STUDIES ON COLLAGEN

I. Structure of the Collagen Group of Proteins

BY G. N. RAMACHANDRAN, F.A.SC. AND GOPINATH KARTHA

(Department of Physics, University of Madras, Madras 25)

1. INTRODUCTION

THE collagen group forms an important subdivision of the fibrous proteins. While the main details of the structure of the keratin group and of silk fibroin are now fairly well understood (Pauling and Corey, 1953; Crick, 1953 *a*; Marsh, Corey and Pauling, 1955), the same cannot be said of the collagen group. Several structures have been proposed for collagen [for a review of work prior to 1952, see the article by Bear (1952)], but none of these are in agreement with the available data on all the various properties of collagen. More recently, structures have been proposed for collagen by Randall, Fraser and North (1953), Crick (1954) and Huggins (1954). These have also not proved to be completely satisfactory.

A structure was put forward from this laboratory about a year ago (Ramachandran and Kartha, 1954). This fits the infra-red and chemical evidence on collagen, but is found to be in disagreement with the latest X-ray data. Cowan, North and Randall (1953) showed that by stretching collagen fibres, the X-ray pattern revealed much more detail than was ever obtained previously. The pattern shows clearly that the structure of collagen is based on a helical arrangement. An approximate measurement of the layer line spacings indicates that the helix may have 7 units in about 20 A along the fibre axis, or 10 units in about 30 A. The former choice was favoured by Cohen and Bear (1953), but the latter was shown to fit the X-ray diagram much better by Ramachandran and Ambady (1954), who also indicated that the radius of the primary helix is of the order of 1 A. It has now been possible to work out the details of the structure satisfying the above specifications demanded by the X-ray data and also taking into consideration evidence from other directions, such as infra-red absorption, chemical composition and reactivity, and so on. This structure turns out to be only a slightly modified form of the earlier structure (Ramachandran and Kartha, 1954) and preserves most of its features. A preliminary account of the structure has been published in *Nature* (Ramachandran and Kartha, 1955).

2. REVIEW OF DATA ON COLLAGEN

In order to appreciate how far the proposed structure helps in understanding the nature and behaviour of collagen, we shall give a brief summary

of the available data on collagen. In view of the excellent review published by Bear in 1952 and the discussion given by Kendrew (1954) in the treatise *The Proteins*, Vol. II B, attention will be directed here only to the salient features which have a bearing on the discussions in later sections.*

(*a*) *Chemical data.*—The amino-acid composition of collagen from different sources exhibits a remarkable uniformity (Tristram, 1953). Approximately one-third the number of residues are glycine (G, say), another one-third are composed of proline and hydroxyproline (P, say) residues, while the remaining one-third comprise the other residues (R, say). In general, the G-residues are slightly more than one-third while the P-residues are less than one-third. The relative proportion among the P-residues of proline and hydroxyproline is variable (from about 1 : 1 to 2 : 1). While mammalian collagen has a high hydroxyproline content (about 14 per cent.), it is as low as 4 per cent. in some fish collagens, and the thermal stability of collagen is found to go hand in hand with the hydroxyproline content (Gustavson, 1954, 1955). Apart from alanine, which forms about 30 per cent. of the R-type residues, there are very few non-polar residues. Most of them are polar, with both acid as well as basic side-groups, the more important ones being arginine and lysine, and aspartic and glutamic acids.

The sequence of amino-acid residues in collagen has been studied by various workers, the latest being those of Schroeder and co-workers (1953, 1954). A number of sequences of two and three residues have been identified in the hydrolysate. They do not appear to support the repeated occurrence of the sequence-G-R-P-. On the other hand, there seems to be definite evidence that the sequence gly-pro-hypro is of frequent occurrence.

There seems to be no definite evidence that *cis*-type of residues occur in collagen. On the other hand, there is reason to believe that they do not occur (Badger and Pullin, 1954).

(*b*) *Infra-red.*—Ambrose and Elliott (1951) observed that the infra-red dichroism of the CO and NH stretching vibrations could be explained satisfactorily if these bonds were at right angles to the fibre axis. Although there is some doubt whether the direction of the induced moment is exactly parallel to the direction of vibration (Price and Fraser, 1952), the angle between the two cannot be large. Thus, the infra-red data indicate that the CO and NH bond directions are nearly perpendicular to the fibre axis. Recently,

* After this paper was prepared in the final form, Symposium No. 9 of the Society for Experimental Biology (Cambridge University Press, 1955) reached the authors. The results reported therein are not considered here in detail.

Sutherland, Tanner and Wood (1954) have recorded data on the dichroism of other bands in the infra-red spectrum of collagen fibres.

(*c*) *X-ray diffraction.*—An important advance in our knowledge of the high angle diffraction pattern was made by the recording of the pattern of stretched collagen by Cowan, North and Randall (1953). It is now clear that the 2·86 A arc found in unstretched collagen is a true meridional reflection, while the 4 A arc is composed of two non-meridional reflections (see also a succeeding paper by Ramachandran and Ambady in this series). A number of other spots have also been recorded, which occur on layer lines, whose spacings are not integral multiples or submultiples of the prominent layer lines at 2·86 A, 4·1 A and 9·5 A. The high angle diagram shows clearly that the structure is helical in nature, and that the resolved component of the length of an amino-acid residue along the fibre axis is 2·86 A in unstretched collagen which may increase to as much as 2·95 on stretching. Several possible values have been given for the details of the helix, such as 7 residues in 2 turns or 10 residues in 3 turns (Cowan, North and Randall, 1953; Ramachandran and Ambady, 1954; Cohen and Bear, 1953; Randall, 1954 *a*). A larger repeat distance along the fibre axis is not ruled out.

The small angle pattern indicates the existence of an axial period of 640 A in dry collagen, which has been confirmed by electron microscopy. The variation of this with stretching does not follow the corresponding variation of the 2·86 A spacing (Cowan, North and Randall, 1954 *a*), but the details are not available. There is also some evidence for the existence of a sub-unit of dimensions 220 A along the fibre axis (North, Cowan and Randall, 1954). The variation of the small angle pattern with moisture content has been studied in detail by Rougvie and Bear (1953), who find in general an increase of the long spacing with increasing humidity. Pronounced changes have been observed in the intensity distribution in the small angle pattern as a result of tanning (Bolduan, Salo and Bear, 1951).

