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

Glycine Rescue of β-Sheets from *cis*-Proline

Madhurima Das[§] and Gautam Basu*[,¶]

[¶]Department of Biophysics and [§]Bioinformatics Centre, Bose Institute, P-1/12 CIT Scheme VIIM, Kolkata 700054, India

Supporting Information

ABSTRACT: Proline is incompatible with ideal β-sheet geometry, and the incompatibility gets magnified when Pro assumes the *cis* peptidyl−prolyl conformation. We show that Gly appears with high propensity at pre-*cis*Pro positions in β-sheets and rescues the β-sheet from severe distortions by assuming a right-handed polyproline conformation (β_{PR}), effectively increasing the local β-sheet register by one residue. The united residue, Gly(β_{PR})-*cis*Pro, is evolutionarily conserved, functionally important, and dynamic in nature.

Canonical protein secondary structures such as α-helices or β-sheets rely on regular hydrogen-bonded scaffolds and show little preference for Pro[1] since it lacks an amide proton. When present in α-helices, Pro can induce kinks.[2] The constrained backbone dihedral angle φ (∼ −60°) of Pro can also distort β-sheets.[3] The structural incompatibility can be aggravated if the peptidyl−prolyl bond assumes the rare[4] and energetically unfavorable[5] *cis* conformation. Despite being incompatible, a small but significant proportion of Pro appears in β-sheets, in both *trans* and *cis* conformations.[6] By exploring local sequence and structural signatures of Pro residues in β-sheets in a non-redundant high-resolution protein structural database, specifically focusing on *cis*Pro and the associated pre-*cis*Pro residue, we identify a new structural united residue, Gly-*cis*Pro.

A pre-culled database consisting of 4922 protein chains (sequence identity: <25%; resolution: <2.0 Å; *R*-factors: <0.25) was obtained from the PISCES[7] server (March 18, 2011 list). The dataset contained 71 *cis*Pro and 4411 *trans*Pro residues in β-sheets (annotated with DSSP[8]). Surprisingly, ∼66% of the pre-*cis*Pro(β) positions were occupied by Gly (Figure S1). The over-representation is statistically significant since the propensity of Gly at pre-*cis*Pro(β) positions (compared to that at pre-*trans*Pro(β) positions) was very high (∼11; Figure 1a); so was the joint propensity of Gly-*cis*Pro in β-sheets (Figure 1b). The propensities of Gly and *cis*Pro in β-sheets are 0.67 and 0.12, respectively (Table S1). Therefore, the expected propensity of β-(Gly-*cis*Pro) is the joint propensity, 0.67 × 0.12 = 0.08. The observed propensity of Gly-*cis*Pro in β-sheets is 0.83 (Table S2), 10 times higher than expected. In comparison, the expected propensity of Val-Val (1.91 × 1.91 = 3.64; Table S1) and the observed propensity of Val-Val (3.3; Table S2) are very similar, indicating that Gly-*cis*Pro is a β-sheet sequence motif. Propensities of Gly to occur in any position in a Pro-containing β-strand also showed that Gly prefers the pre-*cis*Pro position (propensity: 7.1; *z*-value: 15.7) and slightly disfavors the pre-*trans*Pro position (propensity: 0.6; *z*-value: −7.8).

