

STRUCTURE OF COLLAGEN

G. N. RAMACHANDRAN AND V. SASISEKHARAN

Department of Physics, University of Madras, Madras-25

A TRIPLE helical structure for collagen was proposed from this laboratory more than six years ago.¹ This consisted of three polypeptide chains, each having a three-fold screw symmetry (3_2 for L-type residues) which were linked together by hydrogen bonds approximately perpendicular to the length of the chains. A very satisfactory feature of this structure was that every third residue in it must necessarily be a glycyl residue, for there was no space in it for a side group (or even the β -carbon atom) to be attached to the corresponding α -carbon atom. This particular feature, namely, that 33% of the residues are glycyl, is universally true of collagen from different sources. The structure could readily accommodate proline and hydroxyproline residues (the latter in particular being known to be characteristic of collagen alone), besides being in agreement with the infra-red dichroism and other properties of collagen.

It soon became apparent that, although the structure was basically correct, the details required modification, for an analysis of the X-ray pattern of stretched collagen² indicated that the number of residues per turn in collagen is nearly $3\frac{1}{3}$ and not 3. The presence of such a non-integral number of residues per turn required that the three chains must all be further coiled around. Consequently, a coiled coil structure was put forward for collagen, which retained the essential features of the earlier one.³ The structure was made to have as many hydrogen bonds as possible, namely, two hydrogen bonds for every three residues, although this led to some short contacts.

This structure has however been criticized because of the occurrence of these few short contacts⁴ and it has been suggested that only structures with one hydrogen bond for every three residues can be built up which are free of short contacts. It appears to be generally taken for granted following the above-mentioned criticism of Rich and Crick⁴ that a two-bonded structure is impossible for collagen. But, as stated by Ramachandran,⁵ such an one-bonded structure is really unsatisfactory as a set of hydrogen bonds are not formed and a number of imino hydrogens are systematically left unbonded. It is possible, however, that these hydrogens may take part in bonding through a water molecule, but a careful study is required

of this question, and such a bond need be postulated only if a direct hydrogen bond is impossible.

The whole question has been re-examined recently in this laboratory, and it has been found that it is not impossible to bring the short contacts to permissible values while still retaining the two hydrogen bonds. The purpose of this communication is to show that a structure with two hydrogen bonds for every three residues is a possible one and is consistent with the physical and chemical properties of collagen.

Such a structure became possible essentially because a close study made in this laboratory of the available data on various organic structures showed that the exigencies of the situation in several cases may demand (interatomic) contact distances much shorter than the sum of the corresponding van der Waals' radii. Many examples can be cited in which the contact distances are much shorter than the criteria stated by Rich and Crick in their paper⁴ (*viz.*, $C \dots C = 3.6$ to 4.0 Å and $C \dots O = 3.2$ to 3.5 Å). Thus an examination of the reported crystal structures shows that not only are $C \dots C$ distances of the order of 3.2 Å (the value found in our earlier structure) observed in a number of cases, but that they can even be smaller, down to 2.9 Å. So also $C \dots O$ contacts are found to be less than 3.1 Å in a number of cases and values down to 2.7 Å have been reported. Also, if Rich and Crick's criteria were adopted, some of the known structures, for example the α -helix, would be unacceptable. (The full data will be given in a detailed publication.)

The other criticism that has been made is that the two-bonded structure (two hydrogen bonds for every three residues) cannot accommodate the sequence gly-pro-hydro, which appears to be of frequent occurrence in hydrolysates of collagen, for at the second α -carbon atom, there is no NH now to be hydrogen bonded. However, this does not mean that one must switch over from a two-bonded to a completely one-bonded structure. It is obvious that in those regions where proline does not occur, a structure in which the imino group is bonded is preferable to one in which it is not. In our latest approach both hydrogen bonds are made, but where the sequence gly-pro-hydro occurs, the residues slightly rotate around to accommodate these side-chains.

