FOR THE RECORD

Preferred proline puckerings in *cis* and *trans* peptide groups: Implications for collagen stability

LUIGI VITAGLIANO,¹ RITA BERISIO,¹ ANTONIO MASTRANGELO,²
LELIO MAZZARELLA,^{1,2,3} AND ADRIANA ZAGARI^{1,2,3}

¹Centro di Studio di Biocristallografia, CNR, I-80134 Napoli, Italy

²Dipartimento di Chimica, Università degli Studi di Napoli "Federico II", I-80126 Napoli, Italy

³CEINGE, Biotecnologie avanzate Scarl, Napoli, Italy

(RECEIVED July 3, 2001; FINAL REVISION September 11, 2001; ACCEPTED September 12, 2001)

Abstract

The interplay between side-chain and main-chain conformations is a distinctive characteristic of proline residues. Here we report the results of a statistical analysis of proline conformations using a large protein database. In particular, we found that proline residues with the preceding peptide bond in the *cis* state preferentially adopt a down puckering. Indeed, out of 178 *cis* proline residues, as many as 145 (81%) are down. By analyzing the 1–4 and 1–5 nonbonding distances between backbone atoms, we provide a structural explanation for the observed trend. The observed correlation between proline puckering and peptide bond conformation suggests a new mechanism to explain the reported shift of the *cis-trans* equilibrium in proline derivatives. The implications of these results for the current models of collagen stability are also discussed.

Keywords: Proline; *cis* peptide; hydroxyproline; statistical analysis; collagen

Proline is unique among the genetically encoded amino acids (Richardson and Richardson 1989). The covalent bond between the side-chain and the nitrogen backbone atom has relevant structural consequences on the properties of the residue. In particular, (1) the cyclic nature of proline imposes severe restrictions to the conformational freedom of this residue; (2) there is a strict dependence between main- and side-chain proline conformations; (3) the *cis* state of the peptide group preceding proline residues is significantly populated; and (4) the lack of a backbone hydrogen-bonding donor limits proline capability to form main-chain hydrogen bonds.

Because of these intrinsic properties, the presence of proline strongly influences the protein structure elements embodying this residue. Indeed, it is well known that proline residues cannot fit into the regular, internal parts of either α -helices or β -sheets (Richardson and Richardson 1989). In contrast, proline residues are easily accommodated in the collagen triple helix (Beck and Brodsky 1998). Finally, the insertion of proline in loop regions may induce the large conformational changes characteristic of three-dimensional domain swapping (Mazzarella et al. 1995; Bergdoll et al. 1997; Liu et al. 2001).

Proline residues have been the object of several statistical analyses using both protein and oligopeptide databases (Balasubramanian et al. 1971; Stewart et al. 1990; MacArthur and Thornton 1991; Milner-White et al. 1992; Weiss et al. 1998). In this context, the analysis of a limited number of protein structures has indicated that there is a correlation between proline puckering and the *cis/trans* peptide bond conformation (Milner-White et al. 1992). Here we report a more extensive investigation regarding this issue by using a large protein structure database. Furthermore, we propose a

Reprint requests to: Luigi Vitagliano, Centro di Studio di Biocristallografia, CNR, Via Mezzocannone 6, I-80134, Napoli, Italy; e-mail: luigiv@chemistry.unina.it; fax: 39-0812536603.

Abbreviations: cisPro and transPro, proline residues with the peptide bond of the preceding residue in the *cis* and *trans* conformations, respectively; upPro and downPro, proline residues with up and down side-chain puckerings, respectively; 4R-Hyp, 4R-hydroxyproline; 4S-Hyp, 4S-hydroxyproline; 4R-Flp, 4R-fluoroproline; 4S-Flp, 4S-fluoroproline

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1101/ps.26601>.

structural explanation for the observed correlation between proline side-chain conformation and the peptide group state. Finally, we discuss these findings in relation to the *cis-trans* equilibrium of proline derivatives (Eberhardt et al. 1996; Bretscher et al. 2001) and to the current collagen stabilization models (Bella et al. 1994; Holmgren et al. 1998; Vitagliano et al. 2001).