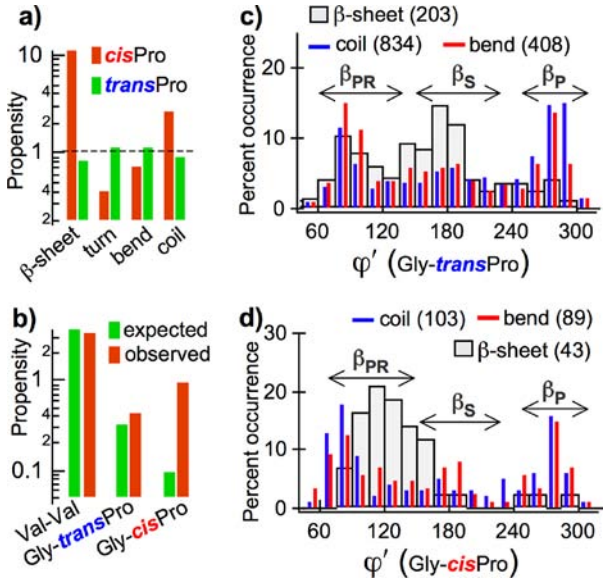

Figure 1. (a) Propensities of Gly at pre-*cis*Pro and pre-*trans*Pro positions as a function of secondary structure of Pro. (b) Expected and observed propensities of Val-Val, Gly-*trans*Pro, and Gly-*cis*Pro in β-sheets. Dihedral angle ($\varphi' = \varphi$ for $\varphi \geq 0°$; $\varphi' = 360° + \varphi$ for $\varphi < 0°$) distributions (bin size 15°) of Gly at (c) pre-*trans*Pro and (d) pre-*cis*Pro positions as a function of Pro secondary structure (total occurrences in each category are shown within parentheses).

The proline backbone strongly prefers the left-handed polyproline II (β_P), while its mirror image, the right-handed polyproline (β_{PR}), is primarily accessible only to Gly.[9] Dihedral angle distributions show that pre-*trans*Pro Gly residues prefer the regular β-strand (β_S) conformation (Figure 1c); in comparison, pre-*cis*Pro Gly residues in β-sheets prefer β_{PR} (Figure 1d). The pre-(*cis*/*trans*)Pro Gly residues, however, show no such differential preference when Pro assumes a bend or coil conformation (Figure 1c,d). Non-Gly residues occurring at pre-(*cis*/*trans*)Pro positions in β-sheets do not exhibit any preferred backbone conformation either (Figure S2). In other words, Gly(β_{PR})-*cis*Pro is not only a sequence motif but also a structural motif, acting like a united residue. We encountered six protein pairs that exhibited structural polymorphism of Gly-Pro containing β-sheets (Gly-*cis*Pro and Gly-*trans*Pro). The Ramachandran map of Gly-Pro in these polymorphs (Figure S3) showed that, irrespective of the Gly conformation in Gly-*trans*Pro, the conformation of Gly in Gly-*cis*Pro is always β_{PR},

Received: August 21, 2012
Published: September 18, 2012

   dx.doi.org/10.1021/ja308110t | *J. Am. Chem. Soc.* 2012, 134, 16536−16539

indicating the robustness of the Gly(β_{PR})-*cis*Pro structural motif.

The amide nitrogen of Pro cannot participate in regular β-sheet hydrogen bonding. One way to cope with this is to form a bulge around the Pro residue. Bulges are often associated with a local disruption of the canonical (i, i+2) β-sheet register. We estimated the local β-strand register around Xaa-Pro motifs in β-sheets using two parameters — d_{ij} (distance between C_β atoms of residues i and j), and θ_{ij} (angle between $C_\alpha(i)\rightarrow C_\beta(i)$ and $C_\alpha(j)\rightarrow C_\beta(j)$ bond vectors of two residues i and j). Four such i,j combinations are shown in Figure 2.