Another factor which helped in working out a satisfactory two-bonded structure is that an examination of the known structures of amino-acids and peptides revealed appreciable deviations either from planarity of the peptide group or from the standard bond lengths and bond angles in the group in almost every case. Obviously, such variations occur because the situation demands it, *e.g.*, some hydrogen bond is made shorter or more straight thereby, or somewhere a short contact is relieved. This suggests that provided the structure proposed has values for the bond distances and angles within the observed range in various known crystal structures, it should be considered satisfactory. In the latest structure, whose co-ordinates are given below, small deviations of the peptide groups from planarity occur, but the bond lengths and angles are all well within the normally observed range.

Another interesting factor revealed by a survey of the literature which helped in working out the two-bonded structure was that the hydrogen bond distance from an amide NH to a carbonyl oxygen ($O=C$) is invariably larger than 2.9 Å. It is found that for hydrogen bond distances lying between 3.0–3.05 Å, the infra-red frequency is about $3,350\text{ cm}^{-1}$ which is nearly the same as found in collagen (actually for collagen it is close to $3,330\text{ cm}^{-1}$). This is rather important, since it is found very difficult to reduce the bond lengths to less than 3.0 Å in the two-bonded structure. On the other hand, it is, in the one-bonded structure, rather difficult to increase the bond length to greater than 2.85 Å if the condition that every third residue should be glycine is assumed. Thus infra-red data show that most of the hydrogen bonds in collagen are weaker than in other proteins and polypeptides and this is in good agreement with the two-bonded structure.

One of the main results of our present study is that, while the two-bonded structure is highly specific and can only occur for helical parameters close to what are actually found in collagen, the one-bonded structure can occur over a very wide range of configurations. Also, a fact which seems to have been overlooked previously is that there is no reason why the exact rational value $10/3$ should occur for the number of residues per turn (n) in the minor helix. In view of this, following Ramachandran's paper,⁶ careful measurements were made on X-ray patterns of both natural and stretched collagen and these showed that the value of n is close to 3.28 and not 3.33 ($10/3$), *i.e.*, the twist for three residues t is only about

30° and not 36° , as has been assumed so far. In addition it was also found that the residue height h for unstretched collagen is 2.95 Å rather than 2.86 Å. These values, *viz.*, $h = 2.95\text{ Å}$, $n = 3.28$, $t = 30^\circ$ should replace the older values $h = 2.86\text{ Å}$, $n = 3.33$, $t = 36^\circ$.

These changes are really significant from the point of view of the stereochemistry of the structure, for it is actually found that a two-bonded structure based on the new values is superior to the one based on the old values with respect to short contacts and hydrogen bond angles.

A projection of the new two-bonded structure, the so-called standard structure, arrived at after a careful study, is shown in Fig. 1 and the co-

