BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

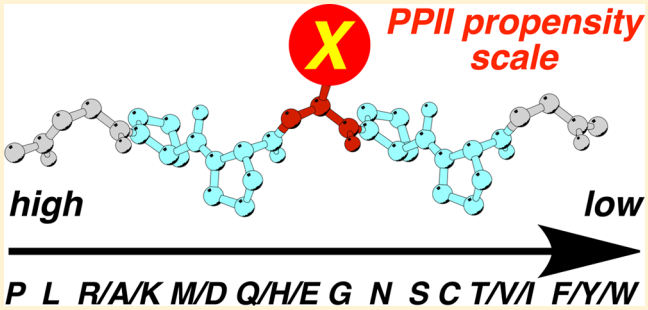
# A Propensity Scale for Type II Polyproline Helices (PPII): Aromatic Amino Acids in Proline-Rich Sequences Strongly Disfavor PPII Due to Proline−Aromatic Interactions

Alaina M. Brown and Neal J. Zondlo*

Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716, United States

Supporting Information

**ABSTRACT:** Type II polyproline helices (PPII) are a fundamental secondary structure of proteins, common in globular and nonglobular regions and important in cellular signaling. We developed a propensity scale for PPII using a host−guest system with sequence Ac-GPPXPPGY-$NH_2$, where X represents any amino acid. We found that proline has the highest PPII propensity, but most other amino acids display significant PPII propensities. The PPII propensity of leucine was the highest of all propensities of non-proline residues. Alanine and residues with linear side chains displayed the next highest PPII propensities. Three classes of residues displayed lower PPII propensities: β-branched amino acids (Thr, Val, and Ile), short amino acids with polar side chains (Asn, protonated Asp, Ser, Thr, and Cys), and aromatic amino acids (Phe, Tyr, and Trp). *tert*-Leucine particularly disfavored PPII. The basis of the low PPII propensities of aromatic amino acids in this context was significant cis−trans isomerism, with proline-rich peptides containing aromatic residues exhibiting 45−60% cis amide bonds, due to Pro−cis-Pro−aromatic and aromatic−cis-Pro amide bonds.




The left-handed type II polyproline helix (PPII) is a fundamental protein secondary structure that is widely observed in globular proteins.[1−3] Outside of globular regions of proteins, PPII appears to be a dominant conformation used in cellular signaling, with more than 400 human proteins containing domains (e.g., SH3, WW, EVH1, and GYF) that recognize protein targets bound as PPII.[4] The cellular targets of PPII-binding proteins include proline-rich domains, which are among the most common domains in eukaryotes and are the most common domain in *Drosophila*.[5] Despite their ubiquitous nature, proline-rich domains generally exist in the biologically important, yet less structurally defined, nonglobular regions of proteins (termed natively disordered regions).[6−9] PPII has been proposed as a major conformation in proline-rich domains and, more generally, of nonglobular regions of proteins.[10]

The PPII conformation is also an important contributor to the "random coil" state of unfolded proteins.[11−13] Indeed, recent studies have shown that even simple alanine peptides display significant PPII populations.[12,14−20] In addition to natural systems, PPII helices have been incorporated as protein design and protein spacer elements and served as the basis for biomimetic oligomers.[21−26] Polyproline helices comprised entirely of proline residues are commonly employed as molecular rulers, spacers, and display elements because of the rigidity, defined geometry, and 3-fold symmetry of the all-proline polyproline helix.[27−37] Nomenclature notwithstanding, non-proline residues are usually significant (or even exclusive) constituents of structures in the PPII conformation. For example, the FHA domain of Rad53 binds phosphopeptide ligands in which six consecutive non-proline residues are bound in a PPII conformation.[14,38]

Because of the importance of PPII in the unfolded state, in natively disordered proteins, in globular proteins, in protein−protein interactions, and in molecular design, there have been multiple efforts to determine the propensities of amino acids for PPII, within stable polyproline helices (or collagen triple helices) or within fully disordered contexts.[2,39−46] Our aim in this study was to address a fundamental question of critical importance to the structure of protein proline-rich domains: which residues favor, and which residues disfavor, PPII within a typical proline-rich context?

## EXPERIMENTAL PROCEDURES

**Peptide Synthesis.** All peptides were synthesized on a Rainin PS3 peptide synthesizer using Rink Amide resin (NovaBiochem) and standard Fmoc/HBTU chemistry, acetylated (10% acetic anhydride in pyridine, 10 min), purified by reverse phase high-performance liquid chromatography (Vydac semipreparative C18 column) using a linear gradient of buffer A (98% $H_2O$/2% acetonitrile/0.06% TFA) and buffer B (20% $H_2O$/80% acetonitrile/0.05% TFA), and lyophilized. Purity was

**Received:** February 29, 2012
**Revised:** June 5, 2012
**Published:** June 5, 2012

determined by the presence of a single peak upon re-injection on either an analytical (Rainin Microsorb MV 100 Å) C18 reverse phase column or the column used for purification. Peptides were characterized by matrix-assisted laser desorption ionization or electrospray ionization mass spectrometry. Characterization data are given in the Supporting Information. All peptides were N-terminally acetylated and contained C-terminal amides.

**Circular Dichroism.** Circular dichroism (CD) data were collected at 25 °C in a 1 or 2 mm cell on an Aviv 400 CD spectrometer. All data are the average of at least three independent trials. Data were background-corrected but not smoothed. Unless otherwise indicated, solution conditions were 150 $\mu$M peptide, 5 mM phosphate buffer (pH 7), and 25 mM KF. Error bars on the data (Supporting Information) indicate the standard error. Control experiments with X = Pro and X = Leu peptides showed no dependence of CD on peptide concentration in the range of 37−500 $\mu$M.

Free energies were approximated on the basis of a two-state model with limiting values of 100% PPII = 5000 deg cm$^2$ dmol$^{-1}$, based on observed mean residue ellipticities for extended stretches of polyproline, and 0% PPII in this context = −2000, the mean residue ellipticity of the valine peptide at 90 °C. Circular dichroism data were converted to energies using a two-state model and calculated via the equation $\Delta\Delta G = -RT \ln[([\theta] - \Theta_U)/(\Theta_F - [\theta])] + RT \ln[(952 - \Theta_U)/(\Theta_F - 952)]$, where mean residue ellipticities of −2000 ($\Theta_U$) and 5000 ($\Theta_F$) represent the limiting values of 0 and 100% PPII in this peptide context, respectively, and 952 is the mean residue ellipticity of glycine (the reference amino acid). Although these limits are estimates, changing their values does not affect the ordering of residues and only moderately affects the relative energies. Notably, these energy differences are similar to those calculated by Kallenbach.[46] Free energies were not calculated for peptides with Phe, Tyr, or Trp because of the high populations of the cis amide bond, which is incompatible with PPII; the unknown aromatic contribution to CD;[47−49] the unknown contribution of the species with the cis amide bond to the CD spectra; and the absence of two-state behavior because of the presence of cis amide bonds.

Temperature-dependent circular dichroism data were collected for the Ac-GPPPPPGY-NH$_2$ peptide at 4, 10, 25, 37, 50, 70, and 90 °C. The solution contained 150 $\mu$M peptide, 5 mM phosphate buffer (pH 7), and 25 mM KF. The experiments were conducted as a single trial with a single peptide solution; thus, these data are internally consistent but have greater error than the data used for the generation of the propensity scale. Data were background-corrected and smoothed. Temperature-dependent CD experiments in which X was Pro, Leu, Gly, Ser, or Val revealed a comparable effect of temperature on CD: all peptides had reduced $[\theta]_{228}$ and increased $\lambda_{max}$ values with increased temperatures, displayed an isodichroic point, and had a noncooperative thermal melt. The relative ordering of these peptides (Pro > Leu > Gly > Ser > Val) was independent of temperature.

**Nuclear Magnetic Resonance (NMR) Spectroscopy.** NMR data were collected on a Brüker 400 or 600 MHz NMR spectrometer. The solution contained 150 $\mu$M peptide, 5 mM phosphate buffer (pH 4 for peptides other than Asp and Glu; pH 2 and 6.8 for Asp and Glu), and 25 mM NaCl in a 90% $H_2O$/10% $D_2O$ mixture. One-dimensional (1D) spectra were collected using watergate water suppression. Amide protons were assigned via TOCSY spectra or by analogy. Coupling constants were determined directly (at pH 4, except for Asp and Glu) from 1D spectra. Populations of peptides with at least one cis amide bond were estimated by the integration of amide protons of the same residue (as determined by TOCSY spectra) among the different species in the 1D spectra, with the major species identified as the all-trans amide peptide and all other species identified as containing at least one cis amide bond. For the Trp-containing peptide, integaration was based on the indole proton. These estimates are lower estimates, as all peaks were not well-resolved, and minor cis-containing species (which could not be assigned via TOCSY) might not be included.

## RESULTS

To improve our understanding of PPII in its disparate roles, we chose a host−guest system to address the conformational preferences of amino acids for PPII. Our peptide design involves a variable residue in the context of a peptide with two prolines each N-terminal and C-terminal (Ac-GPP**X**PPGY-NH$_2$). Previous studies have demonstrated that two prolines are insufficient for stable polyproline helix formation, whereas sequences of three or more consecutive prolines spontaneously adopt PPII.[50,51] Thus, within this context, if guest residue X favors PPII formation, PPII is expected to be observed throughout the central pentapeptide; if residue X disfavors PPII, only modest PPII is expected to be observed from the isolated diproline segments.

Peptides were synthesized with all canonical amino acids in the guest (X) position. Peptides were analyzed by circular dichroism at pH 7, observing the positive band at 228 nm characteristic of PPII.[41,42,52−56] All peptides were also examined by NMR spectroscopy. Tabulated data (Table 1) and characteristic spectra (Figure 1) are given. Temperature-dependent CD data were also collected for the proline-containing peptide, which indicated a two-state equilibrium, based on the presence of an isodichroic point at 213 nm and a linear correlation of a dual-wavelength parametric test of mean residue ellipticity at 228 and 204 nm (Figure 2).[57] The two-state equilibrium was interpreted as an equilibrium between PPII and disordered states.

