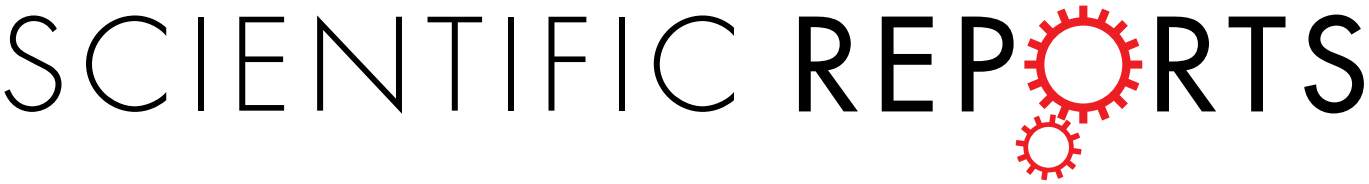

OPEN

Populations of the Minor α-Conformation in AcGXGNH$_2$ and the α-Helical Nucleation Propensities

Received: 26 January 2016
Accepted: 16 May 2016
Published: 03 June 2016

Yanjun Zhou, Liu He, Wenwen Zhang, Jingjing Hu & Zhengshuang Shi

Intrinsic backbone conformational preferences of different amino acids are important for understanding the local structure of unfolded protein chains. Recent evidence suggests α-structure is relatively minor among three major backbone conformations for unfolded proteins. The α-helices are the dominant structures in many proteins. For these proteins, how could the α-structures occur from the least in unfolded to the most in folded states? Populations of the minor α-conformation in model peptides provide vital information. Reliable determination of populations of the α-conformers in these peptides that exist in multiple equilibriums of different conformations remains a challenge. Combined analyses on data from AcGXPNH$_2$ and AcGXGNH$_2$ peptides allow us to derive the populations of PII, β and α in AcGXGNH$_2$. Our results show that on average residue X in AcGXGNH$_2$ adopt PII, β, and α 44.7%, 44.5% and 10.8% of time, respectively. The contents of α-conformations for different amino acids define an α-helix nucleation propensity scale. With derived PII, β and α-contents, we can construct a free energy-conformation diagram on each AcGXGNH$_2$ in aqueous solution for the three major backbone conformations. Our results would have broad implications on early-stage events of protein folding.

Protein sequence-structure relationships are of fundamental importance to the field of protein physical chemistry[1–4]. Intrinsic backbone conformational preferences of 20 amino acids determine the local structure of unfolded protein chains; these intrinsic preferences might guide the folding processes at early stages of protein folding. From this respect, the intrinsic backbone conformational preferences of different amino acids are part of the "folding mechanism" that remains poorly understood after more than 50 years since the protein folding question was raised[1–4]. Currently, to predict protein structure from amino acid sequences, database-based strategies are more successful than the physics-based algorithms. Advances in the physics-based algorithms demand continuous improvements in force field accuracy. The intrinsic backbone conformational preference data are crucial for this purpose.

Among three major backbone conformations, α-structure is relatively minor compared to polyproline II (PII) and β-conformations in model unfolded peptides as demonstrated by recent lines of independent evidence[5–29]. Reliable derivation of populations of the minor α-conformers in model peptides that exist in multiple equilibriums of different backbone conformations remains a challenge[11,18,19,23–25,27]. NMR measurements can only be carried out on a slow time scale as compared to backbone conformers' lifetimes which lie in the range of 10–200 ps, conformational averaging over different conformers occurs during NMR measurements. The optical spectroscopy results are measured on a fast time scale and various optical spectra can be used to detect different backbone conformations[6,8,9,16,17,23]. However, most optical techniques suffer from their resolutions: band overlapping in the measured spectra generally cause significant uncertainties during quantitative analysis, particularly for accurate derivations of minor conformers. In our previous work, quantitative account of the sampled conformation in AcGGXGGNH$_2$ and XAO by NMR $^3J(H_\alpha\text{-}H_N)$ ($^3J_{\alpha N}$) coupling constants was carried out through a two-state analysis for an equilibrium mainly between PII and β conformations; α-population was ignored completely as an approximation[11,18]. Here we have designed two series of peptides: AcGXPNH$_2$ and AcGXGNH$_2$ (X ≠ Gly, Pro).

School of Chemistry and Chemical Engineering, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan 430074, P.R. China. Correspondence and requests for materials should be addressed to Z.S. (email: zs_shi@hust.edu.cn)

Combined analyses on data from both series allow us to derive the populations of three major conformers including PII, β and α in $AcGXGNH_2$.

Proline is unique among the amino acids in that it has a five-membered ring which has a dramatic effect on the conformational preferences of the preceding residue. In $AcGXPNH_2$ peptides, X can only adopt PII or β conformations as steric clashes between the C^δ of proline and both the C^β and amide nitrogen of residue X make α-conformation inaccessible to residue X[30–32]. With the measured $^3J_{\alpha N}$ coupling constants of X, previous procedure through a two-state analysis for the equilibrium between PII and β is justified for $AcGXPNH_2$ peptides[11,18]. PII to β population ratio for each of $AcGXPNH_2$ can be determined; assuming the ratio for X in $AcGXPNH_2$ and $AcGXGNH_2$ is approximately the same, we can derive the population of α-conformer in $AcGXGNH_2$ peptides through equation (1), see Supplementary Information for derivation of the equation in which x_α(GXG) denotes the percentage of α-conformer in $AcGXGNH_2$; $^3J_{\alpha N}$(GXP) and $^3J_{\alpha N}$(GXG), measured $^3J_{\alpha N}$ coupling constants of X in $AcGXPNH_2$ and $AcGXGNH_2$; $^3J_{\alpha N}(\alpha)$, standard $^3J_{\alpha N}$ coupling constant of a residue in α-helices.

