JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

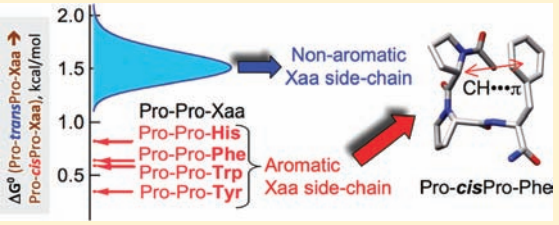
Direct Evidence for CH···π Interaction Mediated Stabilization of Pro-*cis*Pro Bond in Peptides with Pro-Pro-Aromatic motifs

Himal K. Ganguly,[†] Barun Majumder,[†] Sarbani Chattopadhyay,[†] Pinak Chakrabarti,[‡] and Gautam Basu*[,†]

[†]Department of Biophysics and [‡]Department of Biochemistry, Bose Institute, P-1/12 CIT Scheme VIIM, Kolkata 700054, India

Supporting Information

ABSTRACT: Although weak interactions play subtle but important roles in dictating protein structures, their experimental detection is nontrivial. From NOE experiments we provide direct evidence for the presence of CH···π interaction, operational between the C^{α}-H of the first Pro and the aromatic (Aro) side chain of Xaa, in a peptide series with the general sequence Ac-Pro-Pro-Xaa-NH_2. Indirect evidence of CH···π interaction is provided from ring current-induced upfield displacement of Pro(1) C^{α}-H chemical shifts and restriction of side-chain (χ1) rotation of Xaa. A consequence of this interaction is the enhanced stability of the Pro-*cis*Pro conformer in Ac-Pro-Pro-Xaa-NH_2 when Xaa is aromatic. The free energies associated with *trans* to *cis* transformation of the Pro-Pro moiety are 0.35, 0.59, 0.64, and 0.82 kcal/mol when Xaa is Tyr, Trp, Phe, and His (pH of 8.4), respectively. In comparison, the corresponding free energy is ∼1.55 kcal/mol when Xaa is nonaromatic. The observed population of Pro-*cis*Pro-His and the pH-induced perturbation of electron density of the His side chain were correlated, providing further evidence for a direct role of CH···π interaction in modulating the stability of Pro-*cis*Pro population in Ac-Pro-Pro-Aro-NH_2. Our study establishes Pro-Pro-Aro to be a new sequence motif that can stabilize Pro-*cis*Pro peptide bonds. This study not only identifies a new structurally biased sequence motif but also directly demonstrates the role played by CH···π interactions in subtly altering conformational preferences of three-residue peptide sequences with implications on the role played by *cis*-peptide bonds in unfolded proteins.



INTRODUCTION

The three-dimensional structure of a protein results from the net effect of a large number of weak interactions. A folded protein is only marginally stable over the unfolded state, underscoring the fact that disruption of only a few weak interactions can unfold a protein. Among weak interactions, CH···π interactions[1] have been implicated to play a subtle but important role. Although CH···π interactions in proteins are typically assessed from high-resolution crystal structures using a geometric criterion,[2] a recent NMR study provided direct evidence of CH···π interactions in proteins.[3] CH···π interactions have also been implicated to play a role in stabilizing *cis* peptide bonds in short peptides in an earlier study from this laboratory.[4]

Peptide bonds in proteins are rarely observed in *cis* conformation[5] due to unfavorable steric interaction between sequence contiguous C^{α} atoms in the *cis* conformation. However *cis* peptide bonds involving Pro are special. For Xaa-Pro sequence motifs, the Xaa-*trans*Pro conformer also acquires unfavorable C^{α}(Xaa)-C^{δ}(Pro) steric clash, decreasing the *cis*−*trans* energy difference.[6] This is consistent with the over representation of Pro at positions that immediately follow a *cis* peptide bond in proteins.[5e] The *cis*−*trans* energy difference is further lowered when an aromatic (Aro) residue precedes the Pro residue.[7] Conservation of Pro in *cis* peptidyl-prolyl unit in proteins indicates that the conformation of a peptide bond is in a way encoded in the local sequence of a large protein.[8a] Results from a number of bioinformatics studies indicate that selected amino acids show significant propensity to occur around a *cis* peptidyl-prolyl[8] as well as *cis* nonprolyl amide bonds,[8c] emphasizing the importance of local control of *cis* peptide bonds.

Despite evidence in support of local control of *cis* peptide bonds, subtle mismatch of local effects and global compulsions, arising from a folded protein structure, decides the final conformational fate of a peptide bond. Mutation of a *cis* prolyl residue to a nonprolyl residue can result in the retention of the *cis* bond,[9a−g] indicating a role of global tertiary interactions. It can also result in a misfolded protein[9f] or a folded protein with a *trans* conformation[9f−h] signifying the importance of the prolyl residue and, in turn, to the role of local effects in stabilizing the *cis* peptide bond in a protein. Although aromatic residues neighboring *cis* prolyl residues can exert influence on stabilization of the *cis* prolyl bond, to our knowledge, there are no reports where an aromatic residue preceding a *cis* prolyl residue has been mutated to a nonaromatic residue to observe the effect on the protein structure.

One way to delineate local versus global compulsions of a peptide bond to be *cis* is to study a short peptide sequence, in isolation, by NMR spectroscopy. A number of NMR studies have been reported on short peptides with the general sequence

Received: October 4, 2011
Published: February 15, 2012

   dx.doi.org/10.1021/ja209334v | *J. Am. Chem. Soc.* 2012, 134, 4661−4669

-Xaa($i-1$)-**Pro(i)**-Yaa($i+1$)-, where either Xaa($i-1$)[7,10,11] or Yaa($i+1$)[4,11b,12] is aromatic, all focused on measuring the *cis* content of the Xaa($i-1$)-Pro(i) bond. Enhanced *cis* content of Xaa-Pro bond in these peptides was found to be associated with Pro···Aro interaction. From NMR studies on Ac-Pro-Pro-Phe-NH_2, a designed peptide mimicking the CH···π interaction of similar sequences in proteins, we had observed that Pro-*cis*Pro conformation is overpopulated in Ac-Pro-Pro-Phe-NH_2 when compared to Ac-Pro-Pro-Ala-NH_2.[4] It was suggested that this originates from CH···π interaction between Pro1 and Phe.

To generalize our earlier observation on Pro-Pro-Phe,[4] we have studied the complete series of peptides with the sequence Ac-Pro-Pro-Xaa-NH_2, where Xaa is varied among all 20 amino acids. The aim of this study is: (i) to confirm that Pro-Pro-Aro, and not just Pro-Pro-Phe, stabilizes the Pro-*cis*Pro motif in Pro-Pro-Xaa peptide series; (ii) to derive a comparative thermodynamic scale for *trans*−*cis* equilibrium of the Pro-Pro bond in Pro-Pro-Xaa containing peptides as a function of Xaa; (iii) to characterize NOE cross-peaks between the Aro side chain and the $C^{\alpha}H$ of the first Pro residue to provide direct evidence for $C^{\alpha}H\cdots\pi$ interaction between them; and (iv) to compare our results with a previous study on -Gly-Aro-Pro-Gly- series of peptides for which only indirect experimental evidence was provided for $C^{\alpha}H\cdots\pi$ interaction.[11c] We also revisit the percent population of Xaa-*cis*Pro motifs in proteins and discuss its correlation with corresponding experimental values obtained from peptides.[7] Finally, significance and conservation of *cis*-stabilizing Xaa-Pro sequence motifs in proteins are discussed.

MATERIALS AND METHODS

Peptide Synthesis and Purification. All peptides were synthesized by a stepwise procedure in solid phase using Fmoc chemistry and Rink amide MBHA resin (substitution 0.69 mmol/g; Novabiochem) and Fmoc-protected amino acids (Novabiochem). The Fmoc-amino acid derivatives (five times excess of the resin substitution) were coupled with benzotriazol-1-yl-oxytripyrrolidino-phosphonium hexafluorophosphate (PyBOP)/hydroxybenzotriazole (HOBT)/diisopropylethylamine (DIPEA) (used as 5:5:10 times excess of the resin, respectively), and Fmoc cleavage was performed with 20% (v/v) piperidine in dimethylformamide (DMF). N-acetylation was achieved using acetic anhydride/DIPEA (1:10 times excess) in DMF. For Ac-Pro-Pro-Trp-NH_2 and Ac-Pro-Pro-His-NH_2, the N-terminal acetyl carbon was labeled with ^{13}C ($^{13}CH_3COOH$ from Sigma was coupled to H-Pro-Pro-Xaa- moiety using PyBOP/HOBT/DIPEA). Peptides with aliphatic side chains were cleaved from the resin with 95% trifluoroacetic acid (TFA) in water, and others with side-chain functional group protections (Table S1, Supporting Information) were cleaved using a cocktail (85% TFA, 5% H_2O, 5% phenol, 2.5% anisole, and 2.5% tri-isopropyl silane). All peptides were kept in the cleavage cocktail for 1 h (for Xaa = Arg, the cleavage reaction was carried on for 3 h). TFA was removed by evaporation in a rotary evaporator, and the crude peptides were dissolved in methanol and purified using reverse phase HPLC using a Phenomenix C18 column and CH_3OH-H_2O gradients (0−80% CH_3OH in 40 min) containing 0.1% TFA. The final products were characterized by NMR and mass spectrometry (Tables S1 and S2, Supporting Information).