We observed, as expected, that proline displayed the highest polyproline helix propensity among all canonical amino acids. However, most residues exhibited positive CD bands at 228 nm and $^3J_{HN\alpha}$ values consistent with PPII, indicating significant PPII propensity and supporting the observations of considerable PPII content for non-proline residues.[41,42,52,53,58,59] We found the strongest PPII stabilization by Pro, Leu, Ala, and residues with long, linear side chains (Arg, Met, and Lys). In this context, Leu had a particularly high PPII propensity, an observation not seen in previous PPII propensity scales.

Glycine exhibited intermediate PPII propensity, in contrast to its low propensity for $\alpha$-helix and $\beta$-sheet. Residues that displayed low PPII propensities, based on the mean residue ellipticity at 228 nm, can be divided into three classes: (a) $\beta$-branched residues (Ile, Val, and Thr), (b) short residues with polar side chains (Cys, Ser, Thr, and Asn), and (c) aromatic residues (Phe, Tyr, and Trp). $\beta$-Branched and short polar residues also exhibited red-shifted maxima in their CD spectra, away from optimal values for PPII.

$\beta$-Branched residues (Ile, Val, and Thr) potentially clash sterically with the adjacent prolyl ring when in an ideal PPII conformation.[39,41,60] To test the role of sterics in PPII propensities, we synthesized the peptide with the highly

**Table 1. Circular Dichroism Data, NMR Coupling Constant Data, and Type II Polyproline Helix Propensities at 25 °C of Amino Acid Residues Using Ac-GPPXPPGY-$NH_2$ as a Host System[a]**

| residue X of GPPXPPGY | $\lambda_{max}$ (nm) | $[\theta]_{228}$ (deg cm$^2$ dmol$^{-1}$) | $^3J_{HN\alpha}$ (Hz) | $\Delta\Delta G^b$ (kcal mol$^{-1}$) |
|---|---|---|---|---|
| Pro | 228 | 2950 | na | −0.71 |
| Leu | 228 | 2230 | 6.8 | −0.44 |
| Arg | 228 | 1630 | 6.9 | −0.23 |
| Ala | 228 | 1570 | 5.7 | −0.21 |
| Lys | 227 | 1520 | 6.8 | −0.19 |
| Hse (homoserine) | 227 | 1390 | 6.8 | −0.15 |
| Met | 227 | 1380 | 7.2 | −0.15 |
| Asp | 227 | 1310 | 6.2 | −0.12 |
| Gln | 228 | 1180 | 7.2 | −0.08 |
| Glu (pH 2) | 226 | 1130 | 7.5 | −0.06 |
| His | 228 | 1090 | 7.2 | −0.05 |
| Glu | 228 | 1030 | 6.3 | −0.03 |
| Gly | 227 | 950 | nd | *0.00* |
| Asn | 228 | 820 | 6.8 | 0.05 |
| Ser | 231 | 460 | 6.9 | 0.18 |
| Asp (pH 2) | 231 | 280 | 7.5 | 0.24 |
| Cys | 231 | 30 | 7.0 | 0.34 |
| Thr | 230 | −120 | 7.2 | 0.41 |
| Val | 230 | −120 | 8.0 | 0.41 |
| Ile | 230 | −200 | 7.9 | 0.44 |
| Tle (*tert*-leucine) | 231 | −940 | 8.3 | 0.83 |
| Phe | 223 | 1880 | nd | |
| Tyr | 227 | 5430 | nd | |
| Trp | 227 | 4800 | nd | |

[a] $\lambda_{max}$ is the wavelength of maximal intensity of the positive PPII band. $[\theta]_{228}$ is the mean residue ellipticity of the peptide at 228 nm. $^3J_{HN\alpha}$ is the backbone coupling constant of guest residue X in the peptide. na means not applicable. nd means not determined because of spectral overlap. All CD data were collected at 25 °C. Unless otherwise indicated, solutions contained 150 μM peptide, 5 mM phosphate buffer (pH 7), and 25 mM KF. [b] Approximate free energy of PPII formation for each residue, relative to glycine: $\Delta\Delta G_{residue} = \Delta G_{residue} - \Delta G_{Gly}$. The free energies of the aromatic residues were not calculated because of anomalous CD behavior and the absence of two-state equilibria. Solely on the basis of the effects on cis−trans isomerism [$K_{trans/cis}$, where $K_{trans/cis} \geq 9$ for all nonaromatic residues and $K_{trans/cis} \leq 1.2$ (Phe), 0.82 (Tyr), and 0.67 (Trp) for the aromatic residues], the aromatic residues were destabilizing to PPII by at least 1.2 (Phe), 1.4 (Tyr), and 1.5 kcal mol$^{-1}$ (Trp).

sterically congested amino acid *tert*-leucine. The *tert*-leucine-containing peptide displayed the lowest PPII content of any peptide investigated, substantially lower than that of canonical β-branched residues, consistent with data for α-helices (Figure 1b).[61−63] Steric repulsion present for β-branched amino acids in the PPII conformation can be relieved by adopting a more extended (i.e., more β-like) conformation. This interpretation is supported by coupling constant data. $^3J_{HN\alpha}$ is significantly smaller for PPII-stabilizing Ala than for Val, Ile, or *tert*-leucine. The $^3J_{HN\alpha}$ of Val correlates, on the basis of a parametrized Karplus equation, with a more extended conformation ($\phi$ = −90° vs $\phi$ = −72° for Ala) compared to that of canonical PPII ($\phi$ = −75°, and $\psi$ = 145°).[58] Significant PPII destabilization by *tert*-leucine was previously observed by us in the proline-rich tau peptide KTPPAPKTPP, where phosphorylation of both Thr residues increased the PPII content while replacement of both Thr residues with *tert*-leucine substantially reduced the PPII content, as determined by CD and NMR ($^3J_{\alpha N}$ = 7.3 and 7.3 Hz for threonine, 3.3 and 4.4 Hz for phosphothreonine, and 8.1 and 8.4 Hz for *tert*-leucine).[64]

Direct steric repulsion is unlikely the primary determinant of low PPII propensities for the short polar residues Cys, Ser, Thr, and Asn. Side chain hydrogen bond donors can potentially hydrogen bond to main chain carbonyl acceptors.[65] Alternatively, side chain hydrogen bond acceptors can interact with intraresidue amide hydrogen bond donors. In these cases, the side chain may not inherently disfavor PPII (e.g., because of steric crowding) but instead may favor additional, non-PPII conformations, thus reducing the population of PPII.[66,67] A role of hydrogen bonding in stabilizing alternative conformations is suggested by comparing peptides containing Asp and Asn. At pH 7, Asp is deprotonated, incapable of acting as a hydrogen bond donor, and displays a high PPII propensity. In contrast, at pH 2, Asp is protonated and displays a significantly lower PPII propensity, lower than that of Asn and more comparable to that of Ser (Figure 1c), consistent with the hydrogen bond donor capabilities of a carboxylic acid being better than those of an amide. Notably, an additional methylene group is sufficient to disfavor alternative conformations: Glu displays comparable PPII propensity at pH 2 and 7. Similarly, serine, capable of an intramolecular hydrogen bond with the main chain forming a favorable six-membered ring, displays low PPII propensity. In contrast, within this context, homoserine, with an additional methylene, displays high PPII propensity, comparable to that of methionine (Figure 1d).

Computational and experimental data clearly demonstrate that short polar residues can adopt conformations with intramolecular or intraresidue hydrogen bonds.[65,68−70] Moreover, calculations on Ser and Thr demonstrate that intraresidue hydrogen-bonded conformations are among the lowest-energy conformations and that these conformations require backbone dihedral angles that are inconsistent with PPII. In summary, these data indicate a significant geometric dependence of PPII propensity for residues with hydrogen bond donor groups, which is consistent with intramolecular hydrogen bonding being a possible determinant in disfavoring PPII. An alternative explanation for these data is ordered solvation around these short polar side chains disrupting PPII (see also discussion of n → π* interactions below).

Analysis of proline-rich sequences indicates that serine and threonine residues are particularly common in proline-rich domains (Table 2), suggesting that many proline-rich domains have inherently relatively lower PPII populations. Interestingly, serine and threonine residues within proline-rich sequences are commonly observed to be phosphorylated, for example, by proline-directed kinases, including MAP kinases that are critically important in signal transduction pathways.[71] Recent data for peptides derived from the proline-rich domain of the microtubule-associated protein tau indicate that phosphorylation of serine and threonine residues induces the polyproline helix, with nonphosphorylated tau peptides exhibiting a low PPII content but phosphorylated peptides exhibiting an increased PPII content.[64] In addition, protein-binding domains that recognize phosphoprotein targets commonly bind phosphorylated proteins in a PPII conformation. In summary, these data suggest that phosphorylation of serine and threonine residues in proline-rich domains might be a general mode to change the structure and conformational preferences to promote protein−protein interactions.