$$x_\alpha(GXG) = [^3J_{\alpha N}(GXP) - {}^3J_{\alpha N}(GXG)]/[^3J_{\alpha N}(GXP) - {}^3J_{\alpha N}(\alpha)] \quad (1)$$

Further, we can derive the populations of PII and β in $AcGXGNH_2$ through equations (2) and (3) in which $^3J_{\alpha N}$(PII),$^3J_{\alpha N}(\beta)$ and $^3J_{\alpha N}(\alpha)$ denote standard $^3J_{\alpha N}$ coupling constant of a residue in PII, β-, and α-conformations, respectively; x_{PII}(GXG), x_β(GXG) and x_α(GXG) denote the percentage of PII, β- and α-conformations in $AcGXGNH_2$, respectively. With the derived percentage values, the free energy-conformation diagrams of $AcGXGNH_2$ in aqueous solution can be constructed for the three major backbone conformations.

$$x_{PII}(GXG) + x_\beta(GXG) + x_\alpha(GXG) = 1 \quad (2)$$

$$^3J_{\alpha N}(PII) \cdot x_{PII}(GXG) + {}^3J_{\alpha N}(\beta) \cdot x_\beta(GXG) + {}^3J_{\alpha N}(\alpha) \cdot x_\alpha(GXG) = {}^3J_{\alpha N}(GXG) \quad (3)$$

Derived results show that on average residue X in $AcGXGNH_2$ adopt PII, β, and α 44.7%, 44.5% and 10.8% of time, respectively. Importantly, minor populated α-conformations of different amino acids in $AcGXGNH_2$ determine their varying α-helix nucleation capabilities[33]. According to Zimm-Bragg theory[34], helix-to-coil transition can be described by a nucleation constant σ and helix propagation constants s, the product σ·s represents the probability of formation of an α-helical segment comprising three residues[34–36]. From our derived vales of x_α, we can estimate the probability for Ala peptides, $\sigma \cdot s = (x_\alpha)^3 = 4.29 \times 10^{-3}$ ($x_\alpha = 0.1625$ for Ala), this value is very close to those reported[37]. Our free energy-conformation diagrams would set a foundation for physics-based algorithmic developments for protein structure predictions[38,39].

Results and Discussions

Model peptides $AcGXPNH_2$ and $AcGXGNH_2$ and their CD spectra. Our previous study on $AcGGXGGNH_2$ peptides showed that these peptides are present predominantly in the extended PII or β structure, around 10% α or turn structures could be present, but the exact percentage of α or turn conformation could not be determined. In $AcGXGNH_2$, X is expected to sample all three major backbone conformations, with PII or β structure being dominant and α basin being minor; in $AcGXPNH_2$, however, X can sample only PII or β conformations. To avoid end and charge effects, two peptide series of this study have both ends blocked[27]. CD spectra for most $AcGXGNH_2$ peptides except those with ring side chains (His, Trp, Tyr, Phe) show the characteristic far-UV CD signature of a mixture of PII and β conformations, with a strong negative band at ≈198 nm and a weak positive band or shoulder at ≈215 nm[18,28,40](Fig. S1). CD spectra of $AcGXGNH_2$ are very similar to those of $AcGGXGGNH_2$[18]. CD spectra of $AcGXPNH_2$ are obscured by the contributions from Pro (Fig. S1). Small populations of Pro could exist in *cis* configurations; typical CD spectra of Pro peptides in PII helix usually shift to a longer wavelength as compared to those of non-Pro peptides. As a result, interpretation of CD spectra for $AcGXPNH_2$ is not very obvious. Differential spectra between $AcGXPNH_2$ and $AcGXGNH_2$ reveal that Pro exists as a mixture of PII and PI (polyproline I) helices in $AcGXPNH_2$[41]; thus CD spectra of $AcGXPNH_2$ reflect contributions from both X and Pro, contributions from X are expected to show the characteristic far-UV CD signature of a mixture of PII and β conformations, similar to those observed for $AcGXGNH_2$.

Contents of α-conformers in $AcGXGNH_2$ correlate with α-helix nucleation capabilities of X. $^3J_{\alpha N}$ coupling constant is directly related to the backbone φ angle by Karplus equations[42,43]. Measured $^3J_{\alpha N}$ values at 25 °C (pH = 4.0) for $AcGXGNH_2$ and $AcGXPNH_2$ peptides are shown in Table 1 (see Fig. S2 for the NMR spectra and results of fitting). In $AcGXPNH_2$, there is a slow *trans*-to-*cis* equilibrium for Pro, $^3J_{\alpha N}$ for both *cis*- and *trans*- species are well resolved in 1D ^{1}H NMR spectra, here only $^3J_{\alpha N}$ values of X corresponding to *trans*-Pro are reported. Measured $^3J_{\alpha N}$ coupling constants for $AcGXGNH_2$ are compared to those for dipeptides (blocked amino acids)[19] at 30 °C (pH = 4.9) in Fig. 1. The plot reveals a good agreement between two sets of coupling constants (R = 0.86).

