyses of the *T*-dependent CD spectra provided quantitative information on *T*-dependent populations and β/P_{II} equilibrium constants. The van't Hoff plot of β/P_{II} equilibrium constant with respect to temperature gives us ΔH and $T\Delta S$ (at 300 K) values associated with the transition from P_{II} to β [see Fig. 4(A,B)]. Positive ΔH and ΔS values for most of the dipeptides verify that the enthalpy(entropy)-favored P_{II} (β) population decreases (increases) as temperature is increased. We next examine the distributions of the enthalpy and entropy differences between P_{II} and β conformers [Fig. 4(C)]. The ΔH ($T\Delta S$) distribution peaks at 3.9 ± 1.3 (3.6 ± 1.3) kcal/mol, which is significantly larger than the average ΔG ($= 0.32 \pm 0.20$ kcal/mol). Furthermore, both the magnitude and width of the ΔG distribution are an order of magnitude smaller than those of ΔH and $T\Delta S$ distributions. This signifies that the small free energy difference between the two conformers results from

the enthalpy–entropy compensation [Fig. 4(D)]. Such narrow distribution of ΔG values is an important observation indicating that the free energy difference associated with the dipeptide $P_{II} \leftrightarrow \beta$ equilibrium are not strongly dependent on the constituent amino-acids.

The average free energy difference (0.32 kcal/mol) is found to be smaller than the thermal energy at 300 K, which is 0.60 kcal/mol. Consequently, the β/P_{II} equilibrium constants are rather close to unity. On the other hand, the corresponding ΔH and $T\Delta S$ values are about 10 times larger than ΔG, which indicates that the enthalpic contribution to the stabilization of the P_{II} conformer is largely compensated by the entropic destabilization of the same P_{II} conformer. Such enthalpy–entropy compensation effects have been found in a number of different cases including protein folding–unfolding equilibria. Whether the enthalpy–entropy compensation is a real effect or not has been a controversial issue.[40–44] Our ΔH versus ΔS plot shown in Figure 4(D) is fitted with a linear line almost perfectly, where the slope and intercept are 1.01 and -0.38 kcal/mol, respectively, and the correlation coefficient is 0.9988. Despite such a perfectly linear correlation between ΔH and $T\Delta S$, one cannot confirm that the enthalpy–entropy compensation is a real effect here in the present cases. However, it is still interesting to note that such $\Delta H - \Delta S$ compensation found in numerous protein folding–unfolding equilibria is also applicable to the conformational transition between P_{II} and β-strand of each dipeptide. This led us to believe that such a simple conformational transition along the peptide backbone can be a key elementary step relating to the much more involved and complicated protein folding–unfolding processes in aqueous solutions.

Configurational entropy

Another interesting revelation is that the thermodynamic aspects of the dipeptide $P_{II} \leftrightarrow \beta$ equilibria are quite similar to those of thermal protein unfolding processes. A familiar argument on the latter is that ΔG_{unfold} between the folded and unfolded populations is about 10 kcal/mol for a protein with about 100 residues.[45] Typically, the ΔH_{unfold} and $T\Delta S_{unfold}$ associated with a protein unfolding have magnitudes an order of magnitude larger than ΔG_{unfold}, that is, they are about 100 kcal/mol. Then, the conformational entropy is about 3.33 entropy units per residue, indicating that the ratio of the numbers of microscopic states per residue is about 5 assuming the Boltzmann entropy formula, $\Delta S_{unfold} = k_B \ln(W_{unfolded}/W_{folded})$. For the present dipeptides, we have $\Delta S \cong 4.2$ entropy units per residue so that $W_\beta/W_{P_{II}}$ is about 8. Despite the differences in sizes, constituent residues, and involved interactions in proteins and dipeptides, the thermodynamic aspects of conformational transitions from (enthalpy-stabilized) folded to (entropy-favored) unfolded proteins are strikingly similar to those associated with such elementary conformational $P_{II} \leftrightarrow \beta$ transitions in dipeptides. This implies that quite possibly such simple local conformational changes constitute thermodynamic basics of protein unfolding processes.