CD spectra from peptides with aromatic guest residues displayed strong positive bands (Figure 3). However, the CD


dx.doi.org/10.1021/bi3002924 | *Biochemistry* 2012, 51, 5041−5051

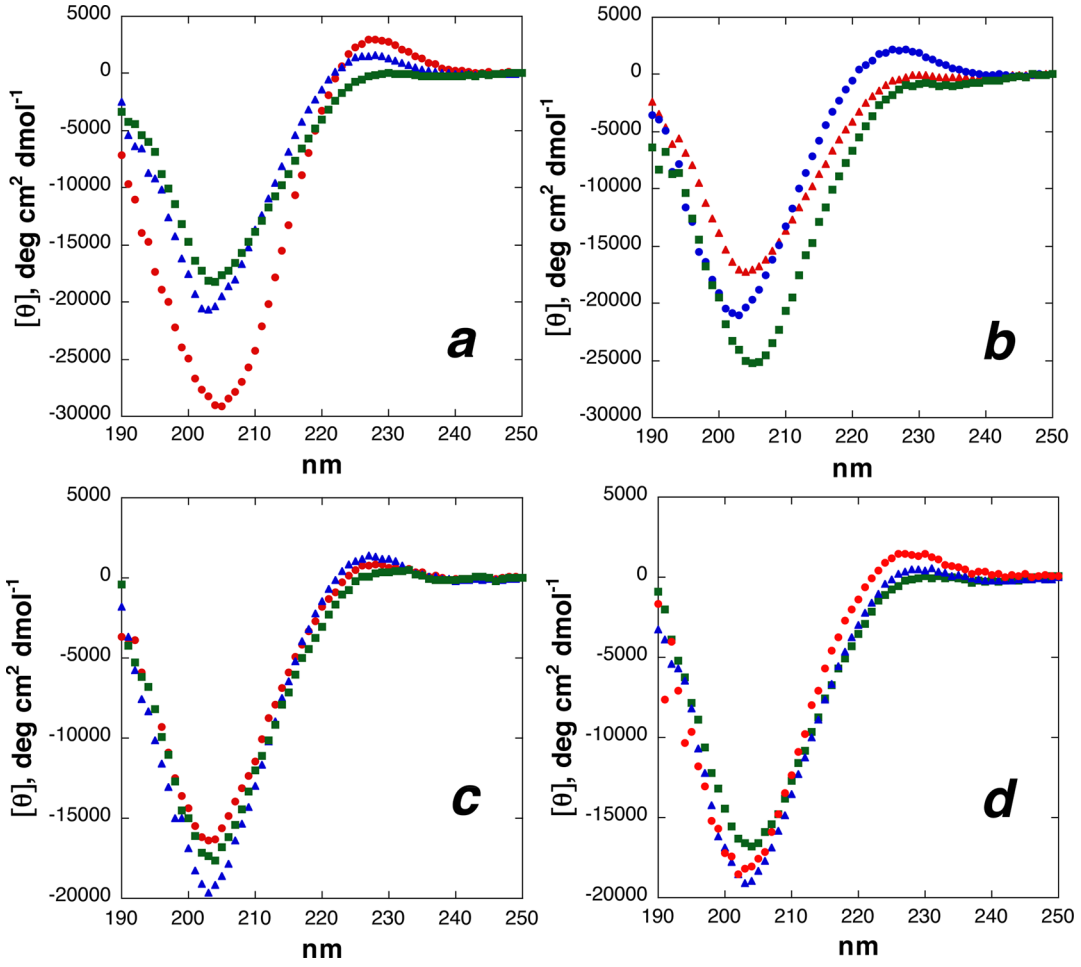



**Figure 1.** Representative circular dichroism data at 25 °C for peptides of the general sequence GPPXPPGY, with substitution of the indicated amino acid residues at position X. The positive band at 228 nm is indicative of PPII. CD data plots for all peptides, with error bars indicating standard errors, are located in the Supporting Information: (a) Pro (red circles), Ala (blue triangles), and Val (green squares); (b) Leu (blue circles), Ile (red triangles), and *tert*-leucine (green squares); (c) Asp at pH 7 (blue triangles), Asn at pH 7 (red circles), and Asp at pH 2 (green squares); and (d) homoserine (red circles), Ser (blue triangles), and Thr (green squares).

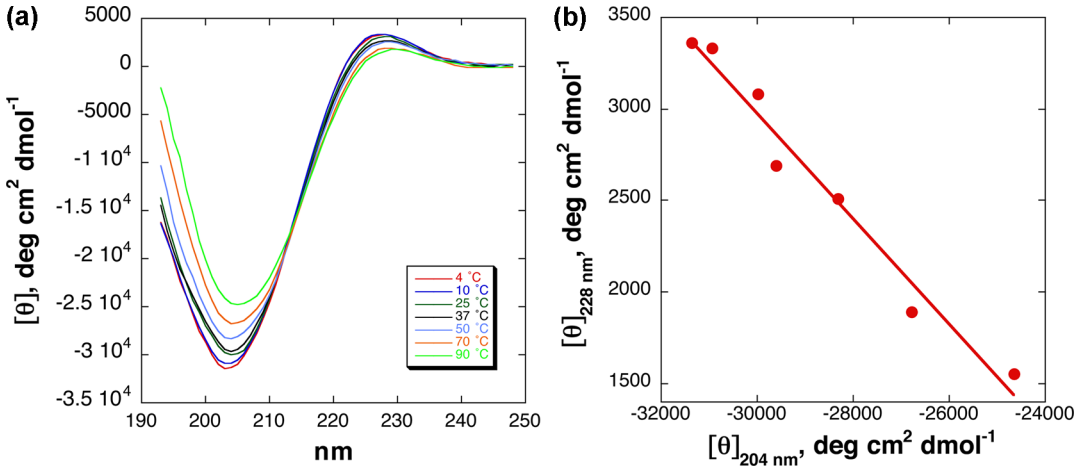



**Figure 2.** Temperature-dependent CD data for Ac-GPPPPPGY-$NH_2$. (a) Temperature-dependent CD of Ac-GPPPPPGY-$NH_2$. (b) Dual-wavelength parametric test (mean residue ellipticity at 228 and 204 nm) for Ac-GPPPPPGY-$NH_2$.

spectra were anomalous in shape, with maxima and minima that were divergent from those of canonical PPII, suggesting alternative structures versus those seen in other peptides with nonaromatic residues and/or a significant aromatic contribution to the CD spectra.[47−49] To identify the effects of aromatic residues, the peptides were analyzed by NMR. NMR spectroscopy of the Phe, Tyr, and Trp peptides revealed a very large population of species containing cis amide bonds


dx.doi.org/10.1021/bi3002924 | *Biochemistry* 2012, 51, 5041−5051

**Table 2. Frequencies of Residues in Proline-Rich Sequences of Human Proteins**[a]

| | no. (n) of proteins with {G}PPXPP{G} | $f$ | $f/f_0$ |
|---|---|---|---|
| A | 726 | 0.097 | 1.25 |
| C | 67 | 0.009 | 0.37 |
| D | 106 | 0.014 | 0.31 |
| E | 213 | 0.028 | 0.44 |
| F | 103 | 0.014 | 0.42 |
| G | 414 | 0.055 | 0.78 |
| H | 126 | 0.017 | 0.63 |
| I | 94 | 0.013 | 0.33 |
| K | 175 | 0.023 | 0.46 |
| L | 806 | 0.107 | 1.08 |
| M | 68 | 0.009 | 0.43 |
| N | 30 | 0.004 | 0.13 |
| P | 2367 | 0.315 | 4.47 |
| Q | 461 | 0.061 | 1.27 |
| R | 441 | 0.059 | 0.96 |
| S | 493 | 0.066 | 0.76 |
| T | 347 | 0.046 | 0.86 |
| V | 406 | 0.054 | 0.90 |
| W | 35 | 0.005 | 0.32 |
| Y | 40 | 0.005 | 0.23 |

[a]Data from PIR-NREF version 1.81 for *Homo sapiens* using a sequence search for {G}PPXPP{G}, where {G} indicates to exclude Gly residues at these positions. Gly was excluded to remove collagen-like sequences. $f$ is the overall frequency of proteins with the residue at position X ($n/\sum n$). $f/f_0$ is the frequency of the residue normalized for the overall frequency ($f_0$) of this residue in human proteins.[86] This analysis, because it is on a per protein basis, avoids biases that would be introduced in proteins with repeat sequences, where the repeat sequences would dominate the results.

[lower estimates of percentages of peptides with at least one cis amide bond, 45% (Phe), 55% (Tyr), and 60% (Trp)] (Figure 4). In contrast, all other peptides showed no significant population of cis amide species (<10% cis). Aromatic−proline (aromatic−cis-Pro and cis-Pro−aromatic) interactions, requiring a cis prolyl amide bond, are known to be stabilizing within a random coil context.[72−83] In addition, outside a PPII context, Pro−Pro and aromatic−Pro amide bonds are the most prone to cis amide conformation.[51,84] These results indicate that the interaction between aromatic rings and proline is particularly stabilizing within a proline-rich context. Consistent with this observation, analysis of protein sequence databases in humans, *Drosophila*, *Caenorhabditis elegans*, and *Arabidopsis* indicated low observed frequencies of aromatic residues in proline-rich regions (Tables 2 and 3).[85−87] Within the GPP(aromatic)-PPGY context, cis amide bonds are possible at four X−Pro amide bonds, with 16 possible permutations of trans and cis X−Pro amide bonds in the peptide. NMR data using Leu or trans-amide bond-favoring (2*S*,4*R*)-hydroxyproline (Hyp) substitution at single proline residues (GPLYPPGY and GPPYLPGY; GPHypYPPGY and GPPYHypPGY) suggest that cis−trans isomerization is important in both the Pro−cis-Pro−aromatic sequence and the aromatic−cis-Pro sequences (data not shown). Similar results were observed by us previously in PYPN, TYPN, TYPP, and PYPP model peptides, as well as by others in model peptides.[78,81,83,88] In the Ac-PYPP-$NH_2$ peptide,[78] 65% of the species observed by NMR exhibited at least one cis amide bond, similar to results herein, where seven of eight possible combinations of trans and cis amide bonds were observed. Interestingly, as was observed previously in aromatic−proline sequences, cis−trans isomerism in GPPXPPGY peptides correlated with aromatic electronics (order of percent cis values, Trp > Tyr > Phe), suggestive of a potentially electronically controllable CH−$\pi$ interaction.[72−74,77−80,89−92]

## DISCUSSION

The data herein indicate that, within the context of proline-rich regions, PPII is a conformation favored by most residues. The highest PPII propensities were observed for Pro, Leu, Ala, and residues with long, linear side chains (Arg, Lys, and Met). These data are broadly consistent with previously described PPII propensities, both within the context of stable polyproline helices and in random coil contexts.[40−42,44−46] Two key differences emerge from our data in a proline-rich context (and as opposed to the context of stable PPII helices, i.e., Creamer's related Ac-PPPXPPPGY-$NH_2$ context) and previous PPII propensity data. First, we observed an unusually high PPII propensity for Leu compared to other non-proline residues, whereas in other contexts, Leu has intermediate PPII propensity;[41,42,46] second, we observed, on the basis of NMR spectroscopy, particularly poor PPII propensities for aromatic amino acids, because of surprisingly high levels of the cis amide bond conformation in this context. Notably, these data constitute the first systematic analysis of NMR data for peptides for all canonical amino acids within a proline-rich context. Interestingly, Brodsky and co-workers observed very poor collagen triple-helix propensities for aromatic amino acids, substantially worse than for nonaromatic amino acids, similar to our observations herein.[40]