$^3J_{\alpha N}$ values for $AcGXGNH_2$ are smaller than those for $AcGXPNH_2$ for most amino acids except for residues Asp (pH = 2.0 and 6.0), Asn and Thr. Excluding Thr, Asn and Asp's, $^3J_{\alpha N}$ values for $AcGXGNH_2$ are on average 0.41 Hz smaller than those for $AcGXPNH_2$. The smaller $^3J_{\alpha N}$ values for $AcGXGNH_2$ are consistent to X samples all three major backbone conformations in $AcGXGNH_2$, while X samples only PII and β conformations in $AcGXPNH_2$ (Thr, Asn and Asp are excluded). For $AcGXPNH_2$ (X = Thr, Asn and Asp), X is expected to form turn structures[44]; it explains smaller observed $^3J_{\alpha N}$ values for these residues in $AcGXPNH_2$ compared to those in $AcGXGNH_2$. For all other amino acids, contents of α conformations in $AcGXGNH_2$ can be calculated from equation (1), in which $^3J_{\alpha N}(\alpha)$ is assigned to be 4.11 Hz, corresponding to a φ value of −60° (Table 1). For Thr, Asn and

Amino acids	$^3J_{\alpha N}$ (Hz) $AcGXGNH_2$	$^3J_{\alpha N}$ (Hz) $AcGXPNH_2$	$^3J_{\alpha N}$ (Hz) dipeptide**	x_α (%) in $AcGXGNH_2$	x_{PII} (%) in $AcGXGNH_2$	x_β (%) in $AcGXGNH_2$
Tyr	6.83	7.30	7.13	14.7%	44.0%	41.3%
Trp	6.48	6.99	6.91	17.6%	49.0%	33.4%
Thr	7.68	7.60	7.35	4.0%*	36.4%	59.6%
Arg	6.96	7.20	6.85	8.0%	49.7%	42.3%
Gln	6.95	7.24	7.14	9.3%	48.2%	42.5%
Asn	7.63	7.34	7.45	2.5%*	39.8%	57.7%
Met	7.03	7.26	7.02	7.2%	48.8%	44.0%
Leu	6.81	7.27	6.88	14.5%	44.7%	40.8%
Lys	6.92	7.19	6.83	8.6%	49.8%	41.6%
Ile	7.24	7.91	7.33	17.6%	29.6%	52.8%
Cys	7.25	7.32	7.31	2.2%	49.8%	48.0%
His	7.29	7.95	7.83	17.0%	29.0%	54.0%
Ala	5.86	6.20	6.06	16.3%	66.9%	16.8%
Phe	7.02	7.36	7.18	10.4%	44.8%	44.8%
Ser	6.86	7.19	7.02	10.7%	48.6%	40.7%
Val	7.20	7.91	7.30	18.7%	29.1%	52.2%
Glu (pH = 2)	6.96	7.54	7.02 (pH = 2.9)	16.9%	37.8%	45.3%
Glu (pH = 6)	6.38	6.71	6.63 (pH = 4.9)	12.7%	58.3%	29.0%
Asp (pH = 2)	7.63	7.54	7.51 (pH = 2.9)	2.0%*	40.2%	57.8%
Asp (pH = 6)	7.11	6.51	6.93 (pH = 4.9)	5.0%*	49.8%	45.2%

Table 1. Experimentally determined $^3J_{\alpha N}$ (298K) of $AcGXGNH_2$, $AcGXPNH_2$ and amino acid dipeptides and derived α, PII and β-contents for X in $AcGXGNH_2$. **$^3J_{\alpha N}$ values of amino acid dipeptides are taken from ref. 19. *Corresponding α content values are taken from ref. 23.

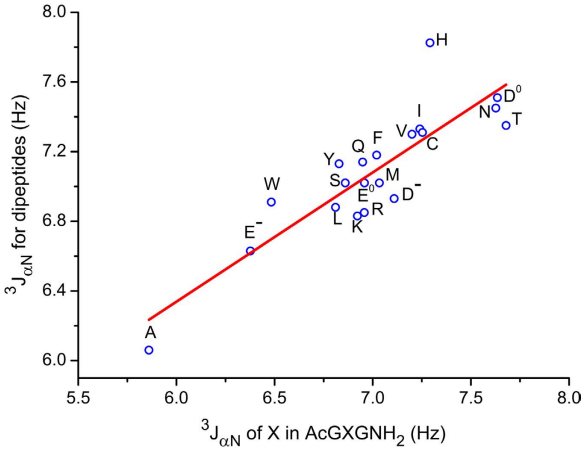

Figure 1. The $^3J_{\alpha N}$ coupling constants measured for $AcGXGNH_2$ peptides are plotted against those for amino acid dipeptides.

Asp (pH = 2.0 and 6.0) in $AcGXGNH_2$, their contents of α conformations cannot be determined. It is a conservative and proximate practice to assign the values to be 0.04, 0.025, 0.02 and 0.05 for Thr, Asn and Asp (pH = 2.0 and 6.0), respectively, corresponding to the values from dipeptides by Grdadolnik *et al.*[23] (Table 1). Contents of α conformations derived from blocked amino acids are significantly smaller than our values, 5.2 % *vs.* 12.6 % on average with Thr, Asn and Asp being excluded.

Our results indicate that x_α values for hydrophobic or aromatic amino acids are significantly larger than those for polar amino acids, 14.9% vs. 7.4% on average. The differences among different non-polar residues are marginal (Table 1). Contents of minor populated α-conformations of different amino acids in $AcGXGNH_2$ determine their varying α-helix nucleation propensities. Our results suggest that: for non-polar amino acids, the nature or the size of side chains, being aromatic ring or β-branching, do not have strong steric impact on helix nucleation, in contrast to their strong effects on helix propagation due to different steric constraints. The x_α values observed show no correlation to any α-helix propensity scales[45–47] that report mainly the propensity of amino acid residues to propagate on a preformed helix; the observation corroborates the conclusion by Miller *et al.*[33] Effects of individual side chains on helix nucleation are difficult to deconvolute from those of helix propagation. Recently, Miller *et al.* have successfully separated the effects through studying a synthetic model and found that amino acid side chains