The unit cell dimensions (*a* and *b*) of collagen in the plane perpendicular to the fibre axis (*c*) have not been definitely established: Two suggestions have been made: (*a*) originally due to Astbury (1940) of taking the spacing of the principal equatorial spot at 11–15 A (varying with moisture content) as one and the diffuse spot at 4·5 A as the other and (*b*) due to Pauling and Corey (1951) of taking the unit cell as hexagonal, identifying the values 11–15 A as the (100) spacing. The principal objection to the latter has been that the spot corresponding to the 110 reflection has never been observed on the equator, even though the value of *a* changes by as much as 50 per cent.; nor has a spot having this ξ-value been previously observed in

any of the other layers. Recently, Ramachandran and Ambady (1954) have recorded such a spot in the layer of spacing 7·2 A, which supports the Pauling-Corey hexagonal cell.

The choice of the unit cell is closely related to the density of collagen. Bear (1952) has concluded that, with Pauling-Corey unit cell, three residues per unit cell occurring in an axial projection of 2·86 A would fit the observed density reasonably well.

Very recently, the cylindrical Patterson function of collagen, computed from its X-ray diagram, has been published (Yakel and Schatz, 1955).

(*d*) *Electron microscopy*.—The spacing of 640 A found in the small angle X-ray pattern has also been observed in electron micrographs of collagen fibrils (for details, see Bear, 1952). The details within this period have been studied, and it is found that in general 6 or 7 bands occur within the main repeat. The widths of these sub-bands have been measured, and although they are not all equal, they roughly subdivide the main spacing into 6 equal parts of spacing approximately 100 A. By a study of the micrographs given by stained and tanned fibres, it has been concluded that the stain is preferentially taken up in certain regions within the main period of 640 A and that other regions are relatively unaffected—called " bands" and " interbands " by Bear. The 6-fold subdivision is thus clearly seen in chrome tanned calfskin, stained with PTA (Schmitt and Gross, 1948).

A number of modifications of collagen have been detected by the electron microscope, mainly by Highberger and associates (Highberger, Gross and Schmitt, 1950; Schmitt, Gross and Highberger, 1953). Two distinct types have been discovered, the so-called fibrous long-spacing or FLS fibrils and the segment long-spacing or SLS fibrils. Both have a period of the order of 2000 A, but while the band system in FLS fibrils are non-polarised and small in number, the bands in the SLS material is polarised, and further the fibrils aggregate side by side, forming sheets or segments. The precise conditions under which these modifications are transformed into one another are not exactly known.

3. DETERMINATION OF THE STRUCTURE

Before describing the proposed structure, it may be worthwhile to mention the steps which led to its formulation. The earlier structure was worked out mainly on the following considerations:—

(*a*) The perpendicular dichroism of the CO and NH bands in the infrared, which indicate that these bonds are nearly at right angles to the fibre axis.

(*b*) The fact that more than one-third the number of residues in the protein are glycine residues.

It was found that these features could be explained by a structure based on a triple chain of residues, each chain being a helix composed of three residues per turn, having the symmetry 3_2. Both 3_1 and 3_2 are permissible; only 3_2 is possible if the amino-acid side-groups should have the L-configuration. The essential basis of the three chain protofibril is shown in Fig. 1, in which the individual chains may be denoted by the symbols A, B and C. The

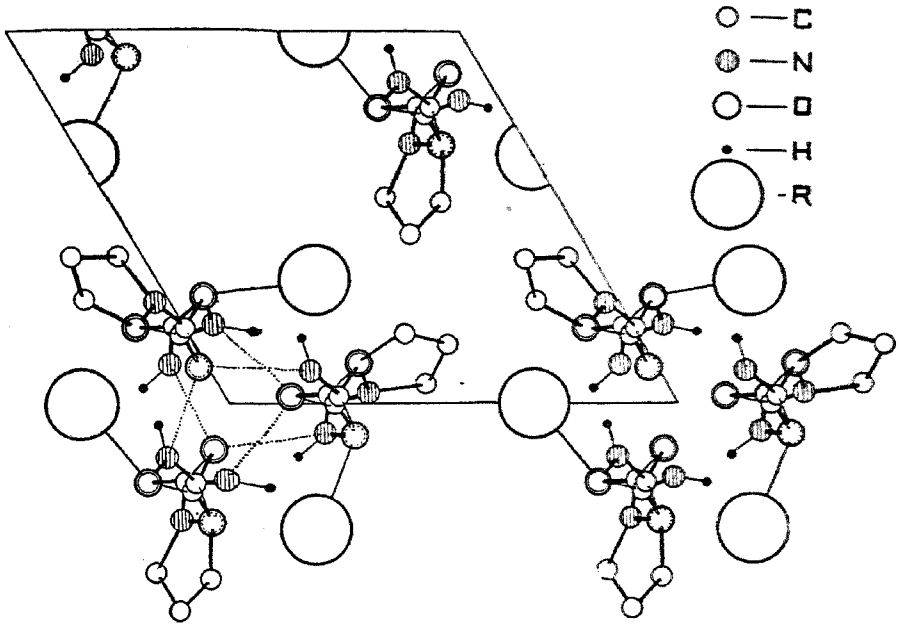

FIG. 1 (*a*) Projection of the structure along the *c*-axis. Two triple-chain rods and the contents of one unit cell are shown. The helices are right-handed (3_1). If the amino-acid residues have the L-configuration, they must be left-handed.

