THE STRUCTURE OF FIBROUS PROTEINS OF THE COLLAGEN-GELATIN GROUP

BY LINUS PAULING AND ROBERT B. COREY

GATES AND CRELLIN LABORATORIES OF CHEMISTRY,* CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIFORNIA

Communicated March 31, 1951

Collagen is a very interesting protein. It has well-defined mechanical properties (great strength, reversible extensibility through only a small range) that make it suited to the special purposes to which it is put in the animal body, as in tendon, bone, tusk, skin, the cornea of the eye, intestinal tissue, and probably rather extensively in reticular structures of cells. During the last thirty years, following the pioneer work of Herzog and Jancke,[1] a number of investigators have attempted to find the structure of collagen (and of gelatin, which gives similar x-ray photographs), but no one has previously proposed any precisely described configuration, nor has attempted to account for the positions and intensities of the x-ray diffraction maxima.

The diffraction pattern of collagen and gelatin is characterized by a meridional arc at 2.86 A. (Good reproductions of x-ray photographs have been published by Astbury.[2]) This arc remains essentially uninfluenced by a change in the source of material or its previous treatment; Bear[3] found that it varied only between the limits 2.82 A and 2.90 A for 26 samples, ranging from demineralized mammoth tusk to plain surgical gut (sheep intestinal submucosa). On the other hand, the principal equatorial reflection, which for thoroughly dried tendon[2] corresponds to the spacing 10.4 A, varies greatly in spacing with source and treatment of the material. At ordinary humidity it is about 11.5 A, and Bear reported the value 15.5 A for kangaroo tail tendon treated with water. It is evident that collagen consists of molecules (polypeptide chains) extending along the fiber axis, and rather loosely packed in parallel orientation. It will be pointed out below that the equatorial reflections correspond to a hexagonal packing of circular cylinders.

The 2.86-A fiber-axis spacing suggests that the amide groups of the polypeptide chain are in the cis configuration, as has been mentioned by Astbury.[2] (Astbury has suggested[2, 4] a structure for collagen which bears little resemblance to our structure.) If we accept the configuration of the amide group in a peptide as predicted from x-ray investigations of related simple substances, and described in our earlier paper[5] (with a single change —we have replaced the value 120° for the C′—N—C angle by the value 123°, which is suggested by general considerations as a reasonable value for the angle between a single bond and a bond with about 50 per cent

double-bond character), we predict that the length of an amide group with the cis configuration is 2.83 A, which is close to the observed fiber-axis spacing. Although the principal meridional arc at 5.1 A observed for proteins with the α-keratin structure, which had always been accepted as representing a fiber-axis distance, has been found[6, 7] to be in fact a diagonal spacing, there seems to be little reason to doubt that the 2.86 A spacing of collagen represents a fiber-axis distance.

We found it impossible to formulate a satisfactory structure for collagen

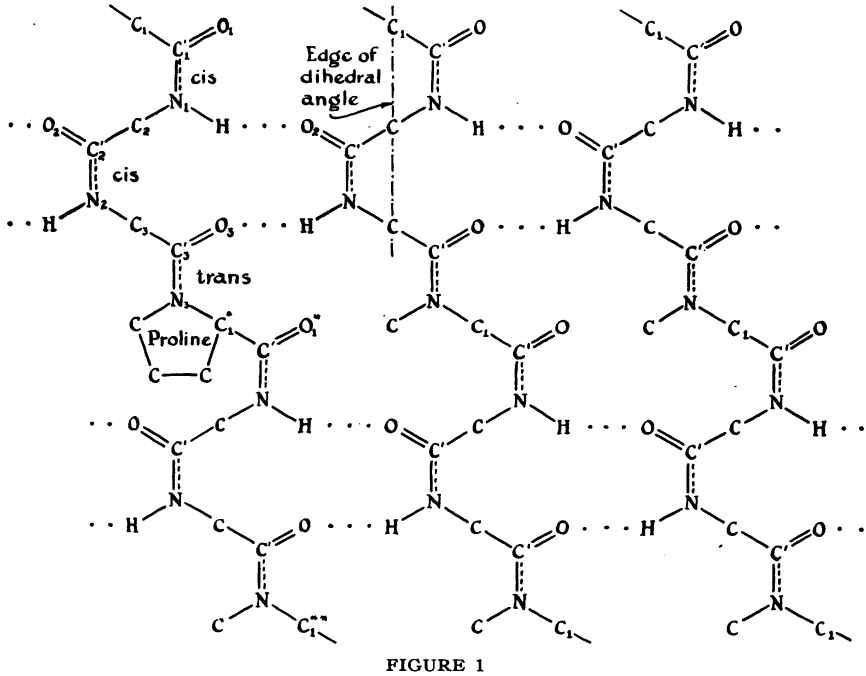

FIGURE 1

Diagrammatic representation of the configuration of polypeptide chains in the collagen-gelatin three-chain helix.

from cis amide groups alone. However, a satisfactory structure, described in detail in the following paragraphs, has been formulated with use of polypeptide chains in which there is an alternation of two cis groups and one trans group, as shown in figures 1 and 2. The angular orientation which the trans group is required to assume by its bonds with the contiguous cis groups is such as to cause its component of length along the fiber axis to be about 2.92 A, the average fiber-axis length per residue for the cis-cis-trans chain thus being 2.86 A, as observed for the collagen fibers.