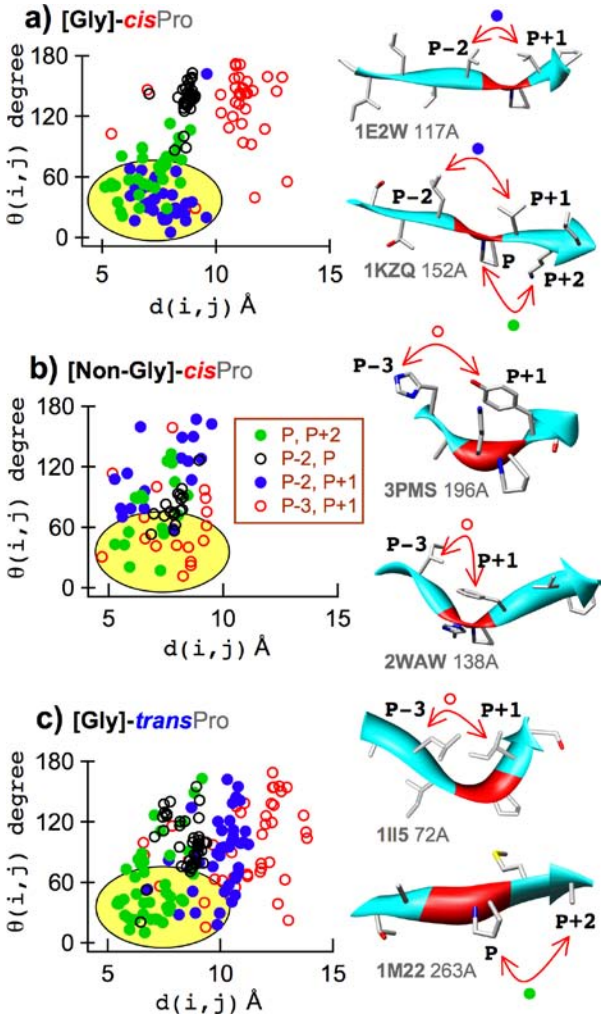

Figure 2. Plots of θ_{ij} versus d_{ij} corresponding to four (i, j) combinations (red, Pro-3, Pro+1; blue, Pro-2, Pro+1; green, Pro, Pro+2; black, Pro-2, Pro) in β-sheets containing (a) Gly-*cis*Pro, (b) non-Gly-*cis*Pro, and (c) Gly-*trans*Pro. Each panel is associated with two representative structures (with PDB codes followed by Pro residue numbers and chain ID). Pro and the pre-Pro residues are colored red.

In an ideal β-strand (φ = −120° and ψ = +120°), $d_{i,i+2} \approx 6.5$ Å and $\theta_{i,i+2} \approx 10°$. With some leeway for less-than-ideal β-sheets, we consider residues i and j to face the same side of the sheet and proximal enough to interact if $d_{ij} < 10$ Å and $\theta_{ij} < 60°$ (yellow highlighted regions of Figure 2). Figure 2a shows that residues X and Y in β-(X_i-Gly-*cis*Pro-Y_{i+3}) face the same side of the β-sheet and are within interacting distance. Thus, a Gly residue preceding *cis*Pro in β-sheets causes a local shift in the β-strand register by one residue, from (i, i+2) to (i, i+3). This brings (Pro − 2) and (Pro + 1) residues on the same side of the β-sheet. The Gly-*cis*Pro-containing β-sheets are also reasonably regular, as indicated by the clear clustering of θ_{ij}/d_{ij} combinations (Figure 2a).

Although $d_{\mathrm{Pro-2,Pro+1}}$ in non-Gly-*cis*Pro motifs (Figure 2b) shows a distribution similar to that in Gly-*cis*Pro motifs (Figure 2a), the corresponding angles are much higher than 60°. This indicates that (Pro−2) and (Pro+1) face away from each other when Gly is replaced by non-Gly in Gly-*cis*Pro motifs. Instead, the distance−angle distribution (Figure 2b) indicates proximity between the ith and (i+4)th residues (Pro−3 and Pro+1), a characteristic of bends. Likewise, Gly-*trans*Pro-containing β-sheets are often distorted or irregular, as indicated by the poor clustering of θ_{ij}/d_{ij} combinations (Figure 2c). For Gly-*trans*Pro units, the only i−j residue combination for which the θ_{ij}/d_{ij} distribution is restricted to the yellow highlighted region is when i = Pro and j = Pro+2, a "normal" (i, i+2) interaction expected in a regular β-sheet.