NMR Spectroscopy. NMR experiments were performed on a Bruker DRX 500 spectrometer. All samples were prepared in 20 mM phosphate buffer pH of 7 at 4 °C (unless otherwise stated), containing 10% ^2H_2O and the sodium salt of 3(trimethylsilyl) propionic-2,2,3,3-d_4 acid (internal standard). Water signals in 1D ^1H NMR spectra were suppressed by standard excitation sculpting procedure. For 2D ^1H NMR experiments, the water signal was suppressed by WATERGATE pulse sequence total correlation spectroscopy and rotating frame overhause effect spectroscopy (TOCSY and ROESY) or by presaturation of the water signal double-quantum filtered-correlated spectroscopy (DQF-COSY). Resonance assignments were achieved using TOCSY and DQF-COSY experiments, while sequential assignments (Pro1 and Pro2) were performed from ROESY experiments (mixing time: 250 ms).[4,13] For each peptide, the four isomers (Figure 1a) were detected from ROESY (Pro-*cis*Pro- and Pro-*trans*Pro- isomers were detected from Pro-Pro C^{α}-H/C^{α}-H and C^{α}-H/C^{δ}-H cross-peaks, while Ac-*cis*Pro- and Ac-*trans*Pro- isomers were detected from Ac-Pro CH_3/C^{α}-H and CH_3/C^{δ}-H cross-peaks). The $^3J_{\alpha\beta}$ coupling constants were measured from 1D ^1H NMR spectra and DQF-COSY spectra. Relative populations of the isomers were estimated from the relative intensities of amide resonances, CH_3 (Ala), or acetyl ^{13}C (Xaa = Trp, His) resonances. For van't Hoff analysis, peptides were equilibrated for at least 30 min at a given temperature prior to obtaining the NMR spectra.

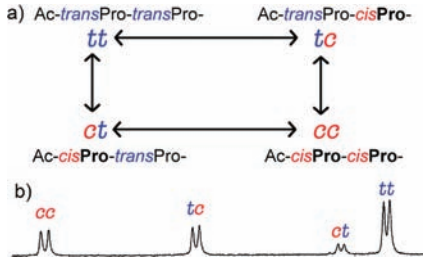


Figure 1. (a) A schematic representation of the four conformational isomers present in Ac-Pro-Pro- motif (*tt*, *tc*, *ct*, and *cc*). (b) Four amide peaks (Tyr) corresponding to the four isomers in panel (a) in Ac-Pro-Pro-Tyr-NH_2.

pH Titrations. The pH dependence of the *cis* populations in Ac-Pro-Pro-His-NH_2 was followed using ^1H NMR (pH < 6) and ^{13}C proton decoupled (pH > 6) NMR spectra (^{13}C NMR signals showed signal overlap below pH 6). The *cis*−*trans* population ratios obtained from NMR spectra were fitted with the eq 1 (see Supporting Information) to yield the *trans* to *cis* equilibrium constants, K_{tc}^0 and K_{tc}^+, corresponding to neutral and positively charged histidine side chains, respectively.

$$\frac{[cis]}{[trans]} = \frac{K_{tc}^{+} + K_{tc}^{0}\left(\frac{1+K_{tc}^{+}}{1+K_{tc}^{0}}\right)10^{(\mathrm{pH}-\mathrm{p}K_a)}}{1+\left(\frac{1+K_{tc}^{+}}{1+K_{tc}^{0}}\right)10^{(\mathrm{pH}-\mathrm{p}K_a)}} \quad (1)$$

Database Analyses. A representative list of nonredundant PDB[14] structures were collected using the culling server PISCES[15] with sequence identity ≤25% and resolution of ≤2 Å. Sequences with Xaa-Pro and Pro-Pro motifs were identified from the structures. A *cis* peptide bond was defined as $-90° \le \omega \le +90°$. Propensities of Xaa to induce Xaa-*cis*Pro motif in Xaa-Pro sequences were calculated as $[N_{X\text{-}cP}/(N_{X\text{-}cP}+N_{X\text{-}tP})]/[N_{cP}/(N_{cP}+N_{tP})]$, and propensities of Xaa to induce Pro-*cis*Pro-Xaa motif in Pro-Pro-Xaa sequences were calculated as $[N_{P\text{-}cP\text{-}X}/(N_{P\text{-}cP\text{-}X}+N_{P\text{-}tP\text{-}X})]/[N_{P\text{-}cP}/(N_{P\text{-}cP}+N_{P\text{-}tP})]$ where N, X, *c*P and *t*P stand for number of occurrences, amino acid Xaa, *cis*Pro, and *trans*Pro, respecively.

MD Simulations. Using the GROMACS 4.0 package,[16] two 20 ns molecular dynamics (MD) simulations were performed on Ac-Pro-*cis*Pro-Phe-NH_2 and Ac-Phe-*cis*Pro-NH_2 (OPLSAA[17] force field) starting from the pdb structures 1vky (residues 277−279A) and 1wbh (134−135A), respectively. Both starting structures were associated with CH···π interaction. The peptides were placed in a cubic box with the box edge adjusted approximately 7.5 Å from the peptides' periphery. The box was filled with suitable number of pre-equilibrated SPC water molecules. After initial energy minimization and pre-equilibrium run, the simulations were carried out at constant temperature (27 °C) and pressure (1 atm) using the Berendsen coupling method.[18] Time constants for temperature and pressure coupling were 0.1 and 0.5 ps, respectively. All bonds were constrained

dx.doi.org/10.1021/ja209334v | J. Am. Chem. Soc. 2012, 134, 4661−4669

using LINCS algorithm.[19] A cutoff of 10 Å was used for nonbonded interactions, and the pairwise neighbor list was updated every 10 steps (integration step: 2 fs). The trajectories were analyzed with in-house FORTRAN programs and visualized with CHIMERA.[20]

RESULTS

Populations of *cc*, *ct*, *tt*, and *tc* in Ac-Pro-Pro-Xaa-NH_2. In an earlier report we demonstrated that the *cis* isomer of the Pro-Pro peptide unit in Ac-Pro-Pro-Phe-NH_2 shows enhanced stability compared to that observed in Ac-Pro-Pro-Ala-NH_2.[4] The origin of this stability was suggested to arise from CH$\cdots\pi$ interaction between the C^{α}-H of the first Pro and the aromatic side chain of Phe (upfield chemical shift changes in C^{α}-H and the restriction of side chain $\chi 1$ of Phe). It was further suggested that such CH$\cdots\pi$ interactions might be a general phenomenon in peptides with Pro-Pro-Aromatic motifs. To fully explore the sequence space of such peptides, we synthesized a series of peptides with the sequence Ac-Pro-Pro-Xaa-NH_2 (PPX) and, taking advantage of the slow *cis*–*trans* isomerization rate, measured the populations of the *cis*/*trans* isomers by integrating appropriate ^{1}H- /^{13}C NMR resonances. As shown in Figure 1, the inherent propensities of Ac-Pro and Pro-Pro moieties to stabilize the *cis* isomer distribute the total population of PPX among four isomers: *cc*, *tc*, *tt*, and *ct*. For most PPX peptides, the amide ^{1}H NMR spectra showed four well resolved peaks (for PPY, see Figure 1b) that could be assigned as one of the four categories from 2D-NMR experiments. For PPW the *tt* amide signal overlapped with the aromatic signals, and for PPH the amide signals were too broad at high pH. The *cis*/*trans* populations for these peptides were measured by integrating the ^{13}C signals of isotope-labeled peptides $^{13}CH_3CO$-Pro-Pro-Xaa-NH_2.