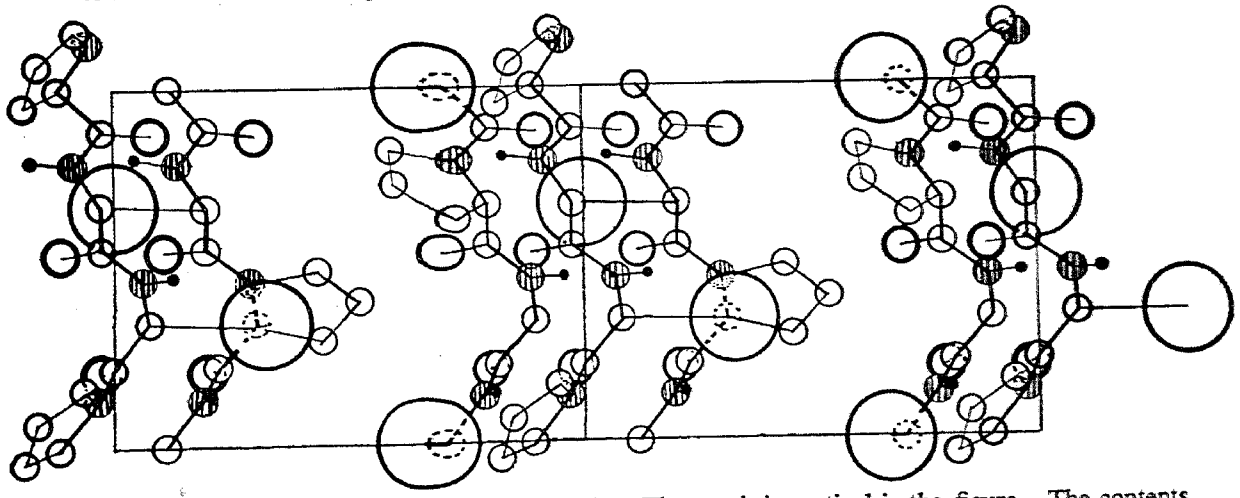

FIG. 1 (*b*). Projection along the *a*-axis. The *c*-axis is vertical in the figure. The contents of two unit cells are shown. Note the almost close packing of the atoms.

three chains are themselves related to each other by the operations of the symmetry element 3_2 parallel to the fibre axis through the centre 0. Thus the α-carbon atoms C_1 in the three chains are displaced successively along the fibre axis through a distance equal to one-third the pitch of the individual chains.

In actually working out the structure, the dimensions of the bonds and bond angles in the planar amide group were taken to be the same as those given by Corey and Pauling (1953). The angle at the α-carbon atom was taken to be equal to the tetrahedral angle and the two planar groups on either side of an α-carbon atom could rotate freely about the bonds joining them to this atom. In view of the rigidity of the backbone linking one α-carbon atom to another and the free rotation possible about the bonds at this atom, it is convenient to take the atoms in a link from one α-carbon atom to the next as a unit and denote them by the same subscript. The three α-carbons in one turn of each chain will be denoted by C_1, C_2, C_3, while the carbonyl carbon atoms will be distinguished by primes as C_1', C_2', C_3'. Thus, the sequence of atoms in the backbone will be:

$$-C_1(HR)-C_1'O_1-N_1H_1-C_2(HR)-C_2'O_2-N_2H_2-$$

If it is necessary to specify the chain to which an atom belongs, this will be done by adding an extra suffix indicating the nature of the chain: thus C_{1A} or C_{2B}. It is obvious that a single residue in the chain will be of the type $-N_1H_1-C_2(HR)-C_2'O_2-$. This residue would be referred to as the residue 2, the index corresponding to the suffix of the α-carbon atom.

The basal projection in Fig. 1 (*a*) shows clearly that the nearest neighbours of an α-carbon atom C_1 are the carbonyl oxygens O_2 of the other two chains. These are so close that there is no space for the hydrogen atom attached to C_1 to be replaced by any other group (not even by a methyl group). Further, the carbon atom C_3 could readily form part of the five-membered ring of a proline or hydroxyproline residue and the ring will be pointing outwards from the rod. It is also possible for the five-membered ring to be formed with the α-carbon atom C_2, with some strain on the bond at this atom. In the earlier note (Ramachandran and Kartha, 1954), this was not mentioned. It was then supposed that both proline and hydroxyproline residues occur only associated with C_3 and that this might explain the fact that these two together approximate to one-third the total number of residues. However, there now appears to be no reason why a proline residue may not occur next to glycine. This will be considered again later. Both NH and CO bonds are very nearly at right angles to the fibre axis in this structure.

An approximate agreement with X-ray data could be obtained by taking the 3 A arc to be non-meridional, and assuming that the length of three residues along the fibre axis is about 10 A. (For convenience of reference, the two principal arcs near the meridian will be referred to as the 3 A and 4 A arcs and the first strong layer line as the 10 A layer line.) However, by the time the details of this structure were worked out, the pattern of stretched collagen published by Cowan, North and Randall (1953) came to the notice of the authors. This showed clearly that the 3 A arc is truly meridional, while the 4 A arc is composed of two non-meridional spots. Attempts were then made in this laboratory to study this pattern in detail and the results are described in another paper in this series by Ramachandran and Ambady. By taking pictures with the stretched fibre set at the appropriate angle to the X-ray beam, the meridional reflection at 3 A could be obtained in great intensity, while the spots in the 4 A layer line clearly broke up into two. It was found that the pattern could be fairly well explained in terms of a helical structure having ten residues in three turns, with the radius of the helix of the order of 1 A (Ramachandran and Ambady, 1954). The order of the radius estimated for the helix was the same as that in the earlier structure, and this was strongly suggestive of the essential correctness of the earlier structure. However, it had to be modified to have three and one-third residues per turn, instead of three residues per turn. The dimensions of the helix, as well as the atomic co-ordinates were calculated for this helix with ten residues per three turns, having a projected length of 2·86 A per residue along the axis. An approximate calculation of the Fourier transform of this helix showed that the strong layer lines would be the third, seventh and tenth, as actually found.

An attempt was then made to put three of these helices side by side so that they are linked together by NH- - -OC′ hydrogen bonds. A number of trials were made, varying (*a*) the orientation of the chains by rotating all the three helices A, B, C simultaneously about their axes and (*b*) shifting them up and down along the common axis, but they were all unsuccessful. However, a satisfactory solution to the problem was found by coiling all the three individual helices about a common axis. This is to be expected, for helices having non-integral number of links per turn will fit only if they are wound into a superhelix. The main difficulty in the earlier attempts was the fact that as many as *ten* different types of hydrogen bonds had to be considered, of which at least six must be of the right length, if most of the free NH groups in the amino-acid residues must form hydrogen bonds. On the other hand, by forming the coiled coils, the number of different types of residues is reduced to three and only two types of NH....O bonds are

formed, assuming of course that the three individual chains are suitably displaced, so that they are symmetrically disposed with respect to one another. The extent of coiling about the axis of the cylindrical rod to achieve this is readily calculated. It turns out to be exactly one-tenth of a turn of every three residues, but in a direction opposite to that of the minor helix. [We shall use the nomenclature of Crick (1953 *b*) and call the helix of radius about 1 A the minor helix and the coil into which this helix is wound the major helix.] Thus in one complete turn of the major helix there are exactly thirty residues per turn, but as a result of the additional coiling, it makes ten turns for the thirty residues. Thus, with respect to the centre of the major helix, every third residue is in an equivalent position (*e.g.*, the links marked 1, 4, 7.... in Fig. 2).

