Proc. Indian Acad. Sci., Vol. 82 A, No. 4, 1975, pp. 152–164

Stabilization of the collagen structure by hydroxyproline residues*

M. BANSAL, C. RAMAKRISHNAN AND G. N. RAMACHANDRAN†, F.A.SC.

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012

MS received 16 June 1975

ABSTRACT

The molecular structure of collagen is now accepted to be based on a triple-stranded coiled-coil, in which the three strands are held together predominantly by hydrogen bonds. Recent experimental evidence has shown that the presence of hydroxyproline residues in the third position of the repeating tripeptide unit lends additional stability to the collagen structure. In this paper, we report a model structure, which is supported by these observations. In a model structure proposed earlier, there are two hydrogen bonds per tripeptide unit, one of which is a direct interchain hydrogen bond, while the second hydrogen bond can be formed *via* a water molecule. It has now been shown that the same water molecule can also form a hydrogen bond with the oxygen of the γ-hydroxyl group of hydroxyproline in the third position in the sequence (Gly-R_2-R_3).

This hydroxyl group can also take part in an inter-triple-helix hydrogen bond. Our studies thus show the role played by hydroxyproline residues in the structure and stability of collagen.

INTRODUCTION

THE molecular structure of the fibrous protein collagen is now generally accepted as corresponding closely to the 'one-bonded' triple helical model[1, 2]. In this model a second set of interchain hydrogen bonds is also possible, with a water molecule acting as an intermediary, so that two of the amino groups per tripeptide are involved in hydrogen bond formation[3]. Various types of hydrogen exchange studies[4] and also the more recent hydrogen-tritium exchange studies[5] support a 'two-bonded' model.

---

* Contribution No. 63 from the Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012.

† Also: Department of Biophysics, University of Chicago, Chicago, Illinois 60637, U.S.A.

In a recent paper[6] it was proposed that hydroxyproline residues, which occur only in collagen, can lend added stability to the collagen triple helix by forming an additional hydrogen bond with the above mentioned water molecule. In this scheme, one of the water protons is pointing outwards from the axis of the triple helix and can be subjected to attack by the medium. Also, since the hydroxyl group of the hydroxyproline acts as the donor for the hydrogen bond, it cannot form a hydrogen bond with a neighbouring triple helix. The present paper describes a more plausible scheme of hydrogen bond formation in which, (*a*) the water molecule, as in the structure of Ramachandran and Chandrasekaran[3], binds two chains of the triple helix through two hydrogen bonds, (*b*) the hydroxyproline oxygen accepts a hydrogen bond from the water molecule and (*c*) the same hydroxy proline oxygen acts as a donor for the formation of an inter-triple-helix hydrogen bond. Thus the hydroxyproline residues can stabilize the collagen structure by forming intermolecular hydrogen bonds, in addition to the intramolecular hydrogen bonds *via* a water molecule. A preliminary report of this has already been published[7] and the fuller details are reported here.

Thermal denaturation studies on collagen samples from different sources have now clearly shown that hydroxylated proline in the third position in the repeating sequence (Gly-R_2-R_3) lends additonal stability to the collagen structure[8,9,10].

Method

Inter-chain hydrogen bond scheme

The structure of Ramachandran and Chandrasekaran[3] has the following scheme of hydrogen bonding:

1. N_4—H_4 (A chain) ... O_2 (B chain)
2. N_2—H_2 (B chain) ... O_1^w
3. O_1^w—H_{11}^w ... O_1 (A chain)

Of these three sets of hydrogen bonds, the first one is a direct interchain hydrogen bond, while the latter two involve a water molecule. In this scheme the amino group N_2H_2 of the B chain is hydrogen bonded to the oxygen of the water molecule which in turn is hydrogen bonded to the carbonyl oxygen O_1 of chain A. Thus, while one of the water protons, say H_{11}^w, is involved in this hydrogen bond, the other proton H_{12}^w is not hydrogen bonded. This type of 'water-bridged structure' can occur

only when the residue in the second position of the triplet (Gly-R_2-R_3) is not an imino acid.

