Origin of the sequence-dependent polyproline II structure in unfolded peptides

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · DECEMBER 2005

Impact Factor: 2.63 · DOI: 10.1002/prot.20655 · Source: PubMed

3 AUTHORS, INCLUDING:

Mihaly Mezei
Icahn School of Medicine at Mount Sinai
173 PUBLICATIONS **4,264** CITATIONS
SEE PROFILE

Roman Osman
Icahn School of Medicine at Mount Sinai
166 PUBLICATIONS **4,367** CITATIONS
SEE PROFILE

PROTEINS: Structure, Function, and Bioinformatics 61:769–776 (2005)

Origin of the Sequence-Dependent Polyproline II Structure in Unfolded Peptides

Alex Kentsis, Mihaly Mezei, and Roman Osman*
Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York University, New York, New York

ABSTRACT **Recent studies have indicated that the unfolded states of polypeptides contain a substantial amount of polyproline type II (P_{II}) structure. This energetically and structurally preorganized state may contribute to the reduction of the folding search, as well as to the recognition of intrinsically unstructured proteins and polyproline ligands. Using Monte Carlo simulations of natively unfolded peptides in the presence of explicit aqueous solvation, we observe that residue-specific P_{II} conformational propensity is the result of the modulation of polypeptide backbone hydration by a proximal side-chain. Such a mechanism may be unique among those that contribute to the modulation of secondary structures in proteins. The calculated conformational propensities should prove useful for the development of a configurational P_{II} scale necessary for the prediction and design of natural-like polypeptides. Proteins 2005;61:769–776.**

Key words: polyproline; unfolded state structure; polypeptide hydration

INTRODUCTION

Spontaneous folding and binding of biological macromolecules in an aqueous environment originate from a nonrandom initial state. Under these conditions, prebound and unfolded states of proteins may contain a substantial amount of polyproline type II (P_{II}) (ϕ, ψ) = ($-75°$, $+145°$) structure[1–4]; for an overview, see Rose.[5] These observations have led to profound revisions of our understanding of the mechanisms of protein folding and binding, diminishing the importance of a combinatorial search of the conformational space. For instance, prebound state structures can reduce the combinatorial search during binding and speed up recognition by 1.6 times.[6] Similarly, in the case of α-helix formation in polyalanine, the segmentally preorganized P_{II} helical unfolded state contributes as much as half to the energetics of the folding search.[7]

Host–guest studies of P_{II} in polyprolyl peptides exhibit residue specific stabilities,[8,9] and analyses of the conformational properties of amino acids found in P_{II} structures in the Protein Data Bank (PDB) reveal residue-specific propensities for the P_{II} conformation.[10–16] It has been suggested that P_{II} structure observed in the unfolded and prebound states of nonprolyl proteins can preorganize folding and binding in a sequence dependent manner, consistent with hierarchical organization of proteins.[17,18] In order to understand the origin of this configurational, sequence-dependent phenomenon, we have undertaken to characterize the P_{II} conformational properties on a single-residue level. Such a conformational, amino acid–dependent representation is required for the development of a configurational partition function that incorporates neighbor effects.

Recently, Kallenbach and colleagues demonstrated that alanine in a blocked GGAGG pentapeptide exists largely in P_{II} conformation.[19,20] Here, we present results of Monte Carlo simulations of blocked GGXGG pentapeptides under the conditions of molecular solvation and near physiological temperature and pressure. We find that residue-specific P_{II} propensities result from the modulation of polypeptide backbone hydration by a proximal side-chain.

METHODS

Molecular Systems

Simulations were performed with the all-atom CHARMM27 force field[21] and the transferable intermolecular potential (TIP3P) water model.[22] Calculations were performed on the blocked pentapeptides Ac-GGXGG-NH_2 [X = M, F, R, Q, K, A, Y, W, E, neutral H (H), protonated H (H′), S, C, I, V, D, G, N, T], as studied experimentally by Kallenbach and colleagues.[19] Initial solute configurations (see below) were generated manually in vacuum, solvated, energy-minimized and equilibrated in periodic boundary conditions (PBCs). The PBC systems were constructed by randomly placing waters into a face-centered cubic cell until appropriate density was reached, using partial specific molecular volume of water of 30 Å^3, with the total number of water molecules adjusted based on the partial molar volume of component amino acids of the solute. Primary hydration shells (PHSs) were constructed from the PBC systems by deleting water molecules outside of the PHS radius as measured from the nearest solute heavy atom. The PHS systems contained on the order of 200 water molecules, corresponding to hydration shells of 6–8 Å or two to three molecular layers in thickness on average.

Grant sponsor: National Institutes of Health MSTP; Grant number: DK 43036 (to R. Osman), Medical Scientist Training Program (to A. Kentsis).

*Correspondence to: Roman Osman, Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York University, New York, NY 10029. E-mail: roman.osman@mssm.edu

Received 16 February 2005; Accepted 25 May 2005

Published online 28 September 2001 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20655