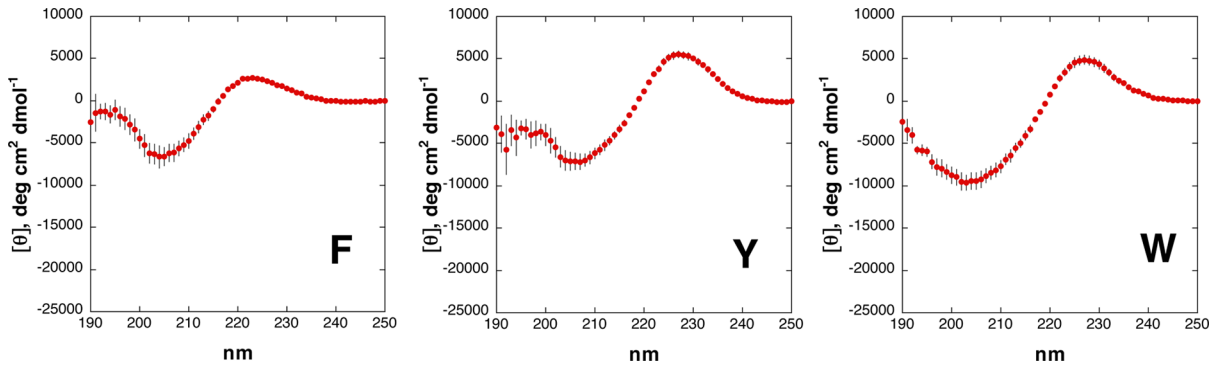



**Figure 3.** CD spectra for GPPXPPGY peptides where X is an aromatic residue.


dx.doi.org/10.1021/bi3002924 | *Biochemistry* 2012, 51, 5041−5051

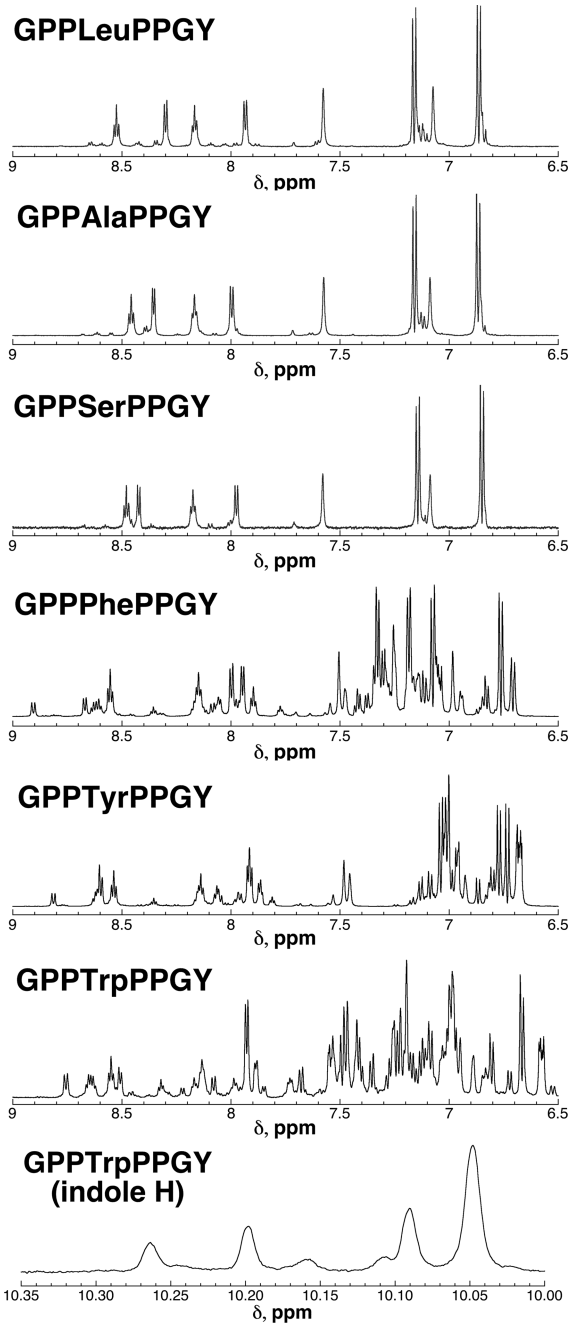



**Figure 4.** Backbone amide and aromatic regions of the NMR spectra of selected GPPXPPGY peptides. In the top panels, X = Leu, Ala, and Ser. For the X = Leu peptide, with four backbone amide protons [$\delta$ 8.52 (Gly1), 8.30 (Leu), 8.16 (Gly7), and 7.93 (Tyr8)], Tyr8 aromatic protons ($\delta$ 7.14 and 6.85), and C-terminal carboxamide ($\delta$ 7.57 and 7.07), the data indicate that the peptide predominantly contains all trans amide bonds (<10% of the species containing cis amide bonds), despite four Xaa–Pro amide bonds. Other peptides with nonaromatic guest X residues exhibited similar NMR spectra. In the middle panels, X = Phe (45% cis amide bond-containing species), Tyr (55% cis), and Trp (60% cis). The percent cis populations are lower estimates, as we were unable to fully assign all amide peaks to discrete species for X = aromatic residue peptides, which showed at least three major species in slow exchange (16 combinations of cis and trans X–Pro amide bonds are possible). In the bottom panel, for X = Trp, the indole proton indicates four major species and at least four minor species.

Compared to Creamer's data,[41,42] we also observed greater differences in PPII propensity between Pro and non-Pro residues and did not observe, even at lower temperatures, the special PPII propensity of Gln that Creamer found in his stable PPII context. Like those of Creamer, our data indicate low PPII propensities in a proline-rich context for $\beta$-branched and short polar residues, including low PPII propensities for Ser and Thr. In contrast, in a glycine-rich context for determining PPII propensities, Kallenbach and co-workers found high PPII propensities for Ala (similar to results in a proline-rich context), Ser, Trp, and Val, with low PPII propensities for Ile, Met, and His.[46] Glu, Gln, and Arg had relatively high PPII propensities in that context, similar to those in proline-rich contexts. These differences are likely due to context-dependent interactions with nearest neighbors: in X-Pro and Pro-X sequences, residues exhibit divergent Ramachandran preferences compared to X-Gly and Gly-X.[45,93]

Polyproline helix formation is driven substantially by local interactions, particularly local sterics and n $\rightarrow \pi^*$ interactions between adjacent carbonyls.[39,45,60,94−99] The high PPII propensity of leucine observed herein is noteworthy in this typical proline-rich context. These data, combined with the data of Creamer within a stable PPII context, suggest that Leu both readily adopts PPII in a proline-rich sequence and strongly promotes the adoption of PPII conformation by adjacent proline residues. The most probable reason for the high PPII propensity of Leu in a proline-rich context is the relative conformational restriction of Leu to a major mt side chain rotamer, which leads to an extended side chain that points away from the polyproline helix backbone to prevent a syn-pentane interaction [see the Protein Data Bank (PDB) analysis below].[100]

The PPII propensity data herein in general correlate modestly with statistical propensities derived from PPII in crystal structures and with experimental propensity data from the collagen triple helix.[2,3,40] In collagen triple helices, the residues with the lowest collagen triple-helical propensities are short polar residues, glycine, and aromatic residues, with Trp having the lowest collagen triple-helical propensity at either the Pro or Hyp position. In general, $\beta$-branched amino acids have higher triple-helix propensities than PPII propensities, and leucine has a significantly lower collagen than PPII propensity. Likewise, in global analysis of residues in polyproline helices of at least three residues in the PDB, $\beta$-branched amino acids are observed with high frequencies (as are R, Q, K, and L).[3] Interestingly, aromatic residues (and Gly) have very low frequencies in polyproline helices in the PDB. The modest correlation is perhaps less surprising given that those data are based on systems with considerable tertiary structure and unique steric and electronic preferences, whereas our model system, like most proline-rich domains, lacks tertiary structure elements.[94] In contrast, our data show more significant correlation with the identity of residues in analogous proline-rich sequences (PPXPP) in human, *Drosophila*, *C. elegans*, and *Arabidopsis* proteins (Tables 2 and 3), where Ala, Gln, Leu, and Arg are commonly observed and Tyr, Trp, Ile, and Asn are infrequent.

Despite their low PPII propensities, serine and threonine are frequently observed residues within proline-rich regions (Tables 2 and 3). Our data indicate that these regions likely exhibit relatively lower populations of PPII. This observation is supported by circular dichroism on the RNA Pol II C-terminal domain repeat (sequence YSPTSPS), which displays small

