STEREOCHEMISTRY OF THE PYRROLIDINE RINGS IN THE COLLAGEN STRUCTURE

G. N. RAMACHANDRAN
Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012 (*India*) *and Department of Biophysics, University of Chicago, Chicago, Illinois* 60637 (*U.S.A.*)
AND
M. BANSAL
Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012 (*India*)

ABSTRACT

In the collagen triple-helical structure, large side groups occurring at location 3 in the repeating triplet sequence $(Gly–R_2–R_3)_n$ are appreciably constrained if a proline residue occurs as R_2 in a neighbouring chain. The severity of the steric hindrance depends on the geometry of the prolyl ring. In this paper we propose two different puckerings for the proline ring, the first one being energetically favourable for most types of residue sequences commonly found in collagen, while the second is preferable when an amino acid residue with a large side group occurs at location 3 in a neighbouring chain. The puckering of the pyrrolidine ring of hydroxyproline, as proposed earlier, is quite favourable from energy, as well as stereochemical considerations.

THE primary structure of the fibrous protein collagen has the repeating tripeptide sequence $(Gly–R_2–R_3)$, where R_2 and R_3 are, quite frequently, the imino acids proline and hydroxyproline respectively[1]. The secondary structure is a coiled-coil triple-helix with one direct inter-chain hydrogen bond per tripeptide, which links the amino group of a glycine residue in one chain and the carbonyl oxygen of the residue R_2 in a neighbouring chain[2]. A second set of hydrogen bonds can be formed, *via* a water molecule as an intermediary, in regions where the residue R_2 is not an imino acid[3]. In what follows, we shall refer to the residues R_2 and R_3 as being in 'locations' 2 and 3, respectively.

We have recently examined the collagen structure in detail to see if there are any stereochemical restrictions on the occurrence of certain amino acids in the positions R_2 and R_3[4]. It was found that the presence of a proline residue as R_2 restricts the freedom of orientation of large side groups occurring as R_3 in a neighbouring chain. However, recent data (from theoretical energy calculations as well as from experiments) have shown that the pyrrolidine ring is not a rigidly fixed entity, but can take up a number of different ring puckerings[6–8]. We have therefore examined the possible geometries of the prolyl rings in the collagen structure, from stereochemical and energy considerations. It is found that the ring puckering for a proline residue at location 2, as given in Ref. 2, is slightly unfavourable if the hydrogen atoms of the methylene groups are explicitly taken into consideration, instead of regarding the methylene groups as single atoms of larger van der Waals radius, as was done earlier. In this paper we propose a slightly different geometry for the proline side chain, which does not have this bad feature. The most favourable ring puckering corresponds to one having type B conformation in the notation of Ref. 5. The C^δ atom is in the plane of the peptide unit and the ring has a C^γ-endo puckering. There are no bad contacts, involving the backbone atoms, less than the extreme limit of the contact criteria[9]. The ring puckering is also very favourable from energy considerations (as can be seen from Table II of Ref. 5). There is only one H ··· H contact of the order of 1·87 Å.

A projection of the collagen triple helix down the helical axis, with a prolyl residue at location 2, is shown in Fig. 1. The various bond angles and dihedral angles describing the geometry of the prolyl ring are given in Table I (first row) and the cylindrical polar coordinates of the atoms of the proline residue (in chain A) are given in Table II. The hydrogen atoms have been fixed in a tetrahedral orientation at the various carbon atoms, with the C—H bond length being fixed at 1·1 Å.

The proline ring with this conformation does not cause any steric hindrance to the occurrence of any other amino acid residue at location 3 in the same chain. But as can be seen from Fig. 1, the side group of a residue at location 3 in chain A is oriented towards the prolyl residue at location 2 in chain B. Hence large side groups occurring in this location in chain A are appreciably constrained in their freedom of orientation. However, all side groups, except those with branching at C^γ (as in leucine and phenylalanine residues) can occur at location 3, though with a limited range of allowed orientations. A side chain, branched at C^γ, can also be accommodated at location 3, in the presence of a prolyl residue at location 2 in a neighbouring chain, if the puckering of the proline ring is slightly altered.