Monte Carlo Simulations

Our implementation of the PHS for Monte Carlo (MC) simulations replaces bulk molecular solvation with a restraining potential at the surface of a molecular solvation layer that maintains proper water structure and energetics by mimicking steric exclusion by the outlying bulk water, and by accurately representing interfacial water structure for waters that are outside of the PHS but within the restraining potential.[23] Usage of a size-independent reference shell energy allows the primary hydration shell to adopt arbitrary shape and size as dictated by conformational sampling of the solute, while capturing principal solvation effects such as density, solvation energy, and molecular water structure.[23] MC-PHS appears to reproduce solvation structure and energetics of both highly polar and nonpolar amino acid solutes, as well as to correctly recapitulate experimental properties of several polypeptide solutes, making it suitable for studies of structure and thermodynamics of flexible polypeptides in the context of explicit aqueous solvation.[7,23] MC simulations were performed using a force-biased Metropolis procedure as implemented in the program MMC (http://inka.mssm.edu/~mezei/mmc). Systems were thermalized for 10,000 sweeps, as judged from energy equilibration, and evolved for 2 million sweeps, saving configurations every 100th sweep, where one sweep represents one step of all the degrees of freedom, including those of the solute and the solvent. Bond lengths and angles were kept constant, and solute torsions were moved one at a time. We utilized the shuffled cyclic procedure for solute torsional moves.[24] Both solute and solvent step sizes were tuned to yield mean acceptance rates of 20–40%. The radius of the primary hydration shell was updated every third sweep, using the normalized reference shell energy of 0.15 kcal/mol/molecule and the restraining force constant of 3.0 kcal/mol/Å^2, as described previously.[23] All nonbonded interactions were included. Interaction energies, radial and orientational distribution functions, and coordination numbers were calculated according to standard methods,[25] referenced to the center of mass of the solute, as implemented in MMC using proximity analysis (http://inka.mssm.edu/~mezei/mmc).

For analysis of geometries of polypeptide conformations, we utilized a self-consistent method for defining conformational basins, using an adaptive resonance theory algorithm based on a self-organizing neural net, as implemented in ART-2.[26] Briefly, the cluster assignment of the dihedral angles extracted from simulation trajectories are optimized subject to a constraint on the cluster radius, such that no member of a cluster is farther than a specified distance from the cluster center. Because the convergence of such minimizations is sensitive to initial conditions, we tested the robustness of assignments to conformational basins by recalculating the cluster assignments using reshuffled simulation trajectories (data not shown). In this manner, we use P_{II} to refer to backbone geometries in the P_{II} conformational basin, and coil to all other conformations.[7] To evaluate sampling efficiency, we calculate the evolution of the apparent self-diffusion coefficient, $\Omega_{a-b}(x) = [\langle f_a(x)\rangle - \langle f_b(x)\rangle]^2$, as a function of simulation length x, where f_a and f_b are phase space variables, such as protein backbone dihedral angles, of two independent simulations starting from different initial conditions a and b. If the sampling of phase space is ergodic, $\Omega_{a-b}(x)/\Omega_{a-b}(0)$ decays to zero at long x. This is a necessary but insufficient condition of ergodicity, since it depends on the choice of initial conditions a and b. To evaluate convergence and ergodicity directly, we carried out two independent simulations using different random number seeds for the initial MC moves and their sizes, and two different initial solute conformations: P_{II} (ϕ, ψ = −75°, +145°) and right-handed P_{II} (ϕ, ψ = +145°, −75°). The right-handed P_{II} conformation was selected because of its maximal displacement in torsional space from P_{II}. Interaction energies, radial and orientational distribution functions, and coordination numbers were calculated according to standard methods,[25] referenced to the center of mass of the solute, as implemented in MMC using proximity analysis.[27]

RESULTS

In order to ensure exhaustive sampling of solute configurational space in the context of explicit aqueous solvation, we employed an MC implementation of the primary hydration shell (MC-PHS) that maintains principal solvation effects such as density, solvation energy, and fine water structure, while efficiently coupling structural rearrangements of the solvent to those of the solute.[23] MC-PHS simulation lengths exceed the apparent computational time for the self-diffusion of the backbone dihedral angles of 7- and 14-residue polyalanine peptides by more than three orders of magnitude.[7] They can be considered converged as judged from the equivalence of results of independent simulations using initial configurations maximally distributed in torsional phase space.[7] Similar sampling efficiency is achieved for much smaller five-residue peptides studied here (data not shown). Results of self-consistent clustering of conformational ensembles of central residues in GGXGG are presented in Figure 1, with the error bars representing $\pm 1\sigma$ of the calculated probabilities from independent simulations using canonical P_{II} (ϕ, ψ) = (−75°, +145°) and right-handed P_{II} (ϕ, ψ) = (−145°, +75°) as initial configurations. As such, they are a heuristic measure of the convergence of the calculations (Fig. 1). Nearly half of the ensemble of GGAGG is in P_{II} conformation with a population-weighted average (ϕ, ψ) = (−81°, +158°), in excellent agreement with circular dichroism (CD) and NMR spectroscopic studies of this peptide (ϕ, ψ) = (−80°, +170°),[19,20] and previous calculations of P_{II} geometry in aqueous solution.[7] Moreover, backbone geometries of P_{II} ensembles formed by various amino acids in GGXGG are similar to each other (Table I).

More importantly, a distribution of P_{II} propensities is observed, with some amino acids predominantly in the coil conformation, with geometries widely distributed in torsional space, while others are predominantly localized in the distinct P_{II} conformation (Fig. 1). Such distribution of residue specific P_{II} propensities is in general agreement with experimentally determined residue specific P_{II} scales.