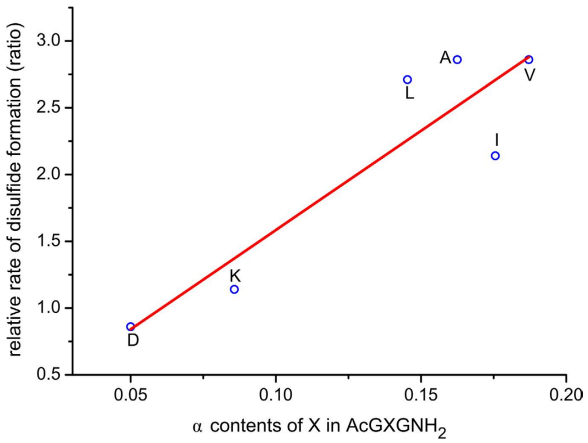

Figure 2. The correlation of determined α-contents for $AcGXGNH_2$ and the relative rates of disulfide formation in a synthetic model.

contribute in a completely different manner to nucleation than to propagation[33]. In this study, the relative rates of disulfide formation serve as indirect indicators for different residues' α-helix nucleation capabilities. Our derived populations of α conformations in $AcGXGNH_2$ are compared to the relative rates of disulfide formation for limited amino acids by Miller *et al.*[33] (Table S1); a good correlation is revealed (Fig. 2, R = 0.88).

From derived vales of x_α, we can calculate the probability of forming an α-helical segment comprising three residues, $\sigma \cdot s = (x_\alpha)^3 = 1.26 \times 10^{-3}$ if we use the average value of x_α for all amino acids. For Ala peptides, we can determine the probability, $\sigma \cdot s = (x_\alpha)^3 = 4.29 \times 10^{-3}$ ($x_\alpha = 0.1625$ for Ala). The value is very close to those reported for Ala-rich peptides (the measured $\sigma = 0.004 \pm 0.002$ with $s_{Ala} = 1.4–1.6$)[37]. As parameters, products of ($x_{\alpha1} \bullet x_{\alpha2} \bullet x_{\alpha3}$) for a combination of three different amino acids would be sensitive indicators to uncover the potential helix nucleation sites within sequences that form α-helices. From the derived x_α values (Table 1), we predict sequences comprised of Val, Trp, Ile, His, Glu (pH = 2.0) and Ala are most likely the nucleation sites at early stages of α-helix formation; whereas sequences comprised of Asp, Cys, Asn and Thr (Pro and Gly are not considered here) are least likely the nucleation sites. Fast folding kinetic studies on model protein/peptides are expected to validate or invalidate our predictions.

Contents of PII and β conformations in $AcGXGH_2$ and construction of free energy-conformation diagrams for three major backbone conformations. Contents of PII and β conformations in $AcGXGNH_2$ can be calculated using equations (2) and (3) (Table 1). We assign standard $^3J_{\alpha N}$ values for PII and β conformations to be 5.42 and 9.30 Hz, respectively. The value of 5.42 Hz for $^3J_{\alpha N}$(PII) corresponds to a ϕ value of −70°; the value of 9.30 Hz for $^3J_{\alpha N}$(β) is the result from fitting measured $^3J_{\alpha N}$ values on blocked dipeptides to their β-populations derived from optical spectroscopic bands[23]. X in $AcGXGNH_2$ adopts predominantly the extended PII or β conformations; on average, X samples about the same amount of time in PII or β basin, 44.7% *vs.* 44.5%. Our analysis indicates that β-contents or ΔG values for corresponding PII to β equilibriums show weak or reasonable correlations with β propensity scales (weak with β-contents and reasonable with ΔG), consistent to the observation in $AcGGXGGNH_2$ peptides[18]. Correlations between ΔG and the β-sheet scale by Kim and Berg[48] are shown in Fig. S3.

A more relevant comparison is between our data to those from blocked amino acids (dipeptides). Grdadolnik *et al.* have determined populations of the three major backbone conformations in 19 amino acid dipeptides (N-acetyl-X-N′-methylamide) by using the amide III region of the peptide infrared and Raman spectra[23]. The work by Grdadolnik *et al.* represents a major advance in band assignments of the peptide infrared and Raman spectra to different backbone conformations[23]. This advance made determination of backbone conformational distribution possible. If we compare our derived ΔG values for PII to β transitions to those derived for dipeptides, we find a reasonably strong correlation (Fig. 3, R = 0.84). Comparison of this correlation to the one in Fig. 1 (R = 0.86) indicates that the correlation between ΔG values is limited to that between $^3J_{\alpha N}$ values. Given totally independent strategies on different systems were used, the correlation provides validations for both methods.

The average length of β-strands in β-sheets is about 6 residues, the probabilities of forming a β-strand of 6 residues is $(x_\beta)^6 = 7.77 \times 10^{-3}$ if we use the average value of x_β for all amino acids. Considering strands of 3–6 amino acids long might all play important roles in the early stages of β-hairpin folding, the population of a preformed β-strand of 3 residues long would reach as high as 20% (corresponding to $x_\beta = 0.585$). Following the procedure for α-helices, products of ($x_{\beta1} \bullet x_{\beta2} \bullet x_{\beta3}$) for a combination of three different amino acids might be used to locate the potential sites that form β-strands at early stages of protein folding. Similarly, from the derived x_β values (Table 1), we predict sequences comprised of Thr, Asp (pH = 2), Asn, His, Ile and Val are most likely the sites that tend to form nascent β-strands; whereas sequences comprised of Ala, Glu (pH = 6) and Trp (Pro and Gly are not considered here) are least likely the sites to form nascent β-strands. Nascent β-strands then initiate a productive or non-productive collision.