The most favourable geometry of the prolyl ring, if large side groups occur at location 3 in a neighbouring chain, is shown in Fig. 1 by dotted lines. The C^δ atom has been moved about 14° out of the plane of the peptide unit and the C^γ atom is in an exo position. This corresponds to a type A conforma

TABLE I

Various parameters which describe the geometry of the pyrrolidine rings of proline and hydroxyproline residues in the collagen structure as described in this paper. The notation follows Ref-5+

Residue	τ^{α}	τ^{β}	τ^{γ}	τ^{δ}	τ^{N}	θ	χ^1	χ^2	χ^3	χ^4
(*a*) Pro	103	107	106	104	113	−9	22	−26	20	−7
(*b*) Pro*	105	107	106	105	111	+5	−18	25	−22	10
(*c*) Hyp†	106	106	105	104	111	−5	−14	26	−29	21

+ The τ's denote the internal bond angles of the pyrrolidine ring, while the dihedral angles correspond to: θ (C^{β}—C^{α}—N—C^{δ}), χ^1 (N—C^{α}—C^{β}—C^{γ}), χ^2 (C^{α}—C^{β}—C^{γ}—C^{δ}), χ^3 (C^{β}—C^{γ}—C^{δ}—N) and χ^4 (C^{γ}—C^{δ}—N—C^{α}).

* This geometry of the proline ring is preferable when large side groups occur in a neighbouring chain (see text).

† These values are for the Hyp ring as given in Ref. 2, and found to be satisfactory in this investigation,

tion of Ref. 5. The various parameters describing its geometry are given in Table I (second row) and the coordinates of the relevant atoms are given in Table II.

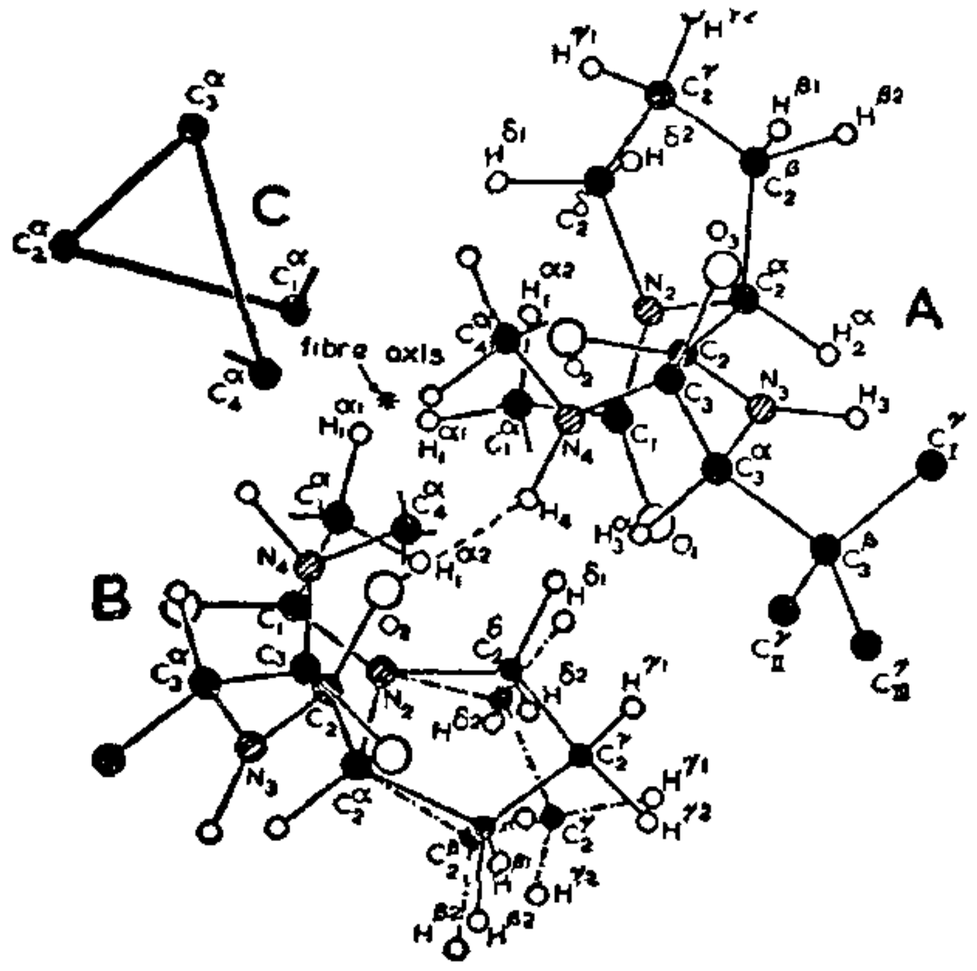

FIG. 1. A projection of the collagen triple-helix down the helical axis. A proline residue is shown at location 2, while C^{γ} atoms, in all the three possible staggered orientations, are shown attached to C_3^{β} in chain A. The puckering of the proline ring, which facilitates the occurrence of large side groups at location 3 in chain A, is shown by broken lines.