The basic backbone for the water-bridged structure is that of the one-bonded structure reported by Ramachandran[2]. Considering a water molecule to be located in the position reported earlier[3], a simple examination of the atomic positions indicated that the hydroxyproline oxygen O_3^γ is at the correct distance from the water oxygen for the formation of a hydrogen bond and the angle C_3^γ—O_3^γ ... $O_1{}^w$ is also quite satisfactory. This had prompted us to propose a structure in which the atom O_3^γ forms a hydrogen bond, with O_1^w acting as an acceptor[6]. But, as mentioned earlier, in this scheme, the second hydrogen atom of the water molecule is left free and the hydrogen atom of the hydroxyl group of hydroxyproline is not available for the formation of an inter-triple-helix bond. An examination of the orientation of the three hydrogen bonds involving the water oxygen clearly revealed that an alternative scheme such as the one given below can also occur:

$$N_2\text{—}H_2\,(\text{B chain}) \ldots O_1^w \begin{cases} H_{11}^w \ldots O_1\,(\text{A chain}) \\ H_{12}^w \ldots O_3^\gamma\,(\text{A chain}). \end{cases}$$

The angle O_1—O_1^w—O_3^γ is about 105°, so that both the hydrogen bonds involving the water oxygen O_1^w are nearly linear. Thus the water molecule receiving one hydrogen bond and acting as a donor for two hydrogen bonds is firmly bound. The directions of these three hydrogen bonds are oriented in approximately tetrahedral directions, emanating from the atom O_1^w. The water oxygen can still accept a hydrogen bond from the medium, or a long polar side chain at C_2^α of the B chain.

The hydrogen atom of the γ-hydroxyl group of hydroxyproline is now free to form any linkage which connects one triple chain to another. In the next section, the possibilities of inter-triple-helical hydrogen bonding system involving the O^γ atom are discussed in detail. Another set of intramolecular hydrogen bonds *via* a second water molecule (as proposed by Ramachandran and Chandrasekaran[3]) can also occur in addition to that given above.

The co-ordinates of all the atoms in one tripeptide unit of the structure are given in table 1, along with the co-ordinates of the atoms of the water molecules linking two of the chains (A and B as per ref. 3). The hydrogen bond parameters are given in table 2. It may be mentioned that the atoms of the water molecules in the positions shown, do not invoke any steric hindrance with the atoms of the polypeptide chains.

The projection of the structure down the helical axis, with the complete hydrogen bond scheme is shown in figure 1.

Inter-protofibrillar hydrogen bond scheme

The possibility of the γ-hydroxyl group of hydroxyproline forming a hydrogen bond between the triple chain helices was first suggested by Ramachandran and Kartha (1955)[11]. From purely theoretical considerations it is found that the γ-hydroxyl group in one protofibril can form a hydrogen bond with a backbone carbonyl oxygen in a neighbouring protofibril. This is found to be possible for a hydroxyproline residue occurring in any of the three chains of a protofibril.

In order to investigate the various possibilities of hydrogen bonding, a suitable computer programme has been written and the method followed is as follows. The relative positions of two protofibrils can be described in terms of the cylindrical polar co-ordinate system, with the central axis of one protofibril as the *z*-axis of the co-ordinate system and the distance