The NMR-derived relative populations of *cc*, *tc*, *tt*, and *ct* isomers of PPX series of peptides measured at 4 °C in aqueous buffer of pH 7.0 (unless otherwise stated) are shown in Table 1. All amino acids with an aromatic side chain (Phe, Tyr and Trp) exhibited enhanced stability of the *tc* state compared to the *tt* state. In addition, PPH also exhibited this enhanced stability at pH 8.4 (neutral His side chain) but not at pH 3.5 (positively charged His side chain). Populations of the *cis* conformers (*tc* and *cc*) were the lowest (≤1%) for Arg, Lys, and Gln, almost at the border of reliable population measurement by NMR. To emphasize the observed tendency of Phe, Tyr, Trp, and His (pH 8.4) to be outliers, we computed the *trans* to *cis* free energy differences for the Ac-Pro bond, $\Delta G^0_{tt/ct}$, and for the Pro-Pro bond, $\Delta G^0_{tt/tc}$, using the relation $\Delta G^0 = -RT \ln K$ (Table 1). As shown in Figure 2, the distribution of $\Delta G^0_{tt/ct}$ for the peptide series shows a Gaussian distribution (mean: 0.91 ± 0.07 kcal/mol) with no outliers, indicating that the nature of X in Ac-PPX-NH_2 does not influence the free energy difference between the Ac-*cis*Pro and Ac-*trans*Pro isomers. In addition to the PPX series of peptides, we also synthesized the peptide Ac-Pro-Ala-NH_2 and measured the ΔG^0 for the *trans* to *cis* isomerization of the Ac-Pro bond. The value of ΔG^0 obtained, 0.88 kcal/mol, is very close to the mean of the $\Delta G^0_{tt/ct}$ distribution, indicating that $\Delta G^0_{tt/ct}$ in PPX remains unaltered even when X is absent.

Except for two groups, X = {R, K, Q} and X = aromatic, $\Delta G^0_{tt/tc}$ values of all PPX peptides clustered around 1.55 ± 0.16 kcal/mol (Figure 2 and Table 1). The first outlier group (Lys, Arg, and Gln) is not considered further, since the data is less reliable due to low sensitivity of NMR resonances of the *cis* populations in these peptides (Table 1). On the other hand, the

Table 1. Populations of *cc*, *tc*, *ct* and *tt* Isomers in PPX Peptide Series

	percent population				ΔG^0 (kcal/mol)[b]	
peptide[a]	*cc*	*tc*	*ct*	*tt*	*tt* → *tc*	*tt* → *ct*
PPF	18.6	17.1	9.5	54.8	0.64	0.97
PPY	21.5	24.3	8.7	45.5	0.35	0.91
PPW	28.5	16.6	6.5	48.4	0.59	1.11
PPH[c]	11.0	14.3	11.3	63.4	0.82	0.95
PPH[d]	3.0	4.8	14.4	77.8	1.54	0.93
PPA	<0.5[e]	5.0	15.5	79.0	1.52	0.90
PPS	1.7	3.6	15.2	79.5	1.71	0.91
PPT	1.2	3.4	15.6	79.8	1.74	0.90
PPI	1.7	4.5	17.2	76.6	1.56	0.82
PPV	1.2	3.0	15.3	80.5	1.81	0.92
PPL	2.0	7.9	16.0	74.1	1.23	0.84
PPG	2.7	4.4	14.0	78.9	1.59	0.95
PPE	1.9	5.1	15.7	77.3	1.50	0.88
PPD	5.9	4.8	15.0	74.3	1.51	0.88
PPN	2.7	7.0	13.7	76.6	1.32	0.95
PPC	2.4	3.8	13.2	80.6	1.68	1.00
PPM	2.3	5.5	16.2	76.0	1.45	0.85
PPQ	0.9[e]	1.1[e]	19.0	79.0	2.36	0.79
PPK	0.5[e]	1.3[e]	16.6	81.6	2.28	0.88
PPR	0.3[e]	0.9[e]	18.8	80.0	2.47	0.80

[a]PPP not studied. [b]All ΔG^0 values correspond to 4 °C and pH = 7 unless otherwise stated. [c]pH = 8.4. [d]pH = 3.5. [e]Unreliable data (signal ~ noise).

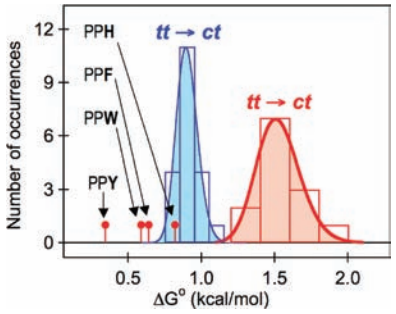


Figure 2. Distributions of free energies for *trans* to *cis* isomerization, $\Delta G^0_{tt/ct}$ (*tt* → *ct*) and $\Delta G^0_{tt/tc}$ (*tt* → *tc*), for peptides with sequence Ac-PPX-NH_2. X = R/K/Q (data not shown) and X = aromatic (Y, W, F, and H (pH 8.4), all highlighted by arrows) data points are excluded from the $\Delta G^0_{tt/ct}$ (*tt* → *tc*) distribution.

$\Delta G^0_{tt/tc}$ values exhibited by all aromatic side chains (see Figure 2) are significantly lower than the rest by several standard deviations (0.82, 0.64, 0.59, and 0.35 kcal/mol for His (pH 8.4), Phe, Trp, and Tyr, respectively). The data are comparable with an earlier NMR study performed on a series of water-soluble peptides with the sequence motif -(Xaa-Pro)-.[7] In that study, while the mean $\Delta G^0_{t/c}$ for nonaromatic Xaa side chains was 1.21 kcal/mol, the $\Delta G^0_{t/c}$ values for aromatic Xaa side chains were lower, 0.98, 0.71, 0.68, and 0.29 kcal/mol for His (pH 8.4), Phe, Tyr, and Trp, respectively).

pH Dependence of *cis*/*trans* Equilibrium of Pro-Pro-His. The *cis*/*trans* isomerization observed for PPH is pH dependent (Table 1). At low pH, when the imidazole side chain is protonated, the relative *cis* content is less than what is observed at high pH, when the imidazole side chain is neutral. This is consistent with a CH$\cdots\pi$ interaction-mediated stabilization of the *cis* conformer, since a neutral imidazole is

dx.doi.org/10.1021/ja209334v | J. Am. Chem. Soc. 2012, 134, 4661–4669

a better π-donor than the protonated species. If this is indeed true, then the *cis*/*trans* populations should be modulated by pH in such a manner that their variation match with the His protonation equilibrium characterized by the His side-chain pK_a. Populations of all the four isomers of PPH were estimated as a function of pH using NMR spectroscopy (^{1}H amide and ^{13}C acetyl resonances). Variation of *cis*/*trans* population ratios (*tc*/*tt* for the Pro-Pro and *ct*/*tt* for the Ac-Pro bond) as a function of pH is shown in Figure 3. Strikingly, the *ct*/*tt* ratio is pH-independent, while the *tc*/*tt* ratio shows the expected pH-dependence.

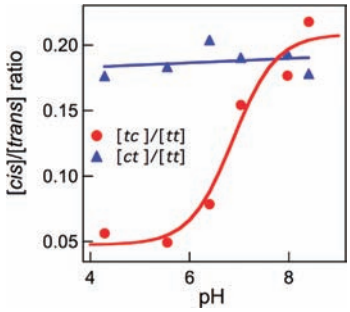


Figure 3. NMR-detected *cis*/*trans* population ratios for Ac-Pro-Pro-His-NH_2 as a function of pH at 4 °C. The [*ct*]/[*tt*] ratio (blue) is unaffected by pH, while the [*tc*]/[*tt*] ratio (red) is strongly modulated by pH. The solid line is the best fit (eq 1) to experimental [*tc*]/[*tt*] ratios.

The pH-dependent [*tc*]/[*tt*] population ratio yielded an excellent fit with eq 1 with pK_a = 6.81 ± 0.22, K^0_{tc} = $[tc]_{PPH^0}/[tt]_{PPH^0}$ = 0.21 ± 0.01 (ΔG^0_{tc0} = 0.86 kcal/mol at 4 °C) and K^+_{tc} = $[tc]_{PPH^+}/[tt]_{PPH^+}$ = 0.047 ± 0.01 (ΔG^0_{tc+} = 1.68 kcal/mol at 4 °C). The pK_a obtained from the fit matches well with the His side-chain pK_a, confirming that indeed the *cis*/*trans* isomerization of the Pro-Pro bond is modulated by the protonation state of the imidazole side chain (which in turn is expected to modulate the CH···π interaction). The net extra stability that PPH gains going from a protonated to an unprotonated His, 0.82 kcal/mol ($\Delta\Delta G^0 = \Delta G^0_{tc+} - \Delta G^0_{tc0}$), arises due to CH···$\pi$ interaction. It is the best estimate for the free energy associated with CH···π interaction, since the two interconverting molecules being considered, PPH^+ and PPH^0, have identical internal degrees of freedom.