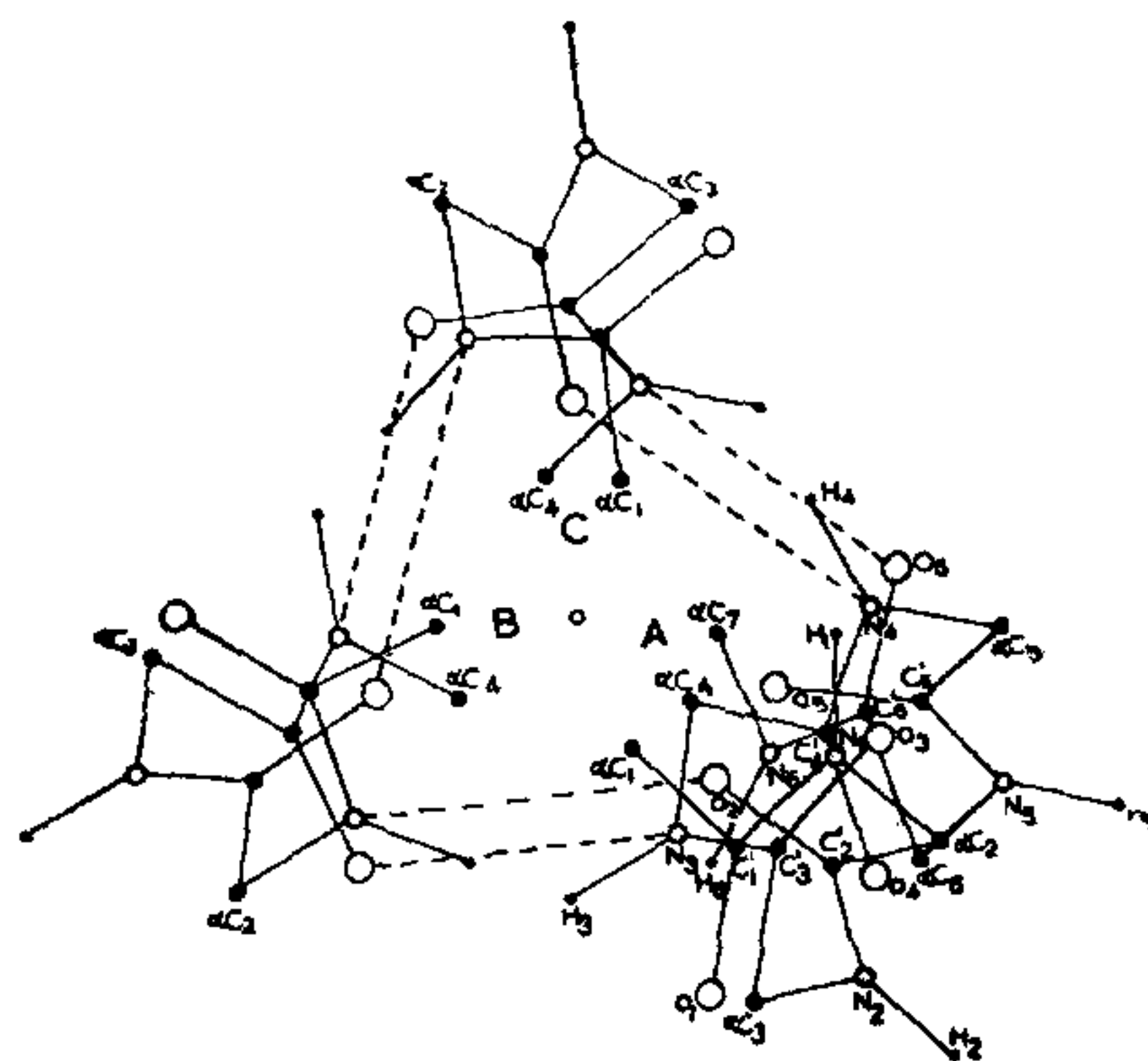


FIG. 1. Projection of the standard structure of collagen along the fibre axis. The two hydrogen bonds, for very three residues are shown by dotted lines.

ordinates are given in Table I. The inter-atomic distances and bond angles for this structure are not significantly different from the Pauling-Corey parameters. The two hydrogen bond lengths, *viz.*, $N_1H_1(B) \dots O_2(A) = 3.00\text{ Å}$ and $N_3H_3(A) \dots O(B) = 3.05$ are agreeing well with the infra-red data and the angle between NH and NO is less than 30° for both the bonds. The only short contacts found in the structure are (1) $O_2(A) \dots O_3(A) = 2.73\text{ Å}$, (2) $C_1(B) \dots O_2(A) = 2.90\text{ Å}$, (3) $C_1(A) \dots O_2(B) = 3.10\text{ Å}$, (4) $C_4(A) \dots O_3(B) = 3.03\text{ Å}$, and (5) $C_1(A) \dots C_1(B) = 3.50\text{ Å}$. The structure is thus seen to be satisfactory in every respect.

The satisfactory nature of this structure explains the occurrence of $h = 2.95\text{ Å}$ in natural collagen which is found to be remarkably constant in specimens from a wide variety of sources.