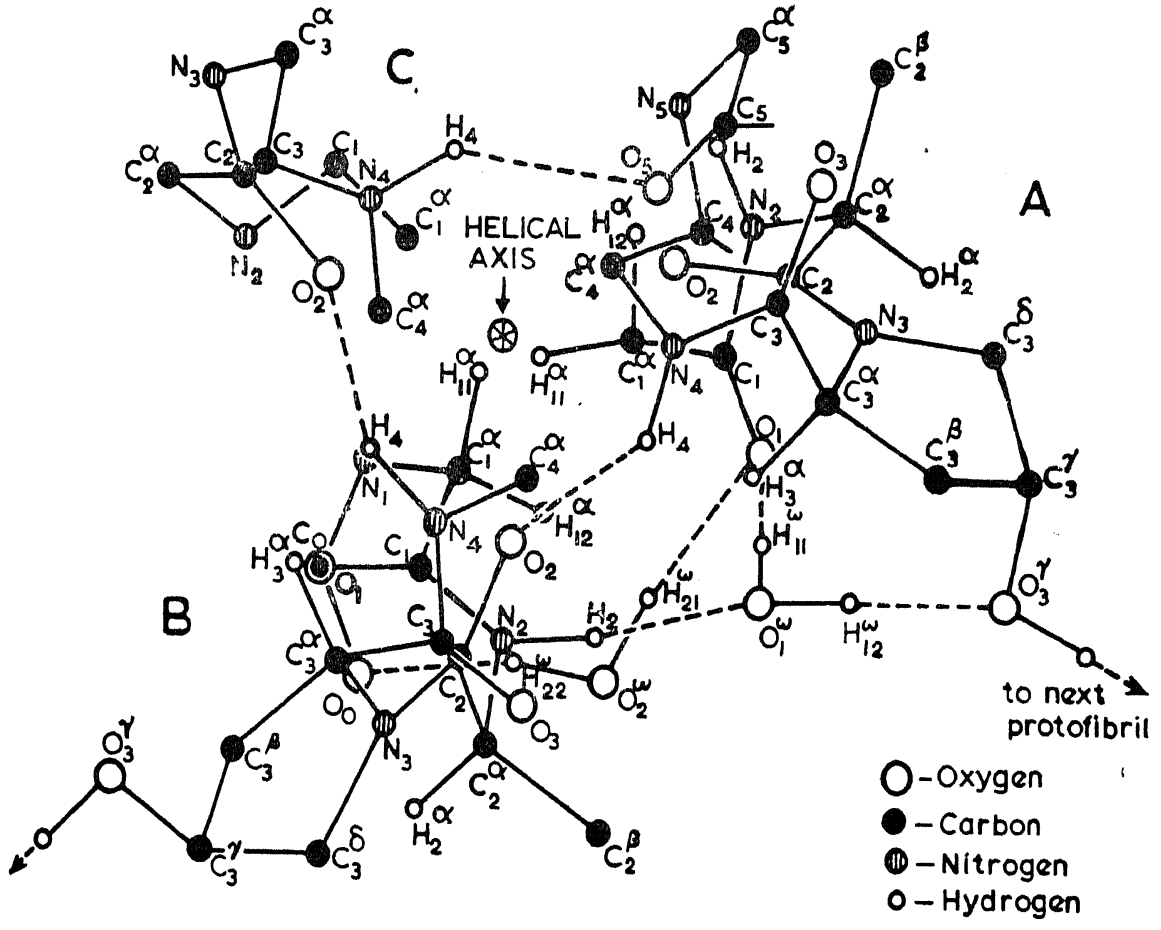

Figure 1. Atomic arrangement in the collagen triple helical structure, containing hydrogen bonds *via* water molecules, shown projected down the helical axis. The hydrogen bonds are shown by dotted lines. Only the water molecules linking the two chains A and B are shown.

Table 1. Cylindrical polar coordinates (in Angstroms) of the atoms in one tripeptide unit of collagen, along with the associated water molecules.

Atom	r	ϕ	z
C_1^{α}	1·40	0·0	0·0
H_{11}^{α}	0·42	− 21·14	0·41
H_{12}^{α}	1·76	34·18	−0·48
C_1	2·44	− 2·00	1·13
O_1	3·11	−22·10	1·53
N_2	2·96	21·90	1·62
H_2	3·04	40·02	1·30
C_2^{α}	3·95	18·20	2·70
H_2^{α}	4·71	8·32	2·44
C_2^{β}	4·92	33·50	2·82
C_2	3·17	11·50	3·96
O_2	2·02	22·00	4·17
N_3	3·99	1·20	4·80
C_3^{δ}	5·44	− 1·00	4·59
C_3^{α}	3·60	-- 9·50	6·03
H_3^{α}	3·05	−25·42	5·78
C_3^{β}	4·97	−15·60	6·56
C_3^{γ}	5·95	−13·00	5·44
O_3^{γ}	6·06	−25·00	4·66
H_3^{γ}	7·04	−25·30	4·46
C_3	3·05	7·50	7·06
O_3	3·82	23·40	7·26
N_4	1·95	− 2·00	7·70