Upfield Shift of C$^\alpha$-H of Pro1 in Pro-Pro-Aro Peptides. CH···π interaction, if operative in the *cis* isomer, is expected to leave its signature in the NMR spectra due to the ring current effect of the π-system on protons that are in its vicinity. In Figure 4, Pro1 C$^\alpha$-H proton chemical shifts of the four isomers are shown for six peptides, three with aromatic side chains (PPW, PPY, and PPF) and three without (PPA, PPD, and PPI). For the *tt* isomer, Pro1 C$^\alpha$-H chemical shifts are close to the 'random coil' chemical shift value for Pro C$^\alpha$-H (4.44 ppm) for all peptides (PPY: 4.65; PPW: 4.57; PPF: 4.67; PPI: 4.72; PPD: 4.71; PPA: 4.71). This is true for all the other three isomers in peptides lacking an aromatic side chain. However, for peptides containing an aromatic side chain, the Pro C$^\alpha$-H exhibits a marked upfield shift (1–2 ppm) for the *tc* and the *cc* isomers, both containing the Pro-*cis*Pro motif and not the *ct* isomer, containing the Ac-*cis*Pro motif only. This strongly suggests that the Pro-*cis*Pro, in PPY, PPW, and PPF, are associated with a conformation where C$^\alpha$-H of Pro1 and the aromatic side chain are proximal, consistent with CH···π interaction. For PPH, the

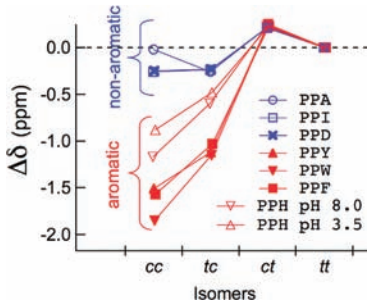


Figure 4. C$^\alpha$-H (Pro1) chemical shift differences, ($\Delta\delta = \delta_{xx} - \delta_{tt}$; xx = *tt*, *ct*, *tc*, and *cc*) in Ac-Pro-Pro-Xaa-NH_2 for all Xaa with aromatic side chains (red) and a selected set of Xaa with nonaromatic side chains (blue).

trend was similar to PPAro peptides, both at low and high pH, indicating that at low pH, even with a protonated side chain, the His chain is in close proximity to C$^\alpha$-H of Pro1; although the stabilization of the CH···π interaction is less than that at high pH, as was shown in the pH dependent study.

Restriction of Aro χ1 Angle in Pro-Pro-Aro Peptides. In addition to the upfield shifts of C$^\alpha$-H of Pro1, we also observed nondegenerate $^3J_{\alpha\beta}$ coupling constants in aromatic residues for the Pro-Pro-Aro peptides. As shown in Figure 5,

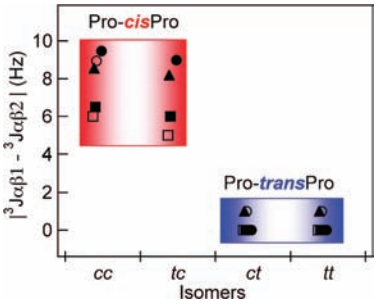


Figure 5. Absolute differences between $^3J_{\alpha\beta1}$ and $^3J_{\alpha\beta2}$ in Aro, as observed in the four isomers of Ac-Pro-Pro-Aro-NH_2 (PPY: filled cricle, PPW: open circle, PPF: filled triangle, PPH (pH 3.5): open square; PPH (8.0): filled square). The Pro-*trans*Pro and the Pro-*cis*Pro isomers are highlighted in blue and red, respectively.

the nondegenerate $^3J_{\alpha\beta}$ coupling constants were observed only for Pro-*cis*Pro isomers (*tc* and *cc*), with an absolute difference of ~6–9 Hz between $^3J_{\alpha\beta1}$ and $^3J_{\alpha\beta2}$. Nondegenerate $^3J_{\alpha\beta}$ coupling constants ($^3J_{\alpha\beta1}$ > 5 Hz and $^3J_{\alpha\beta2}$ < 10 Hz or vice versa; $\Delta^3J_{\alpha\beta}$ > 5 Hz) can arise only when the χ1 dihedral angle is restricted to the *g*- (χ1 = −60°) or the *t*- (χ1 = 180°) rotamer; on the other hand, degenerate $^3J_{\alpha\beta}$ values can either indicate the presence of the g_+ (χ1 = +60°) rotamer ($\Delta^3J_{\alpha\beta1} \sim \Delta^3J_{\alpha\beta2}$ < 5 Hz) or fast rotation along χ1 ($^3J_{\alpha\beta1} \sim {}^3J_{\alpha\beta2} \sim$ 7 Hz).[13] In the Pro-Pro-Aro peptide series, $^3J_{\alpha\beta1}$ and $^3J_{\alpha\beta2}$ values were similar (7 ± 1 Hz) for the *tt* and the *ct* conformations, indicating fast rotation of the aromatic side chain. On the other hand for the *tc* and the *cc* isomers, the observed spread in nondegenerate $^3J_{\alpha\beta}$ values indicated the presence of restricted rotation along χ1 (either 180° or −60°). Analyzing the relative ROE cross-peak intensities (α-βA, α-βB, N-βA, N-βB)[13] it was further found (Table 2) that the preferred χ1 angle is −60° (*g*-) and not 180° (*t*). Restriction of side-chain dihedral angle χ1 of a terminal amino acid in a three-residue peptide is remarkable and is fully compatible with a model where the side chains of the aromatic

dx.doi.org/10.1021/ja209334v | J. Am. Chem. Soc. 2012, 134, 4661–4669

Table 2. Relative ROE Intensities of Aro β-Protons[a]

	$^3J_{\alpha\beta}$ (Hz)	relative ROE intensities			
peptides	α-βA/α-βB	α-βA	α-βB	N-βA	N-βB
PPF (*cc*)	?/12.50	?	?	±	+++
PPF (*tc*)	3.95/12.13	+++	+	±	+++
PPY (*cc*)	3.38/12.84	+++	+	±	+++
PPY (*tc*)	3.85/12.83	+++	+	±	+++
PPW (*cc*)	3.75/12.69	+++	+	±	+++
PPW (*tc*)	?/?	?	?	?	?

[a] βA/βB: downfield/upfield-shifted β-protons; ?: overlap; +++: strong ROE; +: weak ROE; ±: weak/absent ROE.

residues are 'frozen' due to their participation in CH···π interaction.

Thermodynamic Analysis. In our earlier work on PPF and PPA,[4] van't Hoff analyses of temperature-dependent NMR spectra yielded entropic and enthalpic contributions for the *tt* → *tc* transition. Both PPF and PPA yielded unfavorable entropies, while the associated enthalpy was favorable for PPF and unfavorable for PPA. Similar analyses were performed on a larger set of PPX peptides with aromatic (W and Y) as well as nonaromatic (I, E, and N) X. As shown in Figure 6, for all

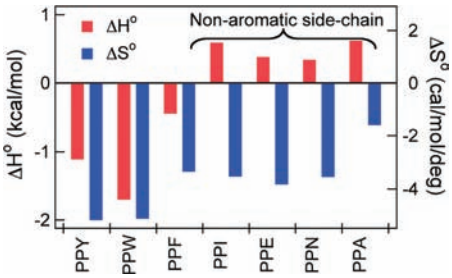


Figure 6. Thermodynamic parameters ($\Delta H°$ and $\Delta S°$) associated with the equilibrium *tt* → *tc* (Figure 1) for a selected set of peptides in the PPX series where X is aromatic (Y, W, and F) and nonaromatic (I, E, N, and A). The $\Delta H°$ and $\Delta S°$ values were obtained from van't Hoff analysis of temperature-dependent NMR experiments (integrals of amide/acetyl resonances).

peptides the entropy change associated with the *tt* → *tc* process is unfavorable, from ∼5 cal/deg/mol (PPW and PPY) to ∼3.5 cal/deg/mol (I, E and N). The associated enthalpy change is favorable only when X is aromatic (−0.45,[4] −1.7, and −1.1 kcal/mol for F, W, and Y, respectively). The observed trend indicates that *tt* → *tc* transition for PPX is enthalpy driven only when X is aromatic.