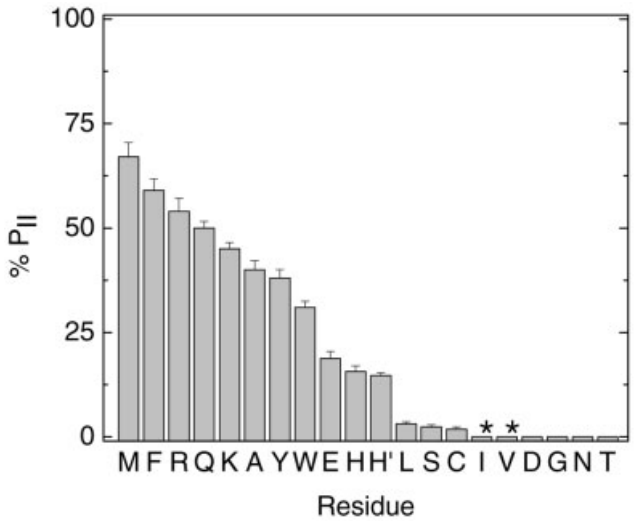

Fig. 1. Fraction of P_{II} sampled by the central residue in GGXGG as assigned using self-consistent clustering. Error bars represent $\pm 1\sigma$ of the calculated probabilities from independent simulations using canonical P_{II} (ϕ, ψ) = ($-75°$, $+145°$) and right-handed P_{II} (ϕ, ψ) = ($-145°$, $+75°$) as initial configurations, and are a heuristic measure of the convergence of the calculations. The right-handed P_{II} conformation serves as a test of convergence of simulations due to its maximal displacement in the torsional phase space from the left-handed P_{II} geometry. Asterisk indicates P_{II} conformations that are continuous with the coil conformational basin, as judged from the contiguity of their conformational basins.[7] Enumeration of P_{II} conformers using geometric binning (http://roselab.jhu.edu/utils/pross.html) produces similar results, as does calculation of P_{II} propensities using molecular dynamics simulations under periodic boundary conditions (data not shown).

TABLE I. Backbone Geometries of P_{II} Clusters Formed by Central Residues in GGXGG, as Expressed by the Centroid (ϕ, ψ) and Its Standard Deviation (σ)

GGXGG	ϕ	ψ	σ
M	−84	167	8.9
F	−84	142	13
R	−93	164	14
Q	−87	160	15
K	−95	158	17
A	−81	158	14
Y	−96	135	16
W	−87	153	13
E	−91	155	11
H	−82	159	10
H′	−86	142	14
L	−78	150	17
S	−93	158	9.1
C	−91	164	11
I	−82	158	—
V	−86	158	—

The calculated geometries are somewhat more extended than the canonical polyprolyl P_{II} geometry of ($-75°$, $+145°$), but are in agreement with the measured geometry of the GGAGG of ($-80°$, $+170°$),[19,20] highlighting the importance of self-consistent microscopic-state clustering.

For example, using P_3XP_3GY peptides, Creamer and colleagues observed relatively high P_{II} content of A- and Q-containing peptides, and relatively low P_{II} content of N- and V-containing peptides, although all peptides contained some P_{II} structure, possibly due to the flanking polyprolines.[9,28] Similarly, Schweitzer-Stenner and colleagues observed that K, A, and M had high P_{II} content in the context of AXA peptides, while V and S formed no significant P_{II} structure. However, F, W, and H failed to form P_{II},[2,29] which differs from our observations (Fig. 1). Discrepancies between P_{II} conformational propensities calculated from simulations of capped GGXGG and those measured above may be due to unique features of P_{II} in polyprolyl peptides such as P_3XP_3GY or uncapped zwitterions such as AXA. More importantly, calculated residue specific P_{II} stabilities of blocked GGXGG (Fig. 1) are in agreement with those measured spectroscopically by Kallenbach and colleagues in the same systems (personal communication).

The calculated conformational P_{II} propensity scale exhibits two general classes: formers (M, F, R, Q, K, A, W, E, H), and nonformers (L, S, C, I, V, D, G, N, T; Fig. 1). Insofar as substitutions of alanine to other amino acids do not significantly increase the apparent conformational P_{II} propensity ($\Delta G < k_B T$), ability to form P_{II} appears to be an inherent property of polypeptides, and residue-specific modulation of this inherent polypeptide backbone property is largely disruptive. In order to delineate the origin of this modulation, we first examined the calculated ensembles of similar residues with opposite P_{II} propensities, such as the isoelectronic (P_{II} former) Q and (P_{II} nonformer) N, which differ only by one methylene side-chain group (Fig. 1). In order to explore whether the side-chain conformation may contribute to the disruption of P_{II} structure in the backbone as a result of a direct side-chain–backbone interaction, we analyzed side-chain orientations in P_{II} and coil ensembles of Q and the coil ensemble of N. As can be seen in Figure 2, orientation of the Q side-chain does not significantly differ between ensembles of P_{II} and coil conformers. Furthermore, the side-chain orientation of P_{II} conformers of Q does not significantly differ from that of the P_{II} nonformer N (Fig. 2). Moreover, clustering of side-chain (χ_1, χ_2) with backbone (ϕ, ψ) of P_{II} and coil ensembles of Q fails to identify any statistically significant clusters that segregate with P_{II} geometry of the backbone (data not shown). A recent survey of the PDB also found no relationship between side-chain conformation and propensity to form backbone P_{II}.[12] Thornton and colleagues reached similar conclusions earlier.[16] An exact comparison with rotamer libraries is not possible due to their binning methods,[30,31] as well as sampling limitations of structural databases.[32,33]

In previous studies, the stability of P_{II} helix in polyalanine was related to the molecular solvation of the backbone, which was entropically most favorable compared to other conformations.[7,34] Thus, we have examined the structure and energetics of hydration of the polypeptide backbone for conformers in the P_{II} and coil ensembles of all simulated GGXGG peptides (Fig. 3; Table II). Analysis of the solute–solvent pair-binding energies of water molecules in proximity to the backbone NH of conformers in P_{II} and coil ensembles demonstrates the increased presence of weakly bound waters in the latter, as reflected by