Table 1—*Contd.*

Atom	r	ϕ	z
H_4	1·84	−32·50	7·50
O_1^w	3·80	−42·00	3·75
H_{11}^w	3·49	−35·50	2·90
H_{12}^w	4·55	−34·10	4·07
O_2^w	3·50	−72·00	0·90
H_{21}^w	2·93	−57·50	0·98
H_{22}^w	3·08	−88·00	0·97
*O^w	8·50	28·20	3·56
*H^w	9·40	26·20	3·86

* Atoms of a water molecule involved in the inter-triple-helix hydrogen bond.

between the two protofibrils as the radial parameter R. Arbitrarily choosing the line joining the centre to the C_1^α atom as the initial position, corresponding to $\theta = 0°$ for the angular parameter, the second protofibril is rotated about the major helix axis of the first protofibril. Thus, as θ varies from $-60°$ to $+60°$, various possible positions of the second protofibril can be obtained and the formation of hydrogen bonds, involving the O^γ-atom of the hydroxyproline residues in the three chains A, B and C, can be examined.

There are two types of hydrogen bonding possible, involving the O^γ-atom. Firstly, the O^γ-atom can form a direct hydrogen bond with an oxygen atom in the neighbouring triple-helix. Secondly it can form a hydrogen bond *via* a water molecule, depending on the distance between the O^γ-atom and carbonyl oxygen. In the actual calculations, the value of R has been varied between 10·5 Å and 14·5 Å, initially at coarse intervals of 0·5 Å and then at finer intervals of 0·1 Å.

For values of R less than 12.0 Å, it has been found that, either hydrogen bonds cannot be formed, because the distance between the γ-oxygen and carbonyl oxygen atoms is too large, or if the distance is of the correct order, there are severe steric hindrances between atoms in other parts of the chains, according to the contact criteria[12].

When the value of R lies between 12·0 Å and 12·5 Å, for a certain range of the angle θ, the γ-hydroxyl group of hydroxyproline in the A chain of the first protofibril can form a good direct hydrogen bond with the carbonyl oxygen atom of the C chain of the second protofibril. In fact, because of the helical symmetry present in the triple chain structure, equivalent positions can be obtained for the second protofibril, such that the O^{γ}-atom in either chain B or C of the first protofibril, can also form a similar type of hydrogen bond with a carbonyl oxygen in a neighbouring protofibril. This scheme of hydrogen bond formation, linking a central protofibril to three nighbouring protofibrils through the γ-hydroxyl group of hydroxyproline, is shown in figure 2 and the hydrogen bond parameters for one typical case are given in table 3 (*a*).

Table 2. Hydrogen bond parameters for the intra-protofibrillar bonds between chains A and B, in one tripeptide unit.

Hydrogen bond XH···Y	Length (Å) X···Y	Angle (°) X-H ∧ X···Y
N_4H_4 (A) ··· O_2 (B)	2·78	20·0
N_2H_2 (B) ··· O_1^w	2·87	3·6
O_1^w H_{11}^w ··· O_1 (A)	2·61	0·0
O_1^w H_{12}^w ··· O_3^{γ} (A)	2·82	0·0
O_2^w H_{21}^w ··· O_1 (A)	2·88	12·7
$O_2^w H_{22}^w$ ··· O_0 (B)	2·85	17·5

Table 3. Interprotofibrillar hydrogen bonds that are possible.