Actually, there is not just one, but three chains (minor helices) and each of them is coiled further along the pattern of the major helix. Given the

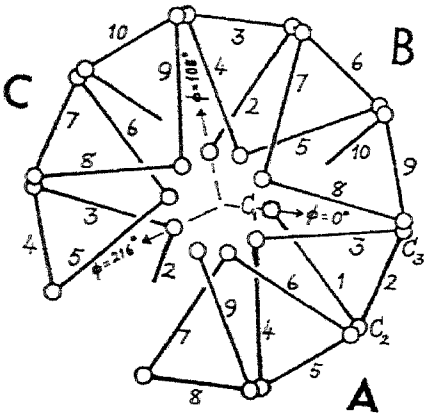

FIG. 2. Arrangement of coiled coils. Only the positions of the α-carbon atoms in each link are shown by open circles. The three chains A, B, C start at $\phi=0$, $z=0$; $\phi=108°$, $z=2{\cdot}86$ A and $\phi=216°$, $z=5{\cdot}72$ A. The major helix is left-handed and the minor helix is right-handed.

disposition of one of the chains, the arrangement of the other two readily follows from the fact that all the three are symmetrically disposed with respect to one another. This can be done by having the B chain to start with the α-carbon atom C_1 at the co-ordinates $\phi = 108°$, $z = 2{\cdot}86$ A and the C chain at $\phi = 216°$, $z = 2\times 2{\cdot}86$ A. Proceeding one more step, the A chain will have an α-carbon atom at $\phi = 324°$, $z = 3\times 2{\cdot}86 = 8{\cdot}58$ A, which is already there.

An approximate attempt at model building showed that good hydrogen bonds could be formed with such an arrangement if the radius of the major helix is close to 2·5 A. Precise drawings were then made with the major helix radius equal to 2·5 A. It is obvious that the length of the minor helix

per residue will no longer be 2·86 A if this is wound into a superhelix such that the projected length per residue along the axis of the major helix is 2·86 A. The former could readily be calculated, if the radius of the major helix is known, and is found to be equal to 2·91 A. when this is 2·5 A. The co-ordinates of the atoms in the minor helix with three and one-third residues per turn and a projected length of 2·91 A per residue were therefore calculated using the Pauling-Corey parameters for the amide group and taking the angle between the N–C and C–C′ bonds at the α-carbon atom close to $109\frac{1}{2}°$. The coiled structure was deduced from this by geometrical methods from the above considerations. The projection along the axis of the major helix was geometrically plotted, while the z-co-ordinate along this axis was calculated from the calculated inclination of the axis of the minor helix to the fibre axis. In both these processes, particularly in the former, it was necessary to assume that the twist induced by the second coiling at a carbonyl oxygen or the imino hydrogen atoms was the same as that at the corresponding carbon or nitrogen atoms. This is reasonable, since it is the latter which form the real backbone of the structure. An alternative was possible, namely, of finding the co-ordinates of the α-carbon atoms in the coiled coil and then putting in the other atoms in the planar Pauling-Corey configuration about the line joining these, as was done for the first non-coiled coil. The computational difficulties were however so large that only the geometrical method indicated above was adopted. The error introduced thereby is not expected to exceed about 0·1 A. In the case of the z-co-ordinates, the errors would be even less, as the NH and CO bonds were nearly perpendicular to the axis of the minor helix.

This still left one more degree of independence, namely, a rotation of the minor helix about its own axis (clearly all the three helices A, B, C have to be rotated by the same amount in the same direction). The trial was first made with the α-carbon atom C_1 as far inside the cylindrical rod as possible (as shown in Fig. 2) and the minor helices were then rotated upto $\pm 15°$ to either side. The best hydrogen bonds were formed (two for every three residues) very close to 0°, so that the first position was adopted. The radius of the major helix was then slightly varied, and the effect on the following were studied:

(*a*) the lengths of the two types of NH....O bonds, and the angles between the corresponding NH and NO directions.

(*b*) The distance between the α-carbon atom C_1 of each chain and the two closest oxygen atoms of the other two chains.

The former was considered to be the more important of the two criteria.

It must be mentioned that the essential feature of the structure is that every third residue in each of the three chains occurs rotated through an angle of 36° about the axis of the major helix and displaced parallel to it through a distance of 2·86 A. It is not essential that the three residues within this unit must conform to the pattern of a true coiled coil. All the essential features of the structure would still be retained even if there be a small deviation from the exact coiled-coil arrangement. Various types of deviations could be imagined, but the one actually tried out was to keep the α-carbon atoms in the same positions as in the exact arrangement, but to rotate the planar residues about the αC- αC-direction through a small angle. This would alter the angle between the N–C and C–C′ bonds at the α-carbon atoms. This angle was allowed to differ from the tetrahedral angle by as much as 3° and the best arrangement for the NH....O hydrogen bonds was studied.

The arrangement finally chosen was one in which the planes of the three amide groups make angles of 70°, 80°, 90° with the vertical plane through the α-carbon atoms at their two ends. Although the above arrangement does not form a true coiled coil, it still has an exact helical distribution of ten repeating units in three turns about the fibre axis, occurring in a distance of 28·6 A along the axis. Strictly, it is a triple helix with ten units occurring in one turn in a projected distance of 85·4 A along the axis, but the occurrence of three symmetrically disposed chains reduces the true repeat distance along the fibre axis to 28·6 A. The variations mentioned earlier were repeated with the new arrangement and the NH....O hydrogen bonds were both made to be approximately 2·8 A long. The major helix radius was finally fixed at 2·5 A.

The two NH....O distances were 2·87 A and 2·88 A. The distances from an α-carbon C_1 to the oxygen atoms nearest to it were 3·5, 3·15, 3·05 and 2·65 A. The last is considerably smaller than the carbon-oxygen Van der Waals contact distance. On the other hand, if the radius of the major helix is made larger, then the NH....O hydrogen bond lengths were increased too much, although the carbon-oxygen distance was satisfactory. If one of the hydrogen bonds is made short, while the other is comparatively long and not at the correct angle, then it is possible to avoid short carbon-oxygen distances; but this arrangement was not considered to be satisfactory.