TABLE I

Atomic co-ordinates of the atoms in one set of three residues of the proposed structure of collagen. The cylindrical polar co-ordinates given are with respect to the axis of the major helix

Atom	r (Å)	(°)	z (Å)
C ₁	1.15	0.0	0.00
C ₁ '	2.27	12.0	1.00
O ₁	3.26	- 3.6	1.15
N ₁	2.43	38.8	1.70
H ₁	2.15	63.1	1.71
C ₂	3.52	35.4	2.72
C ₂ '	2.93	23.0	3.96
O ₂	1.75	17.5	4.16
N ₂	3.79	15.8	4.82
H ₂	4.78	18.1	4.70
C ₃	3.47	2.0	6.00
C ₃ '	2.50	18.7	6.86
O ₃	2.68	45.5	6.50
N ₃	1.95	1.8	7.87
H ₃	2.29	-24.4	8.04
C ₄	1.15	30.0	8.85

So also attempts were made to build structures corresponding to twists t of 10° and 50° for three residues keeping $h = 2.95$ Å, but these were distinctly bad since one of the hydrogen bonds became very long when short contacts were avoided. This explains why collagen forms a coiled coil structure, and also why the coiling corresponds to a twist close to 30° for three residues.

On the other hand for an one-bonded structure which can occur over a very wide range of configurations, no coiled coiling is needed and the residue height can also vary. It is not specific and it cannot explain why collagen has the parameters for its structure which are actually found.

Thus, in short, the two-bonded structure could explain several of the observed facts about collagen such as

(a) The observed height of about 2.95 Å per residue along the fibre axis and its relative inextensibility; (b) the occurrence of a coiled coil instead of a simple triple chain and the actual magnitude of the second coiling and

TABLE II

The co-ordinates of the atoms in a set of three residues of the three chains A, B₁ and B₂ of the structure of collagen. For the chain A which accommodates the sequence gly-pro-hypro, the co-ordinates of the side chain atoms of the proline and hydroxyproline residues are also given. The co-ordinates given are with respect to the axis of the major helix

Atom	r (Å)	(°)	z (Å)	Atom	r (Å)	(°)	z (Å)
Chain A				Chain B ₁ —(Contd.)			
α C ₋₂	1.15	80.0	-2.95	C' ₂	2.93	23.0	3.96
C' ₋₂	2.19	84.7	-1.83	O ₂	1.75	17.5	4.16
O ₋₂	2.90	64.8	-1.38	N ₂	3.79	15.8	4.82
N ₋₂	2.69	112.0	-1.41	H ₂	4.78	18.1	4.70
δ C ₋₂	3.12	139.1	-1.81	α C ₃	3.47	2.0	6.00
γ C ₋₂	4.58	136.9	-1.37	C' ₃	2.50	18.7	6.86
β C ₋₂	4.65	125.5	-0.14	O ₃	2.68	45.5	6.50
α C ₋₁	3.71	109.1	-0.33	N ₃	1.95	1.8	7.87
C' ₋₁	2.99	99.1	0.89	H ₃	2.29	- 24.4	8.04
O ₋₁	1.77	97.8	1.05	α C ₄	1.15	30.0	8.85
N ₋₁	3.83	93.5	1.83	Chain B ₂			
δ C ₋₁	5.28	95.2	1.80	α C ₀	3.47	-138.0	0.10
γ C ₋₁	5.87	85.0	2.78	C' ₀	2.50	-121.3	0.96
O ₋₁	6.28	72.8	2.09	O ₀	2.68	- 94.5	0.60
β C ₋₁	4.90	81.0	3.78	N ₀	1.95	-138.2	1.97
α C ₀	3.46	82.0	3.05	H ₀	2.29	-164.4	2.14
C' ₀	2.50	88.7	3.91	α C ₁	1.15	-110.0	2.95
O ₀	2.68	125.5	3.55	C' ₁	2.27	- 98.0	3.95
N ₀	1.95	81.8	4.92	O ₁	3.26	-113.0	4.10
H ₀	2.29	55.6	5.09	N ₁	2.43	- 71.2	4.65
α C ₁	1.15	110.0	5.90	H ₁	2.15	- 46.9	4.66
Chain B ₁				α C ₂	3.52	- 74.6	5.67
α C ₂	1.15	0.0	0.00	C' ₂	2.93	- 87.0	6.94
C' ₂	2.27	12.0	1.00	O ₂	1.74	- 91.0	7.14
O ₂	3.26	- 3.6	1.15	N ₂	3.78	- 94.8	7.77
N ₂	2.43	38.8	1.70	H ₂	4.70	- 92.5	7.65
H ₂	2.15	63.1	1.71	α C ₃	3.47	-108.0	8.95
α C ₃	3.52	35.4	2.72				

(c) the occurrence of 33% of glycine and a large proportion of proline and hydroxyproline.