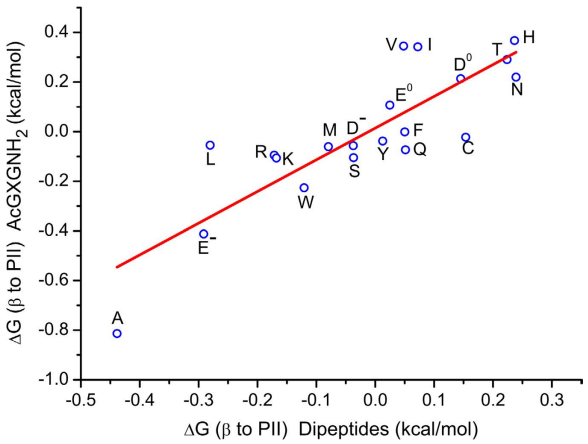

Figure 3. The correlation of ΔG(β to PII) derived for $AcGXGNH_2$ and that for dipeptides.

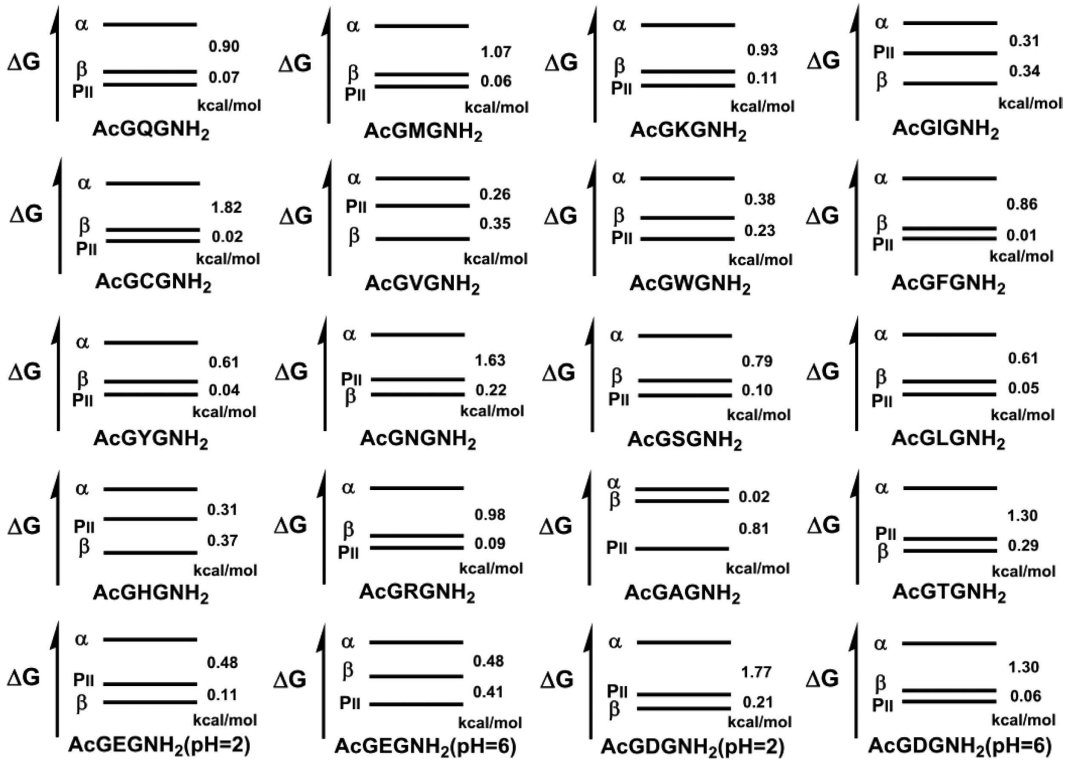

Figure 4. Derived free energy-conformation diagrams for $AcGXGNH_2$.

With the derived PII, β and α-contents, we can construct a free energy-conformation diagram on each $AcGXGNH_2$ in aqueous solution for the three major backbone conformations (Fig. 4). The diagrams clearly show that the free energy level for α-basin is the highest among three for all amino acids; the free energy level for PII basin is the lowest for most amino acids except for Ile, Val, Asn, His, Thr, Glu (pH = 2.0) and Asp (pH = 2.0). Together with the results on 19 amino acid dipeptides from the optical spectroscopic data[23], it is our believe that the derived free energy-conformation diagrams would provide a bench mark for testing predicting calculations of conformational energy maps of flexible model peptides[38,39].

Turn conformations in $AcGNPNH_2$, $AcGTPNH_2$ and $AcGDPNH_2$ (pH = 2 and 6) and effects of different $^3J_{\alpha N}$(PII) and $^3J_{\alpha N}$(β) values on data analysis. We have detected significant turn structures in $AcGNPNH_2$, $AcGTPNH_2$ and $AcGDPNH_2$ (pH = 2 and 6) as shown in Table 2. This observation is consistent with the findings by Hagarman *et al.*[44] In this study, we assign standard $^3J_{\alpha N}$ values for PII and β conformations to be 5.42 and 9.30 Hz, respectively. In our previous study on $AcGGXGGNH_2$ peptides, a set of residue-specific $^3J_{\alpha N}$ reference values for PII and β conformations were used[18,49] (See Table 1 of reference 18). If we use the previous set of reference values to analyze the data in this study, slightly different PII, β and α-contents are obtained.

$AcGXPNH_2$	x_{turn} (%)	x_{PII} (%)	x_{β} (%)
Thr	6.1%	35.6%	58.3%
Asn	10.4%	36.6%	53.0%
Asp (pH = 2)	4.7%	39.1%	56.2%
Asp (pH = 6)	24.0%	39.8%	36.2%

Table 2. Derived turn, PII and β-contents of X in $AcGXPNH_2$ for Thr, Asn and Asp (pH = 2 and 6).