Hydrogen bond XH ··· Y	Length (Å) X ··· Y	Angle (°) X H ∧ X ··· Y
[a] O_3^{γ} H_3^{γ} (A) ··· O_0 (C)	2·88	10·3
[b] O_3^{γ} H_3^{γ} (A) ··· O^w	2·70	15·0
[b] O^w H^w ··· O_0 (C)	2·64	0·0

[a] R = 12·5 Å, θ = − 30°

[b] R = 14·5 Å, θ = − 30°

As the value of R is increased from 12·5 Å the distance between the oxygen atoms in the two protofibrils becomes too large for the formation of a direct hydrogen bond. However when R is about 14.5 Å this distance is of the right order for the formation of a hydrogen bond through a water molecule as an intermediary. A suitable position has been obtained for such a water molecule, by trial and error, so that it forms good hydrogen bonds with the γ-hydroxyl group in one protofibril as well as with the carbonyl oxygen in the neighbouring protofibril. This sheme of hydrogen bonding *via* a water molecule is shown in figure 3 and the hydrogen bond parameters are given in table 3 (*b*). The atoms of the water molecule do not make any short contacts with the other atoms in the two protofibrils.

Discussion

The structural model that has been proposed in this paper is consistent with the experimental observations that hydroxylated proline in the third position in the sequence (Gly-R_2-R_3) gives added stability to the collagen triple helix and that the physical properties of collagen fibres depend, to a