The energy of the triple-helical structure for this geometry of the prolyl ring is only about 2 kcal/mole per tripeptide higher than for the other geometry (but this will be made up by the packing energy of the large side chains). The values for the intra- and inter-chain energy (nonbonded + electrostatic) for the tripeptide sequence (Gly–Pro-Ala)$_n$ for the two different ring puckerings are given in Table III, along with the short contacts which are less than the extreme limits. The energy values for the conformations of the triple-helix, as given by other workers[10, 11], are also given in Table III. They are similar to the values for the Ramachandran one-bonded model. In these models also, the proline ring has the type B conformation and causes appreciable steric hindrance to the occurrence of large side groups at location 3 in the neighbouring chain.

TABLE II

The atomic coordinates, in a cylindrical polar system for a proline residue at position 2 in chain A are given for the two proposed geometries of the proline ring (listed in Table I). The coordinates for a proline residue in chain B (shown in Fig. 1) can be obtained by a transformation using the helical parameters

(*unit height* = 2·91 Å, *unit twist* = − 110)°

Geometry	Atom	*r* (Å)	ϕ (°)	*z* (Å)
Common atoms	N_2	2·96	21·90	1·62
	C_2^{α}	3·95	18·20	2·70
	C_2	3·17	11·50	3·96
As in row (*a*) of Table I	C_2^{β}	4·83	34·55	2·79
	C_2^{γ}	4·40	50·60	2·20
	C_2^{δ}	3·27	47·60	1·20
	$H^{\beta 1}$	5·10	36·70	3·85
	$H^{\beta 2}$	5·75	31·70	2·24
	$H^{\gamma 1}$	4·20	60·80	2·95
	$H^{\gamma 2}$	5·32	54·40	1·65
	$H^{\delta 1}$	2·65	65·50	1·26
	$H^{\delta 2}$	3·64	46·60	0·17
As in row (*b*) of Table I	C_2^{β}	4·85	34·28	2·80
	C_2^{γ}	4·86	43·97	1·57
	C_2^{δ}	3·45	44·59	1·00
	$H^{\beta 1}$	4·66	42·07	3·65
	$H^{\beta 2}$	5·88	29·90	2·96
	$H^{\gamma 1}$	5·26	55·52	1·82
	$H^{\gamma 2}$	5·57	38·76	0·85
	$H^{\delta 1}$	3·12	62·27	1·29
	$H^{\delta 2}$	3·51	42·77	−0·08

TABLE III

The intra- and inter-chain energy values (nonbonded + electrostatic) in kcal/mole per tripeptide for the repeating tripeptide sequence (Gly–Pro–Ala)$_n$, with the triple-helical conformation proposed by various workers. The constants used to calculate the nonbonded and electrostatic contributions have been taken from Refs. 9 and 13 respctively

Structure	Energy (nonbonded + electrostatic) in kcal/mole			Short contacts	
	Intra-chain	Interchain	Total	Type	Distance (Å)
Ramachandran 1-bonded, as in Ref. 2, but with new proline geometry, (*a*) of Table I	−17·14	−13·58	−30·72	H_3^{α} (A) ··· $H^{\gamma 1}$ (B)	1·87*
Same as above, but with proline geometry, (*b*) of Table I	−16·18	−12·39	−28·57	$H_1^{\alpha 2}$ (A) ··· $H^{\delta 2}$ (A)	1·83
Yonath and Traub as in Ref. 11, but with C—H bond length equal to 1·1 Å	−17·10	−11·59	−28·69	H_3^{α} (A) ··· $H^{\delta 1}$ (B)	1·77*
Rich and Crick (II) as in Ref. 10, but with hydrogens also attached	−17·90	−13·36	−31·32	..	..

* Denotes short contacts involving the proline ring atoms in a neighbouring chain.