Results and Discussion

Statistical analyses on proline residues

Statistical analyses on proline conformations were performed on a database containing nonredundant protein chains sharing sequence homologies <25%. Four hundred seventy-eight protein chains were selected from the Protein Data Bank (Berman et al. 2000) by considering only structures refined to a resolution better than 2.0 Å and an R-factor <0.23. Because reliable estimates of proline side-chain conformations require accurate determinations of the local region embodying the residue, additional selecting criteria, based on B-factors and puckering amplitude of each proline, were also applied (see Materials and Methods). The final database contained 3424 prolines, which were classified according to their conformational properties. In particular, for 3246 and 178 prolines, the peptide bond preceding the residue was in *trans* (transPro) and *cis* (cisPro) state, respectively. The percentage of cisPro (5.2%) is in close agreement with some previous surveys performed on protein structure databases (MacArthur and Thornton 1991; Weiss et al. 1998), although slightly higher values have been reported in other earlier studies (Stewart et al. 1990; Milner-White et al. 1992). The proportion of cisPro found in the present study is, however, significantly different from that derived from analyses performed on peptide structure databases. Indeed, it has been reported that the percentage of cisPro may be as high as 35% in oligopeptides (Milner-White et al. 1992). To identify possible causes for this discrepancy, we searched the Cambridge Structural Database (Allen et al. 1979) for *cis* prolines in cyclic and acyclic peptides. The results of this analysis show that the proportion of cisPro is 57.4% and 5.6% in cyclic and acyclic peptides, respectively. As expected, this finding indicates that the high proportion of cisPro is characteristic of cyclic peptides, in which the conformational restrictions imposed by the ring closure enhance the occurrence of the *cis* state.

Several statistical and energetic studies have shown that proline side-chains adopt two different conformational states usually denoted as up (upPro) and down (downPro; Némethy et al. 1992). UpPro and downPro can be identified by considering the distribution of proline side-chain dihedral angles. In particular, upPro is characterized by negative values of χ^1 and χ^3 and positive values of χ^2 and χ^4 . The distribution of upPro and downPro in our database shows

that the two states are equally populated. Indeed, out of 3424 prolines, 1738 (51%) are downPro, whereas 1686 are upPro. In the Cremer and Pople (1975) definition, upPro and downPro adopt phase angles clustered around 282° and 93°, respectively. These values are similar to those adopted by a regular five-membered cycle in the C $^{\gamma}$ -exo envelope (288°) and twisted C $^{\gamma}$ -endo-C $^{\beta}$ -exo (90°) conformations. In agreement with previous reports (Némethy et al. 1992; Vitagliano et al. 2001), upPro and downPro also show significantly different main-chain dihedral angles (Fig. 1A,B). Indeed, the average values of φ are -58.7° and -69.8° for upPro and downPro, respectively.

Fig. 1. φ versus χ_1 (A) and ψ versus χ_1 (B) in transPro (open circles) and cisPro (solid squares). It should be noted that only ψ angles >100° have been plotted.

Subsequently, we analyzed the distribution of *cis*Pro and *trans*Pro among upPro and downPro. As many as 145 *cis*Pro adopt a down conformation, whereas only 33 *cis*Pro are upPro (Fig. 2). In contrast, the populations of down and up conformers are comparable for *trans*Pro. Similar preferences were obtained in previous studies performed on a more limited database constituted of 68 protein structures (Milner-White et al. 1992). It is worth noting that *cis*Pro adopt φ angles characterized by more negative values (Figs. 1A, 2). Interestingly, conformations with $\varphi > -60^\circ$ are widely populated among *trans*Pro and essentially forbidden for *cis*Pro. Because φ angles of downPro are significantly more negative than φ angles of upPro (Fig. 1A), it is not surprising that *cis*Pro preferentially adopts the down state.