The local β-sheet register displacement around Gly-*cis*Pro and the lack (often) of hydrogen bonding by the Gly-*cis*Pro motif (Figure S4) indicates that these motifs may actually participate in β-bulges. The β-bulge has been extensively studied[10] and classified. We identified a total of 4201 bulges in our database using the program PROMOTIF 3.0,[11] of which 48 bulges contained *cis*Pro and 600 bulges contained *trans*Pro. A comparative analysis (Figure 3) showed that the majority

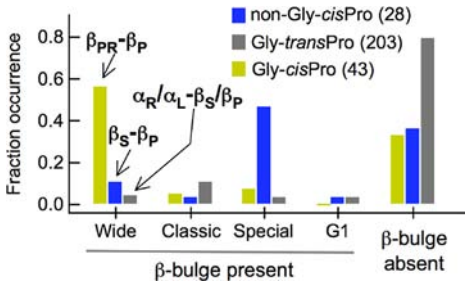

Figure 3. Relative populations of different classes of β-bulges formed by non-Gly-*cis*Pro, Gly-*trans*Pro, and Gly-*cis*Pro motifs in β-sheets (total occurrences given in parentheses).

(~79%) of Gly-*trans*Pro motifs do not participate in any bulges, while the majority of Gly-*cis*Pro (~67%) and non-Gly-*cis*Pro (~64%) motifs participate in bulges. However, the types of bulges in which they participate are very different — Gly-*cis*Pro motifs mostly participate in wide-type bulges, while most non-Gly-*cis*Pro motifs participate in special bulges. The wide-type bulges involving Gly-*cis*Pro do not belong to either of the two known subtypes,[10b] α_L-β or β-α_L. Instead, they comprise a new subtype of wide bulges, β_{PR}-*cis*β_P (Figure S5a). Similarly, the wide-type bulges formed by non-Gly-*cis*Pro motifs (Figure S5b) also belong to a new subtype, β_S-*cis*β_P.

Having established β-(Gly-*cis*Pro) to be a sequence and structural motif, we explored the extent of sequence conservation of Gly and Pro in homologous proteins. Our analysis is based on ConSurf-DB,[12] a repository for evolutionary conservation analysis of known protein structures based on phylogenetic relationships between closely related proteins that ranks individual residues in a 1−9 conservation index (color index) scale, where 9 corresponds to the highest degree of conservation. We focus on sequence conservation of Gly and Pro, both as individual residues and as a pair.

dx.doi.org/10.1021/ja308110t | J. Am. Chem. Soc. 2012, 134, 16536−16539

Histograms of summed (Gly + Pro) conservation indices of Gly-*cis*Pro and Gly-*trans*Pro motifs, with Pro present in β-sheets, bends, and coils, are shown in Figure 4a. Histograms

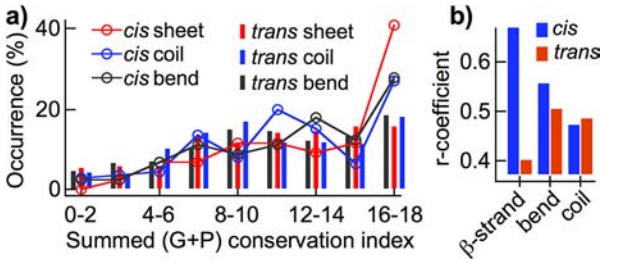

Figure 4. (a) Distribution of summed (Gly + Pro) conservation indices of Gly-(*cis*/*trans*)Pro as a function of Pro secondary structure. (b) Pearson's correlation coefficient (r) between conservation indices of Gly and Pro in Gly-(*cis*/*trans*)Pro as a function of Pro secondary structure.

corresponding to Gly-*trans*Pro are very similar, irrespective of *trans*Pro secondary structure. On the other hand, histograms corresponding to Gly-*cis*Pro show a spike at conservation index 16−18 (most conserved Gly-Pro pair), indicating that the fraction of the most conserved Gly-*cis*Pro pairs is larger than the fraction of Gly-*cis*Pro pairs that are less conserved. The percentage of the most conserved Gly-*cis*Pro pairs is also dependent on the secondary structure of *cis*Pro — 28% when the secondary structure of Pro is bend/coil and ∼41% when Pro is present in a β-sheet. The percentage of the most conserved Gly-*trans*Pro motifs, on the other hand, is much less (∼18%) and independent of the secondary structure of *trans*Pro. The results indicate that the Gly-*cis*Pro motif, when present in β-sheets, exhibits the highest degree of evolutionary conservation, significantly higher than that of Gly-*cis*Pro present in other secondary structures.