Comparison of two sets of results indicates they are matched to each other overall with derived conclusions being the same. (See Supplementary Information for details). Regardless, the choice of different $^3J_{\alpha N}$(PII) and $^3J_{\alpha N}$(β) values has no effects on our derived x_α values for X in $AcGXGNH_2$ as implied by equation (1) (see Supplementary Information for derivation of the equation).

NOE data and error analysis. NOEs can be used to analyze the conformations. Amide region of NOESY spectra for $AcGXGNH_2$ peptides are shown in Fig. S10. Strong $d_{\alpha N}(i, i+1)$ NOE cross peaks are observed for X residues in $AcGXGNH_2$ peptides, while the intensities of $d_{\alpha N}(i, i)$ NOEs are weakened by about two- to fourfold relative to those of $d_{\alpha N}(i, i+1)$ NOEs; the $d_{NN}(i, i+1)$ NOEs are not measurable due to their weak intensities and being very close to the diagonal peaks. These results indicate that $AcGXGNH_2$ peptides are present predominantly in the extended PII or β-conformations that are consistent with our conclusion through analyzing coupling constant data. Figure S2 shows the amide region of 1D NMR spectra for all $AcGXPNH_2$ and $AcGXGNH_2$ peptides. The coupling constants were measured by a peak-fitting procedure to Lorentzian line shape, the fitting results are also shown in the figure. The derived coupling constants can be reproduced within 0.02 Hz if we fit a certain spectrum multiple times independently. In this and our previous studies, we used the Karplus equation by Vuister and Bax[43] with coefficients: A = 6.51, B = −1.76 and C = 1.60; another parametrization for the Karplus equation with A = 6.98, B = −1.38 and C = 1.72 by Wang and Bax[50] is believed to be more accurate. Calculated $^3J_{\alpha N}(\alpha)$ values for $\phi = -60°$ are coefficient dependent: 4.11 *vs.* 4.16 Hz for two sets of parameters; as a result, the derived α-population differs by ~2%. Given the average difference between $^3J_{\alpha N}$ of $AcGXPNH_2$ and $AcGXGNH_2$ is about 0.41 Hz, plus a maximal uncertainty of 0.2 Hz on $^3J_{\alpha N}(\alpha)$ due to the uncertainties on the Karplus equation coefficients, we estimate the error of the derived α-population being around 10% for the majority of residues with non-overlapping amide signals, the estimated error could reach to 15–20% for those residues with overlapping peaks.

The relative population ratio between PII and β for $AcGXGNH_2$ and $AcGXPNH_2$. In this study, we assume that the population ratio between PII and β is approximately the same for $AcGXGNH_2$ and $AcGXPNH_2$. It is a known fact that there are secondary neighboring residue effects; we consider the effects from the side chain of residue X itself the primary effects. To our knowledge, Pro as a neighboring residue will make X favoring PII as compared to other neighboring residues. As a result, the population ratio between PII and β cannot be exactly the same for $AcGXGNH_2$ and $AcGXPNH_2$; it is most likely that the ratio for $AcGXPNH_2$ is relatively larger than that for $AcGXGNH_2$. Unfortunately, our current understanding on neighboring residue effects remains poor. To investigate the effects, first we define a parameter for the ratio of ratios, $RR = GXG_{PII/\beta}/GXP_{PII/\beta}GXG_{PII/\beta} = [x_{PII}(GXG)/x_\beta(GXG)]/[x_{PII}(GXP)/x_\beta(GXP)]$, then we analyze our data systematically with the parameter RR setting from 0.80–1.10 in a step function of 0.05. (Table S2). It is clear that the derived content values shift in the same direction for all residues upon changing the value of RR. Specifically, average contents of PII increase by 1.8%, while average contents of β and α decrease by 0.5% and 1.3%, respectively, upon increasing the parameter RR by 0.05. To our gratification, the correlations and the conclusions hold really well upon changing the value of the parameter RR from 0.80–1.10 (Figs S11–S13).

Conclusion

We have determined the populations of three major conformers in $AcGXGNH_2$ through analyzing $^3J_{\alpha N}$ coupling constants of $AcGXPNH_2$ and $AcGXGNH_2$; the free energy-conformation diagrams are constructed for $AcGXGNH_2$ peptides in aqueous solution. Our derived results show that on average residue X in $AcGXGNH_2$ adopt PII, β, and α 44.7%, 44.5% and 10.8% of time, respectively. Minor populated α-conformations of different amino acids in $AcGXGNH_2$ determine their varying α-helix nucleation capabilities. The contents of α-conformations for different amino acids define an α-helix nucleation propensity scale. There are no correlations observed between the x_α values and any α-helix propensity scales[45–47]. Based on our derived β-contents, ΔG values for the corresponding PII to β equilibriums show a reasonable correlation with the β-sheet scale by Kim and Berg[48], consistent to the observation in $AcGGXGGNH_2$ peptides[18]. Derived ΔG values for PII to β transitions show a good correlation to those derived for dipeptides[23]. We have detected significant turn structures in $AcGNPNH_2$, $AcGTPNH_2$ and $AcGDPNH_2$ (pH = 2 and 6)[44]. Results from this study have broad implications on the early-stage events of protein folding. Together with the results on 19 amino acid dipeptides[23], our results would provide a bench mark for force field developments and for testing predicting calculations of conformational energy maps of flexible model peptides[38,39].

Methods

Equation (1) was derived by assuming the PII to β population ratio of X in $AcGXPNH_2$ and $AcGXGNH_2$ being approximately the same. Peptides were synthesized and characterized as described[27], by using an automated peptide synthesizer with standard Fmoc chemistry. CD spectra were recorded on a J-810 spectrometer with about 100–500 μM peptides in 10 mM phosphate buffer at 25 °C. The concentrations of peptides were determined from a combination of UV absorbance and NMR peak integration[27]. 1D and 2D (TOCSY and NOESY) ^1H NMR spectra were collected on Bruker AVANCE 400/600 MHz spectrometers at 25 °C. $^3J_{\alpha N}$ coupling constants were determined from high resolution 1D spectra. Details are described in Materials and Methods of Supplementary Information.