Why should more negative values of the φ angle be required in *cis* prolines? As expected on the bases of pure geometrical considerations, several 1–4 and 1–5 distances between backbone atoms strongly depend on the value of the φ angle (Fig. 3A,B). In particular, we focused our attention on the distances in the $C_{i-1}-C_i$, $C^{\alpha}_{i-1}-C_i$, and $O_{i-1}-C_i$ pairs (Figs. 3, 4). As shown in Figure 4, all of these distances increase as φ assumes more negative values in both *cis* and *trans* conformations. Therefore, because downPro adopts more negative φ angles, these nonbonding distances are usually larger in downPro than in upPro (Fig. 4). The larger distances found in downPro may attenuate the severe steric hindrance occurring particularly in the *cis* state between the C^{α} atom (and its substituents) of the residue that precedes the proline and the carbonyl oxygen of the proline (Fig. 4D). On the other hand, the interaction between O_{i-1} and C_i occurring in *trans*Pro has a potential stabilizing effect (Bretscher et al. 2001). This allows rather short distances and thus large values of φ (Fig. 4C). Con-

Fig. 2. φ versus ω in downPro (open triangles) and upPro (solid triangles).

Fig. 3. Schematic representation of 1–4 and 1–5 nonbonding backbone distances occurring in *trans*Pro (A) and *cis*Pro (B).

sequently, both up and down states are compatible with *trans* conformations.

These observations do not necessarily imply that the presence of bulky residues preceding proline disfavors *cis* conformations, as specific interactions involving side-chains may occur. In fact, in our database ~20% of residues preceding *cis*Pro are aromatic, in accordance with previous reports (Reimer et al. 1998; Jabs et al. 1999). The high frequency of aromatic residues preceding *cis*Pro may be attributed to the $CH\cdots\pi$ interactions between the proline and the aromatic rings, as reported by Jabs et al. (1999).

A new mechanism to explain the cis-trans equilibrium observed in proline derivatives

The *cis-trans* equilibrium of substituted pyrrolidine rings has been the object of several recent investigations for its implications on collagen stability. In a series of interesting studies, Raines and coworkers have shown that the presence of electron-withdrawing atoms on proline ring may significantly affect the *cis-trans* equilibrium of the preceding peptide group (Eberhardt et al. 1996; Bretscher et al. 2001).

Fig. 4. Dependence of 1–4 and 1–5 nonbonding distances in downPro (open triangles) and upPro (solid triangles) on φ : C_{i-1} – C_i distances in transPro (A), C_{i-1} – C_i distances in cisPro (B), O_{i-1} – C_i distances in transPro (C), and C_{i-1}^α – C_i distances in cisPro (D).

Indeed, they have shown that this equilibrium is more shifted toward the *trans* state in 4R-hydroxyproline (4R-Hyp) and 4R-fluoroproline (4R-Flp) than in proline (Eberhardt et al. 1996). Interestingly, in 4S-hydroxyproline (4S-Hyp) and 4S-fluoroproline (4S-Flp) the equilibrium is shifted toward the *cis* state (Bretscher et al. 2001). These experimental data have been interpreted considering the interactions of the proline carbonyl group with the carbonyl group of the residue preceding the proline and with the ring substituent (Bretscher et al. 2001).

We propose an additional mechanism that explains the experimental data on the bases of the above described correlation between proline puckering and peptide bond con-

formation. Several experimental reports and statistical analyses have indicated that 4R-Hyp and 4R-Flp preferentially adopt the up state (Holmgren et al. 1999; Vitagliano et al. 2001). Because we have shown that up conformers strongly adverse the *cis* state, the shift observed for 4R-Hyp and 4R-Flp toward the *trans* state (Eberhardt et al. 1996) is straightforwardly explained. On the other hand, there are indications that 4S-Hyp and 4S-Flp preferentially adopt the down state (Gerig and McLeod 1973; Shamala et al. 1976; Holmgren et al. 1999). Therefore, they are more prone to assume the *cis* state.