The one-bonded structure cannot explain any of these properties except (c) since a wide range is possible for it. The infra-red data also completely support the occurrence of long hydrogen bonds 3.0–3.05 Å, found in the two-bonded structure.

Even when the sequence gly-pro-hypro occurs in one of the three chains, or locally in all the three chains, the standard structure can be slightly modified to incorporate these. The co-ordinates of the atoms in the three chains in the former case are given in Table II and the structure is shown in Fig. 2. Five hydrogen

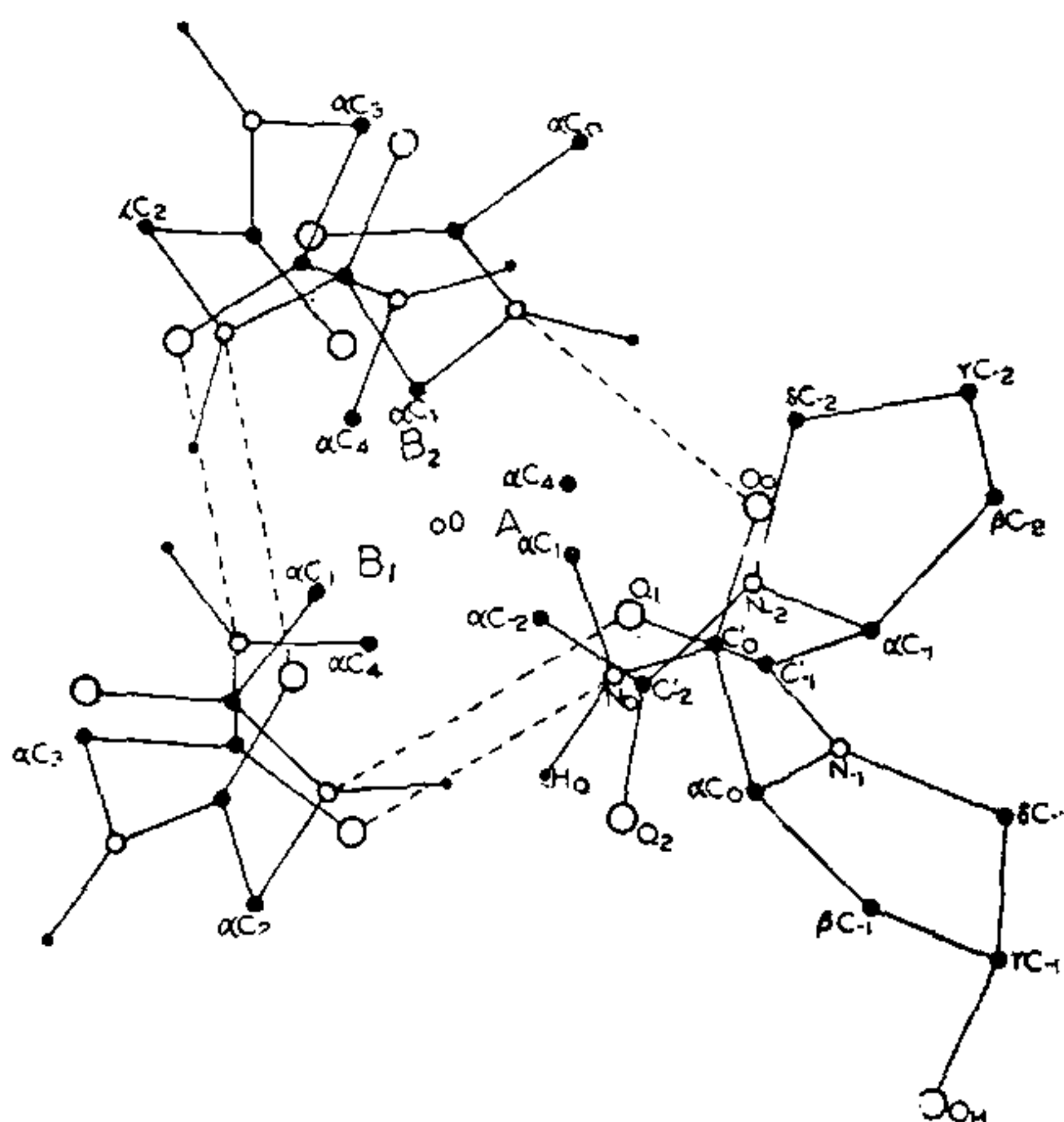


FIG. 2. Projection along the fibre axis of the structure of collagen with five hydrogen bonds for nine residues. For the chain 'A' which accommodates the sequence gly-pro-hypro, the positions of all the atoms of the proline and hydroxyproline residues in the projection are also shown.

bonds are formed for every nine residues. The atoms in two backbones of the chains are practically in the same positions as in the standard structure, while in the third chain none of them are displaced by more than 0.5 Å. If all the

three chains have the sequence gly-pro-hypro, then only an one-bonded structure is possible, but even in this case, the shifts of individual atoms are less than 0.5 Å.

Finally it may be mentioned that the calculated Fourier transform (F.T.) of the standard two-bonded structure is in good agreement with observation. Also, the two-bonded structure is distinctly in better agreement with the observed X-ray pattern than the one-bonded structure. Thus, its F.T. has a broad belt of intensity centred at $\xi = 0.35$, as actually observed, while this belt is distinctly nearer the meridian ($\xi = 0.30$) for the one-bonded structure. So also, the first maximum on the third layer is much stronger with the two-bonded structure, as actually observed. The position of the first maximum on