4. DESCRIPTION OF THE STRUCTURE

The finally chosen co-ordinates of the various atoms in the backbone of the structure is given in Table I. The portion of the structure between the level $z = 0$ and $z = 8·58$ A is shown in Figs. 3 and 4. Figure 3 contains

TABLE I

Co-ordinates of atoms in one set of three residues. Cylindrical polar co-ordinates are given with respect to centre of major helix as origin

Atom	r in A	θ in degrees	z in A
Backbone			
C_1	0·98	0·0	0·00
C_1'	2·04	− 9·5	1·05
O_1	2·91	−10·5	1·41
N_1	2·51	−37·5	1·57
H_1	2·69	−50·1	1·30
C_2	3·50	−33·5	2·62
C_2'	2·89	−20·7	3·53
O_2	1·71	−12·0	3·95
N_2	3·76	−15·0	4·74
H_2	4·75	−17·0	4·73
C_3	3·50	− 1·6	5·95
C_3'	2·50	−16·5	6·78
O_3	2·66	−45·0	6·47
N_3	1·91	0·5	7·22
H_3	2·41	23·2	8·05
C_4	0·98	− 36·0	8·58
Hydroxyproline side-chain			
C_5	4·83	− 1·9	6·71
C_6	5·90	−10·0	5·92
C_7	5·25	−17·5	4·74
O_4	6·93	− 3·0	5·48
H_4 (C_6)	6·35	−12·0	6·51

all the various atoms, while Fig. 4 is drawn with the purpose of showing the essential features of the structure and contains only the α-carbon atoms and the N and O atoms which are linked by hydrogen bonds. The co-ordinates are not considered to be final, but it is unlikely that they will have to be appreciably altered. Further refinement is difficult to make by calculation; precise models may be more useful for this purpose.

It will be noticed that the three chains together form a cylindrical rod, which may be considered to be the protofibril of collagen. As in the earlier model, every third residue in each chain is in a similar position as far as the backbone of the structure is concerned. In fact, every third α-carbon atom (type C_1) occurs on the surface of a cylinder of radius 0·5 A. It is not possible for the hydrogen attached to this atom to be replaced by a larger group and so every third residue must be a glycyl residue. The hydrogen bond formation in the new structure is slightly different from and more satisfactory

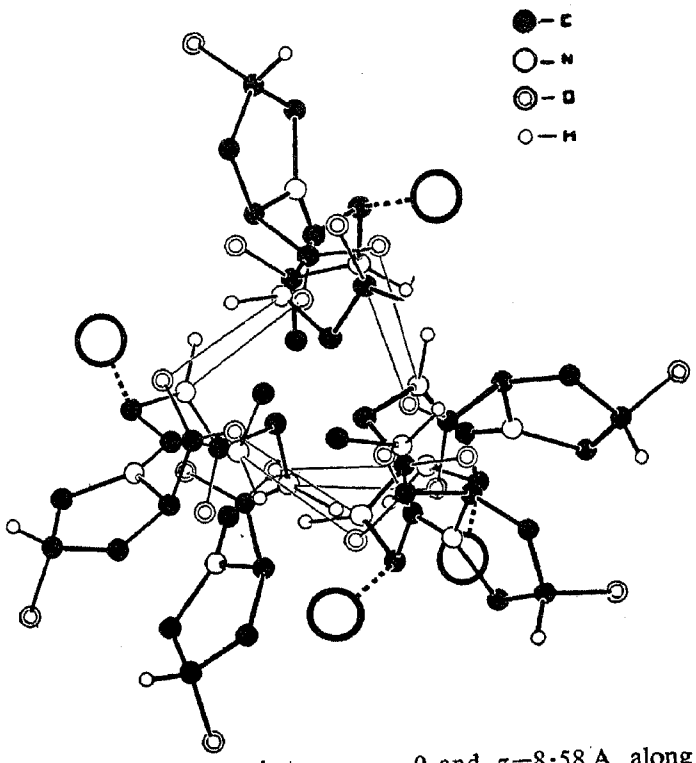

FIG. 3. Projection of the contents between $z=0$ and $z=8{\cdot}58$ A along the c-axis. The arrangement of the coils follows the scheme shown in Fig. 2. The hydrogen bonds are shown by thin lines. Refer to Table I for the co-ordinates of the atoms.

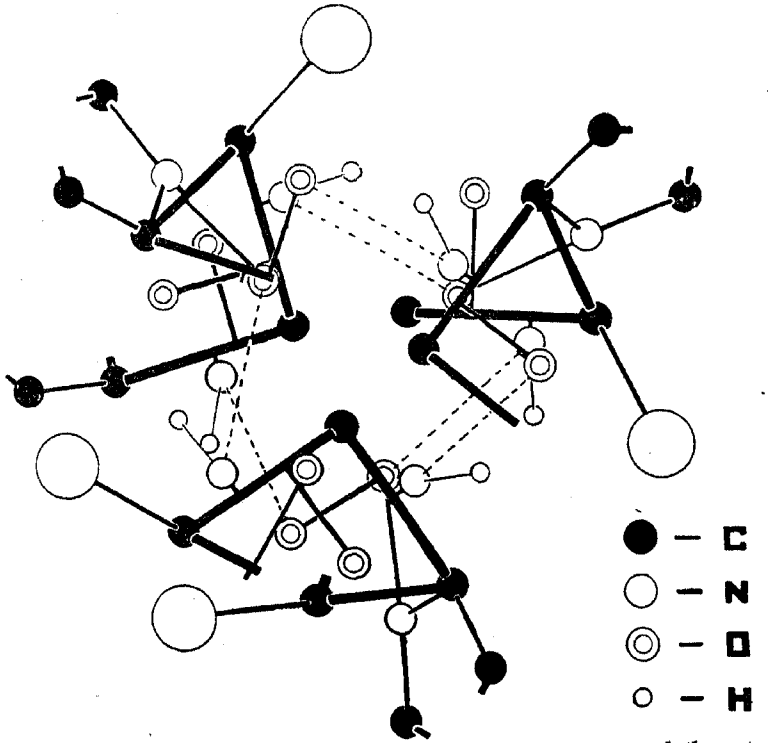

FIG. 4. Same as Fig. 3 except that only the α-carbon atoms and the atoms involved in hydrogen-bonding are shown. The links are indicated by straight lines, and their arrangement may be compared with Fig. 2.

than in the earlier model. The hydrogens in two of the three NH groups are linked to different oxygen atoms. The angles between the NH and N....O directions for the two types of bonds $N_1 \ldots O_1$ and $N_3 \ldots O_3$ are about 30°. These angles will be reduced further if the hydrogen atom is not coplanar with the rest of the atoms in the amide group. The third NH bond points out from the cylindrical rod and the side-chain of the corresponding residue will also extend outwards from the centre of the cylindrical unit. The pyrrolidine ring of a proline or hydroxyproline residue could conveniently occur in this position, and is marked as such in Fig. 3. The ring has been assumed to be planar and the C–OH bond is taken to be at the tetrahedral angle to the other two bonds of the same carbon atom in the plane of the ring. Although every third residue has been shown to be a hypro residue in this diagram, actually only part of them may be of this type.