Although other effects such as interactions of backbone atoms and pyrrolidine substituents may play a role in the

cis-trans equilibrium, the above described mechanism represents a reliable explanation for the observed experimental data (Eberhardt et al. 1996; Bretscher et al. 2001). Furthermore, the mechanism highlights the interplay between the conformational properties of backbone and side-chain in imino acids. Because the conformational preferences of the substituted imino acid side-chains are likely driven by stereo-electronic effects (Holmgren et al. 1999; Bretscher et al. 2001), our mechanism emphasizes the importance of inductive effects in directing the conformational properties of proteins.

Implications for collagen stability

The structural role of hydroxyproline residues on collagen stability is a long-standing issue that recently has received much attention. Earlier hypotheses on the role of the hydroxyproline were essentially based on its interaction with the solvent (Ramachandran et al. 1973; Privalov 1982). Although high-resolution X-ray structures of collagen-like polypeptides have indeed shown that hydroxyproline residues strongly interact with the solvent (Bella et al. 1994, 1995; Berisio et al. 2001), recent findings have shifted the attention toward other effects. Indeed, Raines and coworkers (Holmgren et al. 1998, 1999) have indicated that collagen stabilization induced by hydroxyproline relies on the shift of the *cis-trans* equilibrium toward the *trans* state, which is the one required in collagen triple helix. This stabilization model does not explain the inability of polypeptides with repeating sequence Hyp-Pro-Gly to fold in triple helix (Inouye et al. 1982), even when the three chains are covalently linked (Henkel et al. 1999). Based on the observation that upPro and downPro φ dihedral angles suited the X and Y positions, respectively, in the repetitive X-Y-Gly sequence of the collagen triple helix and that 4R-Hyp preferentially adopts an up state, we have recently proposed a new stabilization model (Vitagliano et al. 2001). This model is able to explain the effects of the presence of 4R-Hyp in both the X and Y positions on the stability of imino acid-rich collagen-like polypeptides. The suitability of φ angles of downPro and upPro to the X and Y positions, respectively, is confirmed in the present statistical analysis. Furthermore, we show that ψ angles typical of imino acids in the X position ($\sim 160^\circ$) are allowed mainly to downPro (Fig. 1B).

It is worth mentioning that very recent comparative studies on (Pro-Thr-Gly)₁₀ and (4R-Hyp-Thr-Gly)₁₀ have shown that 4R-Hyp in the X position may have stabilizing effects on the triple helix (Bann and Bächinger 2000). Based on this observation, it can be surmised that intrinsic conformational properties of imino acids may play a crucial role in the stabilization of imino acid-rich collagen polypeptides, in which even small differences in main-chain dihedral angles may not be tolerated. In contrast, main-chain distor-

tions may be more easily accommodated in less rigid collagen-like polypeptides as (Pro-Thr-Gly)₁₀. Therefore, it is likely that for these polypeptides, other effects such as the shifting of the *cis-trans* equilibrium and networking through water molecules may explain the stabilizing role of 4R-Hyp. Because real collagen molecules consist of both regions with low and high content of imino acids, this observation leads to the suggestive hypothesis that stabilization effects of 4R-Hyp may be different in the distinct regions of the collagen molecule. This idea is supported by the finding that the triple helical parameters are different in regions with high and low content of imino acids (Kramer et al. 1999) and that Gly-Pro-Hyp and Gly-X-Hyp triplets contribute differently to the triple helix stability (Burjanadze 2000).

Materials and methods

The May 1999 release of the Protein Data Bank (Berman et al. 2000) was searched for proline residues. In particular, 478 nonredundant protein chains having sequence homology <25% and belonging to structures refined to a resolution better than 2.0 Å and to an R-factor <0.23 were considered. These chains contained 5168 proline residues, which were further selected according to their B-factors and puckering amplitudes. Proline residues showing average B-factors >1.2 times the average B-factor of the protein structure they belonged to were rejected. In addition, the Cremer and Pople (1975) algorithm was used to exclude proline residues with an unusual puckering (Hooft et al. 1996). Indeed, only proline residues with puckering amplitudes in the range of 0.20–0.45 were included in the final database, which contained 3424 proline residues.

