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.

© 2005 Wiley-Liss, Inc.

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; for an overview, see Rose. 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. 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.

Host–guest studies of P$_{II}$ in polyproly peptides exhibit residue specific stabilities, 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. Recently, Kallenbach and colleagues demonstrated that alanine in a blocked GGAGG pentapeptide exists largely in P$_{II}$ conformation. 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 and the transferable intermolecular potential (TIP3P) water model. 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. 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

© 2005 WILEY-LISS, INC.
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. 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. 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. 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. 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/Å2, as described previously. All nonbonded interactions were included. Interaction energies, radial and orientational distribution functions, and coordination numbers were calculated according to standard methods, referenced to the center of mass of the solute, as implemented in MMC using proximity analysis.

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. 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. They can be considered converged as judged from the equivalence of results of independent simulations using initial configurations maximally distributed in torsional phase space. 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 ±1σ of the calculated probabilities from independent simulations using canonical PII (φ, ψ) = (−75°, +145°) and right-handed PII (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 PII 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°), and previous calculations of PII geometry in aqueous solution. Moreover, backbone geometries of PII ensembles formed by various amino acids in GGXGG are similar to each other (Table I).

More importantly, a distribution of PII 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 PII conformation (Fig. 1). Such distribution of residue specific PII propensities is in general agreement with experimentally determined residue specific PII scales.
CONFORMATIONAL PII PROPENSITY SCALE

For example, using P3XP3GY peptides, Creamer and colleagues observed relatively high PII content of A- and Q-containing peptides, and relatively low PII content of N- and V-containing peptides, although all peptides contained some PII structure, possibly due to the flanking polyprolines. Similarly, Schweitzer-Stenner and colleagues observed that K, A, and M had high PII content in the context of AXA peptides, while V and S formed no significant PII structure. However, F, W, and H failed to form PII, which differs from our observations (Fig. 1). Discrepancies between PII conformational propensities calculated from simulations of capped GGXGG and those measured above may be due to unique features of PII in polypropyl peptides such as P3XP3GY or uncapped zwitterions such as AXA. More importantly, calculated residue specific PII stabilities of blocked GGXGG (Fig. 1) are in agreement with those measured spectroscopically by Kellenbach and colleagues in the same systems (personal communication).

The calculated conformational PII 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 PII propensity (ΔG < k_BT), ability to form PII 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 PII propensities, such as the isoelectronic (PII former) Q and (PII 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 PII structure in the backbone as a result of a direct side-chain—backbone interaction, we analyzed side-chain orientations in PII 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 PII and coil formers. Furthermore, the side-chain orientation of PII formers of Q does not significantly differ from that of the PII nonformer N (Fig. 2). Moreover, clustering of side-chain (χ1, χ2) with backbone (φ, ψ) of PII and coil ensembles of Q fails to identify any statistically significant clusters that segregate with PII 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 PII.Thornton and colleagues reached similar conclusions earlier. An exact comparison with rotamer libraries is not possible due to their binning methods, as well as sampling limitations of structural databases.

In previous studies, the stability of PII helix in polyalanine was related to the molecular solvation of the backbone, which was entropically most favorable compared to other conformations. Thus, we have examined the structure and energetics of hydration of the polypeptide backbone for formers in the PII 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 formers in PII and coil ensembles demonstrates the increased presence of weakly bound waters in the latter, as reflected by
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 PII 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 PII and coil conformers were observed for solvent distribution functions of water in proximity to the backbone CO and Ca groups (data not shown). Consistent with such preferential PII backbone hydration, radial and orientational distribution functions of water molecules in proximity to the backbone NH reveal a PII backbone hydration structure that is more stable and preferentially oriented than watersolvating 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 III]. This suggests that residue-specific backbone solvation, and not side-chain orientation per se, is responsible for the relative stabilization of observed PII structure compared to the PII nonformers (Fig. 1). It follows from these findings that the disruption of the water structure around the backbone in the PII nonformers may be the reason for their reduced propensity to form PII.

In order to test this hypothesis directly, we performed calculations of PII nonforming GGNGG peptide with the backbone constrained in PII geometry. If backbone hydration is responsible for the stability of PII, then backbone constraint in the context of a PII-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 NH₂...
in PII-constrained GGN

ggG 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 PII-stabilizing backbone hydration structure. Thus, insofar as alanine reflects the inherent properties of the peptide backbone and exhibits a significant PII propensity (Fig. 1), sequence-dependent PII structure appears to be related to a perturbation of backbone hydration by a proximally placed side-chain.