Thus, by allowing small variations in the puckering of the proline ring, it is possible to accommodate any amino acid at location 3 even in the presence of a prolyl residue at location 2 in a neighbouring chain. Hence any polytripeptide with a sequence of the type (Gly-Pro-R_3)$_n$ can take up a collagen-like triple-helical structure, even if residue R_3 has a bulky side chain, as is found to be the case from studies on synthetic polytripeptides[11, 12].

The geometry of the pyrrolidine ring of the hydroxyproline residue, as given in Ref. 2, is quite favourable energetically. It has a type A conformation with C^{γ} exo, the ring geometry being very similar to that found in crystal structures of 4-hydroxy-L-proline[14, 15] and in solution for poly (L-hydroxyproline)[16]. The bond angles and dihedral angles describing the ring geometry are given in Table I, row (*c*). There is only one bad contact involving a hydrogen of the δ-methylene group ($H_3^{\delta} \cdots H_2^{\alpha} = 1{\cdot}76$ Å). The presence of a hydroxyproline residue at location 3 slightly restricts the freedom of orientation of a side group R_2 in the same chain, if the latter occurs with its C^{γ} atom *trans* to the N atom, about the bond C^{α}—C^{β}. However these restrictions are not very severe. The hydroxyproline residue also does not cause any appreciable steric hindrance to the occurrence of any amino acid residues in the neighbouring chains. It may be mentioned that the hydroxyproline ring, with this puckering, is suitably oriented so as to form a hydrogen bond with the water molecule linking the two chains of the triple-helix, as well as an inter-protofibrillar hydrogen bond with a neighbouring triple-helix, thus giving additional stability to the collagen structure[17].

We wish to acknowledge grants from SERC (DST) in Bangalore and from U.S. Public Health Service (Grant No. AM 11493) in Chicago. MB is grateful to CSIR, India, for a scholarship.

1. Hulmes, D. J. S., Miller, A., Parry, D. A. D. Piez, K. A. and Woodhead-Galloway, J., *J. Mol. Biol.*, 1973, 79, 137.
2. Ramachandran, G. N., In *Treatise on Collagen* (G. N. Ramachandran ed.), Academic Press New York. 1967, **1**, 130.
3. — and Chandrasekaran, R., *Biopolymers*, 1968, **6**, 1649.
4. Bansal, M., *Int. J. Pept. Prot. Res.*, 1976 (in press).
5. Ramachandran, G. N., Lakshminarayanan, A. V., Balasubramanian, R. and Tegoni, G., *Biochim. Biophys. Acta*, 1970, **221**, 165.
6. Venkatachalam, C. M., Price, B. J. and Krimm, S., *Macromolecules*, 1974, **7**, 212.
7. Balasubramanian, R., Lakshminarayanan, A. V., Sabesan, M. N., Tegoni, G., Venkatesan, K. and Ramachandran, G. N., *Int. J. Protein Res.*, 1971, **3**, 25–33.
8. Torchia, D. A. and Lyerla, J. R., Jr., *Biopolymers*, 1974, **13**, 97.
9. Ramachandran, G. N. and Sasisekharan V., In *Advances in Protein Chemistry* (Anfinsen, C. B. Jr., Anson, M. L., Edsall, J. T. and Richards, F. M., eds.), Academic Press, New York, 1968, **23**, 284.
10. Rich, A. and Crick, F. H. C., *J. Mol. Biol.*, 1961, **3**, 483.
11. Traub, W. and Piez, K. A., In *Advances in Protein Chemestry* (Anfinsen, C. B. Jr., Edsall, J. T. and Richards, F. M., eds.), Academic Press, New York, 1971, **25**, 243.
12. Scatturin, A., Tamburro, A. M., Del Pra, A. and Bordingnon, E., *Int. J. Pept. Prot. Res.*, 1975, **7**, 425.
13. Poland, D. and Scheraga, H. A., *Biochemistry*, 1967, **6**, 3791.
14. Donohue, J. and Trueblood, K. N., *Acta Cryst.*, 1952, **5**, 419.
15. Koetzle, T. F., Lehmann, M. S. and Hamilton, W. C., *Ibid.*, 1973, **29 B**, 231.
16. Torchia, D. A., *Macromolecules*, 1972, **5**, 566.
17. Bansal, M., Ramaksrihnan, C. and Ramachandran, G. N., *Proc. Indian Acad. Sci.*, 1975, **82 A**, 152.