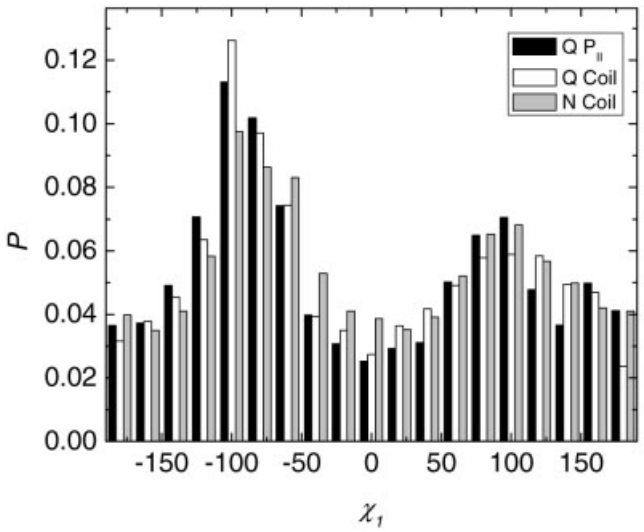

Fig. 2. Histograms of χ_1 side-chain orientations of N and Q in P_{II} and coil ensembles. The observed major population of g^- is in agreement with backbone dependent rotamer geometries,[30,31] although an exact comparison is not possible. We note that the calculated distributions deviate from one expected based solely on steric considerations, where g^- and t conformers predominate, as expressed by the χ_1 torsional potential of the force field.[21] Nevertheless, the apparently flat distribution of central χ_1 (energy difference between favored and unfavored conformers is within thermal range) may be due to the flanking glycines, as molecular dynamics simulations of the same peptides under periodic boundary conditions yield similarly flat distributions (data not shown).

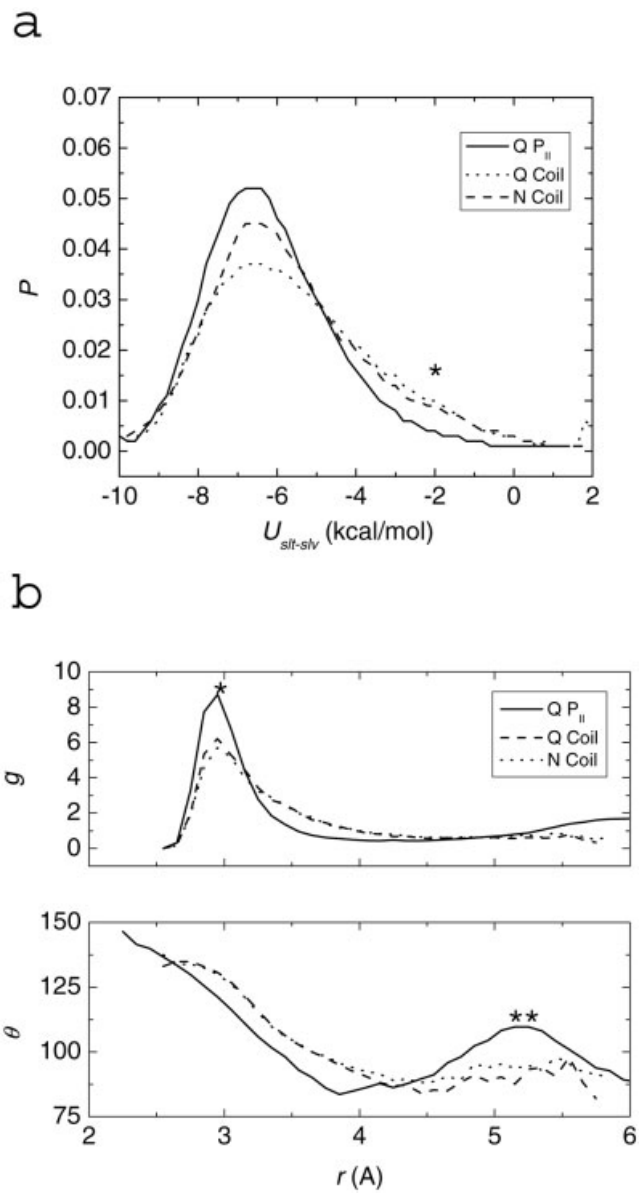

Fig. 3. (**a**) Solute–solvent pair binding energy $U_{slt\text{-}slv}$ probability distributions in proximity to the backbone NH of N and Q in P_{II} (solid) and coil (dashed, dotted) ensembles. Asterisk indicates the increased population of weakly bound waters in the ensembles of coil conformers. (**b**) Solute–solvent radial $g(r)$ (top) and mean water dipole orientational $\Theta(r)$ (bottom) water distribution functions in proximity to the backbone NH of N and Q in P_{II} (solid line) and coil (dashed, dotted lines) ensembles. Asterisk and double asterisk indicate the increased stability of the first water shell, and increased orientation of the second water shell, respectively, in the ensembles of the P_{II} conformers.

the long tail of the binding energy distributions in the coil ensembles of Q and N [Fig. 3(a)]. This is consistent with our previous observations that P_{II} conformers are best suited to accommodate hydration of the polypeptide backbone with the least disruption of the bulk water structure.[7,34] No significant differences between P_{II} and coil conformers were observed for solvent distribution functions of water in proximity to the backbone CO and Cα groups (data not shown). Consistent with such preferential P_{II} backbone hydration, radial and orientational distribution functions of water molecules in proximity to the backbone NH reveal a P_{II} backbone hydration structure that is more stable and preferentially oriented than water-solvating backbone geometries in the coil ensemble [Fig. 3(b), Table II], in agreement with water binding energetics [Fig. 3(a)]. This hydration structure is reflected in the enhanced stability of the first water shell (r = 2.8 Å), as reflected by its greater peak in the radial distribution function, and enhanced organization of the second water shell (r = 5.3 Å), as reflected by its more anisotropic mean water dipole orientational distribution function [Fig. 3(b), Table II]. This suggests that residue-specific backbone solvation, and not side-chain orientation per se, is responsible for the relative stabilization of observed P_{II} structure compared to the P_{II} nonformers (Fig. 1). It follows from these findings that the disruption of the water structure around the backbone in the P_{II} nonformers may be the reason for their reduced propensity to form P_{II}.