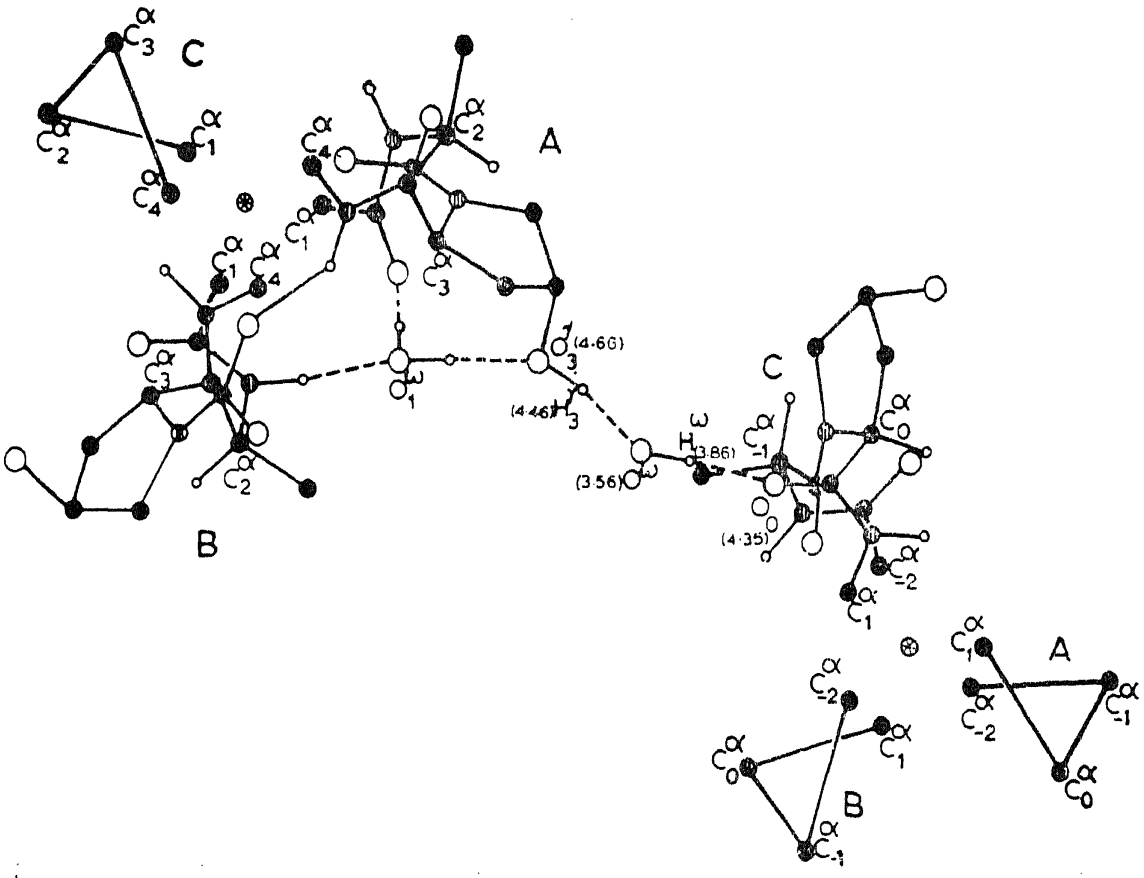

Figure 3. The same hydrogen bond scheme as in Fig. 2, but *via* a water molecule. The interprotofibrillar distance in this case is 14·5 Å. Only the water molecule linking the chain A of the central protofibril to the chain C of a neighbouring protofibril has been shown. The heights of the relevant atoms along the helical axis are also marked.

large extent, on its water content. For example, the characteristic X-ray diffraction pattern of collagen is destroyed by the removal of its water content[13]. Also, infrared dichroism studies[14], as well as calorimetric studies[15] indicate that water in collageneous tissue is organized in a specific way and, at least a part of it is strongly bound to the collagen molecule. In the proposed model, the water molecules involved in the intramolecular hydrogen bonds are bound quite firmly to the polypeptide chains and if a hydroxyproline residue is present, then there is no free proton present, and hence these water molecules cannot be easily disturbed by the surrounding medium. The water molecules are further protected if long side chains occur attached to C_2^{α}, because they are then fully encased inside the triple helix.

The inter-triple-helix hydrogen bond involving the γ-hydroxyl group of hydroxyproline can lend added stability to the collagen fibrils, which have previously been shown to be stabilized by intermolecular covalent crosslinks[16], and electrostatic and hydrophobic interactions[17,18,19].

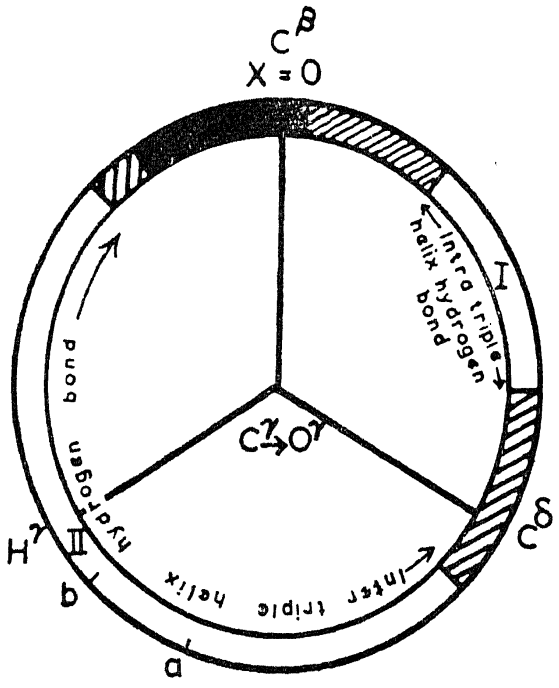

Figure 4. Orientation of the proton in the γ-hydroxyl group with respect to the hydroxyproline ring. The dark region is disallowed, and the shaded only partially allowed, by the contacts criteria. Of the two allowed regions, the smaller region I is favourable for the formation of an intra-triple-helix hydrogen bond, while region II is favourable for an inter-triple-helix hydrogen bond formation. The point *a* is the experimentally observed position for this proton in the crystal structure of 4-hydroxyproline from neutron diffraction data (Ref. 21). The point *b* indicates the location calculated by us for the minimum hydrogen bond angle.

According to our scheme this hydrogen bond can be formed directly in the near-dry state, while, in the wet state, it can be fomed through water molecules as intermediaries. This scheme agrees very well with the *x*-ray diffraction data.