P_{II} and β populations: MD results

In addition to NMR and CD, MD simulations of polypeptides in solutions have provided intricate details on their conformational distributions. In the present work, using NMR and CD methods, we showed that the two preponderant conformations are PII and β. However, to elucidate the underlying driving forces stabilizing such conformations, we carried out MD simulations for all the dipeptides. More specifically, we wanted to estimate the $n \rightarrow \pi^*$ interaction energies that are known to be important for stabilizing P_{II} conformation and to find possible correlations of solvent accessible surface area, molecular volume change, number of local water molecules with the determined β/P_{II} equilibrium constants, associated ΔG, ΔH, ΔS, and conformational propensities of amino-acids. Without MD trajectories, it is impossible to calculate such mechanical and molecular properties. We believe that this work is the first attempt trying to study such correlations between molecular (mechanical) properties of dipeptides and the thermodynamic properties associated with the $P_{II} \rightarrow \beta$ conformational transitions. It is believed that such comparative investigation is important in understanding the origin of the conformational propensities of amino-acids.

Even though there are certain discrepancies between the MD and experimental results on the backbone conformations of short peptides, Graf *et al.*[35] successfully demonstrated that a combination of MD and NMR methods is useful to estimate the reference $^3J_{HN\alpha}$ coupling constants of P_{II} and β conformers of alanine-based peptides. In the present MD simulations, the force fields used are known to have a tendency to overestimate the right-handed α-helical propensity of short oligopeptides. Therefore, we rather focus on the north-west quadrant in the ϕ,ψ-distributions of the dipeptides (Supporting Information Fig. S3 for the Ramachandran plots[46] of a few representative dipeptides) and calculated the relative P_{II} populations assuming that $P_{P_{II}} + P_\beta = 1$ [Fig. 5(A)]. Examining the conformational distributions of P_{II} and β-strand of all the 400 dipeptides, we found that these two conformers are simultaneously present in most of the dipeptides, except for those containing Pro. The dipeptides containing a proline residue have definite preferences of the P_{II} conformation with negligible β population, which is presumably because of the conformational constraint imposed by the pyrrolidine ring in Pro.

In Figure 5(B), the $\Delta G_{MD}/RT$ values using the calculated $P_{P_{II}}$ and P_β are directly compared with those from the NMR studies, but they show no notable correlation even though the average ΔG_{MD} ($=0.25 \pm 0.32$ kcal/mol)