In order to test this hypothesis directly, we performed calculations of P_{II} nonforming GGNGG peptide with the backbone constrained in P_{II} geometry. If backbone hydration is responsible for the stability of P_{II}, then backbone constraint in the context of a P_{II}-disrupting side-chain should disrupt its hydration. In agreement with this mechanism, imposing a backbone constraint leads to a profound disruption of backbone hydration, with marked disruption of water binding energetics [Fig. 4(a)], and destabilization of water structure in proximity to the backbone NH [Fig. 4(b)]. Furthermore, the binding energy distribution of waters in proximity to the side-chain NH_2

TABLE II. Characteristics of Solute–Solvent Radial $g(r)$ and Mean Water Dipole Orientational $\theta(r)$ Water Distribution Functions in Proximity to the Backbone NH of Central Residues in GGXGG Ensembles of P_{II} and Coil Conformers [Figs. 1, 3(b)]

	P_{II}		Coil	
X	$g_{max}{}^1$	$\theta_{max}{}^2$	$g_{max}{}^1$	$\theta_{max}{}^2$
M	12.1	129	6.1	93
F	11.8	124	5.6	96
R	11.4	120	5.3	91
Q	8.7	121	5.7	94
K	12.2	133	6.3	96
A	10.1	126	5.9	95
Y	12.5	128	6.6	98
W	12.1	125	6.1	94
E	9.1	129	5.0	91
H	10.9	124	6.1	97
H′	11.2	125	6.4	92
L	11.8	120	5.8	98
S	9.8	124	6.4	94
C	9.5	124	5.9	97
I	—	—	5.8	93
V	—	—	6.9	98
D	—	—	6.4	95
G	—	—	6.2	92
N	—	—	5.6	93
T	—	—	6.8	97

For clarity, only the maximum values of the radial $g_{max}{}^1$ and mean water dipole orientational $\theta_{max}{}^2$ distribution functions of the first and second water shells, respectively, are listed. Regardless of the chemical properties of sidechains of P_{II}-forming residues, hydration structures around their polypeptide backbones exhibit enhanced stabilities of the first water shells (mean $g_{max}{}^1$ = 10.9 vs 5.9 for P_{II} versus coil conformers, respectively) and enhanced organization of the second water shell (mean $\theta_{max}{}^2$ = 125 vs 95 for P_{II} vs coil conformers, respectively). Backbone hydration structures of coil conformers of P_{II}-forming residues are similar to the hydration structures of P_{II}-nonforming residues in coil conformation (mean $g_{max}{}^1$ and $\theta_{max}{}^2$ = 6.3 and 95, respectively).

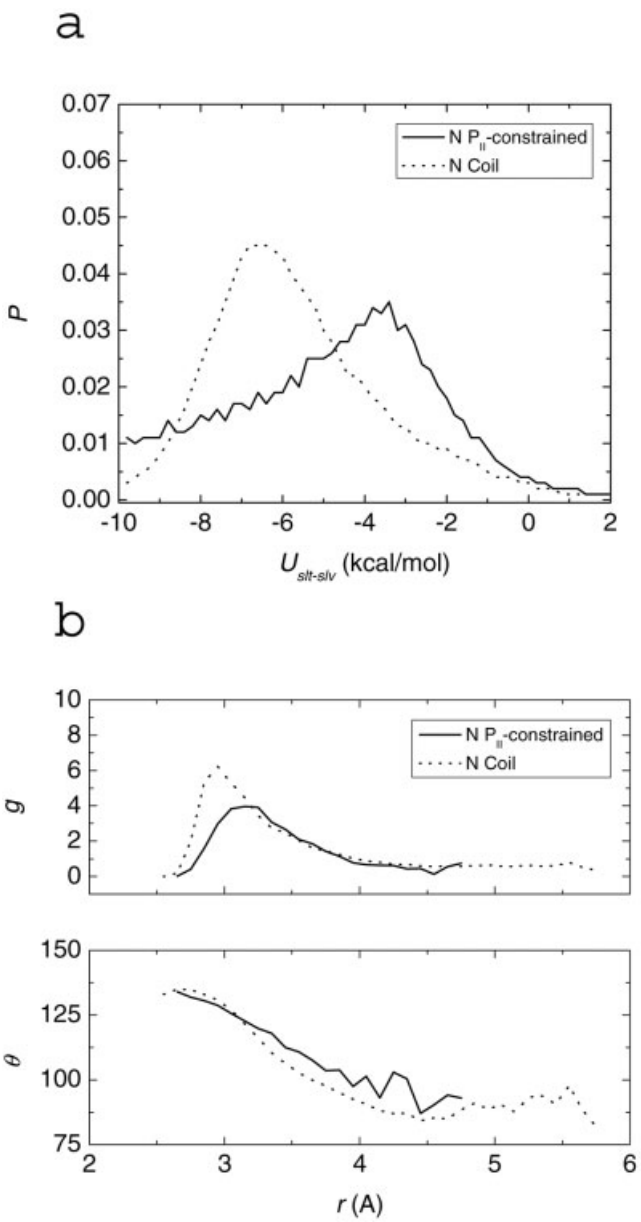

Fig. 4. (**a**) Solute–solvent pair-binding energy distributions in proximity to the backbone NH of coil (dotted line) and P_{II}-constrained (solid line) N. (**b**) Solute-solvent radial $g(r)$ (top) and mean water dipole orientational $\Theta(r)$ (bottom) water distribution functions distributions in proximity to the backbone NH of coil (dotted) and P_{II}-constrained (solid) N.

in P_{II}-constrained GGNGG exhibits a strongly bound population (Fig. 5), suggesting that disruption of backbone hydration occurs as a result of proximal side-chain placement, which competes with and/or disrupts P_{II}-stabilizing backbone hydration structure. Thus, insofar as alanine reflects the inherent properties of the peptide backbone and exhibits a significant P_{II} propensity (Fig. 1), sequence-dependent P_{II} structure appears to be related to a perturbation of backbone hydration by a proximally placed side-chain.