References

1. Dill, K. A. & MacCallum, J. L. The protein-folding problem, 50 years on. *Science* **338,** 1042–1046 (2012).
2. Dobson, C. M. Protein folding and misfolding. *Nature* **426,** 884–890 (2003).
3. Fersht, A. R. From the first protein structures to our current knowledge of protein folding: delights and scepticisms. *Nat. Rev. Mol. Cell Biol.* **9,** 650–654 (2008).
4. Shea, J. & Brooks, C. L. III From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.* **52,** 499–535 (2001).
5. Poon, C., Samulski, E. T., Weise, C. F. & Weisshaar, J. C. Do bridging water molecules dictate the structure of a model dipeptide in aqueous solution? *J. Am. Chem. Soc.* **122,** 5642–5643 (2000).
6. Woutersen, S. & Hamm, P. Structure determination of trialanine in water using polarization sensitive two-dimensional vibrational spectroscopy. *J. Phys. Chem. B* **104,** 11316–11320 (2000).
7. Blanch, E. W. *et al.* Is polyproline II helix the killer conformation? a raman optical activity study of the amyloidogenic prefibrillar intermediate of human lysozyme. *J. Mol. Biol.* **301,** 553–563 (2000).
8. Schweitzer-Stenner, R., Eker, F., Huang, Q. & Griebenow, K. Dihedral angles of trialanine in D_2O determined by combining FTIR and polarized visible raman spectroscopy. *J. Am. Chem. Soc.* **123,** 9628–9633 (2001).
9. Woutersen, S. & Hamm, P. Isotope-edited two-dimensional vibrational spectroscopy of trialanine in aqueous solution. *J. Chem. Phys.* **114,** 2727–2737 (2001).
10. Kelly, M. A. *et al.* Host-guest study of left-handed polyproline II helix formation. *Biochemistry* **40,** 14376–14383 (2001).
11. Shi, Z., Olson, C. A., Rose, G. D., Baldwin, R. L. & Kallenbach, N. R. Polyproline II structure in a sequence of seven alanine residues. *Proc. Natl. Acad. Sci. USA* **99,** 9190–9195 (2002).
12. Ding, L., Chen, K., Santini, P. A., Shi, Z. & Kallenbach, N. R. The pentapeptide GGAGG has PII conformation. *J. Am. Chem. Soc.* **125,** 8092–8093 (2003).
13. Rucker, A. L., Pager, C. T., Campbell, M. N., Qualls, J. E. & Creamer, T. P. Host-guest scale of left-handed polyproline II helix formation. *Proteins* **53,** 68–75 (2003).
14. Asher, S. A., Mikhonin, A. V. & Bykov, S. UV raman demonstrates that α-helical polyalanine peptides melt to polyproline II conformations. *J. Am. Chem. Soc.* **126,** 8433–8440 (2004).
15. Eker, F., Griebenow, K., Cao, X., Nafie, L. A. & Schweitzer-Stenner, R. Preferred peptide backbone conformations in the unfolded state revealed by the structure analysis of alanine-based (AXA) tripeptides in aqueous solution. *Proc. Natl. Acad. Sci. USA* **101,** 10054–10059 (2004).
16. McColl, I. H., Blanch, E. W., Hecht, L., Kallenbach, N. R. & Barron, L. D. Vibrational raman optical activity characterization of poly(l-proline) II helix in alanine oligopeptides. *J. Am. Chem. Soc.* **126,** 5076–5077 (2004).
17. Kim, Y. S., Wang, J. & Hochstrasser, R. M. Two-dimensional infrared spectroscopy of the alanine dipeptide in aqueous solution. *J. Phys. Chem. B* **109,** 7511–7521 (2005).
18. Shi, Z. *et al.* Polyproline II propensities from GGXGG peptides reveal an anticorrelation with β-sheet scales. *Proc. Natl. Acad. Sci. USA* **102,** 17964–17968 (2005).
19. Avbelj, F., Grdadolnik, S. G., Grdadolnik, J. & Baldwin, R. L. Intrinsic backbone preferences are fully present in blocked amino acids. *Proc. Natl. Acad. Sci. USA* **103,** 1272–1277 (2006).
20. Graf, J., Nguyen, P. H., Stock, G. & Schwalbe, H. Structure and dynamics of the homologous series of alanine peptides: a joint molecular dynamics/NMR study. *J. Am. Chem. Soc.* **129,** 1179–1189 (2007).
21. Chen, K., Liu, Z., Zhou, C., Bracken, W. C. & Kallenbach, N. R. Spin relaxation enhancement confirms dominance of extended conformations in short alanine peptides. *Angew. Chem. Int. Edit.* **46,** 9036–9039 (2007).
22. Hagarman, A., Measey, T. J., 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.* **132,** 540–551 (2010).
23. Grdadolnik, J., Mohacek-Grosev, V., Baldwin, R. L. & Avbelj, F. Populations of the three major backbone conformations in 19 amino acid dipeptides. *Proc. Natl. Acad. Sci. USA* **108,** 1794–1798 (2011).
24. Oh, K. *et al.* A comprehensive library of blocked dipeptides reveals intrinsic backbone conformational propensities of unfolded proteins. *Proteins* **80,** 977–990 (2012).
25. Oh, K., Jung, Y., Hwang, G. & Cho, M. Conformational distributions of denatured and unstructured proteins are similar to those of 20 × 20 blocked dipeptides. *J. Biomol. NMR* **53,** 25–41 (2012).
26. Brown, A. M. & Zondlo, N. J. A propensity scale for type II polyproline helices (PPII): aromatic amino acids in proline-rich sequences strongly disfavor PPII due to proline-aromatic interactions. *Biochemistry* **51,** 5041–5051 (2012).
27. He, L., Navarro, A. E., Shi, Z. & Kallenbach, N. R. End effects influence short model peptide conformation. *J. Am. Chem. Soc.* **134,** 1571–1576 (2012).
28. Shi, Z., Woody, R. W. & Kallenbach, N. R. Is polyproline II a major backbone conformation in unfolded proteins? *Adv. Protein Chem.* **62,** 163–240 (2002).
29. Shi, Z., Chen, K., Liu, Z. & Kallenbach, N. R. Conformation of the backbone in unfolded proteins. *Chem. Rev.* **106,** 1877–1897 (2006).
30. MacArthur, M. W. & Thornton, J. M. Influence of proline residues on protein conformation. *J. Mol. Biol.* **218,** 397–412 (1991).
31. Hurley, J. H., Mason, D. A. & Matthews, B. W. Flexible-geometry conformational energy maps for the amino acid residue preceding a proline. *Biopolymers* **32,** 1443–1446 (1992).
32. Maigret, M., Pullman, B. & Caillet, J. The conformational energy map of an alanyl residue preceding proline: a quantum-mechanical approach. *Biochem. Biophys. Res. Commun.* **40,** 808–813 (1970).
33. Miller, S. E., Watkins, A. M., Kallenbach, N. R. & Arora, P. S. Effects of side chains in helix nucleation differ from helix propagation. *Proc. Natl. Acad. Sci. USA* **111,** 6636–6641 (2014).
34. Zimm, B. H. & Bragg, J. K. Theory of the phase transition between helix and random coil in polypeptide chains. *J. Chem. Phys.* **31,** 526–535 (1959).
35. Qian, H. & Schellman, J. A. Helix-coil theories: a comparative study for finite length polypeptides. *J. Phys. Chem.* **96,** 3987–3994 (1992).
36. Lifson, S. & Roig, A. On the theory of helix-coil transition in polypeptides. *J. Chem. Phys.* **34,** 1963–1974 (1961).