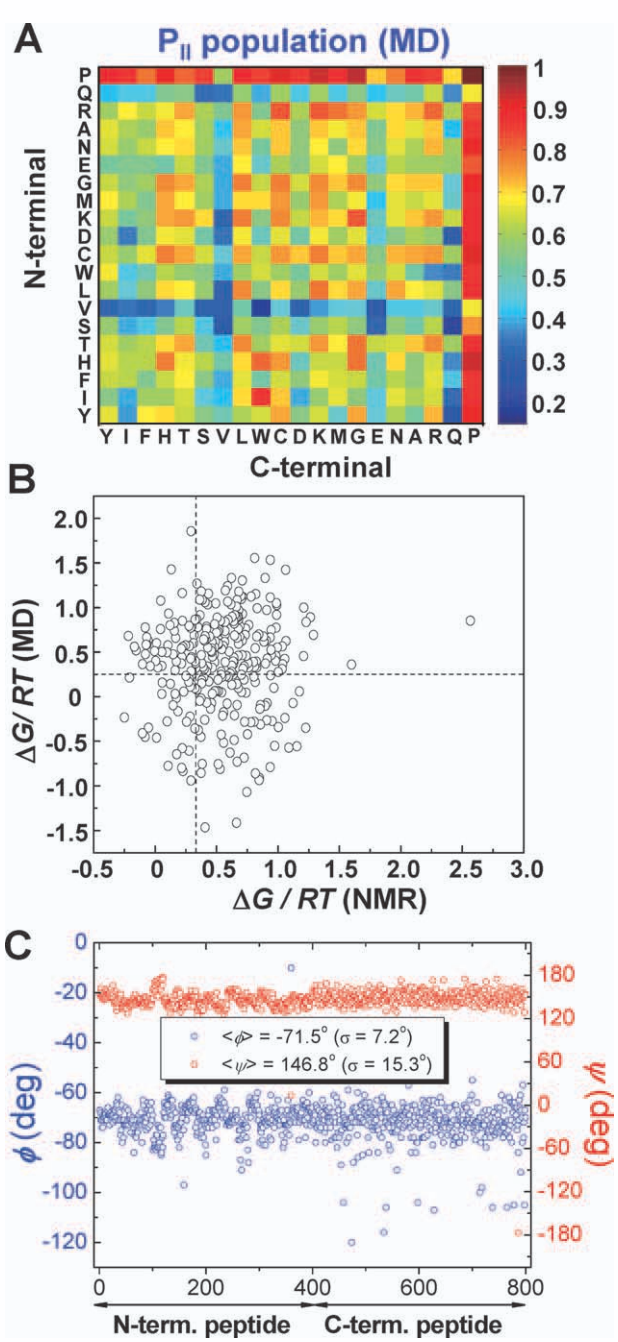

Figure 5

P_{II} populations obtained from the conformational probability analyses using MD trajectories (at 300 K) (**A**). The $\Delta G/RT$ ($= -\ln(P_\beta/P_{P_{II}})_{MD}$) values from the MD simulations are compared with those from the NMR studies (**B**). From the P_{II} trajectories of 400 dipeptides, the calculated ϕ (blue open circles) and ψ (red open circles) angles at the N- and C-terminal peptides are plotted (**C**). The average ϕ and ψ angles are found to be -71.5° ($\pm 7.2^\circ$) and 146.8° ($\pm 15.3^\circ$), respectively.

is coincidentally close to the experimental result ($=0.32 \pm 0.20$ kcal/mol). Despite the failure of the MD simulation method in reproducing the β/P_{II} equilibrium constants, still the MD trajectories can be useful to extract quantitative information about solvent accessible surface areas (SASA), solvent accessible volumes (SAV), numbers of surrounding water molecules, solvation energies, and *ab initio* electronic energies of the dipeptide P_{II} and β conformers. Often, the conformational transition of a biomolecule involves changes of molecular volume and surface area. The latter was found to be an important factor for describing the extent of such conformational change of protein. Also, the molecular volume change is known to be related to the associated reaction entropy, since the solvent configurational entropy is determined by the solute volume. First, to calculate SASA's of P_{II} and β conformers of each dipeptide, we used a method similar to the rolling-a-ball algorithm developed by Richards.[47,48] The entire space was divided by three-dimensional grid with 0.2 Å space. The cubic boxes (pixels) within the distance (=van der Waal radius + water solvent radius) from each atomic center are checked, where the water solvent radius was assume to be 1.4 Å. The SASA is obtained by counting the checked cubic boxes on its surface. The SAV can be simultaneously obtained by counting the total number of checked cubic boxes inside that boundary. In addition to SASA and SAV, we calculated the numbers of water molecules (more specifically water oxygen atoms) within the solvent shell whose boundary is located at 3.4 Å away from the van der Waals surface of each dipeptide. These water molecules in this distance range effectively participate in the hydrogen-bonding interactions with the backbone peptides as well as with the side-chains. For the P_{II} and β structures, we calculated the average numbers of water molecules in the shells, and then the difference numbers associated with the $P_{II} \rightarrow \beta$ conformational transition for all the dipeptides were calculated.

Such calculations enabled us to examine to what extent the changes of these mechanical properties are related to (i) relative P_{II} populations, (ii) enthalpy and entropy differences, (iii) NMR chemical shifts, and (iv) hydrophobicities of dipeptides. Various cross-correlation analyses by comparing numerically calculated SASA, SAV, and number of solvating water molecules with experimentally determined results such as β/P_{II} equilibrium constants, $\Delta\bar{G}(Xaa)$, and so on, were performed. However, we could not find any statistically significant correlations that are applicable to all the dipeptides considered here. This is mostly because the free energy differences among dipeptide conformers are very small and the force fields may not have such a high chemical accuracy yet. In literatures, the quantum mechanical/molecular mechanical (QM/MM) MD simulations, where the solute peptides are treated quantum mechanically and the surrounding solvent water molecules are considered as classical particles with an appropriate set of force field parameters, were found to be reliable in quantitatively predicting the conformational preferences of a few short peptides. However, our recent simulation studies

for a blocked dialanine in water showed that the semiempirical AM1[49] or PM3/MM MD[50,51] simulation methods are not quantitatively reliable in reproducing the experimentally measured vibrational CD spectra.[52] Therefore, unless extensive computational studies utilizing Hartree-Fock or DFT/MM MD methods are performed for short peptides, it is premature to conclude that the QM/MM MD methods are always better than the classical MD methods fully utilizing properly parameterized force fields.