Direct Evidence for CH···π Interaction from ROE. The indirect evidence for CH···π interaction operative in the Pro-*cis*Pro conformations in PPX series where X is aromatic was further confirmed from direct experimental evidence. Direct evidence for CH···π interaction was obtained from ROE cross peaks between Pro1 C^α-H and Tyr side-chain protons in PPY. As shown in Figure 7b, the ROESY spectrum of PPY shows clear cross-peaks between all the aromatic protons of Tyr (C^ε-H and C^δ-H) and Pro1 C^α-H for the Pro-*cis*Pro isomers (*cc* and *tc*) only. Models of PPY in *cc* and in *tc* conformations are shown in Figure 7a for clarity. For PPY, ROESY cross-peaks were also observed between the methyl protons of the acetyl group and the Tyr side chain (C^ε-H/C^δ-H) for the *cc* isomer.

ROESY cross-peaks were observed between Pro1 C^α-H and aromatic side-chain protons for PPF and PPW as well. As shown in Figure 7c, cross-peaks between Pro1 C^α-H and Phe C^δ-H were observed for the *cc* as well as the *tc* isomer in PPF. The acetyl methyl protons also exhibited a cross peak with the aromatic side chain (either C^ε-H or C^ζ-H but not C^δ-H) for the *cc* isomer. For PPW (Figure 7d) all aromatic protons of the *cc* isomer exhibited ROE cross-peaks with Pro1 C^α-H. This is true for the methyl protons of the acetyl group as well. For the *tc* isomer the C^α-H of Pro1 (3.40 ppm) overlaps with C^β-H (3.45 and 3.31 ppm). Thus even if ROE cross-peaks were present between Pro1 C^α-H and Trp aromatic protons, they would be buried under a strong C^β-H cross-peak.

DISCUSSION

CH···π Interaction and the Pro-Pro-Aro Motif. Weak interactions can play an important role in dictating protein structures. Among them the CH···π interaction is notable.[1] First proposed by Perutz,[21] CH···π interactions have been extensively studied in crystal structures of proteins[2] and peptides.[22] In some proteins, even the disruption of a single CH···π interaction can alter the stability and functional robustness of a protein.[23] However, due to its weak nature, a direct experimental observation of CH···π interaction is tricky. A recent report describes the direct detection of CH···π interactions in proteins from carefully designed NMR experiments.[3] In this work we report the direct observation of CH···π interactions in a series of three-residue peptides with the general sequence Pro-Pro-Aro, operating between the C^α-H of the first Pro residue and the π-electrons from the side chain of the third aromatic residue. The CH···π interaction stabilizes the *cis* conformation of the tertiary amide (^{t}amide) bond between the first two Pro residues.

As demonstrated earlier by a number of studies on peptides,[7,10−12,24] the ^{t}amide bond (Xaa-Pro) can be stabilized by the presence of aromatic residues in the vicinity of Pro in the amino acid sequence. Inspired by the observation of local CH···π interactions in protein structures associated with the *cis* ^{t}amide bond,[5e] we had also demonstrated that the presence of Phe instead of Ala, following a Pro-Pro fragment, increases the *cis* population of the ^{t}amide bond.[4] None of these studies, including ours, could provide direct evidence for the Pro···Aro interaction, although indirect evidence from chemical shift anomaly and restricted rotation of the aromatic side chain was provided in some of the studies.[4,10,11c,12] In another study, not related to *cis* peptide bonds, direct evidence of C^ε-H(Lys)···π interaction was provided from NOE crosspeaks in a designed β-hairpin peptide.[25] Interestingly, an early NMR study of cyclic somatostatin analog by Kessler et al.,[26] where evidence was provided for interaction between sequence contiguous Phe (aromatic protons) and D-Pro (C^δ-H), but the nature of the interaction not explored, might have been one of the earliest observation on the manifestation of CH···π interaction in peptides.

The presence of ROE cross-peaks between Pro1 C^α-H of and aromatic protons in Ac-Pro-*cis*Pro-Aro-NH_2 (Aro = Tyr, Phe, Trp) is a direct evidence CH···π interaction between Pro1 and Aro3, separated by Pro2. This is supported by the observed restriction of Aro χ1 angle and upfield shift of Pro1 C^α-H. The corresponding Ac-Pro-*trans*Pro-Aro-NH_2 conformers did not exhibit any evidence for CH···π interaction. An enhanced stability of the -Pro-*cis*Pro-Xaa- unit (compared to -Pro-*trans*Pro-Xaa) was associated with this CH···π interaction. When Xaa changes from nonaromatic to aromatic, the corresponding ΔG^0 (*tt* → *tc*) also becomes less unfavorable

dx.doi.org/10.1021/ja209334v | J. Am. Chem. Soc. 2012, 134, 4661−4669

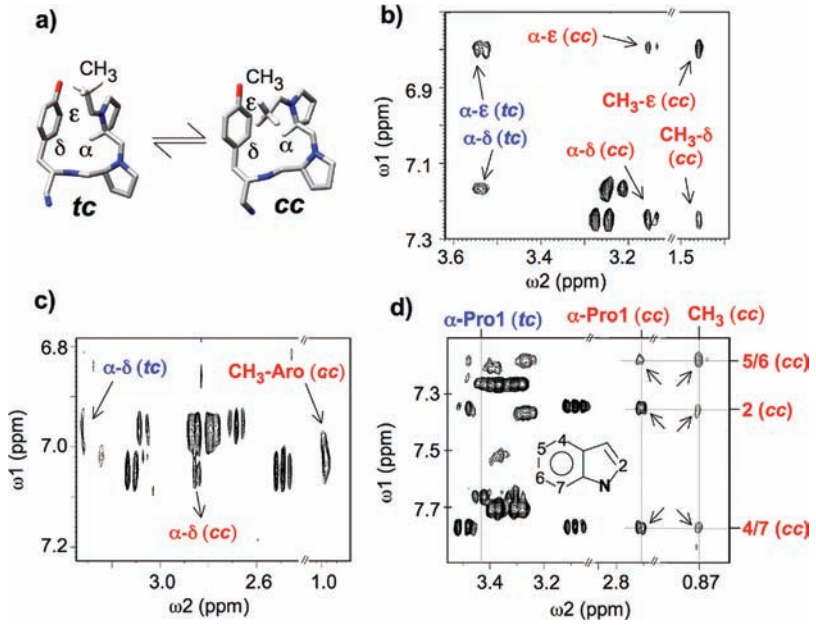


Figure 7. (a) Molecular models of the *tc* (Ac-*t*Pro-*c*Pro) and the *cc* (Ac-*c*Pro-*t*Pro) isomers in Ac-Pro-Pro-Tyr-NH_2. ROE cross peaks between Pro1 C^{α}-H/acetyl methyl and aromatic ring protons in: (b) Ac-Pro-Pro-Tyr-NH_2, (c) Ac-Pro-Pro-Phe-NH_2, and (d) Ac-Pro-Pro-Trp-NH_2. Relevant cross peaks are indicated by arrows and colored according to isomer type (*cc*: red; *tc*: blue).

(1.55 ± 0.16 to 0.82, 0.64, 0.59, and 0.35 kcal/mol for His (pH 8.4), Phe, Trp, and Tyr, respectively). Therefore, the CH···π interaction stabilizes the Pro-*cis*Pro-Aro conformer by about 0.95 kcal/mol. The pH-dependent modulation of the Pro-*cis*Pro conformation in Pro-Pro-His yielded a ΔG^0 value of 0.82 kcal/mol for the *trans*/*cis* isomerization of the Pro-Pro bond, demonstrating how protonating the His side chain, which directly affects the CH···π interaction, can make Pro-Pro-His behave like Pro-Pro-Xaa (Xaa not aromatic) series of peptides. Our results establish Pro-Pro-Aro to be a new sequence motif that can stabilize Pro-*cis*Pro peptide bonds along with a Xaa-dependent thermodynamic scale of the *cis*/*trans* isomerization of Pro-Pro-Xaa motif (Table 1).

The thermodynamic scale could be useful in fine-tuning the design of water-soluble peptides. Balaram and co-workers[27] studied the effect of N-terminal diproline templates in helix nucleation and had observed significant *cis* population of Pro-Pro-Xaa moiety from NMR studies in nonaqueous solvents. Hilser and co-worker[28] probed the denatured state of polyproline II conformation by studying the binding of SH3 domain of SEM-5 to a series of peptides with Pro-Pro-Xaa motif. In both cases, Xaa-mediated *cis*/*trans* isomerization can potentially affect the conformational ensemble of the designed peptides. To account for such effect, a comprehensive thermodynamic scale as in Table 1 will be helpful.