the fourth layer is closer to the meridian for the two-bonded structure and occurs close to the observed position. In the sixth layer, there is a clear maximum in the F.T. at $\xi = 0.66$, as observed for this structure, while there is no such maximum in the F.T. of the one-bonded structure. Lastly, the strong blob of intensity observed on the equator corresponding to $d = 4.4$ Å is explained by the occurrence of nearly parallel planar residues in the backbones of neighbouring chains at this distance apart. The corresponding distance in the one-bonded structure is 4.8 to 4.9 Å, and does not agree with observation.

Details of these studies are given in a series of papers in the Report of the Collagen Symposium held at Madras in November 1960 (under publication). We wish to acknowledge the assistance of Messrs. Y. T. Thathachari, B. R. Lakshmanan and C. Ramakrishnan in part of this work.

1. Ramachandran, G. N. and Kartha, G., *Nature*, 1954, **174**, 269.
2. — and Ambady, G. K., *Curr. Sci.*, 1954, **23**, 345.
3. — and Kartha, G., *Proc. Ind. Acad. Sci.*, 1959, **42 A**, 215.
4. Crick, F. H. C. and Rich, A., *Nature*, 1955, **176**, 915.
5. Ramachandran, G. N., *Ibid.*, 1956, **177**, 710.
6. — *Proc. Ind. Acad. Sci.*, 1960, **52 A**, 240.

LONG WAVELENGTH SPECTRUM OF SOLAR CORONA

A SPECTRUM of the solar corona in the near infra-red region 1μ , was obtained by the staff members of the Sternberg Institute of Astronomy, USSR, during the solar eclipse of February 15, 1961. A detailed study of the emission line of wavelength 1.0747 micron (10747 Å) has been made in this region. This

line is emitted by strongly ionized atoms of iron which have 12 electrons knocked out of them under extremely high temperature. It is reported that the obtained information warrants definite conclusions concerning the physical conditions of matter in the outer envelope of the solar atmosphere.—(USSR News).