To genuinely qualify as a sequence motif, the presence of Gly and Pro in evolutionarily related proteins must also be correlated. Pearson's correlation coefficient r between conservation indices of Gly and Pro in Gly-Pro motifs, computed as a function of the *cis*−*trans* conformation and the secondary structure of Pro, is highest (r = 0.67) for Gly-*cis*Pro and lowest (r = 0.4) for Gly-*trans*Pro when Pro is present in β-sheets. In comparison, Gly-*cis*Pro and Gly-*trans*Pro are characterized by intermediate r-values (Figure 4b) that are not significantly different from each other when the Pro assumes the bend or the coil secondary structure. This clearly indicates that the Gly-*cis*Pro pair acts as a united residue in β-sheets.

In proteins for which the β-(Gly-*cis*Pro) summed conservation index is not in the highest range (for example, PDB IDs 1mya and 2jlq), the Gly-*cis*Pro motif was often found to be conserved in a clade-specific manner in the phylogenetic tree, emphasizing structural and/or functional importance of the motif (Figures S6 and S7). The ease of *cis*−*trans* isomerization of Pro in β-(Gly-Pro) motifs (Figure S3a), without any major changes in global structure, suggests that the motif might play an important role in local rearrangement of binding interfaces such that a single protein can bind multiple partners in a context-dependent manner. For example, the Gly-Pro motif on the surface of ribosomal S10 protein assumes a Gly-*trans*Pro conformation (PDB ID 2avy)[13a] when bound to ribosomal RNA and a Gly-*cis*Pro conformation (PDB ID 3d3b)[13b] when bound to NusB, with an overlapping binding site that includes the Gly-Pro motif. Interestingly, the Gly backbone dihedral angles are similar in the two structures, while backbone dihedral angles of three residues preceding the Gly-Pro motif are markedly different in 3d3b (compared to 2avy; Figure S3b). A non-isomerizing Gly-*cis*Pro motif may also play important roles. For example, a Gly-*cis*Pro motif, completely conserved and present at the dimer interface in D-Tyr-tRNA^{Tyr} deacylase, not only is important in inter-monomer interactions:[14a] the Pro residue also has been implicated to play an important role in tRNA interaction[14b] and as a chaperone[14c] for a conserved Met residue with invariant conformation that probably plays a role in substrate chiral selectivity. Another example is the flavivirus NS3 protein (PDB ID 2jlq),[15] in which the *cis*Pro residue of the Gly-*cis*Pro motif has been implicated in stacking interactions with RNA bases, responsible for RNA unwinding. Interestingly, the Gly-*cis*Pro motif in the viral protein is conserved in a clade-specific manner (Figure S7), implying further subtleties in structure−function relationship.

Although Pro and Gly are the least preferred residues in β-sheets, this study shows that together, as a united residue, the β-sheet propensity of Gly-*cis*Pro is increased 10-fold over the value expected on the basis of individual propensities of Gly and *cis*Pro. A Gly residue, preceding *cis*Pro in a β-sheet, rescues an otherwise regular β-sheet from *cis*Pro-induced distortions by assuming the β_{PR} conformation and shifts the local β-sheet register by one residue. The "rescue" is reminiscent of aromatic rescue of glycine in β-sheets.[16] The β-(Gly-*cis*Pro) motif, the first example of a united residue in proteins, is highly conserved, functionally important, and dynamic in nature.