Of the three types of carbonyl oxygen atoms, O_2 and O_3 are involved in internal hydrogen bond formation, while O_1 occurs on the outer surface of the cylindrical rod and cannot serve for this purpose. It is therefore available for forming external hydrogen bonds with other cylindrical rods by linking with a side-group.

For the same reason, it is obvious that the sequence of residues should exhibit a sort of periodicity, with every third residue being of the same type. One such regularity has already been noted, namely, the occurrence of a G-type residue once in every three residues. It had been assumed in the earlier model that the sequence of residues was of the type -G-R-P-G-. This however is not necessary. We shall now modify it to be of the type $-G-R_2-R_3-G-$. As mentioned earlier, the five-membered ring of a proline or hydroxyproline residue could be attached to either the C_2 or C_3 carbon atom. However C_3 will be favoured, as the corresponding NH group is not hydrogen-bonded. The choice of the location of the proline or hydroxyproline residues as well as the distribution of the other residues cannot be determined from a study of the backbone of a protofibril, but rather by a discussion of their cross-linkages. As will be shown below, the hydroxyl group of a hypro residue will be in the right position to form a hydrogen bond with the carbonyl oxygen O_1 of the neighbouring rod, if it occurs as the residue R_3. On the other hand, a proline residue occurring in this position serves no such purpose. Consequently, the proline may occur in position R_2.

Thus, the next step is to determine the relationship of the triple chain cylindrical rods to the lattice of the structure. As a first approximation, the lattice may be taken to be hexagonal, with the helix axis parallel to c,

which is the most reasonable arrangement of approximately cylindrical rods. The value of a for well-dried collagen has been found to be 12·0 A (= 10·4/0·866, Rougvie and Bear, 1953). Now, the oxygen of the hypro hydroxyl group extends to a distance of 6·93 A from the centre of the rod and its z-co-ordinate (5·48 for the residue A 2) is nearly the same as that of the non-hydrogen-bonded carbonyl oxygen at $z = 4·27$ (O_1 of chain B). It is reasonable to suppose that these two atoms in neighbouring rods would be linked by a hydrogen bond and that the orientation of the cylindrical rod with reference to the lattice will be such as to make the length of this bond a minimum. If the hydrogen bond distance is made equal to 2·85 A, then the distance between the two cylindrical rods comes out as 11·4 A. This is seen to correspond roughly to the minimum value of a found for thoroughly dry kangaroo tail tendon.*

However, this simple arrangement does not ensure that a large number of hypro residues will be hydrogen bonded *via* their hydroxyl groups to their neighbouring rods. This is because the screw symmetry of the coiled coil does not fit in with the hexagonal symmetry of the lattice. However, a very slight alteration in the second coiling will be sufficient to satisfy this condition. The positions of the G-type α-carbon atoms in one chain are shown by dots in Fig. 5 in the natural position. The atom marked 37 has now to be moved anticlockwise through 12°, so that it has a value $\phi = 60°$. Then, if the hypro side-chain attached to the R_3 type residue belonging to the series starting from 1 can form readily a hydrogen bond with a carbonyl oxygen atom of its neighbour, so can the corresponding hypro side-chain belonging to the series beginning from 37. As will be seen from the figure, this only involves a very slight untwisting of the coiled coil. The twist for three residues is changed only from 36 to 35° and it is reasonable to assume that the structure could readily accommodate this change in order to form a large number of cross-links.

This condition, namely, that a large number of cross-linking hydrogen bonds are formed between the hydroxyl groups of hypro residues and carbonyl oxygens of a neighbouring rod, automatically introduces also a long-spacing for collagen. As shown in Fig. 5, the α-carbon 37 is at an angle $\phi = 60°$, so that α-carbon atom 217 will be exactly above the atom numbered 1. Thus, the repeat distance along the fibre axis is $72 \times 8·58 = 618$ A. This is close to the value found for thoroughly dried unstretched kangaroo tail tendon (Rougvie and Bear, 1953).

* A value of 11·75 A, corresponding to the value of 10·15 A for the (100) spacing has been reported by Cowan, North and Randall (1955) with rat tail tendon.

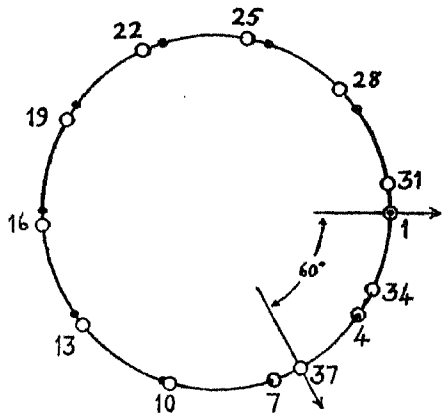

FIG. 5. Change produced in the coiling of the major helix in fitting the structure in a hexagonal lattice. Carbon atom 37 occurs at $\phi = 360° + 60°$ instead of $\phi = 360° + 72°$.

The distances from the hypro oxygen atom to a carbonyl oxygen in a neighbouring rod were measured for all possible positions for the hypro side-chain. The distances were measured graphically, corresponding to a value $a = 11{\cdot}4$ A for the lattice. It was found that in 12 out of the 36 possible positions the value was between 2·85 A and 3·15 A, while for the rest, it is longer. Thus, if we assume that a hypro residue occurs wherever it can readily form a hydrogen bond cross-link, then the corresponding percentage of hypro residues in collagen comes to be $100 \times 1/9$ or 11 per cent.