(*a*) *Comparison with x-ray diffraction data.*—Rougvie and Bear (1953)[20] had found that the equatorial spacing for dry collagen is of the order of 10.6 ~ 10·7 Å, which corresponds to an interprotofibrillar distance of about 12·0 Å, However in this state the triple helices are not prefectly aligned—they seem to have a crinkled structure, hence the interprotofibrillar distance is not constant throughout the length of the fibrils. In our studies on the packing of protofibrils, we also find that in those regions of the protofibrils where no bulky sidechains are present (we have considered a side chain with only C^{β} being present), the protofibrils can approach as close as 10·7 Å. But if the side chains at C_2^{α} and C_3^{α} are bulky and rigid, like those of imino acids, then the closest distance of approach is about 12·5 Å. This is in agreement with the observation of Rougvie and Bear that when the equatorial spacing increases to about 12·0 Å the triple helix aggregation is relatively perfect. When the specimen was further hydrated, the triple chains moved apart and the maximum spacing observed by them was 14·5 Å.

(*b*) *Orientation of the hydrogen atom of the γ-hydroxyl group of hydroxyproline.*—In our calculations the atom H^{γ} was located so as to minimize the direct hydrogen bond angle ($O^{\gamma}\ H^{\gamma}\ \Lambda\ O^{\gamma}\ O_3$). In this position, it is oriented in a direction very similar to that reported for it by neutron diffraction studies on 4-hydroxy-L-proline[21]. As shown in figure 4, this proton has a greater freedom of orientation when it is pointing away from the triple helix so as to link two neighbouring triple helices than when it is oriented so as to form an intra-triple-helical hydrogen bond.

(*c*) *Packing of the protofibrils in collagen.*—Studies on model compounds of collagen, *viz.*, poly (Gly-Pro-Hyp) and poly (Gly-Hyp-Hyp) indicate that the packing of these molecules cannot be described by one definite type of cell but instead seems to have small regions of orderliness in which different types of packing occur[22]. Hence we have not attempted the packing of protofibrils in an extensive way. Also the hydroxyproline residues in collagen do not seem to occur in any regular pattern, unlike glycine which occurs at every third position. However in local regions of the protofibrils, where hydroxyproline occurs at the third position in the repeating triplet (Gly-R_2-R_3), the protofibrils may be arranged as shown in figure 2. This arrangement is consistent with the near hexagonal

packing scheme proposed by Katz and Li (1973)[23] from consideration of the geometry of the intermolecular space as obtained from an x-ray diffraction method and a new physical chemical technique in which polymer molecules of different sizes are used as probes to characterize the intermolecular space in collagenous tissues.

We also repeated the above calculations assuming a hydroxyproline residue to be present at the second position in the triplet (Gly-R_2-R_3), though no evidence is available for its presence in this position. It was very satisfying to note that the hydroxyproline residues, if present in this position, are unable to form the above type of hydrogen bonded structure, which is free of short contacts.+

CONCLUSIONS

The suggestion that hydroxyproline residues stabilize the collagen structure by virtue of their γ-hydroxyl group was first put forward by Gustavson in 1955[27]. As mentioned earlier, this suggestion is supported by recent experimental observations from thermal denaturation studies on collagen samples in the unhydroxylated and the hydroxylated state. These studies clearly indicate that hydroxylated proline in the third position of the repeating triplet (Gly-R_2-R_3) lends additional stability to the collagen structure.

The structural model proposed by us, from theoretical calculations, shows that the hydroxyproline residues, in addition to stabilizing the triple-helical structure by virtue of the stereochemical properties of the pyrrolidine ring, can also lend added stability to the collagen structure by taking part in both intra as well as inter triple-helical hydrogen bonds through their γ-hydroxyl group.

It may be mentioned that even if the sequence-Gly-Pro-Hyp-occurs in chain B in figure 1, making the amino group N_2H_2 in chain B unavailable, the two water molecules O_1^w and O_2^w can exist (but at slightly displaced locations) and make the hydrogen bonds $O_1^wH_{11}^w \ldots O_1$ (A), $O_1^wH_{12}^w \ldots O_3^\gamma$ (A), $O_2^wH_{21}^w \ldots O_1$ (A) and $O_2^wH_{22}^w \ldots O_0$ (B). The interchain hydrogen bond N_4H_4 (A) ... O_2 (B) is, of course, unaffected and continues to exist.