End-Capping Effects. The peptide series we studied, capped by an acetyl and an amide group at the two termini, were only three-residues long, essentially eliminating end-capping effects that could arise from the presence of full amino acid residues at the termini. A preliminary study of end-capping effects was undertaken with two peptides with an aromatic and a nonaromatic residue following the Pro-Pro unit, PPY and PPN. Four peptides corresponding to PPY and PPN were synthesized, each capped by an Ala residue at the either termini: Ac-APPY-NH_2, Ac-APPN-NH_2, Ac-PPYA-NH_2, and Ac-PPNA-NH_2. NMR investigation (data not shown) performed on the peptides showed that the *cis*/*trans* population ratios of APPY and PPYA were comparable to PPY, while that in APPN and PPNA were comparable to PPN, demonstrating that the enhanced preference for the Pro-*cis*Pro conformation in Pro-Pro-Aro is not affected by Ala capping of the two termini. Although only a thorough investigation with a variety of amino acids capping the Pro-Pro-Xaa motif can result in a more robust conclusion about end-capping effects on modulating the *cis*/*trans* equilibrium of Pro-Pro-Xaa sequence motif, an earlier published result shows that the PPY motif's preference for the Pro-*cis*Pro remains unaffected even when placed in a longer peptide. Kemmink and Creighton studied two peptides with the general sequence RPDFSLE-(PPX)-TGPSK (X = A and Y).[10] The populations of the Pro-*cis*Pro isomer in their 'peptide-capped' PPA and PPY peptides were 5% and 20%, respectively, matching well with our 'uncapped' PPA and PPY peptides (Table 1).

CH···π Interaction in Pro-*cis*Pro-Aro and Aro-*cis*Pro. Another characteristic of the *trans*/*cis* isomerization of the Pro-Pro motif in Pro-Pro-Xaa series of peptides is: (i) unfavorable ΔS^0 (3.5−5 cal/mol/deg) independent of the nature of Xaa and (ii) favorable ΔH^0 only when Xaa is aromatic. Wu and Raleigh have reported ΔH^0 and ΔS^0 values (van't Hoff analysis) associated with *trans* → *cis* isomerization of a related peptide series, containing the -Aro-Pro- motif: Gly-Aro-Pro-Gly (Aro = Phe, Tyr, Trp).[11c] The ΔS^0 values for the Aro-Pro series of peptides are near zero, while for the Pro-Pro-Xaa series it is unfavorable. This difference is understandable since only one (Aro-ψ) backbone and two (Aro-χ1, Aro-χ2) side-chain dihedral angles get restricted in the Aro-*cis*Pro conformer, while three (Pro1-ψ, Pro2-ψ, Aro-ϕ) backbone and two (Aro-χ1, Aro-χ2) side-chain dihedral angles get restricted in the Pro-*cis*Pro isomer (assuming the Pro-ϕ angles are constrained). The ΔH^0 values reported for the Aro-Pro series of peptides are all unfavorable, 0.53, 0.65, and 0.75 kcal/mol for Phe, Tyr, and Trp, respectively. The ΔH^0 values strongly contrast with favorable ΔH^0 values for the same (*trans* → *cis*) isomerization process in our peptide series (−0.45,[4] −1.7, and −1.1 kcal/mol

dx.doi.org/10.1021/ja209334v | *J. Am. Chem. Soc.* 2012, 134, 4661−4669

for PPF, PPW, and PPY, respectively). Notwithstanding the enthalpy difference trend, there was indirect evidence for Aro1···Pro2/Pro1···Aro3 interaction in the *cis* isomer for both the Aro-Pro as well as the Pro-Pro-Aro series of peptides: (i) upfield shifted Pro C^{α}-H resonances in NMR and (ii) restriction of Aro $\chi 1$ angle. Yet, the *trans* isomer is enthalpically favored (compared to the *cis* isomer) for the Aro-Pro series, while the *cis* isomer is enthalpically favored (compared to the *trans* isomer) for the Pro-Pro-Aro series. The observed ΔH^0 is the enthalpy difference between the *cis* and the *trans* states. Therefore, either the nature of Aro···Pro interaction in Aro-*cis*Pro and Pro-*cis*Pro-Aro is different or one of the two *trans* conformers is characterized by some unique interaction absent in the other *trans* state that gets disrupted in the *cis* form.

If differences in the Pro-*cis*Pro and Aro-*cis*Pro isomers are responsible for the different ΔH^0 trends in the two peptide series, the underlying cause can either be the different nature of CH···Aro interaction or the different residence times of CH···Aro interactions in the two *cis* isomers (or both). The NMR-derived dominant Aro $\chi 1$ angle for the *cis* conformer is different in both *t*- ($\chi 1 = 180°$) in Aro-*cis*Pro and *g*- ($\chi 1 = -60°$) in Pro-*cis*Pro-Aro. Analyses of high-resolution protein structures from pdb also showed that the Aro···Pro interacting geometries are different for the two cases (Figure 8). For **Pro-***cis*Pro-**Aro** motifs, the most populated $\chi 1$ angle in Aro is −60°. On the other hand, for **Aro-***cis***Pro** motifs, the most populated $\chi 1$ value is 180°. These two $\chi 1$ angles correspond to favorable CH···Aro interaction in the two peptides. We compared the nature of CH···π interactions in the two peptides by analyzing 20 ns MD trajectories of Ac-Pro-*cis*Pro-Phe-NH_2 and Ac-Phe-*cis*Pro-NH_2 in water (starting from two pdb coordinates that exhibited CH···Aro interaction, 1vky and 1wbh). Distances between Pro1 C^{α}-H and Phe ring center and the angle between Pro1 C^{α}, H^{α} and Phe ring center were calculated for those structures for which $\chi 1$ was 180° (Phe-Pro) or −60° (Pro-Pro-Phe). These values were similar for the two peptides: 4.0 ± 0.3 Å and 137 ± 13° for Phe-Pro and 4.1 ± 0.3 Å and 132 ± 9°, both compatible with favorable CH···π interaction geometry. In other words, the two *cis* conformations are very similar as far as CH···π interaction is concerned.

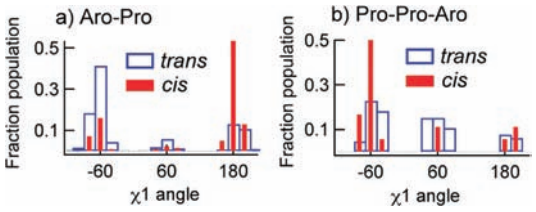


Figure 8. Side-chain $\chi 1$ dihedral angle distributions of Aro side chain in a) Aro-Pro and b) Pro-Pro-Aro sequence motifs in high-resolution protein structures as a function of *cis*/*trans* isomeric state of the Aro-Pro or the Pro-Pro ᵗamide bond.

Next we investigated the residence times of CH···π interactions in the two *cis* conformations. Using a methodology suggested by Kemmink and Creighton,[10] relative populations of Pro-***cis***Pro and Pro-*cis*Pro were calculated where Pro-***cis***Pro and Pro-*cis*Pro are two types of *cis* isomers, with and without the CH···Aro interaction, respectively. Let K_{ARO} and K_{NAR} be the *trans* → *cis* equilibrium constants for two peptides with sequence Pro-Pro-Xaa, where Xaa is aromatic (K_{ARO}) and nonaromatic (K_{NAR}), and K_0 = [Pro-***cis***Pro-Aro]/[Pro-*cis*Pro-Aro]. K_0 is given by $K_0 = K_{NAR}/K_{ARO} - 1$. Using K_{NAR} (*tt* → *tc* of Pro-Pro-Ala) = 0.06, K_0 values (ratio of *tc* isomers with and without CH···Aro interaction) were calculated for the Pro-Pro-Aro series. Similarly, using K_{NAR} (*ct* → *cc* of Pro-Pro-Ser) = 0.11, K_0 values (ratio of *cc* isomers with and without CH···Aro interaction) were calculated for the Pro-Pro-Aro series. The ratios were translated into percent Pro-***cis***Pro populations (Table 3). About 90−95% of the *cc* isomers and 70−85% of the *tc* isomers were found to exhibit the CH···Aro interaction. These values are larger by ∼20−30% than the corresponding *cis* populations in the Aro-Pro series that exhibited the CH···Aro interaction,[11c] estimated by an identical method (Trp-*cis*Pro: 76%, Tyr-*cis*Pro: 67%, Phe-*cis*Pro: 60%).

Table 3. Percent *cis* Population Associated with CH···π Interaction

	percent population	
peptides	*cc* isomer	*tc* isomer
PPF	94 (91)[a]	78 (87)
PPY	96 (94)	88 (94)
PPW	97 (93)	81 (-)
PPH (pH 8.0)	89 (77)	71 (77)

[a]Numbers within parentheses are $^3J_{\alpha\beta}$-based populations.