Since all P_{II}-forming amino acids exhibit similar hydration properties of the polypeptide backbone (Table II), we have sought to ascertain the degree of similarity in the mechanism of P_{II} backbone disruption by the different P_{II}-forming amino acids studied here (M, F, R, Q, K, A, W, E, H; Fig. 1). To that end, we calculated the difference in excess binding potential for waters hydrating the backbone NH of P_{II} and coil conformers, as expressed by the difference in logarithms of their solute–solvent binding energy probabilities, and averaged over all P_{II}-forming residues (Fig. 6). Strongly bound waters are found to interact preferentially with P_{II} conformers, and weakly bound waters with coil conformers, consistent with the preferential hydration of P_{II} as compared to β backbone conformations.[34] Moreover, there exists an apparent gradient of binding potential for water molecules hydrating the backbone P_{II} and coil conformers, such that the hydrations of P_{II} and coil are structurally and energetically distinct (Figs. 3 and 6). In contrast, a gradient and excess binding potential difference of near zero is observed for waters hydrating the backbone NH of the terminal glycine (Fig. 6), consistent with the independence of its hydration from the conformation of the central residue, and the use of GGXGG peptides to identify residue specific P_{II} prefer-

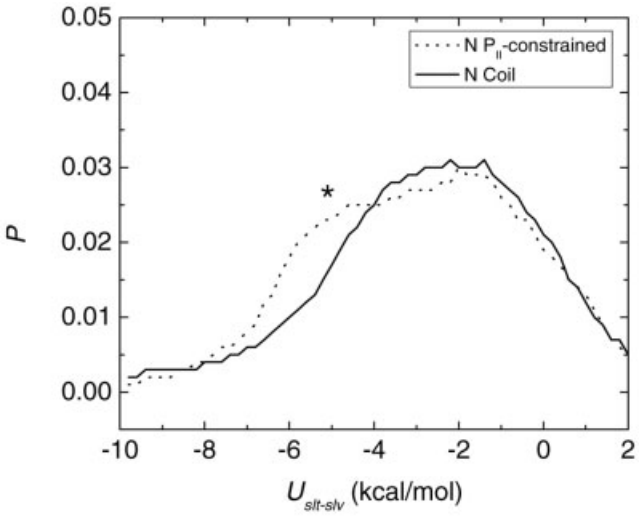

Fig. 5. Solute-solvent pair binding energy $U_{slt\text{-}slv}$ probability distributions in proximity to the side-chain NH_2 of coil (dotted) and P_{II}-constrained (solid) N. Asterisk indicates the increased population of strongly bound waters in the ensemble of P_{II}-constrained conformers.

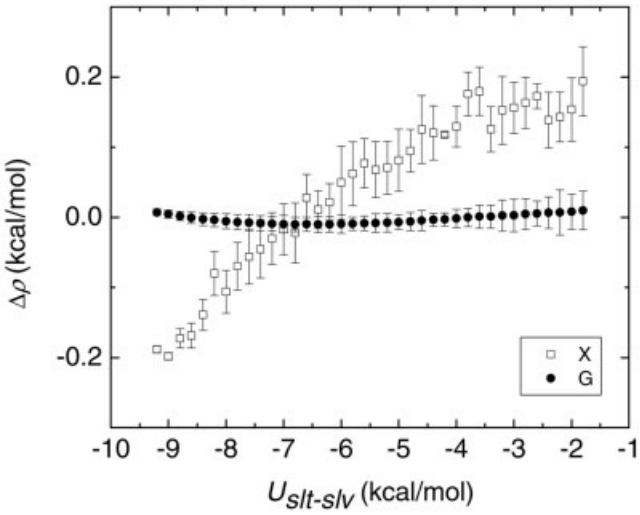

Fig. 6. Excess potential $\Delta\rho$ of solute–solvent pair-binding $U_{slt\text{-}slv}$ in proximity to the backbone NH of central X (□) and terminal G (●) in P_{II} conformation of GGXGG as referenced to that in the coil conformation for P_{II} forming amino acids (X = M, F, R, Q, K, A, E, H). Error bars represent $\pm 1\sigma$ of the calculated excess pair binding potentials, and reflect the dissimilarity of conformational backbone hydration effects exerted by different P_{II} forming amino acids in GGXGG.

ences.[19,20,35,36] Most importantly, insofar as the different P_{II}-forming amino acids (M, F, R, Q, K, A, E, H; Fig. 1) exhibit small differences in the excess binding potentials of waters hydrating the backbone NH of their P_{II} and coil conformers, as reflected by their relatively small standard deviations (Fig. 6), the mechanism of their P_{II} backbone disruption is relatively similar.

DISCUSSION

Surveys of P_{II} helices in folded proteins noted that P_{II} residues are more surface exposed than those in other regular secondary structure elements.[10,12,37] The stability of P_{II} in unfolded peptides is due to a combination of increased backbone entropy, as probed by geometry sampling using a soft-sphere repulsion potential in vacuum,[38] and increased hydration entropy and reduced enthalpy, as studied using thermodynamic integration in explicit water[34] and energy decomposition analysis.[7] Hydration of the unfolded P_{II} polypeptide backbone approximates the hydration of water itself, whereby hydrating waters form a channel and a collection of sites surrounding the backbone of a P_{II} helix whose interaction with the bulk phase is minimally disrupted.[14,34,39,40] Unique features of P_{II} hydration have been noted,[41–43] but their contribution to the energetic origin of P_{II} and its sequence dependence have remained unclear.