Since all PII-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 PII backbone disruption by the different PII-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 PII and coil conformers, as expressed by the difference in logarithms of their solute–solvent binding energy probabilities, and averaged over all PII-forming residues (Fig. 6). Strongly bound waters are found to interact preferentially with PII conformers, and weakly bound waters with coil conformers, consistent with the preferential hydration of PII as compared to β backbone conformations. Moreover, there exists an apparent gradient of binding potential for water molecules hydrating the backbone NH of coil conformers of PII-forming residues similar to the hydration structures of PII-forming residues in coil conformation (mean gmax1 and θmax2 = 6.3 and 95, respectively).

For clarity, only the maximum values of the radial gmax1 and mean water dipole orientational θmax2 distribution functions of the first and second water shells, respectively, are listed. Regardless of the chemical properties of sidechains of PII-forming residues, hydration structures around their polypeptide backbones exhibit enhanced stabilities of the first water shells (mean gmax1 = 10.9 and 5.9 for PII versus coil conformers, respectively) and enhanced organization of the second water shell (mean θmax2 = 125 and 95 for PII versus coil conformers, respectively). Backbone hydration structures of coil conformers of PII-forming residues are similar to the hydration structures of PII-forming residues in coil conformation (mean gmax1 and θmax2 = 6.3 and 95, respectively).

For clarity, only the maximum values of the radial gmax1 and mean water dipole orientational θmax2 distribution functions of the first and second water shells, respectively, are listed. Regardless of the chemical properties of sidechains of PII-forming residues, hydration structures around their polypeptide backbones exhibit enhanced stabilities of the first water shells (mean gmax1 = 10.9 and 5.9 for PII versus coil conformers, respectively) and enhanced organization of the second water shell (mean θmax2 = 125 and 95 for PII versus coil conformers, respectively). Backbone hydration structures of coil conformers of PII-forming residues are similar to the hydration structures of PII-forming residues in coil conformation (mean gmax1 and θmax2 = 6.3 and 95, respectively).

**Table II. Characteristics of Solute–Solvent Radial g(r) and Mean Water Dipole Orientational θ(r) Water Distribution Functions in Proximity to the Backbone NH of Central Residues in GGXGG Ensembles of PII and Coil Conformers [Figs. 1, 3(b)]**

<table>
<thead>
<tr>
<th>X</th>
<th>gmax1</th>
<th>θmax2</th>
<th>gmax1</th>
<th>θmax2</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>12.1</td>
<td>129</td>
<td>6.1</td>
<td>93</td>
</tr>
<tr>
<td>F</td>
<td>11.8</td>
<td>124</td>
<td>5.6</td>
<td>96</td>
</tr>
<tr>
<td>R</td>
<td>11.4</td>
<td>120</td>
<td>5.3</td>
<td>91</td>
</tr>
<tr>
<td>Q</td>
<td>8.7</td>
<td>121</td>
<td>5.7</td>
<td>94</td>
</tr>
<tr>
<td>K</td>
<td>12.2</td>
<td>133</td>
<td>6.6</td>
<td>90</td>
</tr>
<tr>
<td>A</td>
<td>10.1</td>
<td>126</td>
<td>5.9</td>
<td>95</td>
</tr>
<tr>
<td>Y</td>
<td>12.5</td>
<td>135</td>
<td>6.6</td>
<td>98</td>
</tr>
<tr>
<td>W</td>
<td>9.1</td>
<td>129</td>
<td>5.0</td>
<td>91</td>
</tr>
<tr>
<td>H</td>
<td>10.9</td>
<td>124</td>
<td>6.1</td>
<td>97</td>
</tr>
<tr>
<td>H’</td>
<td>11.2</td>
<td>125</td>
<td>6.4</td>
<td>92</td>
</tr>
<tr>
<td>L</td>
<td>11.8</td>
<td>120</td>
<td>5.8</td>
<td>98</td>
</tr>
<tr>
<td>S</td>
<td>9.8</td>
<td>124</td>
<td>6.4</td>
<td>94</td>
</tr>
<tr>
<td>C</td>
<td>9.5</td>
<td>124</td>
<td>5.9</td>
<td>97</td>
</tr>
<tr>
<td>I</td>
<td>—</td>
<td>—</td>
<td>5.8</td>
<td>93</td>
</tr>
<tr>
<td>V</td>
<td>—</td>
<td>—</td>
<td>6.9</td>
<td>98</td>
</tr>
<tr>
<td>D</td>
<td>—</td>
<td>—</td>
<td>6.4</td>
<td>95</td>
</tr>
<tr>
<td>G</td>
<td>—</td>
<td>—</td>
<td>6.2</td>
<td>92</td>
</tr>
<tr>
<td>N</td>
<td>—</td>
<td>—</td>
<td>5.6</td>
<td>93</td>
</tr>
<tr>
<td>T</td>
<td>—</td>
<td>—</td>
<td>6.8</td>
<td>97</td>
</tr>
</tbody>
</table>