**Table 3. Frequencies of Proteins with Residues in PPXPP Sequences in Model Eukaryotes Based on the PIR UniProtKB Database (http://pir.georgetown.edu)[a]**

| | human *n* | human $f/f_0$ | *Drosophila n* | *Drosophila* $f/f_0$ | *C. elegans n* | *C. elegans* $f/f_0$ | *Arabidopsis n* | *Arabidopsis* $f/f_0$ |
|---|---|---|---|---|---|---|---|---|
| A | 675 | 1.11 | 242 | 1.08 | 96 | 1.23 | 108 | 0.80 |
| C | 199 | 1.04 | 19 | 0.34 | 6 | 0.23 | 8 | 0.22 |
| D | 92 | 0.26 | 14 | 0.09 | 5 | 0.08 | 18 | 0.16 |
| E | 252 | 0.50 | 74 | 0.39 | 13 | 0.16 | 24 | 0.17 |
| F | 128 | 0.50 | 5 | 0.05 | 4 | 0.06 | 27 | 0.31 |
| G | 728 | 1.31 | 119 | 0.64 | 161 | 2.42 | 56 | 0.41 |
| H | 123 | 0.59 | 37 | 0.46 | 13 | 0.45 | 21 | 0.36 |
| I | 83 | 0.28 | 87 | 0.60 | 36 | 0.46 | 19 | 0.18 |
| K | 162 | 0.41 | 85 | 0.51 | 38 | 0.47 | 48 | 0.37 |
| L | 709 | 0.91 | 294 | 1.07 | 100 | 0.91 | 205 | 1.06 |
| M | 112 | 0.67 | 65 | 0.91 | 29 | 0.89 | 23 | 0.45 |
| N | 50 | 0.20 | 37 | 0.26 | 18 | 0.29 | 31 | 0.35 |
| P | 2288 | 4.12 | 1143 | 6.97 | 475 | 8.21 | 904 | 8.98 |
| Q | 481 | 1.26 | 140 | 0.90 | 36 | 0.70 | 76 | 1.06 |
| R | 483 | 1.00 | 143 | 0.86 | 59 | 0.92 | 81 | 0.73 |
| S | 458 | 0.68 | 178 | 0.95 | 54 | 0.72 | 289 | 1.57 |
| T | 344 | 0.81 | 167 | 1.00 | 50 | 0.69 | 58 | 0.55 |
| V | 398 | 0.84 | 127 | 0.72 | 53 | 0.68 | 35 | 0.25 |
| W | 87[b] | 0.76[b] | 0 | 0.00 | 1 | 0.07 | 4 | 0.16 |
| Y | 33 | 0.18 | 6 | 0.07 | 10 | 0.25 | 29 | 0.50 |
| total | 7885 | | 2982 | | 1257 | | 2064 | |

[a]Peptide match search of PPXPP for the organisms indicated. *f* is the overall frequency of proteins with the residue at position X ($n/\sum n$). $f/f_0$ is the frequency of the residue normalized for the overall frequency ($f_0$) of this residue in the organism indicated.[86] In this table, all PPXPP sequences were included, resulting in an increased level of representation of collagen repeat sequences (ProHypGly repeats, which at the gene level are ProProGly repeats) compared to those of Table 2. [b]Although the UniProtKB database is a nonredundant sequence database, the human Trp (PPWPP) data include 32 almost fully identical copies of extraembryonic spermatogenesis homeobox-1 (ESX1) protein (sequence of RMAPVPPWPP-MAPVPPWPPMAPV, which interestingly has two PPWPP sequences), 15 Bcl-Abl protein isoforms (identical repeat of NGGGSRPPWPPLEYQP), and eight titin proteins (sequence of KDPIDPPWPPGKPTV), anomalies that we did not see in other residues or organisms. They are included here because we did not analyze all sequences in all organisms individually to identify all duplicate sequences but are a likely explanation for the differences between Trp data for humans in this search and those of Trp in other organisms, in the Protein Data Bank, or in the PIR-NREF search (Table 2). Notably, the residue frequency numbers for mice are generally similar to those for humans (A 617, C 45, D 97, E 172, F 61, G 539, H 87, I 83, K 145, L 644, M 56, N 31, P 1831, Q 337, R 311, S 432, T 261, V 270, W 20, and Y 31) other than the substantially lower frequencies for Trp (and Cys), consistent with data for other organisms and in PIR-NREF of low Trp frequencies in PPXPP sequences.

populations of PPII in the absence of the Pin1 WW domain.[101] Notably, while Ser and Thr are relatively favored in proline-rich sequences, the acidic residues Asp and Glu, both of which strongly favor PPII, are highly disfavored in proline-rich sequences, in contrast to most PPII-favoring residues. It is noteworthy in this context that Ser/Thr residues may be phosphorylated within proline-rich contexts, and that phosphorylated Ser/Thr residues induce a polyproline helix conformation, both in solution and in observed protein−protein interactions, where PPII appears to be the preferred conformation for binding ligands containing phosphoserine and phosphothreonine residues.[38,64,102−104] The observation of reduced Asp/Glu content in proline-rich sequences is consistent with the concept that in eukaryotes phosphorylation modulates protein structure and/or function in proline-rich domains, with selection against Asp/Glu residues serving to maximize the impact of phosphorylation.

The correlation of PPII propensity with sequence abundance was also observed within the context of complete analyses of proteomes. Pe'er et al. analyzed the frequencies of all possible dipeptides and tripeptides in *Homo sapiens*, *Escherichia coli*, and *Pyrococcus furiosus*, and in broader comparisons among eukaryotes, eubacteria, and archea, to identify overabundant and underabundant peptide sequences, normalized for the frequency of use of the residues.[105] They observed that in humans, PP sequences are the fourth most overabundant dipeptides of 400 possible sequences. In humans and across all eukaryotes, PPP, PLP, and PAP are all among the 10 most overabundant tripeptides among the 8000 possible sequences. Interestingly, DPP sequences were the most underabundant tripeptide sequence of all in humans and the fifth most underabundant across eukaryotes. Across all eukaryotes, PP dipeptides were the fourth most overabundant, while EP dipeptides were the second most underabundant. These data indicate a strong selection for PP, PPP, PLP, and PAP sequences, and a selection against DP, EP, and DPP sequences, in proline-rich sequences in eukaryotes.

Similarly, analysis of the PDB reveals that sequences containing PPXPP are relatively common for some residues where X strongly favors PPII (when X = P, 118 structures; when X = R, 66 structures; when X = A, 48 structures) or is the Gly normally present in ProHypGly collagen repeat sequences (82 structures).[106] In contrast, there are substantially fewer examples of PPXPP sequences in structures with residues that disfavor PPII or are acidic (3 structures with D, 17 structures with E, 4 structures with N, 12 structures with S, and 8 structures with T; for aromatic, see below; for 18 structures with I, 15 with K, 22 with Q, and 25 with V frequencies are intermediate). This correlation, while imperfect and subject to the statistical anomalies of small numbers as well as sampling bias in structure determination, exists despite a strong bias


dx.doi.org/10.1021/bi3002924 | *Biochemistry* 2012, 51, 5041−5051

against proline-rich and intrinsically disordered sequences in the PDB.

Of the 39 examples of PPLPP sequences in the PDB, in all except one (entry 2J9L), the sequence is at least in part in a PPII conformation, and the Leu side chain has side chain torsion angles of approximately −60° ($\chi_1$) and 180° ($\chi_2$) (mt rotamer) in 32 of 33 crystal structures of PPLPP (Figure 5a).[107,108] Again subject to the caveat of the limited number of

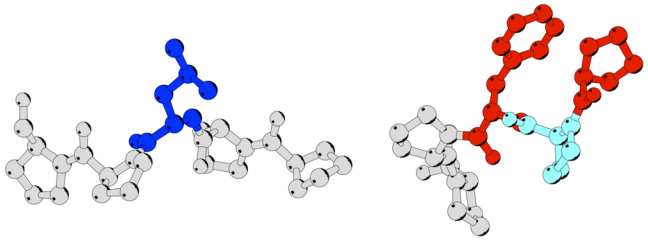


**Figure 5.** X-ray crystal structures of (left) the PPLPP sequence in a PPII conformation, from PDB entry 2ZB3[108] [for Leu (blue), $\phi$ = −55°, $\psi$ = 153°, $\chi_1$ = −63°, and $\chi_2$ = −172°] (2.00 Å resolution) residues 267−271, and (right) the PPFPP sequence with a Pro−cis-Pro−Phe conformation, from PDB entry 2CUL[109] (1.65 Å resolution) residues 48−52. Pro48 and Phe50, which exhibit an $i-2/i$ prolyl−aromatic interaction with Pro48 $H_\alpha$ interacting with the phenyl ring, are colored red. The cis amide Pro49 is colored cyan.[107]

these structures in the PDB, these data strongly suggest that Leu in a PPII conformation strongly prefers the mt rotamer, which is the more populated of the two major Leu rotamers. These data are consistent with data for $\alpha$-helices, where side chain conformational preferences consistent with secondary structure cause Leu to have a high secondary structure ($\alpha$-helix or PPII) propensity.[66] It is also noteworthy that, as in the case with $\alpha$-helices, residues with long side chains (Leu, Met, Arg, and Lys) that have conformational preferences for avoiding syn−pentane interactions strongly favor PPII.[61]

There are only six proteins in the PDB with PPXPP (X = Phe, Tyr, or Trp) sequences, consistent with overall sequence data suggesting selection against aromatic residues in proline-rich sequences (Tables 2 and 3). No protein in the PDB contains a PPWPP sequence. Three proteins (2V1C, 1W3S, and 1U5K) in the PDB contain PPYPP sequences; in each of these proteins, the sequence adopts a PPII conformation with all-trans amide bonds. In contrast, in two of the proteins in the PDB with PPFPP sequences (2L3J and 2YVS), the Phe residue is in an extended conformaiton. In the other PPFPP sequence (PDB entry 2CUL), the structure contains a Pro−cis-Pro−Phe cis amide bond, which was observed to be a major conformation in similar sequences herein.[109] Interestingly, in this structure, the Phe aromatic ring interacts with the Pro two residues prior in the sequence, in contrast to the aromatic interaction with the immediately preceding Pro in turn structures.[69] This $i-2/i$ proline−aromatic interaction, in addition to the more appreciated $i-1/i$ proline−aromatic interaction and $i/i+1$ aromatic−proline interactions that promote cis amide bonds, might explain the particularly high population of cis amide bonds for aromatic residues in this proline-rich context. A similar preference for cis-Pro−aromatic interactions was also observed in Ala-Pro-Tyr and Pro-Pro-Phe sequences.[75,81,88,110]

The very high percentage of cis amide bonds for peptides with aromatic residues in this proline-rich context is noteworthy as it is particularly high for a short disordered peptide. The observation of large populations of cis amide bonds when combining Tyr and Trp with proline-rich sequences also has implications for the application of the polyproline ruler to these sequences. Stryer and Haugland's initial spectroscopic ruler experiments involved formation of amide bonds between dye molecules and the N- and C-termini of polyproline oligomers.[35] Several investigators who have used the polyproline ruler with Tyr and Trp residues at the termini (as opposed to directly conjugated dyes) have observed anomalous FRET behavior, and some authors have suggested that the polyproline ruler is not reliable on the basis of those FRET data.[83,111−114] Indeed, when investigated by NMR, these Pro-rich peptides, especially with smaller numbers of Pro residues, contain significant populations of cis amide bonds. Our data suggest that aromatic−proline and proline−aromatic interactions are particularly favorable for the promotion of cis amide bonds in a proline-rich context and suggest these interactions form the basis for anomalous FRET behavior when the fluorophore is associated with an aromatic amino acid side chain, where these interactions are favorable at the N- and C-termini.