The residence times of the side-chain rotamers ($\chi 1$) of the Aro moiety in Xaa-Pro series have also been reported for the Aro-Pro series of peptides using $^3J_{\alpha\beta}$ values.[11c] The reported values are not much different than those obtained from equilibrium constant-based values. We also calculated NMR-derived residence times of Aro side-chain rotamers in the Pro-*cis*Pro-Aro peptide series using a methodology suggested by Pachler.[28] The estimated values for the most populated rotamer (Pro-***cis***Pro-Aro) are shown in Table 3. The NMR-derived populations of the most populated rotamers are consistent with corresponding estimates from equilibrium constants, confirming that Pro-*cis*Pro-Aro is characterized by a very tight CH···Aro interaction. This was also confirmed from MD simulations. While the Pro1-Phe3 interaction in Pro-*cis*Pro-Aro remained intact for the entire 20 ns trajectory, Phe1-Pro2 interaction in the Phe-*cis*Pro peptide survived for only about 50% of the 20 ns trajectory.

Pro-*cis*Pro-Xaa Populations in Peptides and Proteins. Using a database of nonredundant protein structures (25 588 Xaa-Pro and 1039 Pro-Pro units), we recalculated propensities of Pro-*cis*Pro as a function of Xaa. Qualitatively the propensity values (Table S3, Supporting Information) matched with an earlier report that used a different database.[4] Tyr showed the highest propensity (3.5), followed by Phe (2.8). The high propensity (1.6) for Trp is not reliable due to sparse data. Overall the trend, that aromatic side chains show a preference to occur following a Pro-*cis*Pro unit, correlated with the NMR results.

Reimer et al.[7] had estimated percent Xaa-*cis*Pro units in proteins and compared them with percent *cis* population of Xaa-Pro moiety in a synthetic peptide series Ac-Ala-Xaa-Pro-Ala-Lys-NH_2 (blue symbols in Figure 9). Xaa = Pro was an outlier in their data in the sense that the NMR-derived *cis* content was markedly lower than the corresponding protein *cis* content, when compared to other amino acids. This prompted us to hypothesize that an isolated -Pro-Pro- motif does not truly reflect major local factors responsible for the observed -Pro-*cis*Pro- population in proteins. We designed and studied

dx.doi.org/10.1021/ja209334v | *J. Am. Chem. Soc.* 2012, 134, 4661−4669

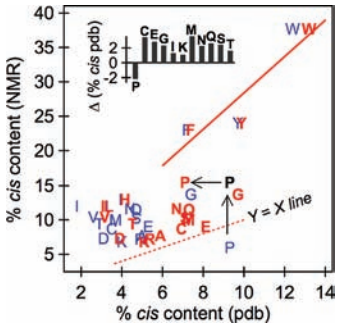


Figure 9. Correlation between percent *cis* population of Xaa-Pro sequence motifs in peptides and proteins. The peptide-based data are from NMR studies on synthetic peptides by Reimer et al.[7] The protein-based data are from analysis of pdb structures, one set from Reimer et al.[7] (blue) and the other from this study (red); differences between the two sets are shown in the inset. For Pro, the third point (P, shown in black) corresponds to protein *cis* population from Reimer et al.[7] and an adjusted peptide *cis* population based on this study (see text for details). The solid line is the linear fit for a subset (F, Y, and W), and the dotted line represents equal populations of Xaa-*cis*Pro conformation in peptides and proteins.

the Pro-Pro-Xaa series of peptides to investigate if the nature of Xaa affects the Pro-*cis*Pro content in Pro-Pro-Xaa. Indeed Xaa was found to affect the *cis* content of Pro-Pro-Xaa. As a follow up, the corrected NMR *cis* content for Pro in Figure 9 was updated as: $[\%cis]_{\text{Pro-Pro}} = \sum_{\text{Xaa}}(F_{\text{Xaa}} \times Y_{\text{Xaa}})$, where F_{Xaa} is the fraction of Pro-*cis*Pro-Xaa in all Pro-Pro motifs in proteins and Y_{Xaa} is the percent *cis* content of the Pro-Pro-Xaa peptide (*cc* + *tc* in Table 1). As expected, the new value, shown in Figure 9 (black letter P, indicated by a vertical arrow), increased the NMR-derived *cis* content of Pro-Pro (from 6% to 15.4%) and Pro was no more an outlier.

We also recalculated the percent *cis* populations of Xaa-Pro motifs in proteins using our database (Table S4, Supporting Information; red alphabets in Figure 9) to update the values reported by Reimer et al.[7] Eleven amino acid types showed more than 1% variation than that reported by Reimer et al.[7] (inset to Figure 9). All amino acids except Pro registered an increase; with the new protein Pro-Pro *cis* content (horizontal arrow, Figure 9), the Pro-Pro data point moved closer to the line representing the line correlating peptide and protein *cis* contents of Phe, Tyr and Trp. However, the new Gly-Pro data seems to be the new exception. Interestingly, analysis of -Gly-*cis*Pro- fragments in protein showed an unusually high number of Gly-*cis*Pro-Tyr fragments. Possibly Gly-Pro-Aro is a sequence motif that is responsible for the unusually high percent *cis* population of Gly-Pro in proteins. Only a series of peptides with the general sequence Gly-Pro-Xaa (with Xaa being aromatic and nonaromatic) can resolve the issue.

Conservation of Pro-*cis*Pro-Xaa in Proteins. *Cis* prolyl residues are conserved more than *trans* prolyl residues in proteins, the former sometimes maintained in proteins with sequence identity as low as 20%.[8] We looked at the sequence conservation of Pro-Pro-Aro motifs in our structural data set. Conservation of Pro-Pro-Aro sequence motifs in sequence homologous proteins of a representative set of protein structures with Pro-*cis*Pro-Aro and Pro-*trans*Pro-Aro fragments were examined. Preliminary analysis showed that Pro-Pro-Aro motifs are more conserved (either as Pro-Pro-Aro or Xaa-Pro-Aro, where Xaa is any amino acid) in sequence homologous proteins when the Pro-Pro ^{t}amide bond in the original protein structure is *cis* rather than when it is *trans*. For example, for tyrosinase (pdb code: 1wxc; containing ^{168}Pro-*cis*Pro-Tyr^{170}), alignment of homologous sequences (up to 37% sequence identity) obtained from a BLAST[30] search showed Pro-Pro-Tyr to be conserved with the general sequence motif Xaa-Pro-Aro (Figure 10a). As shown in Figure 10b, NMR spectra of Ac-Ala-

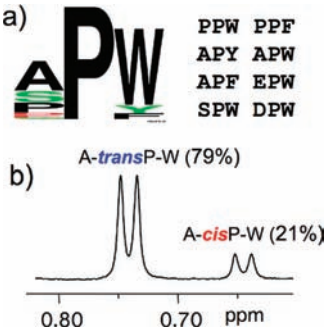


Figure 10. (a) Sequence logo plot[31] of a three-residue fragment of tyrosinase (^{168}Pro-*cis*Pro-Tyr^{170}; pdb code: 1wxc) and its sequence homologues (up to 37% sequence identity). Major sequence motifs are listed separately. (b) ^{1}H NMR subspectrum of Ac-Ala-Pro-Trp-NH_2 depicting the methyl (Ala) resonances in Ac-Ala-*trans*Pro-Trp-NH_2 and Ac-Ala-*cis*Pro-Trp-NH_2 isomers.

Pro-Trp-NH_2 (Xaa = Ala; Aro = Trp) showed the population of Ala-*cis*Pro to be about 21%, compatible with the *cis* content of Pro-Pro-Aro (Table 1), while the corresponding population of Ac-Ala-*cis*Pro-Ala-NH_2 was only 5%, also compatible with *cis* content of Pro-Pro-nonAro (Table 1). It has been previously shown that the Ala-*cis*Pro conformation is about 20% populated in Ac-Ala-Pro-Tyr-$NHCH_3$, another peptide with Ala-Pro-Aro motif.[12b] In other words, the Xaa-*cis*Pro ^{t}amide bond is conserved, sometimes as Pro-Pro-Aro but more generally as Xaa-Pro-Aro. It is not that the Pro-Pro-Aro sequence motif is conserved (as Pro-Pro-Aro or as Xaa-Pro-Aro) only in proteins containing Pro-*cis*Pro ^{t}amide bonds. Even when the Pro-Pro is *trans*, the Pro-Pro-Aro sequence motif is sometimes conserved among homologous proteins, probably implying either a functional role of the sequence motif or that the *cis* conformer is accessed by the protein transiently for proper folding or function.