■ ASSOCIATED CONTENT

Supporting Information

Figures S1−S7, Tables S1 and S2, and details of propensity calculations. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

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

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by DST, India. M.D. acknowledges financial assistance from DBT, India. G.B. thanks George Rose, Jöel Janin, Gourishankar Ghosh, Akinori Kidera, Saumya Dasgupta and Raghavan Varadarajan for a critical reading of a preliminary draft.

■ REFERENCES

(1) Chou, P.; Fasman, G. *Biochemistry* **1974**, *13*, 222−245.
(2) MacArthur, M.; Thornton, J. *J. Mol. Biol.* **1991**, *218*, 397−412.
(3) Daffner, C.; Chelvanayagam, G.; Argos, P. *Protein Sci.* **1994**, *3*, 876−882.
(4) (a) Stewart, D.; Sarkar, A.; Wampler, J. *J. Mol. Biol.* **1990**, *214*, 253−260. (b) Pal, D.; Chakrabarti, P. *J. Mol. Biol.* **1999**, *294*, 271−288. (c) Ganguly, B.; Chattopadhyay, S.; Chakrabarti, P.; Basu, G. *J. Am. Chem. Soc.* **2012**, *134*, 4661−4669.
(5) (a) Ramachandran, G. N.; Mitra, A. *J. Mol. Biol.* **1976**, *107*, 85−92. (b) Zimmerman, S. S.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 408−416.
(6) Pahlke, D.; Freund, C.; Leitner, D.; Labudde, D. *BMC Struct. Biol.* **2005**, *5*, 8.
(7) Wang, G.; Dunbrack, R. *Bioinformatics* **2003**, *19*, 1589−1591.
(8) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577−2637.

(9) Ho, B. K.; Brasseur, R. *BMC Struct. Biol.* **2005**, *5*, 14.
(10) (a) Richardson, J. S.; Getzoff, E. D.; Richardson, D. C. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 2574−2578. (b) Chen, A. W. E.; Hutchinson, E. G.; Harris, D.; Thornton, J. M. *Protein Sci.* **1993**, *2*, 1574−1590.
(11) Hutchinson, E. G.; Thornton, J. M. *Protein Sci.* **1996**, *5*, 212−220.
(12) Goldenberg, O.; Erez, E.; Nimrod, G.; Ben-Tal, N. *Nucleic Acids Res.* **2009**, *37*, D323−327.
(13) (a) Schuwirth, B. S.; Borovinskaya, M. A.; Hau, C. W.; Zhang, W.; Vila-Sanjurjo, A.; Holton, J. M.; Cate, J. H. *Science* **2005**, *310*, 827−834. (b) Luo, X.; Hsiao, H. H.; Bubunenko, M.; Weber, G.; Court, D. L.; Gottesman, M. E.; Urlaub, H.; Wahl, M. C. *Mol. Cell* **2008**, *32*, 791−802.
(14) (a) Ferri-Fioni, M.; Schmitt, E.; Soutourina, J.; Plateau, P.; Mechulam, Y.; Blanquet, S. *J. Biol. Chem.* **2001**, *276*, 47285−47290. (b) Lim, K.; Tempczyk, A.; Bonander, N.; Toedt, J.; Howard, A.; Eisenstein, E.; Herzberg, O. *J. Biol. Chem.* **2003**, *278*, 13496−502. (c) Bhatt, T. K.; Yogavel, M.; Wydau, S.; Berwal, R.; Sharma, A. *J. Biol. Chem.* **2010**, *285*, 5917−5930.
(15) Luo, D.; Xu, T.; Watson, R. P.; Scherer-Becker, D.; Sampath, A.; Jahnke, W.; Yeong, S. S.; Wang, C. H.; Lim, S. P.; Strongin, A.; Vasudevan, S. G.; Lescar, J. *EMBO J.* **2008**, *27*, 3209−3219.
(16) Merkel, J. S.; Regan, L. *Fold. Des.* **1998**, *3*, 449−455.