Proline-rich domains are employed biologically as loci for protein−protein interactions.[4,9,10,115] Separately, aromatic residues are commonly central recognition residues in ligands at protein−protein interfaces, because of their large hydrophobic surface area and the possibility of specific interactions with the aromatic ring.[116−118] On the basis of the significant independent roles of both proline-rich domains and aromatic rings in recognition, it is naively surprising that aromatic residues are uncommon in proline-rich domains and are not commonly employed for protein binding in these contexts. The data herein provide a basis for the nonincorporation of aromatic residues in proline-rich domains: the substantial conformational heterogeneity that would result due to multiple cis−trans isomerization events in slow exchange. The slow exchange is relevant because presumably the time scale of cis−trans isomerism is normally incompatible with induced fit and additionally could potentially form a locus for protein misfolding or nonspecific interactions.

We have presented a complete experimental set of amino acid propensities for the PPII conformation within a proline-rich context, using CD and NMR to characterize the extent of polyproline helix within this proline-rich context. We expect these data to be useful in understanding the roles of the PPII conformation in protein structure and proline-rich domains and in the design of novel ligands targeting protein−protein interfaces involving polyproline helix recognition elements.[119]

## ■ ASSOCIATED CONTENT

### Supporting Information

Characterization data and CD data for all peptides. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION


### Corresponding Author

*E-mail: zondlo@udel.edu. Phone: (302) 831-0197.



### Funding

This work was supported by funds from the University of Delaware, the American Heart Association, the Howard Hughes Medical Institute Undergraduate Science Education Program (fellowship support to A.M.B.), and the National Science Foundation (CAREER, CHE-0547973).

## Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank Anil Pandey for additional NMR analysis of peptides containing aromatic guest residues. We thank Steve Brown for helpful discussions about data analysis.

## ■ REFERENCES

(1) Adzhubei, A. A., and Sternberg, M. J. E. (1993) Left-Handed Polyproline-II Helices Commonly Occur in Globular-Proteins. *J. Mol. Biol. 229*, 472−493.

(2) Stapley, B. J., and Creamer, T. P. (1999) A survey of left-handed polyproline II helices. *Protein Sci. 8*, 587−595.

(3) Cubellis, M. V., Caillez, F., Blundell, T. L., and Lovell, S. C. (2005) Properties of polyproline II, a secondary structure element implicated in protein-protein interactions. *Proteins 58*, 880−892.

(4) Zarrinpar, A., Bhattacharyya, R. P., and Lim, W. A. (2003) The Structure and Function of Proline Recognition Domains. *Sci. STKE 2003*, re8.

(5) Rubin, G. M., Yandell, M. D., Wortman, J. R., Miklos, G. L. G., Nelson, C. R., Hariharan, I. K., Fortini, M. E., Li, P. W., Apweiler, R., Fleischmann, W., Cherry, J. M., Henikoff, S., Skupski, M. P., Misra, S., Ashburner, M., Birney, E., Boguski, M. S., Brody, T., Brokstein, P., Celniker, S. E., Chervitz, S. A., Coates, D., Cravchik, A., Gabrielian, A., Galle, R. F., Gelbart, W. M., George, R. A., Goldstein, L. S. B., Gong, F. C., Guan, P., Harris, N. L., Hay, B. A., Hoskins, R. A., Li, J. Y., Li, Z. Y., Hynes, R. O., Jones, S. J. M., Kuehl, P. M., Lemaitre, B., Littleton, J. T., Morrison, D. K., Mungall, C., O'Farrell, P. H., Pickeral, O. K., Shue, C., Vosshall, L. B., Zhang, J., Zhao, Q., Zheng, X. Q. H., Zhong, F., Zhong, W. Y., Gibbs, R., Venter, J. C., Adams, M. D., and Lewis, S. (2000) Comparative Genomics of the Eukaryotes. *Science 287*, 2204−2215.

(6) Wright, P. E., and Dyson, H. J. (1999) Intrinsically Unstructured Proteins: Re-assessing the Protein Structure-Function Paradigm. *J. Mol. Biol. 293*, 321−331.

(7) Dunker, A. K., Brown, C. J., and Obradovic, Z. (2002) Identification and functions of usefully disordered proteins. *Adv. Protein Chem. 62*, 25−49.

(8) Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovic, Z. (2002) Intrinsic disorder and protein function. *Biochemistry 41*, 6573−6582.

(9) Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2005) Showing your ID: Intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit. 18*, 343−384.

(10) Rath, A., Davidson, A. R., and Deber, C. M. (2005) The structure of "unstructured" regions in peptides and proteins: Role of the polyproline II helix in protein folding and recognition. *Biopolymers 80*, 179−185.

(11) Shi, Z. S., Woody, R. W., and Kallenbach, N. R. (2002) Is polyproline II a major backbone conformation in unfolded proteins? *Adv. Protein Chem. 62*, 163−240.

(12) Mezei, M., Fleming, P. J., Srinivasan, R., and Rose, G. D. (2004) Polyproline II helix is the preferred conformation for unfolded polyalanine in water. *Proteins 55*, 502−507.

(13) Grdadolnik, J., Mohacek-Grosev, V., Baldwin, R. L., and Avbelj, F. (2011) Populations of the three major backbone conformations in 19 amino acid dipeptides. *Proc. Natl. Acad. Sci. U.S.A. 108*, 1794−1798.

(14) Shi, Z. S., Olson, C. A., Rose, G. D., Baldwin, R. L., and Kallenbach, N. R. (2002) Polyproline II structure in a sequence of seven alanine residues. *Proc. Natl. Acad. Sci. U.S.A. 99*, 9190−9195.

(15) Ding, L., Chen, K., Santini, P. A., Shi, Z., and Kallenbach, N. R. (2003) The Pentapeptide GGAGG Has PII Conformation. *J. Am. Chem. Soc. 125*, 8092−8093.

(16) Chen, K., Liu, Z. G., and Kallenbach, N. R. (2004) The polyproline II conformation in short alanine peptides is non-cooperative. *Proc. Natl. Acad. Sci. U.S.A. 101*, 15352−15357.

(17) Ramakrishnan, V., Ranbhor, R., and Durani, S. (2004) Existence of specific "folds" in polyproline II ensembles of an "unfolded" alanine peptide detected by molecular dynamics. *J. Am. Chem. Soc. 126*, 16332−16333.

(18) Gokce, I., Woody, R. W., Anderluh, G., and Lakey, J. H. (2005) Single peptide bonds exhibit poly(pro)II ("random coil") circular dichroism spectra. *J. Am. Chem. Soc. 127*, 9700−9701.

(19) Eker, F., Griebenow, K., Cao, X. L., Nafie, L. A., and Schweitzer-Stenner, R. (2004) 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. U.S.A. 101*, 10054−10059.

(20) Hagarman, A., Measey, T. J., Mathieu, D., Schwalbe, H., and Schweitzer-Stenner, R. (2010) 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.

(21) Neidigh, J. W., Fesinmeyer, R. M., and Andersen, N. H. (2002) Designing a 20-residue protein. *Nat. Struct. Biol. 9*, 425−430.

(22) Zondlo, N. J., and Schepartz, A. (1999) Highly specific DNA recognition by a designed miniature protein. *J. Am. Chem. Soc. 121*, 6938−6939.

(23) Arora, P. S., Ansari, A. Z., Best, T. P., Ptashne, M., and Dervan, P. B. (2002) Design of Artificial Transcriptional Activators with Rigid Poly-L-proline Linkers. *J. Am. Chem. Soc. 124*, 13067−13071.

(24) Tremmel, P., and Geyer, A. (2002) An Oligomeric Ser-Pro Dipeptide Mimetic Assuming the Polyproline II Helix Conformation. *J. Am. Chem. Soc. 124*, 8548−8549.

(25) Golemi-Kotra, D., Mahaffty, R., Footer, M. J., Holtzman, J. H., Pollard, T. D., Theriot, J. A., and Schepartz, A. (2004) High Affinity, Paralog-Specific Recognition of the Mena EVH1 Domain by a Miniature Protein. *J. Am. Chem. Soc. 126*, 4−5.

(26) Daniels, D. S., and Schepartz, A. (2007) Intrinsically cell-permeable miniature proteins based on a minimal cationic PPII motif. *J. Am. Chem. Soc. 129*, 14578−14579.

(27) Schimmel, P. R., and Flory, P. J. (1967) Conformational Energy and Configurational Statistics of Poly-L-Proline. *Proc. Natl. Acad. Sci. U.S.A. 58*, 52−59.

(28) Schimmel, P. R., and Flory, P. J. (1968) Confromational energies and configurational statistics of copolypeptides containing L-proline. *J. Mol. Biol. 34*, 105−120.

(29) Crespo, L., Sanclimens, G., Montaner, B., Pérez-Tomás, R., Royo, M., Pons, M., Albericio, F., and Giralt, E. (2002) Peptide Dendrimers Based on Polyproline Helices. *J. Am. Chem. Soc. 124*, 8876−8883.

(30) Schuler, B., Lipman, E. A., Steinbach, P. J., Kumke, M., and Eaton, W. A. (2005) Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. *Proc. Natl. Acad. Sci. U.S.A. 102*, 2754−2759.

(31) Fillon, Y. A., Anderson, J. P., and Chmielewski, J. (2005) Cell penetrating agents based on a polyproline helix scaffold. *J. Am. Chem. Soc. 127*, 11798−11803.

(32) Kumin, M., Sonntag, L. S., and Wennemers, H. (2007) Azidoproline containing helices: Stabilization of the polyproline II structure by a functionalizable group. *J. Am. Chem. Soc. 129*, 466−467.

(33) Erdmann, R. S., and Wennemers, H. (2010) Functionalizable Collagen Model Peptides. *J. Am. Chem. Soc. 132*, 13957−13959.