■ CONCLUSION

From a systematic analysis of a peptide series with the general sequence Ac-Pro-Pro-Xaa-NH_2, we show that the populations of the Pro-*cis*Pro structural motif are significantly higher in peptides when Xaa is aromatic. This establishes Pro-Pro-Aro as a sequence motif that can stabilize the Pro-*cis*Pro conformation. As supported by direct (NOE) as well as indirect (ring-current effect on Pro1 C^{α}-H chemical shifts and $\chi 1$ restriction) evidence, the origin of this stabilization arises due to interaction between C^{α}-H of the first proline and the side chain of the third aromatic residue. The nature of the interaction is CH$\cdots\pi$ type, as evident from the pH-dependent variation of *cis*/*trans* population ratio in the peptide Ac-Pro-Pro-His-NH_2. When compared to another sequence motif that also stabilizes *cis*Pro, Aro-*cis*Pro, the CH$\cdots\pi$ interaction in Pro-Pro-*cis*Aro sequence motif was shown to be tighter, allowing direct observation of NOE cross peaks between C^{α}-H of the first proline residue and the side-chain atoms of the third aromatic residue. A complete thermodynamic scale, quantifying the effect of Xaa in stabilizing

dx.doi.org/10.1021/ja209334v | J. Am. Chem. Soc. 2012, 134, 4661−4669

Pro-*cis*Pro conformation in Pro-Pro-Xaa sequence motifs, is presented, and the relevance of this sequence motif in stabilizing *cis*Pro bonds in proteins is explored.

■ ASSOCIATED CONTENT

Supporting Information

Derivation of eq 1 and four tables (Tables S1−S4). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

gautamda@gmail.com or gautam@boseinst.ernet.in

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

G.B. acknowledges financial support from Department of Science and Technology, Government of India (SR/SO/BB/-11/2008).

■ REFERENCES

(1) Nishio, M.; Hirota, M.; Umezawa, Y. *The CH/π Interaction. Evidence, Nature and Consequences*, Wiley-VCH: New York, 1988.

(2) Brandl, M.; Weiss, M. S.; Jabs, A.; Suhnel, J.; Hilgenfeld, R. *J. Mol. Biol.* **2001**, *307*, 357−377.

(3) Plevin, M. J.; Bryce, D. L.; Boisbouvier, J. *Nat. Chem.* **2010**, *2*, 466−471.

(4) Dasgupta, B.; Chakrabarti, P.; Basu, G. *FEBS Lett.* **2007**, *581*, 4529−32.

(5) (a) Stewart, D. E.; Sarkar, A.; Wampler, J. E. *J. Mol. Biol.* **1990**, *214*, 253−260. (b) MacArthur, M. W.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 397−412. (c) Weiss, M. S.; Jabs, A.; Hilgenfeld, R. *Nat. Struct. Biol.* **1998**, *5*, 676. (d) Jabs, A.; Weiss, M. S.; Hilgenfeld, R. *J. Mol. Biol.* **1999**, *286*, 291−304. (e) Pal, D.; Chakrabarti, P. *J. Mol. Biol.* **1999**, *294*, 271−288.

(6) Maigret, B.; Perahia, D.; Pullman, B. *J. Theor. Biol.* **1970**, *29*, 275−291.

(7) Reimer, U.; Scherer, G.; Drewello, M.; Kruber, S.; Schutkowski, M.; Fischer, G. *J. Mol. Biol.* **1998**, *279*, 449−460.

(8) (a) Lorenzen, S.; Peters, B.; Goede, A.; Preissner, R.; Frommel, C. *Proteins* **2005**, *58*, 589−595. (b) Exarchos, K. P.; Exarchos, T. P.; Papaloukas, C.; Troganis, A. N.; Fotiadis, D. I. *BMC Bioinformatics* **2009**, *10*, 113. (c) Wathen, B.; Jia, Z. *J. Proteome Res.* **2008**, *7*, 145−153.

(9) (a) Tweedy, N. B.; Nair, S. K.; Paterno, S. A.; Fierke, C. A.; Christianson, D. W. *Biochemistry* **1993**, *32*, 10944−10949. (b) Mayr, L. M.; Willbold, D.; Rosch, P.; Schmid, F. X. *J. Mol. Biol.* **1994**, *240*, 288−293. (c) Dodge, R. W.; Scheraga, H. A. *Biochemistry* **1996**, *35*, 1548−1559. (d) Jin, L.; Stec, B.; Kantrowitz, E. R. *Biochemistry* **2000**, *39*, 8058−8066. (e) Wu, Y.; Matthews, C. R. *J. Mol. Biol.* **2002**, *322*, 7−13. (f) Guan, R. J.; Xiang, Y.; He, X. L.; Wang, C. G.; Wang, M.; Zhang, Y.; Sundberg, E. J.; Wang, D. C. *J. Mol. Biol.* **2004**, *341*, 1189−1204. (g) Birolo, L.; Malashkevich, V. N.; Capitani, G.; De Luca, F.; Moretta, A.; Jansonius, J. N.; Marino, G. *Biochemistry* **1999**, *38*, 905−913. (h) Forstner, M.; Muller, A.; Rognan, D.; Kriechbaum, M.; Wallimann, T. *Protein Eng.* **1998**, *11*, 563−568.

(10) Kemmink, J.; Creighton, T. E. *J. Mol. Biol.* **1993**, *234*, 861−878.

(11) (a) Grathwohl, C.; Wuthrich, K. *Biopolymers* **1976**, *15*, 2025−2041. (b) Yao, J.; Feher, V. A.; Espejo, B. F.; Reymond, M. T.; Wright, P. E.; Dyson, H. J. *J. Mol. Biol.* **1994**, *243*, 736−753. (c) Wu, W. J.; Raleigh, D. P. *Biopolymers* **1998**, *45*, 381−394.

(12) (a) Kemmink, J.; Creighton, T. E. *J. Mol. Biol.* **1995**, *245*, 251−260. (b) Nardi, F.; Kemmink, J.; Sattler, M.; Wade, R. C. *J. Biomol. NMR* **2000**, *17*, 63−77.

(13) Wüthrich, K. *NMR of Protiens and Nucleic Acids*; Wiley-Interscience Publication: New York, 1986.

(14) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(15) Wang, G.; Dunbrack, R. L. Jr. *Nucleic Acids Res.* **2005**, *33*, W94−98.

(16) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. *J. Comput. Chem.* **2005**, *26*, 1701−18.

(17) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.

(18) Berendsen, H. J. C.; Postma, J. P. M; DiNola, A.; Haak, J. R. *J. Chem. Phy.s* **1984**, *81*, 3684−3690.

(19) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 6933−6939.

(20) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M. *J. Comput. Chem.* **2004**, *24*, 1605−1612.

(21) Perutz, M. F. *Philos. Trans. R. Soc. London A* **1993**, *345*, 105−112.

(22) Umezawa, Y.; Tsuboyama, S.; Takahashi, H.; Uzawa, J.; Nishio, M. *Bioorg. Med. Chem.* **1999**, *7*, 2021−2026.

(23) Harigai, M.; Kataoka, M.; Imamoto, Y. *J. Am. Chem. Soc.* **2006**, *128*, 10646−10647.

(24) Meng, H. Y.; Thomas, K. M.; Lee, A. E.; Zondlo, N. J. *Biopolymers* **2006**, *84*, 192−204.

(25) Tatko, C. D.; Waters, M. L. *J. Am. Chem. Soc.* **2004**, *126*, 2028−2034.

(26) Kessler, H.; Bats, J. W.; Griesinger, C.; Koll, S.; Will, M.; Wagner, K. *J. Am. Chem. Soc.* **1988**, *110*, 1033−1049.

(27) (a) Rai, R.; Aravinda, S.; Kanagarajadurai, K.; Raghothama, S.; Shamala, N.; Balaram, P. *J. Am. Chem. Soc.* **2006**, *128*, 7916−28. (b) Chatterjee, B.; Saha, I.; Raghothama, S.; Aravinda, S.; Rai, R.; Shamala, N.; Balaram, P. *Chem.—Eur. J.* **2008**, *14*, 6192−204.

(28) Ferreon, J. C.; Hilser, V. J. *Protein Sci.* **2003**, *12*, 447−57.

(29) Pachler, K. G. R. *Spect. Chim. Acta* **1963**, *19*, 2085−2092.

(30) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**, *215*, 403−410.

(31) Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. *Genome Res.* **2004**, *14*, 1188−1190.