Solute–solvent interaction energies: MD studies

Despite the limited accuracy of the MD force fields in predicting P_{II} and β populations, it was expected that the solute–solvent interaction energy calculations would provide a clue about how P_{II} and β conformers are specifically stabilized in aqueous solutions. Thus, using the MD trajectories of 400 dipeptides, we calculated the average solute–solvent Lennard-Jones (LJ) and Coulomb interaction energies, denoted as E_{LJ} and E_C, of the P_{II} and β conformers separately. In addition, the intrinsic electronic energies of the P_{II} and β structures without solvent molecules should also be different from each other. Using the P_{II} and β configurations sampled from the MD trajectories, we obtained the ensemble-average P_{II} and β structures and carried out quantum chemistry calculations with B3LYP/6-311++G** method for these two representative structures of each dipeptide. In total, we carried out 728 DFT calculations (364 P_{II} forms and 364 β forms)—note that the dipeptides containing Pro were excluded. The energy differences, denoted as ΔE_{LJ}, ΔE_C, and ΔE_{elec} associated with the conformational transition from P_{II} to β were obtained by subtracting E_{LJ}, E_C, and E_{elec} of β from those of P_{II}. We found that the average ΔE_{LJ}, ΔE_C, ΔE_{elec}, and ΔE_{total} over the 364 dipeptides are −0.42, 4.02, −1.64, and 1.96 kcal/mol, respectively, where the dipeptides having zero β-strand population are excluded in the averaging. As expected, the solute–solvent LJ interaction contribution, ΔE_{LJ}, to the total energy difference ΔE_{total} is comparatively smaller than the Coulomb interaction and intrinsic electronic energy contributions. The fact that the solute–solvent Coulomb interaction contribution to ΔE_{total} is the largest among the three indicates that the P_{II} conformation is Coulomb-energetically more stable than the β-strand conformation by about 4.02 kcal/mol. On the other hand, the electronic energy of the β-strand is lower than that of the P_{II}. Overall, the average energy difference, ΔE_{total}, calculated by analyzing the MD trajectories appears to be quantitatively similar to the experimental ΔH of 2.9 kcal/mol.

ϕ and ψ angles of P_{II} and β from MD trajectories

Analyzing the ϕ,ψ-distribution of each dipeptide obtained from MD trajectory, the peak ϕ and ψ angles of P_{II} and β conformers were determined and the resulting histograms were examined (Supporting Information Fig. S4). The distribution of ϕ angles of P_{II} conformers peaks at around −70° on average, whereas that of β-strand conformers at −155°. Interestingly, the ϕ angle distribution is narrow in comparison to the ψ angle distribution. This is fully consistent with the observation made by Avbelj and Baldwin in Ref. 39, where they carried out detailed analyses of coil library and examined the ϕ angle distributions of various amino-acids. The peak position of the P_{II} band is relatively constant for varying amino-acids, but that of the β band strongly depends on amino-acids. Indeed, the present MD simulation results show that the peak position of the P_{II} band in the ϕ angle distribution does not vary for different dipeptides. On the other hand, the ϕ angles of β-strand conformers are highly dependent on the constituent amino-acids of each dipeptide. This suggests that the potential energy well at the P_{II} basin is stiff in comparison to that at the β-strand basin. It will be highly interesting to carry out extensive quantum chemistry calculation studies on the potential energy surface of model dipeptide with respect to the ϕ angle with implicit or explicit solvent models in the future.

$n \rightarrow \pi^*$ electronic interaction energy

The energetic origin on the stability of the P_{II} conformation has not been fully understood yet, despite prolonged investigations to elucidate the stabilizing factors for the P_{II} conformation. It has been shown that the P_{II} structure has particularly low intramolecular steric repulsive interaction and that the backbone peptide solvation by surrounding water molecules may be an important factor. However, recently Raines and coworkers[20,53] suggested that the so-called $n \rightarrow \pi^*$ interaction between the non-bonding orbital of the $(i+1)$th peptide carbonyl oxygen atom and the π^*-orbital of the carbonyl carbon atom of the ith peptide bond can be an important stabilization factor that has not been taken into consideration in the previous molecular mechanics calculation studies. Since this $n \rightarrow \pi^*$ interaction is a short-range property in nature, they emphasized that such interactions should be important in determining the stability of various secondary protein structures. However, the β-sheet (or β-strand) conformation is not favorable for the $n \rightarrow \pi^*$ interaction not only because the relative distance between the two atoms (the $(i+1)$th peptide carbonyl oxygen atom and the carbonyl carbon atom of the ith peptide bond) is too large (>3.0 Å) but also because the angle between the neighboring peptide bonds is not close to 90°. Therefore, the $n \rightarrow \pi^*$ interaction energy for the dipeptide in the β-strand conformation is believed to be zero.