37. Yang, J., Zhao, K., Gong, Y., Vologodskii, A. & Kallenbach, N. R. α-Helix nucleation constant in copolypeptides of alanine and ornithine or lysine. *J. Am. Chem. Soc.* **120,** 10646–10652 (1998).
38. Ponder, J. W. & Case, D. A. Force fields for protein simulations. *Adv. Protein Chem.* **66,** 27–85 (2003).
39. Lindorff-Larsen, K., Piana, S., Dror, R. O. & Shaw, D. E. How fast-folding proteins fold. *Science* **334,** 517–520 (2011).
40. Woody, R. W. Circular dichroism and conformation of unordered polypeptides. *Adv. Biophys. Chem.* **2,** 37–79 (1992).
41. Kuemin, M., Engel, J. & Wennemers, H. Temperature-induced transition between polyproline I and II helices: quantitative fitting of hysteresis effects. *J. Pept. Sci.* **16,** 596–600 (2010).
42. Karplus, M. Contact electron-spin coupling of nuclear magnetic moments. *J. Chem. Phys.* **30,** 11–15 (1959).
43. Vuister, G. W. & Bax, A. Quantitative J correlation: a new approach for measuring homonuclear three-bond $J_{HNH\alpha}$ coupling constants in ^{15}N-enriched proteins. *J. Am. Chem. Soc.* **115,** 7772–7777 (1993).
44. Hagarman, A. *et al.* Amino acids with hydrogen-bonding side chains have an intrinsic tendency to sample various turn conformations in aqueous solution. *Chem.-Eur. J.* **17,** 6789–6797 (2011).
45. O'Neil, K. T. & DeGrado, W. F. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* **250,** 646–651 (1990).
46. Lyu, P. C., Liff, M. I., Marky, L. A. & Kallenbach, N. R. Side chain contributions to the stability of alpha-helical structure in peptides. *Science* **250,** 669–673 (1990).
47. Rohl, C. A., Chakrabartty, A. & Baldwin, R. L. Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Protein Sci.* **5,** 2623–2637 (1996).
48. Kim, C. A. & Berg, J. M. Thermodynamic beta-sheet propensities measured using a zinc-finger host peptide. *Nature* **362,** 267–270 (1993).
49. Avbelj, F. & Baldwin, R. L. Role of backbone solvation and electrostatics in generating preferred peptide backbone conformations: distributions of phi. *Proc. Natl. Acad. Sci. USA* **100,** 5742–5747 (2003).
50. Wang, A. C. & Bax, A. Determination of the backbone dihedral angles φ in human ubiquitin from reparametrized empirical Karplus equations. *J. Am. Chem. Soc.* **118,** 2483–2494 (1996).

Acknowledgements

The authors are grateful to the Analytical and Testing Center of HUST. Particularly, we thank Ping Liang for the support on NMR experiments. We gratefully acknowledge helpful discussions from Robert Baldwin, Tobin Sosnick, Neville Kallenbach and Sunney Chan. This work was supported by National Natural Science Foundation of China (21272081). Z.S. acknowledges HUST for a start-up support.

Author Contributions

Z.S. conceived and designed the experiments; Y.Z., L.H., W.Z. and J.H. performed the experiments; Z.S., Y.Z. and L.H. analyzed the data; Z.S. wrote the main manuscript text; Y.Z. prepared all the figures and tables. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Zhou, Y. *et al.* Populations of the Minor α-Conformation in $AcGXGNH_2$ and the α-Helical Nucleation Propensities. *Sci. Rep.* **6**, 27197; doi: 10.1038/srep27197 (2016).