5. Comparison with Known Data on Collagen

(i) *Amino-acid composition and sequence of residues.*—The structure explains in a natural manner the fact that slightly more than one-third the total number of residues are glycyl residues. The structure demands further that every third residue shall be a glycyl residue. This is found to be in good accord with observation, since no sequence of three residues has been observed in the hydrolysate of collagen (or gelatin), which does not contain a glycyl residue (Schroeder *et al.*, 1954 and the earlier results quoted therein).

Further, the hypro residue* must occur in the position R_3 in the sequence $-G-R_2-R_3-$ according to the structure. Thus, the sequence hypro-gly must be of very frequent occurrence. It is interesting to note that of all the peptides isolated by Schroeder *et al.* (1954) which contain hydroxyproline, more than 95 per cent. has this sequence. So also, a good percentage of peptides having other hydroxy-amino-acid residues such as threonine and serine have the sequence threo-gly and ser-gly (90 and 85 per cent.). Thus it appears that all the residues having a terminal OH group in their side-chain occur in the position R_3. This again is reasonable, for they could

* We shall use the standard abbreviations for the amino-acid residues.

then readily link with a CO group of an adjacent rod by forming hydrogen bonds. The lengths of the side-chains in threonine and serine are much less than in hydroxyproline, but in the wet state (in which the collagen fabrils are built up), these groups could readily link to a carbonyl oxygen *via* a water molecule. Tyrosine residues should also occur in the position R_3 for the same reason, but no definite chemical evidence appears to be available. Further evidence that the hydroxylic residues occur as a group in the position R_3 is provided by a comparison of the amino-acid composition of fish skin and mammalian collagen (Tristram, 1953, p. 222). In fish skin, the hydroxyproline content is appreciably less, about 9 per cent. as compared with 14 per cent. for mammalian collagen, but the percentage of serine and threonine both go up appreciably, enough to make the total number of hydroxylic residues to be practically the same in both.

The position of proline residues is more difficult to fix from a consideration of the structure. As mentioned earlier, there is no difficulty in accommodating these residues in the position R_2 with a small strain and the data of Schroeder *et al.*, which unmistakably indicate a large probability (92 per cent.) for the occurrence of the sequence gly-pro suggest that they in fact occur in this position.

As mentioned earlier, if we assume that a hypro residue occurs in every position where it can form a strong linkage through a hydrogen bond with a neighbouring rod, then the percentage of hypro residues comes to be 11 per cent. It is indeed gratifying to note that the maximum percentage of hypro residues observed in mammalian collagen (Gustavson, 1954) is 10·9.*

(ii) *Infra-red absorption.*—It has been shown already (Ramachandran, 1954) that in a simple 3–10 helix, the directions of the NH and CO vibrations are in broad agreement with infra-red data. In Table II below are listed the angles made by these directions for the three types of residues in the coiled coil structure. It will be seen that the expected dichroism is again in good agreement with observation.

(iii) *X-ray pattern.*—It is found that the structure agrees fairly well with the observed X-ray pattern. The equatorial transform has been published in the preliminary report (Ramachandran and Kartha, 1955) and is not included in this paper. It shows a large peak at the origin which drops

* We are grateful to Dr. Gustavson for pointing out the correct value in a personal communication to G.N.R. This value seems to be in even better agreement with the structure than the value 10 per cent. mentioned in the preliminary report. However, it must be stated that this exact agreement must be considered somewhat fortuitous, as the theoretical value will change if the parameters in the structure are altered even slightly.

TABLE II

Orientation of NH and CO bonds in the coiled coil structure

Residue type	Angle made with axis of major helix by	
	NH bond	CO bond
1	75°	74°
2	88°	86°
3	80°	76°

ro corresponding to a spacing of 7·4 A. There is a second peak at and a third one at 2·2 A. Thus, the occurrence of the first and second reflections of the (100) planes in dry collagen and the absence of the eflections are explained. Calculations (made by Y. T. Thathachari) that the minimum at 7·4 A shifts to longer spacings with increasing content, thus explaining the continued absence of the 110 reflections. ccurrence of a diffuse spot at 4·4 A is to be attributed to the second in the Fourier transform along the equator. So also, the cylindrical son diagram of collagen, calculated by Yakel and Schatz (1955), is od agreement with the backbone of the structure, as described in the r report.

he optical transform of the structure reported in *Nature* was obtained r. Cowan and Mr. North of King's College, London and is shown in *a*, Plate IV, along with the X-ray diffraction pattern obtained by Ambady of this laboratory with the fibre inclined at an angle of 75° X-ray beam (Fig. 6 *b*). The remarkable similarity of the two patterns icative of the fact that the structure is essentially correct. It must be oned that the optical transform was obtained with measurements taken ie projection of the structure, so that it is not very accurate. The ving features of optical transform may be noted:—

(*a*) The prominent layer lines are third, seventh and tenth.

(*b*) The maximum intensity in the fourth layer line occurs at about $\xi = 0{\cdot}20$, as found by experiment.

(*c*) Belts of strong intensity occur near about the row lines corresponding to $\xi = 0{\cdot}35$ and $0{\cdot}7$, again as is found in the X-ray pattern.

n the X-ray pattern yielded by wet collagen fibres, the diffraction pattern ead out along the layer lines, so that the intensity distribution in this

pattern is a much better representation of the Fourier transform of the collagen protofibril than the pattern given by dry specimens. All the features (*a*), (*b*), (*c*) are exhibited by the pattern of wet collagen also (Ramachandran and Ambady, to be published).

(iv) *Electron micrographs.*—The existence of a long-spacing of 618 A along the fibre axis has been shown to follow from the condition that the protofibrils are packed in a hexagonal array. If the arrangement is exactly hexagonal, it also follows that the distribution of side-groups will repeat six times within this spacing, although the three chains make a complete number of revolutions, (namely 7) only in 618 A. Such an approximately equal six-fold subdivision has been observed in stained specimens by Schmitt and Gross (1948), and also by others. Actually, the subdivision is not exact. This would happen if the lattice is not truly hexagonal, but monoclinic, as is found to be the case with native fibres (North, Cowan and Randall, 1954). However, a full discussion of this feature appears to be premature at this stage of our understanding of the collagen structure.

So also, the existence of longer spacings of the order of 2000 A may be explained on the basis of a different scheme of fitting the helix of the protofibril in a hexagonal lattice, but this suggestion is purely tentative.