For clarity, only the maximum values of the radial gmax1 and mean water dipole orientational θmax2 distribution functions of the first and second water shells, respectively, are listed. Regardless of the chemical properties of sidechains of PII-forming residues, hydration structures around their polypeptide backbones exhibit enhanced stabilities of the first water shells (mean gmax1 = 10.9 vs 5.9 for PII versus coil conformers, respectively) and enhanced organization of the second water shell (mean θmax2 = 125 vs 95 for PII versus coil conformers, respectively). Backbone hydration structures of coil conformers of PII-forming residues are similar to the hydration structures of PII-forming residues in coil conformation (mean gmax1 and θmax2 = 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 PII-constrained (solid line) N. (b) Solute-solvent radial g(r) (top) and mean water dipole orientational θ(r) (bottom) water distribution functions distributions in proximity to the backbone NH of coil (dotted) and PII-constrained (solid) N.
Most importantly, insofar as the different PII-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 PII and coil conformers, as reflected by their relatively small standard deviations (Fig. 6), the mechanism of their PII backbone disruption is relatively similar.

DISCUSSION

Surveys of PII helices in folded proteins noted that PII residues are more surface exposed than those in other regular secondary structure elements.\footnote{\textsuperscript{10,12,37}} The stability of P\textsubscript{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,\footnote{\textsuperscript{58}} and increased hydration entropy and reduced enthalpy, as studied using thermodynamic integration in explicit water\footnote{\textsuperscript{14}} and energy decomposition analysis.\footnote{\textsuperscript{7}} Hydration of the unfolded P\textsubscript{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\textsubscript{II} helix whose interaction with the bulk phase is minimally disrupted.\footnote{\textsuperscript{14,34,39,40}} Unique features of P\textsubscript{II} hydration have been noted,\footnote{\textsuperscript{41–43}} but their contribution to the energetic origin of P\textsubscript{II} and its sequence dependence have remained unclear.

Here, we have shown that conformational P\textsubscript{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\textsubscript{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\textsubscript{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\textsubscript{II} conformations in aqueous solution is due to backbone NH hydration,\footnote{\textsuperscript{44}} as demonstrated in this work as well (Figs. 3 and 6). Waters hydrating backbone NH of residues in P\textsubscript{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\textsubscript{II}.\footnote{\textsuperscript{34,38}} Consistent with this inherently molecular basis of P\textsubscript{II} stability, ab initio and semiempirical calculations of alanine peptides recapitulate experimentally observed P\textsubscript{II} conformations only when explicit molecular hydration is included.\footnote{\textsuperscript{45,46}}

Inasmuch as properties of polyalanine reflect the physics of the polypeptide backbone, the preference for P\textsubscript{II} observed in this work and in previous studies of polyalanine,\footnote{\textsuperscript{7,34,38}} suggests that the ability to form P\textsubscript{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\textsubscript{II} propensity relative to alanine (\textbf{Δ}G < k\textsubscript{B}T; Fig. 1). Moreover, the sequence specific modulation of P\textsubscript{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\textsubscript{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 PII propensities similar to that of alanine (Fig. 1, Tables I and II). In spite of chemical differences among the side-chains of examined residues (amicde, carboxylate, hydroxyl, alkyl, phenyl, etc.), the physical mechanism of their PII 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 PII stability is not necessarily exclusive of other contributions, such as amphiphatic hydration of the polypeptide backbone with its differential polar and apolar hydration, electronic hyperconjugation, and steric hindrance.

Inherent PII propensity and its modulation by amino acid side-chains distinguish PII 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, whereas propensity to form β-sheet conformations is due to solvent shielding of the backbone by the side-chain. On the other hand, amino acid propensity to form PII 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 PII structure emphasizes the role of dehydration in the folding and binding of PII-containing proteins. Insofar as unfolded proteins contain significant amounts of PII structure, hydration-mediated, sequence-dependent PII 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 PII 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

4. Wright PE, Dyson HJ. Intrinsically unstructured proteins, may rely on 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).

CONFORMATIONAL PII PROPENSITY SCALE