Here, we have shown that conformational P_{II} propensity of amino acids in the context of an otherwise unstructured GGXGG pentapeptide can be divided into two groups and is distinguished by backbone hydration. Among P_{II}-forming amino acids, the hydration of the backbone is characterized by a well-defined radial organization of the first shell and an orientational organization of the second shell (Figs. 1 and 6, Table II). In contrast, residues that do not appreciably form P_{II} fail to do so due to a disruption of this backbone hydration structure by a proximally placed side-chain (Figs. 1 and 6, Table II). Although most of the enthalpy of solvation of alanine is due to the backbone CO:water interaction, the enthalpic preference for P_{II} conformations in aqueous solution is due to backbone NH hydration,[44] as demonstrated in this work as well (Figs. 3 and 6). Waters hydrating backbone NH of residues in P_{II} conformation exhibit higher interaction energy and excess binding potential (Figs. 3 and 6). This enthalpic contribution is concomitant with the enhanced free energy of hydration and conformational entropy of P_{II}.[34,38] Consistent with this inherently molecular basis of P_{II} stability, ab initio and semiempirical calculations of alanine peptides recapitulate experimentally observed P_{II} conformations only when explicit molecular hydration is included.[45,46]

Inasmuch as properties of polyalanine reflect the physics of the polypeptide backbone, the preference for P_{II} observed in this work and in previous studies of polyalanine[7,34,38] suggests that the ability to form P_{II} is an inherent property of peptides and its origin is in the hydration of the peptide backbone. Indeed, substitutions in GGXGG tend not to cause significant increases in conformational P_{II} propensity relative to alanine ($\Delta G < k_BT$; Fig. 1). Moreover, the sequence specific modulation of P_{II} propensity on a single-residue level is largely determined by the disruptive effects on the backbone solvation by the proximal side-chain (Figs. 3 and 6, Table II).

For example, methyl to hydroxymethyl conversion of alanine to serine, and shortening of butyl-carboxamide to propyl-carboxamide in glutamine versus asparagine lead to disruption of P_{II} in GGXGG as a result of disruption of backbone hydration by a proximally placed side-chain (Figs. 3 and 6, Table II). On the other hand, amino acids that are sufficiently long to separate their side-chain functional groups from the backbone, such as lysine and

arginine, exhibit conformational P_{II} propensities similar to that of alanine (Fig. 1, Tables I and II). In spite of chemical differences among the side-chains of examined residues (amide, carboxylate, hydroxyl, alkyl, phenyl, etc.), the physical mechanism of their P_{II} disruption is apparently similar and stems from disruption of peptide backbone hydration by a proximal side-chain (Fig. 6, Table II). Such a water-mediated mechanism of residue-specific P_{II} stability is not necessarily exclusive of other contributions, such as amphipathic hydration of the polypeptide backbone with its differential polar and apolar hydration,[14] electronic hyperconjugation,[47] and steric hindrance.[38]

Inherent P_{II} propensity and its modulation by amino acid side-chains distinguish P_{II} from other secondary structures of polypeptides. For example, α-helical propensity of amino acids is due largely to the effect of side-chain on the conformational entropy of the backbone,[13,48,49] whereas propensity to form β-sheet conformations is due to solvent shielding of the backbone by the side-chain.[50–52] On the other hand, amino acid propensity to form P_{II} conformations appears to be linked directly to the hydration properties of the polypeptide backbone, and to be modulated by the disruption of this hydration by a proximal side-chain (Figs. 1 and 6; Table II).

Such a mechanism of sequence-dependent P_{II} structure emphasizes the role of dehydration in the folding and binding of P_{II}-containing proteins. Insofar as unfolded proteins contain significant amounts of P_{II} structure, hydration-mediated, sequence-dependent P_{II} stability implies that the hydrophobic effect can act early in the folding process to preorganize unfolded state structure in a sequence-dependent manner. Similarly, specific recognition of P_{II} ligands such as binding of short peptides by SH3 and WW domains, antigenic peptides by class II major histocompatibility complexes, as well as binding of numerous intrinsically unstructured proteins, may rely on hydration properties of these molecules. In all, both the conformational preferences of polypeptides and their sequence dependence appear to be intimately linked to their aqueous hydration under physiological conditions.

ACKNOWLEDGMENTS

We are grateful to Neville Kallenbach, George Rose, and Tobin Sosnick for useful discussions and sharing of unpublished manuscripts.

REFERENCES

1. Tiffany ML, Krimm S. New chain conformations of poly(glutamic acid) and polylysine. Biopolymers 1968;6:1379–1382.
2. Eker F, Cao X, 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.
3. Shi Z, 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. Wright PE, Dyson HJ. Intrinsically unstructured proteins: reassessing the protein structure-function paradigm. J Mol Biol 1999;293:321–331.
5. Rose GD, Editor. Unfolded proteins. New York: Academic Press; 2002.
6. Shoemaker BA, Portman JJ, Wolynes PG. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci USA 2000;97:8868–8873.
7. Kentsis A, Mezei M, Gindin T, Osman R. Unfolded state of polyalanine is a segmented polyproline II helix. Proteins 2004;55: 493–501.
8. Ferreon JC, Hilser VJ. The effect of the polyproline II (PPII) conformation on the denatured state entropy. Protein Sci 2003;12: 447–457.
9. Kelly MA, Chellgren BW, Rucker AL, Troutman JM, Fried MG, Miller AF, Creamer TP. Host–guest study of left-handed polyproline II helix formation. Biochemistry 2001;40:14376–14383.
10. Adzhubei AA, Sternberg MJ. Left-handed polyproline II helices commonly occur in globular proteins. J Mol Biol 1993;229:472–493.
11. 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.
12. Stapley BJ, Creamer TP. A survey of left-handed polyproline II helices. Protein Sci 1999;8:587–595.
13. Jha A, Colubri A, Zaman MH, Koide S, Sosnick TR, Freed KF. Helix, sheet, and polyproline propensities and strong nearest neighbor effects in a restricted coil library. Biochemistry 2005;44: 9691–9702.
14. Fleming PJ, Fitzkee NC, Mezei M, Srinivasan R, Rose GD. A novel method reveals that solvent water favors polyproline II over beta-strand conformation in peptides and unfolded proteins: conditional hydrophobic accessible surface area (CHASA). Protein Sci 2005;14:111–118.
15. 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.
16. 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.
17. Rose GD. Hierarchic organization of domains in globular proteins. J Mol Biol 1979;134:447–470.
18. Baldwin RL, Rose GD. Is protein folding hierarchic?: I. Local structure and peptide folding. Trends Biochem Sci 1999;24:26–33.
19. Ding L, Chen K, Santini PA, Shi Z, Kallenbach NR. The pentapeptide GGAGG has PII conformation. J Am Chem Soc 2003;125:8092–8093.
20. Chen K, Liu Z, Kallenbach NR. The polyproline II conformation in short alanine peptides is noncooperative. Proc Natl Acad Sci USA 2004;101:15352–15357.
21. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prod'hom B, Reiher IWE, Roux B III, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.
22. Jorgensen W, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
23. Kentsis A, Mezei M, Osman R. MC-PHS: A Monte Carlo implementation of the primary hydration shell for protein folding and design. Biophys J 2003;84:805–815.
24. Mezei M. On the selection of the particle to be perturbed by the Monte Carlo method. J Comp Phys 1981;39:128–136.
25. Ben-Naim A. Statistical thermodynamics for chemists and biochemists. New York: Plenum Press; 1992.
26. Karpen ME, Tobias DJ, Brooks CL III. Statistical clustering techniques for the analysis of long molecular dynamics trajectories: analysis of 2.2-ns trajectories of YPGDV. Biochemistry 1993;32:412–420.
27. Mezei M. Modified proximity criteria for the analysis of solvation of a polyfunctional solute. Mol Simulations 1988;1:327–332.
28. Rucker AL, Pager CT, Campbell MN, Qualls JE, Creamer TP. Host–guest scale of left-handed polyproline II helix formation. Proteins 2003;53:68–75.
29. Eker F, Griebenow K, Cao X, Nafie LA, 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 2004;101:10054–10059.