Here, we found that the average ΔH between P_{II} and β is about 3.9 kcal/mol experimentally, which was attributed to the backbone peptide solvation and electronic

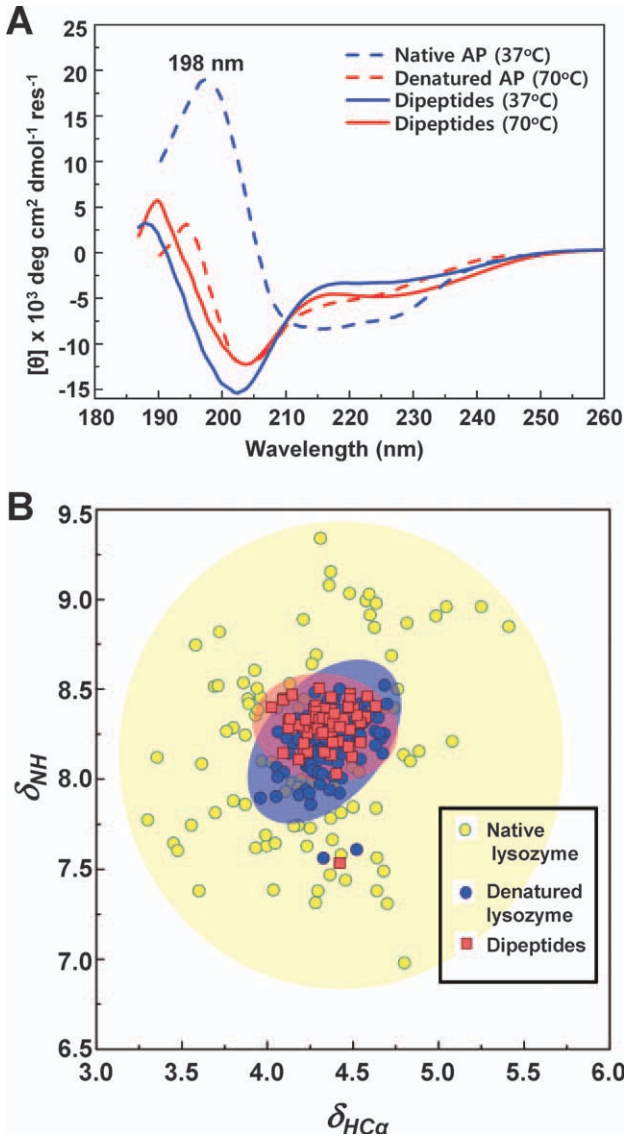

Figure 6

A: CD spectra of native (37°C) and heat-denatured (70°C) acylphosphatase(AP)[26] and simulated CD spectra using Eq. (2) with our dipeptide CD library. The simulated CD spectra are red-shifted by 6.91 nm and the negative peak intensity of the simulated CD spectrum at 70°C is matched to that of the denatured AP (see Method for the amino-acid sequence of AP). The NMR chemical shifts δ_{NH} and $\delta_{HC\alpha}$ of native[55] (yellow circles) and 8 M-urea-denatured[27] (blue circles) lysozymes are compared with those of the present dipeptides (red circles) in (**B**).

$n \rightarrow \pi^*$ interaction preferentially stabilizing the P_{II} over β. Using the MD trajectories, we determined 800 ϕ and ψ angles of the dipeptide P_{II} conformers [Fig. 5(C)]. Then, the ensemble average P_{II} structure ($\langle\phi\rangle = -71.5^\circ$ and $\langle\psi\rangle = 146.8^\circ$) and the $n \rightarrow \pi^*$ interaction energy map reported by Bartlett *et al.*[20] were used to estimate the $n \rightarrow \pi^*$ interaction energy of each dipeptide P_{II} conformer to be about 0.6 kcal/mol. This shows that the $n \rightarrow \pi^*$ interaction may not be the dominant stabilization factor for P_{II} compared with β in the present cases and that the residual 3.3 kcal/mol should be attributed to the backbone peptide solvation energy difference, which is in good agreement with the average electrostatic screening model calculation result from 18 blocked oligopeptides.[39]

Structural database for unfolded proteins

Here, we propose that the CD and NMR data of the present comprehensive set of 400 dipeptides can serve as structural database for unfolded proteins. Despite a number of papers suggesting that the short peptides could be excellent models representing underlying structural motifs of unfolded proteins, there was no direct evidence supporting this assumption before. Let us assume that a given unfolded protein can be viewed as a collection of dipeptide units. Since we have shown that the neighbor interaction effects on the dipeptide conformation may not be significant, the per-residue CD spectra of heat-denatured protein with N residues can be approximately given as a sum of the CD spectra of the constituent dipeptides, that is,

$$\Delta A_{\text{protein}}(X_1, \cdots, X_N) \approx \frac{N}{2(N-1)} \sum_{j=1}^{N-1} \Delta A_{\text{dipeptide}}(X_j, X_{j+1}) \tag{7}$$

where the factor of 2 is used to avoid the double-counting error. Here, $\Delta A_{\text{dipeptide}}(X_j, X_{j+1})$ is our CD spectrum of the X_jX_{j+1} dipeptide.