(v) *Shrinkage temperature and gelatin.*—It has been reported that the shrinkage temperature of collagen from different sources is directly correlated to the hydroxyproline content (Gustavson, 1954, 1955). This fact is readily explicable on the basis of the present structure, in which the important cross-links between different protofibrils are produced by the hydrogen bonds occurring between hydroxyproline OH groups and the carbonyl oxygen atoms. Thus, the greater the number of such linkages, the higher should be the shrinkage temperature.

It is suggested that in thermally shrunk collagen, the protofibrils remain intact, although they may take a folded configuration. Consequently, one would expect to find the X-ray pattern to be a highly disoriented representation of the pattern in the normal state. This is indeed found to be the case, for the pattern contains a fairly sharp ring corresponding to 2·86 A in addition to the diffuse ring at about 4·4 A. Also, it follows that an oriented arrangement could be recovered by cooling and stretching the specimens.

In the diffraction pattern of gelatin (Hermann, Gerngross and Abitz, 1930), the 2·86 A ring is comparatively weak, but the rest of the pattern is practically the same as for thermally shrunk collagen. This again shows that the protofibrillar structure is practically undisturbed in gelatin, but that the fibrils are highly distorted and folded in gelatin. The occurrence

of a ring corresponding to 12 to 17 A in gelatin shows that even in this state, there is an approximately parallel aggregation of the cylindrical rods at the protofibrillar level. Excepting this ring, the diffraction pattern of gelatin should be considered as the spherical transform of the collagen protofibril.

ACKNOWLEDGEMENTS

We wish to thank Prof. P. S. Sarma, whose gift of a specimen of shark ray collagen started our interest in this field, and Dr. H. Anderson of the Australian Leather Research Association for a generous gift of kangaroo tail tendon. We are much indebted to Prof. J. D. Bernal for the stimulating discussions we had with him during his visit to Madras. We are also grateful to Dr. P. M. Cowan and Mr. A. C. T. North of King's College, University of London, for kindly providing us with their optical transform, with permission to publish the same. Lastly, we wish to thank the Council of Scientific and Industrial Research for financial assistance.

SUMMARY

The paper reports the details of the revised structure of collagen. It is composed of three helical polypeptide chains, each of which has ten residues in three turns of a left-handed helix. The three chains are further wound into a superhelix in the form of intertwined coiled coils. The major helix is right-handed and makes one turn in thirty residues. The structure has reasonable hydrogen bonds, two for every three residues. The three-chain cylindrical rods are stacked together in hexagonal array and stabilised by cross-linkages through hydroxyproline side-groups. In trying to fit these in the lattice, a slight uncoiling of the major helix is required, resulting in a repeat of 6·8 A along the fibre axis. The proposed structure is in good agreement with the infra-red and X-ray data and also fits in broadly with the amino-acid composition and other properties of collagen.

REFERENCES

Ambrose, E. J. and Elliott, A. — *Proc. Roy. Soc.*, 1951, **206A**, 206.

Astbury, W. T. — *J. Intern. Soc. Leath. Trades' Chemists*, 1940, **24**, 69.

Badger, R. M. and Pauling, L. — *J. Chem. Phys.*, 1954, **22**, 1142.

Bear, R. S. — *Adv. Protein Chem.*, 1952, **7**.

Boldman, O. F. A., Salo, T. P. and Bear, R. S. — *J. Am. Leath. Chem. Assocn.*, 1951, **46**, 124.

Cohen, C. and Bear, R. S. — *J. Am. Chem. Soc.*, 1953, **75**, 2783.

Corey, R. B. and Pauling, L. — *Proc. Roy. Soc.*, 1953, **141B**, 10.

Cowan, P. M., North, A. C. T. and Randall, J. T. — *The Nature and Structure of Collagen*, London, Butterworth, 1953, 241.

——— — *Acta Cryst.*, 1954, **7**, 655.

——— — *Symp. Soc. Exp. Biol.*, 1955, **9**, 115.

Crick, F. H. C. .. *Acta Cryst.*, 1953 *a*, **6**, 689.

———— .. *Ibid.*, 1953 *b*, **6**, 685.

———— .. *J. Chem. Phys.*, 1954, **22**, 347.

Gustavson, K. H. .. *Acta Chem. Scand.*, 1954, **8**, 1298.

———— .. *Nature*, 1955, **175**, 70.

Hermann, K., Gerngross, O and Abitz, W. *Z. Phys. Chem.* 1930, **10B**, 371.

Highberger, J. H., Gross, J. and Schmitt, F. O. *J. Am. Chem. Soc.*, 1950, **72**, 3321.

Huggins, M, L. *Ibid.*, 1954, **76**, 4045.

Kendrew, J. C. *The Proteins*, New York, Academic Press, 1954, **2B**, 909.

Marsh, R. E,, Corey, R. B. and Pauling, L. *Biochim. Biophys. Acta.*, 1955, **16**, 1.

North, A. C. T., Cowan, P. M. and Randall, J. T. *Nature*, 1954, **174**, 1142.

Pauling, L. and Corey, R. B. .. *Proc. Nat. Acad. Sci. U.S.*, 1951, **37**, 272.

———— .. *Nature*, 1953, **171**, 59.

Price, W. C. and Fraser, R. D. B. .. *Ibid.*, 1952, **170**, 490.

Ramachandran, G. N. .. *J. Chem. Phys.*, 1954, **23**, 601.

——— and Kartha, G. .. *Nature*, 1954, **174**, 269.

——— and Ambady, G. K. .. *Curr. Sci.*, 1954, **23**, 349.

——— and Kartha, G. .. *Nature*, 1955, **176**, 593.

Randall, J. T., Fraser, R. D. B. and North, A. C. T. *Proc. Roy. Soc.*, 1953, **141B**, 62.

Rougvie, M. A. and Bear, R. S. .. *J. Am. Leath. Chem. Assocn.*, 1953, **48**, 735.

Schmitt, F. O. and Gross, J. .. *Ibid.*, 1948, **43**, 658.

——— and Highberger, J. H. .. *Proc. Nat. Acad. Sci. U.S.*, 1953, **39**, 459.

Schroeder, W. A., Honnen, L. and Green, F. C. *Ibid.*, 1953, **39**, 23.

———, Kay, L. M., LaGette, J., Honnen, L. and Green, F. C. *J. Am. Chem. Soc.*, 1954, **76**, 3556.

Tristram, G. R. .. *The Proteins*, New York, Academic Press, 1953, **1A**, 272.

Yakel, H. L. and Schatz, P. N. .. *Acta Cryst.*, 1954, **8**, 22.