Up or down puckerings were assigned depending on the combination of pyrrolidine ring dihedral angles. Namely, proline rings with $(\chi^1 + \chi^3 - \chi^2 - \chi^4) > 40^\circ$ were regarded as down, whereas those with $(\chi^1 + \chi^3 - \chi^2 - \chi^4) < -40^\circ$ were regarded as up.

Acknowledgments

We are indebted to Mr. Luca De Luca for technical assistance. This work was financially supported by the Agenzia Spaziale Italiana and by the CNR Agenzia2000.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Allen, F.H., Bellard, S., Brice, M.D., Cartwright, B.A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B.G., Kennard, O., Motherwell, W.D.S., Rodgers, J.R., and Watson, D.G. 1979. The Cambridge Crystallographic Data Centre: Computer-based search, retrieval, analysis and display of information. *Acta Crystallogr. B* **35**: 2331–2339.
- Balasubramanian, R., Lakshminarayanan, A.V., Sabesan, M.N., Tegoni, G., Venkatesan, K., and Ramachandran, G.N. 1971. Studies on the conformation of amino acids, VI: Conformation of the proline ring as observed in crystal structures of amino acids and peptides. *Int. J. Protein Res.* **3**: 25–33.
- Bann, J.G. and Bächinger, H.P. 2000. Glycosylation/hydroxylation-induced stabilization of the collagen triple helix: 4-*trans*-hydroxyproline in the Xaa position can stabilize the triple helix. *J. Biol. Chem.* **275**: 24466–24469.
- Beck, K. and Brodsky, B. 1998. Supercoiled protein motifs: The collagen triple-helix and the α -helical coiled coil. *J. Struct. Biol.* **122**: 17–29.