To test this assumption in Eq. (7), we compared the predicted CD spectrum [red solid line in Fig. 6(A)] obtained using our dipeptide CD (70°C) library with that (red dashed line in this figure) of the heat-denatured (70°C) acylphosphatase(AP) containing 98 residues. Here, for the sake of comparison, the calculated CD spectra [solid lines in Fig. 6(A)] were red-shifted by 6.9 nm and the negative peak intensity (red-solid line) was matched to that of experimental spectrum (red-dashed line). The (+, −, −) peak pattern from short to long wavelength and the overall CD line shape are in excellent agreement with those of heat-denatured AP at 70°C. This is one of the strongest evidences supporting the notion that the dipeptides can be elementary building blocks for structural analyses of unfolded proteins. Second, in Figure 6(B) we compare the distribution of δ_{NH} and $\delta_{HC\alpha}$ values of the dipeptides with those of the constituent residues in native and urea-denatured lysozymes. Because of the backbone conformational heterogeneity in native lysozyme, the δ_{NH} and $\delta_{HC\alpha}$ values are broadly distributed, whereas their distributions of the denatured one are significantly narrow and particularly overlap with those of the present dipeptides. This is independent evidence supporting the assumption that the spectroscopic library of the

complete set of dipeptides serves as a database for analyzing unfolded protein structural distributions.

SUMMARY

In this article, we present a comprehensive spectroscopic library of 400 dipeptides, which will be of critical use in future studies of unfolded proteins. As a matter of fact, the 400 dipeptides constitute the entire library (combinations) of all the possible pairs of naturally abundant amino-acids found in unfolded proteins. Spectroscopic investigations of this library revealed that the dipeptides preferentially adopt P_{II} and β conformations and provided statistics on their ΔG, ΔH, and ΔS. In contrast to native protein structures governed by Anfinsen hypothesis,[21] unfolded protein structures are probably determined by conformational propensities of individual amino-acids with no strong dependences on side-chains and peptide sequence. These experimental findings point to the important restriction of conformational space of denatured or unfolded proteins in aqueous solutions and led us to propose the hypothesis that the chemical diversity of amino-acid side-chains, which is undoubtedly a prerequisite for native protein's structure and function, is an undesired complexity for unfolded proteins that must fold on a biological timescale. In water, the various amino-acid side-chains have little effect on the backbone conformation, leading to unfolded peptide chains with a narrower range of conformations than assumed by Levinthal.[54] We further anticipate that the NMR, CD, and MD results can also serve as a reference database useful to improve the molecular mechanical parameters for future investigations of protein dynamics and structures.