(34) Owens, N. W., Stetefeld, J., Lattova, E., and Schweizer, F. (2010) Contiguous O-Galactosylation of 4(R)-Hydroxy-L-proline Residues Forms Very Stable Polyproline II Helices. *J. Am. Chem. Soc. 132*, 5036−5042.

(35) Stryer, L., and Haugland, R. P. (1967) Energy transfer: A spectroscopic ruler. *Proc. Natl. Acad. Sci. U.S.A. 58*, 719−726.

(36) Sahoo, H., Roccatano, D., Hennig, A., and Nau, W. M. (2007) A 10-Å Spectroscopic Ruler Applied to Short Polyprolines. *J. Am. Chem. Soc. 129*, 9762−9772.

(37) Best, R. B., Merchant, K. A., Gopich, I. V., Schuler, B., Bax, A., and Eaton, W. A. (2007) Effect of flexibility and cis residues in single-molecule FRET studies of polyproline. *Proc. Natl. Acad. Sci. U.S.A. 104*, 18964−18969.

(38) Durocher, D., Taylor, I. A., Sarbassova, D., Haire, L. F., Westcott, S. L., Jackson, S. P., Smerdon, S. J., and Yaffe, M. B. (2000)

The molecular basis of FHA domain: Phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms. *Mol. Cell 6*, 1169−1182.
(39) Creamer, T. P. (1998) Left-handed polyproline II helix formation is (very) locally driven. *Proteins 33*, 218−226.
(40) Persikov, A. V., Ramshaw, J. A. M., Kirkpatrick, A., and Brodsky, B. (2000) Amino acid propensities for the collagen triple-helix. *Biochemistry 39*, 14960−14967.
(41) Kelly, M. A., Chellgren, B. W., Rucker, A. L., Troutman, J. M., Fried, M. G., Miller, A. F., and Creamer, T. P. (2001) Host-guest study of left-handed polyproline II helix formation. *Biochemistry 40*, 14376−14383.
(42) Rucker, A. L., Pager, C. T., Campbell, M. N., Qualls, J. E., and Creamer, T. P. (2003) Host-Guest Scale of Left-Handed Polyproline II Helix Formation. *Proteins 53*, 68−75.
(43) Vila, J. A., Baldoni, H. A., Ripolli, D. R., Ghosh, A., and Scheraga, H. A. (2004) Polyproline II Helix Conformation in a Proline-Rich Environment: A Theoretical Study. *Biophys. J. 86*, 731−742.
(44) Jha, A. K., Colubri, A., Zaman, M. H., Koide, S., Sosnick, T. R., and Freed, K. F. (2005) Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library. *Biochemistry 44*, 9691−9702.
(45) Chen, K., Liu, Z. G., Zhou, C. H., Shi, Z. S., and Kallenbach, N. R. (2005) Neighbor effect on PPII conformation in alanine peptides. *J. Am. Chem. Soc. 127*, 10146−10147.
(46) Shi, Z., Chen, K., Liu, Z., Ng, A., Bracken, W. C., and Kallenbach, N. R. (2005) Polyproline II propensities from GGXGG peptides reveal an anticorrelation with $\beta$-sheet scales. *Proc. Natl. Acad. Sci. U.S.A. 102*, 17964−17968.
(47) Woody, R. W. (1994) Contributions of Tryptophan Side-chains to the Far-Ultraviolet Circular-Dichroism of Proteins. *Eur. Biophys. J. 23*, 253−262.
(48) Sreerama, N., Manning, M. C., Powers, M. E., Zhang, J. X., Goldenberg, D. P., and Woody, R. W. (1999) Tyrosine, phenylalanine, and disulfide contributions to the circular dichroism of proteins: Circular dichroism spectra of wild-type and mutant bovine pancreatic trypsin inhibitor. *Biochemistry 38*, 10814−10822.
(49) Manning, M. C., and Woody, R. W. (1989) Theoretical study of the contribution of aromatic side chains to the circular dichroism of basic bovine pancreatic trypsin inhibitor. *Biochemistry 28*, 8609−8613.
(50) Helbecque, N., and Loucheux-Lefebvre, M. H. (1982) Critical chain length for polyproline-II structure formation in H-Gly-(Pro)n-OH. *Int. J. Pept. Protein Res. 19*, 94−101.
(51) Saha, I., and Shamala, N. (2012) Investigating Diproline Segments in Proteins: Occurrences, Conformation and Classification. *Biopolymers 97*, 54−64.
(52) Sreerama, N., and Woody, R. W. (1994) Poly(Pro)II Helices in Globular Proteins: Identification and Circular Dichroic Anaysis. *Biochemistry 33*, 10022−10025.
(53) Woody, R. W. (2009) Circular Dichroism Spectrum of Peptides in the Poly(Pro)II Conformation. *J. Am. Chem. Soc. 131*, 8234−8245.
(54) Whittington, S. J., and Creamer, T. P. (2003) Salt bridges do not stabilize polyproline II helices. *Biochemistry 42*, 14690−14695.
(55) Chellgren, B. W., and Creamer, T. P. (2004) Effects of $H_2O$ and $D_2O$ on polyproline II helical structure. *J. Am. Chem. Soc. 126*, 14734−14735.
(56) Whittington, S. J., Chellgren, B. W., Hermann, V. M., and Creamer, T. P. (2005) Urea promotes polyproline II helix formation: Implications for protein denatured states. *Biochemistry 44*, 6269−6275.
(57) Wallimann, P., Kennedy, R. J., Miller, J. S., Shalongo, W., and Kemp, D. S. (2003) Dual Wavelength Parametric Test of Two-State Models for Circular Dichroism Spectra of Helical Peptides: Anomalous Dichroic Properties of Alanine-Rich Peptides. *J. Am. Chem. Soc. 125*, 1203−1220.
(58) Vuister, G. W., and Bax, A. (1993) 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.
(59) Manning, M. C., and Woody, R. W. (1991) Theoretical CD Studies of Polypeptide Helices: Examination of Important Electronic and Geometric Factors. *Biopolymers 31*, 569−586.
(60) Creamer, T. P., and Campbell, M. N. (2002) Determinants of the polyproline II helix from modeling studies. *Adv. Protein Chem. 62*, 263−282.
(61) Lyu, P. C., Sherman, J. C., Chen, A., and Kallenbach, N. R. (1991) $\alpha$-Helix Stabilization by Natural and Unnatural Amino-Acids with Alkyl Side-Chains. *Proc. Natl. Acad. Sci. U.S.A. 88*, 5317−5320.
(62) Cornish, V. W., Kaplan, M. I., Veenstra, D. L., Kollman, P. A., and Schultz, P. G. (1994) Stabilizing and Destabilizing Effects of Placing $\beta$-Branched Amino Acids in Protein $\alpha$-Helices. *Biochemistry 33*, 12022−12031.
(63) Mendel, D., Ellman, J. A., Chang, Z., Veenstra, D. L., Kollman, P. A., and Schultz, P. G. (1992) Probing Protein Stability with Unnatural Amino Acids. *Science 256*, 1798−1802.
(64) Bielska, A. A., and Zondlo, N. J. (2006) Hyperphosphorylation of tau induces local polyproline II helix. *Biochemistry 45*, 5527−5537.
(65) Hagarman, A., Mathieu, D., Toal, S., Measey, T. J., Schwalbe, H., and Schweitzer-Stenner, R. (2011) 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.
(66) Creamer, T. P., and Rose, G. D. (1992) Side-Chain Entropy Opposes $\alpha$-Helix Formation but Rationalizes Experimentally Determined Helix-Forming Propensities. *Proc. Natl. Acad. Sci. U.S.A. 89*, 5937−5941.
(67) Chellgren, B. W., and Creamer, T. P. (2006) Side-Chain Entropy Effects on Protein Secondary Structure Formation. *Proteins 62*, 411−420.
(68) Vijayakumar, M., Qian, H., and Zhou, H. X. (1999) Hydrogen bonds between short polar side chains and peptide backbone: Prevalence in proteins and effects on helix-forming propensities. *Proteins 34*, 497−507.
(69) Eswar, N., and Ramakrishnan, C. (2000) Deterministic features of side-chain main-chain hydrogen bonds in globular protein structures. *Protein Eng. 13*, 227−238.
(70) Gronert, S., and O'Hair, R. A. J. (1995) Ab Initio Studies of Amino Acid Conformations. 1. The Conformers of Alanine, Serine, and Cysteine. *J. Am. Chem. Soc. 117*, 2071−2081.
(71) Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z., and Dunker, A. K. (2004) The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res. 32*, 1037−1049.
(72) Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A., and Wright, P. E. (1988) Folding of Immunogenic Peptide-Fragments of Proteins in Water Solution. 1. Sequence Requirements for the Formation of a Reverse Turn. *J. Mol. Biol. 201*, 161−200.
(73) Yao, J., Dyson, H. J., and Wright, P. E. (1994) 3-Dimensional Structure of a Type-VI turn in a Linear Peptide in Water Solution: Evidence for Stacking of Aromatic Rings as a Major Stabilizing Factor. *J. Mol. Biol. 243*, 754−766.
(74) Yao, J., Feher, V. A., Espejo, B. F., Reymond, M. T., Wright, P. E., and Dyson, H. J. (1994) Stabilization of a Type-VI Turn in a Family of Linear Peptides in Water Solution. *J. Mol. Biol. 243*, 736−753.
(75) Kemmink, J., and Creighton, T. E. (1995) The Physical Properties of Local Interactions of Tyrosine Residues in Peptides and Unfolded Proteins. *J. Mol. Biol. 245*, 251−260.
(76) Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M., and Fischer, G. (1998) Side-chain effects on peptidyl-prolyl cis/trans isomerization. *J. Mol. Biol. 279*, 449−460.
(77) Wu, W.-J., and Raleigh, D. P. (1998) Local Control of Peptide Conformation: Stabilization of cis Proline Peptide Bonds by Aromatic Proline Interactions. *Biopolymers 45*, 381−394.
(78) Meng, H. Y., Thomas, K. M., Lee, A. E., and Zondlo, N. J. (2006) Effects of *i* and *i*+3 Residue Identity on Cis-Trans Isomerism of the Aromatic$_{i+1}$-Prolyl$_{i+2}$ Amide Bond: Implications for Type VI $\beta$-turn Formation. *Biopolymers 84*, 192−204.