- Bella, J., Eaton, M., Brodsky, B., and Berman, H.M. 1994. Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science* **266**: 75–81.
- Bella, J., Brodsky, B., and Berman, H.M. 1995. Hydration structure of a collagen peptide. *Structure* **3**: 893–906.
- Bergdoll, M., Remy, M.H., Cagnon, C., Masson, J.M., and Dumas, P. 1997. Proline-dependent oligomerization with arm exchange. *Structure* **5**: 391–401.
- Berisio, R., Vitagliano, L., Mazzarella, L., and Zagari, A. 2001. Crystal structure determination of the collagen-like polypeptide with repeating sequence Pro-Hyp-Gly: Implications for hydration. *Biopolymers* **56**: 8–13.
- Berman, H.M., Bhat, T.N., Bourne, P.E., Feng, Z., Gilliland, G., Weissig, H., and Westbrook, J. 2000. The Protein Data Bank and the challenge of structural genomics. *Nat. Struct. Biol.* **7**: 957–959.
- 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.
- Burjanadze, T.V. 2000. New analysis of the phylogenetic change of collagen thermostability. *Biopolymers* **53**: 523–528.
- Cremer, D. and Pople, J.A. 1975. General definition of ring puckering coordinates. *J. Amer. Chem. Soc.* **97**: 1354–1358.
- Eberhardt, E.S., Panisik, N.J., and Raines, R.T. 1996. Inductive effects on the energetics of prolyl peptide bond isomerization: Implications for collagen folding and stability. *J. Am. Chem. Soc.* **118**: 12261–12266.
- Gerig, J.T. and McLeod, R.S. 1973. Conformations of *cis*- and *trans*-4-fluoro-L-proline in aqueous solution. *J. Am. Chem. Soc.* **95**: 5725–5729.
- Henkel, W., Vogl, T., Echner, H., Voelter, W., Urbanke, C., Schleuder, D., and Rauterberg, J. 1999. Synthesis and folding of native collagen III model peptides. *Biochemistry* **38**: 13610–13622.
- Holmgren, S.K., Bretscher, L.E., Taylor, K.M., and Raines, R.T. 1999. A hyperstable collagen mimic. *Chem. Biol.* **6**: 63–70.
- Holmgren, S.K., Taylor, K.M., Bretscher, L.E., and Raines, R.T. 1998. Code for collagen's stability deciphered. *Nature* **392**: 666–667.
- Hoof, R.W., Vriend, G., Sander, C., and Abola, E.E. 1996. Errors in protein structures. *Nature* **381**: 272.
- Inouye, K., Kobayashi, Y., Kyogoku, Y., Kishida, Y., Sakakibara, S., and Prockop, D. J. 1982. Synthesis and physical properties of (hydroxyproline-proline-glycine)₁₀: Hydroxyproline in the X-position decreases the melting temperature of the collagen triple helix. *Arch. Biochem. Biophys.* **219**: 198–203.
- Jabs, A., Weiss, M.S., and Hilgenfeld, R. 1999. Non-proline *cis* peptide bonds in proteins. *J. Mol. Biol.* **286**: 291–304.
- Kramer, R.Z., Bella, J., Mayville, P., Brodsky, B., and Berman, H.M. 1999. Sequence dependent conformational variations of collagen triple-helical structure. *Nat. Struct. Biol.* **6**: 454–457.
- Liu, Y., Gotte, G., Libonati, M., and Eisenberg, D. 2001. A domain-swapped RNase A dimer with implications for amyloid formation. *Nat. Struct. Biol.* **8**: 211–214.
- MacArthur, M.W. and Thornton, J.M. 1991. Influence of proline residues on protein conformation. *J. Mol. Biol.* **218**: 397–412.
- Mazzarella, L., Vitagliano, L., and Zagari, A. 1995. Swapping structural determinants of ribonucleases: An energetic analysis of the hinge peptide 16–22. *Proc. Natl. Acad. Sci.* **92**: 3799–3803.
- Milner-White, E.J., Bell, L.H., and Maccallum, P.H. 1992. Pyrrolidine ring puckering in *cis* and *trans*-proline residues in proteins and polypeptides: Different puckers are favoured in certain situations. *J. Mol. Biol.* **228**: 725–734.
- Némethy, G., Gibson, K.D., Palmer, K.A., Yoon, C.N., Paterlini, G., Zagari, A., Rumsey, S., and Scheraga, H.A. 1992. Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in ECEPP/3 algorithm, with application to proline-containing peptides. *J. Phys. Chem.* **96**: 6472–6484.
- Privalov, P.L. 1982. Stability of proteins: Proteins which do not present a single cooperative system. *Adv. Protein Chem.* **35**: 1–104.
- Ramachandran, G.N., Bansal, M., and Bhatnagar, R.S. 1973. A hypothesis on the role of hydroxyproline in stabilizing collagen structure. *Biochim. Biophys. Acta* **322**: 166–171.
- Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M., and Fischer, G. 1998. Side-chain effects on peptidyl-prolyl *cis/trans* isomerisation. *J. Mol. Biol.* **279**: 449–60.
- Richardson, J.S. and Richardson, D.C. 1989. Principles and patterns of protein conformation. In *Prediction of protein structure and the principles of protein conformation* (ed. G.D. Fasman), pp. 1–98. Plenum Press, New York.
- Shamala, N., Guru Row, T.N., and Venkatesan, K. 1976. The crystal and molecular structure of allo-4-hydroxy-L-proline dihydrate. *Acta Crystallogr. B* **32**: 3267–3270.
- 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.
- Vitagliano, L., Berisio, R., Mazzarella, L., and Zagari, A. 2001. Structural bases of collagen stabilization induced by proline hydroxylation. *Biopolymers* **36**: 459–464.
- Weiss, M.S., Jabs, A., and Hilgenfeld, R. 1998. Peptide bonds revisited. *Nat. Struct. Biol.* **5**: 676.