+ In rare cases of vertebrate collagen [*e.g.*, residue 999 in calf α_1 (I)] Hyp occurs in the position R_2, but it is 3-hydroxyproline and not 4-hydroxyproline, as in the position R_3. As mentioned by Piez[24] even in Type IV collagen it is probably true that only 3-Hyp succeeds Gly in the second position. However, in earthworm cuticle, Goldstein and Adams[25,26] apparently have found 4-hydroxyproline to succeed glycine in the amino acid sequence. This requires further study.

ACKNOWLEDGEMENTS

This work was supported by U. S. PHS Grants AM-15964 in Bangalore and AM-11493 in Chicago. M. B. is grateful to the CSIR, India, for a scholarship.

REFERENCES

1. Rich, A. and Crick, F. H. C., *J. Mol. Biol.* **3** 483 (1961).
2. Ramachandran, G. N., In: *Treatise on Collagen*, Vol. 1, ed. G. N. Ramachandran, Academic Press, New York, p. 103 (1967).
3. Ramachandran, G. N. and Chandrasekaran, R., *Biopolymers* **6** 1649 (1968).
4. Harrington, W. F., Josephs, R. and Segal, D. M., *Ann. Rev. Biochem.* **35** 599 (1966).
5. Yee, R. Y., Englander, S. W. and von Hippel, P. H., *J. Mol. Biol.* **83** 1 (1974).
6. Ramachandran, G. N., Bansal, M. and Bhatnagar, R. S., *Biochim. Biophys. Acta* **322** 166 (1973).
7. Ramachandran, G. N., Bansal, M. and Ramakrishnan, C., *Curr. Sci.* **44** 1 (1975).
8. Jimenez, S., Harsch, M. and Rosenbloom, *J. Biochem. Biophys. Res. Commun.* **52** 106 (1973).
9. Berg, R. A. and Prockop, D. J., *Biochem. Biophys. Res. Commun.* **52** 115 (1973).
10. Ward, A. R. and Mason, P., *J. Mol. Biol.* **79** 431 (1973).
11. Ramachandran, G. N. and Kartha, G., *Nature* **176** 593 (1955).
12. Ramachandran, G. N. and Sasisekharan, V., *Adv. Protein Chem.* **23** 283 (1968).
13. Esipova, N. G. and Chirgadze, Yu. N., In *Water in Biological System*, ed. L. P. Kayushia, Translation Consulatants Bureau, New York, p. 42 (1969).
14. Fraser, R. D. B. and MacRae, T. P., *Nature* **183** 179 (1959).
15. Haly, A. R. and Snaith, J. W., *Biopolymers* **10** 1681 (1971).
16. Piez, K. A., *Ann. Rev. Biochem.* **37** 547 (1968).
17. Cooper, A., *J. Mol. Biol.* **55** 123 (1971).
18. Finch, A. and Ledward, D. A., *Biochim. Biophys. Acta* **278** 433 (1972).
19. Hulmes, D. J. S., Miller, A., Parry, D. A. D., Piez, K. A. and Woodhead-Galloway, J., *J. Mol. Biol.* **9** 137 (1973).
20. Rougvie, M. A. and Bear, R. S., *J. Am. Leather Chemists Assoc.* **48** 735 (1953).
21. Koetzle, T. F., Lehmann, M. S. and Hamilton, W. C., *Acta Cryst.* **29B** 231 (1973).
22. Millinova, M. I., *Biofizika* **18** 416 (1973).

23. Katz, E. P. and Li, ShuTung, *J. Mol. Biol.* **3** 351 (1973).

24. Piez, K. A., In: *Biochemistry of Collagen*, G. N. Ramachandran, ed., Plenum Press (In Press).

25. Goldstein, A. and Adams, E., *J. Biol. Chem.* **243** 55 (1968).

26. Goldstein, A. and Adams, E., *J. Biol. Chem.* **245** 5478 (1970).

27. Gustavson, K. H., *Nature* **175** 70 (1955).