(79) Thomas, K. M., Naduthambi, D., and Zondlo, N. J. (2006) Electronic control of amide cis-trans isomerism via the aromatic-prolyl interaction. *J. Am. Chem. Soc. 128*, 2216−2217.
(80) Forbes, C. R., and Zondlo, N. J. (2012) Synthesis of Thiophenylalanine-Containing Peptides via Cu(I)-mediated Cross-Coupling. *Org. Lett. 14*, 464−467.
(81) Nardi, F., Kemmink, J., Sattler, M., and Wade, R. C. (2000) The cisproline(i-1)-aromatic(i) interaction: Folding of the Ala-cisPro-Tyr peptide characterized by NMR and theoretical approaches. *J. Biomol. NMR 17*, 63−77.
(82) Toth, G., Murphy, R. F., and Lovas, S. (2001) Stabilization of local structures by $\pi$-CH and aromatic-backbone amide interactions involving prolyl and aromatic residues. *Protein Eng. 14*, 543−547.
(83) Poznanski, J., Ejchart, A., Wierzchowski, K. L., and Ciurak, M. (1993) $^{1}$H- and $^{13}$C-NMR investigations on cis-trans isomerization of proline peptide bonds and conformation of aromatic side chains in H-Trp-(Pro)$_n$-Tyr-OH peptides. *Biopolymers 33*, 781−795.
(84) Stewart, D. E., Sarkar, A., and Wampler, J. E. (1990) Occurrence and role of cis peptide-bonds in protein structures. *J. Mol. Biol. 214*, 253−260.
(85) Wu, C. H., Huang, H., Arminski, L., Castro-Alvear, J., Chen, Y., Hu, Z.-Z., Ledley, R. S., Lewis, K. C., Mewes, H.-W., Orcutt, B. C., Suzek, B. E., Tsugita, A., Vinayaka, C. R., Yeh, L.-S. L., Zhang, J., and Barker, W. C. (2002) The Protein Information Resource: An integrated public resource of functional annotation of proteins. *Nucleic Acids Res. 30*, 35−37.
(86) Echols, N., Harrison, P., Balasubramanian, S., Luscombe, N. M., Bertone, P., Zhang, Z., and Gerstein, M. (2002) Comprehensive analysis of amino acid and nucleotide composition of eukaryotic genomes, comparing genes and pseudogenes. *Nucleic Acids Res. 30*, 2515−2523.
(87) http://pir.georgetown.edu/.
(88) Dasgupta, B., Chakrabarti, P., and Basu, G. (2007) Enhanced stability of cis Pro-Pro peptide bond in Pro-Pro-Phe sequence motif. *FEBS Lett. 581*, 4529−4532.
(89) Wright, P. E., Dyson, H. J., and Lerner, R. A. (1988) Conformation of Peptide-Fragments of Proteins in Aqueous-Solution: Implications for Initiation of Protein Folding. *Biochemistry 27*, 7167−7175.
(90) Pal, D., and Chakrabarti, P. (1999) Cis Peptide Bonds in Proteins: Residues Involved, Their Conformations, Interactions and Locations. *J. Mol. Biol. 294*, 271−288.
(91) Halab, L., and Lubell, W. D. (2002) Effect of sequence on peptide geometry in 5-tert-butylprolyl type VI $\beta$-turn mimics. *J. Am. Chem. Soc. 124*, 2474−2484.
(92) Thomas, K. M., Naduthambi, D., Tririya, G., and Zondlo, N. J. (2005) Proline Editing: A Divergent Strategy for the Synthesis of Conformationally Diverse Peptides. *Org. Lett. 7*, 2397−2400.
(93) Ting, D., Wang, G., Shapovalov, M., Mitra, R., Jordan, M. I., and Dunbrack, R. L. (2010) Neighbor-Dependent Ramachandran Probability Distributions of Amino Acids Developed from a Hierarchical Dirichlet Process Model. *PLoS Comput. Biol. 6*, No. e1000763.
(94) Bretscher, L. E., Jenkins, C. L., Taylor, K. M., DeRider, M. L., and Raines, R. T. (2001) Conformational Stability of Collagen Relies on a Stereoelectronic Effect. *J. Am. Chem. Soc. 123*, 777−778.
(95) Hinderaker, M. P., and Raines, R. T. (2003) An electronic effect on protein structure. *Protein Sci. 12*, 1188−1194.
(96) Hodges, J. A., and Raines, R. T. (2006) Energetics of an n $\rightarrow \pi^*$ interaction that impacts protein structure. *Org. Lett. 8*, 4695−4697.
(97) Horng, J. C., and Raines, R. T. (2006) Stereoelectronic effects on polyproline conformation. *Protein Sci. 15*, 74−83.
(98) Bartlett, G. J., Choudhary, A., Raines, R. T., and Woolfson, D. N. (2010) n $\rightarrow \pi^*$ interactions in proteins. *Nat. Chem. Biol. 6*, 615−620.
(99) Zondlo, N. J. (2010) Fold Globally, Bond Locally. *Nat. Chem. Biol. 6*, 567−568.
(100) Lovell, S. C., Word, J. M., Richardson, J. S., and Richardson, D. C. (2000) The Penultimate Rotamer Library. *Proteins 40*, 389−408.
(101) Bienkiewicz, E. A., Moon Woody, A.-Y., and Woody, R. W. (2000) Conformation of the RNA Polymerase II C-Terminal Domain: Circular Dichroism of Long and Short Fragments. *J. Mol. Biol. 297*, 119−133.
(102) Verdecia, M. A., Bowman, M. E., Lu, K. P., Hunter, T., and Noel, J. P. (2000) Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nat. Struct. Biol. 7*, 639−643.
(103) Yaffe, M. B., and Eila, A. E. H. (2001) Phosphoserine/threonine-binding domains. *Curr. Opin. Cell Biol. 13*, 131−138.
(104) Manke, I. A., Lowery, D. M., Nguyen, A., and Yaffe, M. B. (2003) BRCT repeats as phosphopeptide-binding modules involved in protein targeting. *Science 302*, 636−639.
(105) Pe'er, I., Felder, C. E., Man, O., Silman, I., Sussman, J. L., and Beckmann, J. S. (2004) Proteomic Signatures: Amino Acid and Oligopeptide Compositions Differ Among Phyla. *Proteins 54*, 20−40.
(106) Jenkins, C. L., and Raines, R. T. (2002) Insights on the conformational stability of collagen. *Nat. Prod. Rep. 19*, 49−59.
(107) Kraulis, P. J. (1991) MOLSCRIPT: A Program to Produce Both Detailed and Schematic Plots of Protein Structures. *J. Appl. Crystallogr. 24*, 946−950.
(108) Wu, Y. H., Ko, T. P., Guo, S. M., Hu, S. M., Chuang, L. M., and Wang, A. H. J. (2008) Structural Basis for Catalytic and Inhibitory Mechanisms of Human Prostaglandin Reductase PTGR2. *Structure 16*, 1714−1723.
(109) Iwasaki, W., Miyatake, H., and Miki, K. (2005) Crystal Structure of the Small Form of Glucose-Inhibited Divisional Protein A from *Thermus thermophilus* HB8. *Proteins 61*, 1121−1126.
(110) Kemmink, J., and Creighton, T. E. (1993) Local Conformations of Peptides Representing the Entire Sequence of Bovine Pancreatic Trypsin Inhibitor and Their Roles in Folding. *J. Mol. Biol. 234*, 861−878.
(111) Chiu, H. C., and Bersohn, R. (1977) Electronic Energy Transfer Between Tyrosine and Tryptophan in the Peptides Trp-(Pro)$_n$-Tyr. *Biopolymers 16*, 277−288.
(112) Wierzchowski, K. L., Majcher, K., and Poznanski, J. (1995) CD Investigations on Conformation of H-X-(Pro)(N)-Y-OH Peptides (X = Trp, Tyr, Y = Tyr, Met) Models for Intramolecular Long-Range Energy-Transfer. *Acta Biochim. Pol. 42*, 259−268.
(113) Bobrowski, K., Poznanski, J., Holcman, J., and Wierczowski, K. L. (1999) Pulse Radiolysis Studies of Intramolecular Electron Transfer in Model Peptides and Proteins. 8. Trp(NH$^{\bullet+}$) $\rightarrow$ Tyr[O$^{\bullet}$] Radical Transformation in H-Trp-(Pro)$_n$-Tyr-OH, n = 3−5, Series of Peptides. *J. Phys. Chem. B 103*, 10316−10324.
(114) Doose, S., Neuweiler, H., Barsch, H., and Sauer, M. (2007) Probing polyproline structure and dynamics by photoinduced electron transfer provides evidence for deviations from a regular polyproline type II helix. *Proc. Natl. Acad. Sci. U.S.A. 104*, 17400−17405.
(115) Li, S. S. C. (2005) Specificity and versatility of SH3 and other proline-recognition domains: Structural basis and implications for cellular signal transduction. *Biochem. J. 390*, 641−653.
(116) Koide, S., and Sidhu, S. S. (2009) The Importance of Being Tyrosine: Lessons in Molecular Recognition from Minimalist Synthetic Binding Proteins. *ACS Chem. Biol. 4*, 325−334.
(117) Meyer, E. A., Castellano, R. K., and Diederich, F. (2003) Interactions with Aromatic Rings in Chemical and Biological Recognition. *Angew. Chem., Int. Ed. 42*, 1210−1250.
(118) Salonen, L. M., Ellermann, M., and Diederich, F. (2011) Aromatic Rings in Chemical and Biological Recognition: Energetics and Structures. *Angew. Chem., Int. Ed. 50*, 4808−4842.
(119) Chen, J. K., Lane, W. S., Brauer, A. W., Tanaka, A., and Schreiber, S. L. (1993) Biased Combinatorial Libraries: Novel Ligands for the SH3 Domain of Phosphatidylinositol 3-Kinase. *J. Am. Chem. Soc. 115*, 12591−12592.