30. Dunbrack RL Jr, Karplus M. Backbone-dependent rotamer library for proteins: application to side-chain prediction. J Mol Biol 1993;230:543–574.
31. Dunbrack RL Jr, Karplus M. Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. Nat Struct Biol 1994;1:334–340.
32. Thomas PD, Dill KA. Statistical potentials extracted from protein structures: How accurate are they? J Mol Biol 1996;257:457–469.
33. Ben-Naim A. Statistical potentials extracted from protein structures: Are these meaningful potentials? J Chem Phys 1997;107: 3698–3706.
34. Mezei M, Fleming PJ, Srinivasan R, Rose GD. The solvation free energy of the peptide backbone is strongly conformation-dependent. Proteins 2004;55:502–507.
35. O'Connell TM, Wang L, Tropsha A, Hermans J. The "random-coil" state of proteins: comparison of database statistics and molecular simulations. Proteins 1999;36:407–418.
36. Plaxco KW, Morton CJ, Grimshaw SB, Jones JA, Pitkeathly M, Campbell ID, Dobson CM. The effects of guanidine hydrochloride on the "random coil" conformations and NMR chemical shifts of the peptide series GGXGG. J Biomol NMR 1997;10:221–230.
37. Adzhubei AA, Sternberg MJ. Conservation of polyproline II helices in homologous proteins: implications for structure prediction by model building. Protein Sci 1994;3:2395–2410.
38. Pappu RV, Srinivasan R, Rose GD. The Flory isolated-pair hypothesis is not valid for polypeptide chains: implications for protein folding. Proc Natl Acad Sci USA 2000;97:12565–12570.
39. Garcia AE. Characterization of non-alpha helical conformations in Ala peptides. Polymer 2004;45:669–676.
40. Drozdov AN, Grossfield A, Pappu RV. Role of solvent in determining conformational preferences of alanine dipeptide in water. J Am Chem Soc 2004;126:2574–2581.
41. Aizenkhaber F, Adzhubei AA, Aizenmenger F, Esipova NG. Hydration of the left spiral of the poly-L-proline type: study by the Monte Carlo method. Biofizika 1992;37:62–67.
42. Tanaka S, Scheraga HA. Theory of the cooperative transition between two ordered conformations of poly(L-proline): III. Molecular theory in the presence of solvent. Macromolecules 1975;8:516–521.
43. Sreerama N, Woody RW. Molecular dynamics simulations of polypeptide conformations in water: a comparison of alpha, beta, and poly(Pro)II conformations. Proteins 1999;36:400–406.
44. Mezei M, Mehrotra PK, Beveridge DL. Monte Carlo computer simulation of the aqueous hydration of the glycine zwitterion at 25°C. J Biomol Struct Dyn 1984;2:1–27.
45. Han WG, Suhai S. Density functional studies on *N*-methylacetamide-water complexes. J Phys Chem 1996;100:3942–3949.
46. Head-Gordon T, Head-Gordon M, Frisch MJ, Brooks CL III, Pople JA. Theoretical study of blocked glycine and alanine peptide analogs. J Am Chem Soc 1991;113:5989–5997.
47. Hinderaker MP, Raines RT. An electronic effect on protein structure. Protein Sci 2003;12:1188–1194.
48. D'Aquino JA, Gomez J, Hilser VJ, Lee KH, Amzel LM, Freire E. The magnitude of the backbone conformational entropy change in protein folding. Proteins 1996;25:143–156.
49. Luque I, Mayorga OL, Freire E. Structure-based thermodynamic scale of alpha-helix propensities in amino acids. Biochemistry 1996;35:13681–13688.
50. Minor DL Jr, Kim PS. Context is a major determinant of beta-sheet propensity. Nature 1994;371:264–267.
51. Street AG, Mayo SL. Intrinsic beta-sheet propensities result from van der Waals interactions between side chains and the local backbone. Proc Natl Acad Sci USA 1999;96:9074–9076.
52. Bai Y, Sosnick TR, Mayne L, Englander SW. Protein folding intermediates: native-state hydrogen exchange. Science 1995;269: 192–197.