REFERENCES

1. Tiffany ML, Krimm S. Circular dichroism of poly-L-proline in an unordered conformation. Biopolymers 1968;6:1767–1770.
2. Shi ZS, Woody RW, Kallenbach NR. Is polyproline II a major backbone conformation in unfolded proteins? Adv Protein Chem 2002;62:163–240.
3. Shi ZS, Olson CA, Rose GD, Baldwin RL, Kallenbach NR. Polyproline II structure in a sequence of seven alanine residues. Proc Natl Acad Sci USA 2002;99:9190–9195.
4. Eker F, Cao XL, Nafie L, Schweitzer-Stenner R. Tripeptides adopt stable structures in water. A combined polarized visible Raman, FTIR, and VCD spectroscopy study. J Am Chem Soc 2002;124: 14330–14341.
5. Avbelj F, Grdadolnik SG, Grdadolnik J, Baldwin RL. Intrinsic backbone preferences are fully present in blocked amino acids. Proc Natl Acad Sci USA 2006;103:1272–1277.
6. Zanni MT, Stenger J, Asplund MC, Hochstrasser RM. Solvent dependent conformational dynamics of dipeptides studied with two-dimensional infrared spectroscopy. Biophys J 2001;80:8A–9A.
7. Hagarman A, Measey TJ, Mathieu D, Schwalbe H, Schweitzer-Stenner R. Intrinsic propensities of amino acid residues in GxG peptides inferred from amide I′ band profiles and NMR scalar coupling constants. J Am Chem Soc 2010;132:540–551.
8. Lee KK, Hahn S, Oh KI, Choi JS, Joo C, Lee H, Han HY, Cho M. Structure of *N*-acetylproline amide in liquid water: Experimentally measured and numerically simulated infrared and vibrational circular dichroism spectra. J Phys Chem B 2006;110:18834–18843.
9. Lee KK, Joo C, Yang S, Han H, Cho M. Phosphorylation effect on the GSSS peptide conformation in water: Infrared, vibrational circular dichroism, and circular dichroism experiments and comparisons with molecular dynamics simulations. J Chem Phys 2007;126: 235102.
10. Hahn S, Lee H, Cho M. Theoretical calculations of infrared absorption, vibrational circular dichroism, and two-dimensional vibrational spectra of acetylproline in liquids water and chloroform. J Chem Phys 2004;121:1849–1865.
11. Dukor RK, Keiderling TA. Reassessment of the random coil conformation—Vibrational Cd study of proline oligopeptides and related polypeptides. Biopolymers 1991;31:1747–1761.
12. Oh KI, Lee KK, Park EK, Yoo DG, Hwang GS, Cho M. Circular dichroism eigenspectra of polyproline II and beta-strand conformers of trialanine in water: singular value decomposition analysis. Chirality 2010;22:E186–E201.
13. Wright PE, Dyson HJ, Lerner RA. Conformation of peptide-fragments of proteins in aqueous-solution—implications for initiation of protein folding. Biochemistry 1988;27:7167–7175.
14. Shortle D, Ackerman MS. Persistence of native-like topology in a denatured protein in 8 M urea. Science 2001;293:487–489.
15. Schweitzer-Stenner R, Measey TJ. The alanine-rich XAO peptide adopts a heterogeneous population, including turn-like and polyproline II conformations. Proc Natl Acad Sci USA 2007;104:6649–6654.
16. Schweitzer-Stenner R, Hagarman A, Mathieu D, Toal S, Measey TJ, Schwalbe H. Amino acids with hydrogen-bonding side chains have an intrinsic tendency to sample various turn conformations in aqueous solution. Chem Eur J 2011;17:6789–6797.
17. Makowska J, Rodziewicz-Motowidlo S, Baginska K, Makowski M, Vila JA, Liwo A, Chmurzynski L, Scheraga HA. Further evidence for the absence of polyproline II stretch in the XAO peptide. Biophys J 2007;92:2904–2917.
18. Zagrovic B, Lipfert J, Sorin EJ, Millettt IS, van Gunsteren WF, Doniach S, Pande VS. Unusual compactness of a polyproline type II structure. Proc Natl Acad Sci USA 2005;102:11698–11703.
19. Jha AK, Colubri A, Freed KF, Sosnick TR. Statistical coil model of the unfolded state: resolving the reconciliation problem. Proc Natl Acad Sci USA 2005;102:13099–13104.
20. Bartlett GJ, Choudhary A, Raines RT, Woolfson DN. n ->pi* interactions in proteins. Nat Chem Biol 2010;6:615–620.
21. Anfinsen CB. Principles that govern folding of protein chains. Science 1973;181:223–230.
22. Swindells MB, Macarthur MW, Thornton JM. Intrinsic Phi,Psi propensities of amino-acids, derived from the coil regions of known structures. Nat Struct Biol 1995;2:596–603.
23. Serrano L. Comparison between the Phi distribution of the amino-acids in the protein database and Nmr data indicates that amino-acids have various phi propensities in the random coil conformation. J Mol Biol 1995;254:322–333.
24. Oneil KT, Degrado WF. A Thermodynamic scale for the helix-forming tendencies of the commonly occurring amino-acids. Science 1990;250:646–651.
25. Kim CWA, Berg JM. Thermodynamic beta-sheet propensities measured using a zinc-finger host peptide. Nature 1993;362:267–270.
26. Saudek V, Wormald MR, Williams RJP, Boyd J, Stefani M, Ramponi G. Identification and description of beta-structure in horse muscle acylphosphatase by nuclear magnetic-resonance spectroscopy. J Mol Biol 1989;207:405–415.
27. Schlorb C, Ackermann K, Richter C, Wirmer J, Schwalbe H. Heterologous expression of hen egg white lysozyme and resonance assignment of tryptophan side chains in its non-native states. J Biomol Nmr 2005;33:95–104.
28. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker RC,

Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Mathews DH, Schafmeister C, Ross WS, Kollman PA. AMBER 9. San Francisco: University of California; 2006.
29. Darden T, York D, Pedersen L. Particle Mesh Ewald—an N.Log(N) method for Ewald sums in large systems. J Chem Phys 1993;98: 10089–10092.
30. Karplus M. Contact electron-spin coupling of nuclear magnetic moments. J Chem Phys 1959;30:11–15.
31. Avbelj F, Kocjan D, Baldwin RL. Protein chemical shifts arising from alpha-helices and beta-sheets depend on solvent exposure. Proc Natl Acad Sci USA 2004;101:17394–17397.
32. Zhang HY, Neal S, Wishart DS. RefDB: a database of uniformly referenced protein chemical shifts. J Biomol Nmr 2003;25:173–195.
33. Shi ZS, Chen K, Liu ZG, Ng A, Bracken WC, Kallenbach NR. Polyproline II propensities from GGXGG peptides reveal an anticorrelation with beta-sheet scales. Proc Natl Acad Sci USA 2005;102: 17964–17968.
34. Patgiri A, Jochim AL, Arora PS. A hydrogen bond surrogate approach for stabilization of short peptide sequences in alpha-helical conformation. Accounts Chem Res 2008;41:1289–1300.
35. Graf J, Nguyen PH, Stock G, Schwalbe H. Structure and dynamics of the homologous series of alanine peptides: a joint molecular dynamics/NMR study. J Am Chem Soc 2007;129:1179–1189.
36. Grdadolnik J, Mohacek-Grosev V, Baldwin RL, Avbelj F. Populations of the three major backbone conformations in 19 amino acid dipeptides. Proc Natl Acad Sci USA 2011;108:1794–1798.
37. Avbelj F, Baldwin RL. Origin of the neighboring residue effect on peptide backbone conformation. Proc Natl Acad Sci USA 2004;101: 10967–10972.
38. Tanford C. Protein denaturation. Adv Protein Chem 1968;23:121–282.
39. Avbelj F, Baldwin RL. Role of backbone solvation and electrostatics in generating preferred peptide backbone conformations: distributions of phi. Proc Natl Acad Sci USA 2003;100:5742–5747.
40. Kimura T, Matubayasi N, Sato H, Hirata F, Nakahara M. Enthalpy and entropy decomposition of free-energy changes for side-chain conformations of aspartic acid and asparagine in acidic, neutral, and basic aqueous solutions. J Phys Chem B 2002;106: 12336–12343.
41. Liu L, Guo QX. Isokinetic relationship, isoequilibrium relationship, and enthalpy-entropy compensation. Chem Rev 2001;101:673–695.
42. Krug RR, Hunter WG, Grieger RA. Statistical interpretation of enthalpy-entropy compensation. Nature 1976;261:566–567.
43. Kauzmann W, Bodanszky A, Rasper J. Volume changes in protein reactions. II. Comparison of ionization reactions in proteins and small molecules. J Am Chem Soc 1962;84:1777–1788.
44. Douglas JF, Dudowicz J, Freed KF. Crowding induced self-assembly and enthalpy-entropy compensation. Phys Rev Lett 2009;103:13.
45. Mezei M, Fleming PJ, Srinivasan R, Rose GD. Polyproline II helix is the preferred conformation for unfolded polyalanine in water. Proteins-Struct Funct Bioinformatics 2004;55:502–507.
46. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. J Mol Biol 1963;7:95–99.
47. Richards FM. Areas, volumes, packing, and protein-structure. Ann Rev Biophys Bioeng 1977;6:151–176.
48. Moore PB, Voss NR, Gerstein M, Steitz TA. The geometry of the ribosomal polypeptide exit tunnel. J Mol Biol 2006;360:893–906.
49. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP. The development and use of quantum-mechanical molecular-models .76. Am1—a new general-purpose quantum-mechanical molecular-model. J Am Chem Soc 1985;107:3902–3909.
50. Stewart JJP. Optimization of parameters for semiempirical methods. II. Applications. J Comp Chem 1989;10:221–264.
51. Stewart JJP. Optimization of parameters for semiempirical methods. I. Method. J Comp Chemistry 1989;10(2):209–220.
52. Kwac K, Lee KK, Han JB, Oh KI, Cho M. Classical and quantum mechanical/molecular mechanical molecular dynamics simulations of alanine dipeptide in water: comparisons with IR and vibrational circular dichroism spectra. J Chem Phys 2008;128:10.
53. Raines RT, Hinderaker MP. An electronic effect on protein structure. Protein Sci 2003;12:1188–1194.
54. Levintha C. Are there pathways for protein folding. J De Chimie Physique Et De Physico-Chimie Biologique 1968;65:44–45.
55. Redfield C, Dobson CM. Sequential H-1-Nmr assignments and secondary structure of hen egg-white lysozyme in solution. Biochemistry 1988;27:122–136.