Let us now consider the folding of the cis-cis-trans chain in such a way as to form satisfactory hydrogen bonds. The ease of lateral swelling of

collagen indicates that hydrogen bonds between molecules of the protein are not present, and hence that a structure involving intramolecular hydrogen bonds is to be sought.

The principal equatorial spacing for thoroughly dried collagen is 10.4 A. For cylindrical molecules in hexagonal packing this value for $d_{10\cdot0}$ leads to 125 A^2 as the basal-plane area per molecule, and with 2.86 A as the fiber-axis length to 358 A^3 as the volume per unit, or 215 cm^3 per mole of units. If the density is taken as 1.35 g cm^{-3} (the reported density of dry gelatin) the mass per unit is 291 g. The average residue weight for collagen and gelatin is found by analysis to be slightly less than 100; it is hence evident that there are about three residues in the unit of the molecule (2.86 A length). We accordingly reach the conclusion that the molecule of collagen and gelatin, essentially cylindrical in shape, is not a single polypeptide chain, but consists of three polypeptide chains. This is pleasing, inasmuch as there is no way in which a single polypeptide chain with a cis-cis-trans sequence of amide groups can be folded to give intramolecular hydrogen bonds and a fiber-axis residue length of 2.86 A. The intramolecular hydrogen bonds in the collagen-gelatin molecule must be lateral bonds between the three polypeptide chains that constitute the molecule.

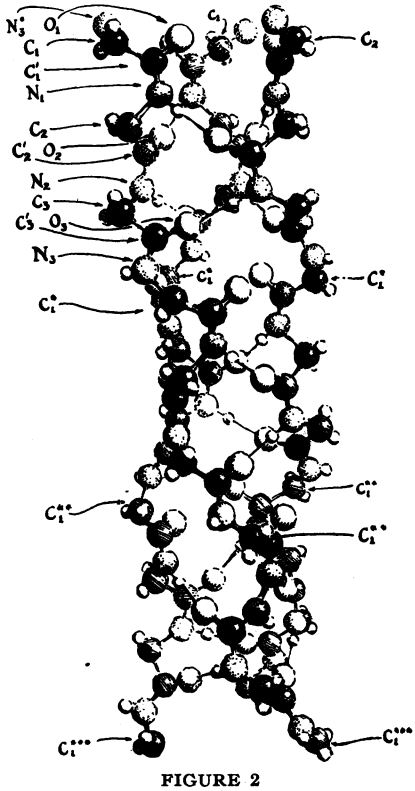

FIGURE 2

Drawing representing the proposed structure of the collagen-gelatin molecule.

A satisfactory structure can be built in which each of the three polypeptide chains is coiled into a helix, the coiling being achieved through the bending of the chain at the positions of the α carbon atoms, and the three helixes having a common axis. For the ideal configuration, in which the three α carbon atoms are similarly oriented directly above and below one another along a line parallel to the fiber axis, the dihedral angle at these carbon atoms is 97°. Some distortion of the structure is required in order that satisfactory hydrogen bonds be formed. This distortion consists in a

rotation around the two single bonds formed by the α carbon atoms, so as to draw the α carbon atom C_2 closer to the axis of the helix, by 0.34 A, than the α carbon atoms C_1 and C_3, as shown in figure 2. The hydrogen bonds are introduced in such a way that the three chains of the molecule are related to one another by a three-fold axis of symmetry. The α carbon atoms C_2 of the three chains lie at the corners of an equilateral triangle of edge 5.60 A, and the atoms C_1 and C_3 at the corners of an equilateral triangle with edge 6.20 A. The two cis amide residues are rotated by about 9° out of the orientation parallel with the fiber axis. The trans residue is rotated through 30° about its $C_3 - C_1^*$ axis. The trace of this axis on the basal plane is 2.44 A. The relation of this trace to the 6.20-A triangle of the C_3 and C_1 atoms is such that the angle of rotation about the fiber axis that converts one three-residue element of a polypeptide chain into the element following it in the chain is 40°. We have found that this angle of rotation can hardly be varied by more than 3° without introducing unsatisfactory structural features. The helix formed by a single chain of the collagen molecule is thus found to have very nearly a 9-fold screw axis of symmetry.

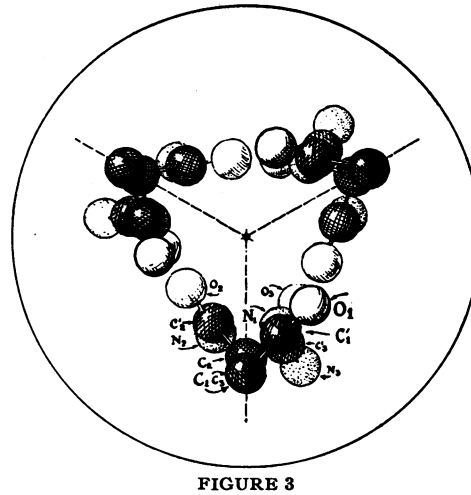

FIGURE 3

Plan of the three-chain configuration proposed for collagen and gelatin.

There is, of course, no reason for an isolated gelatin molecule to have exactly the angle 40° for the rotation of its rotatory translation, or to have exactly a 9-fold screw axis. In an aggregate of molecules in hexagonal packing, however, the influence of adjacent molecules on a given molecule might well be such as to introduce a small torque that would cause it to assume exactly the value 40° for this angle, and thus to assume a true 9-fold screw axis, in addition to the 3-fold symmetry axis that converts one chain into another, as shown in basal-plane projection in figure 3. We have found this phenomenon to occur in the hexagonal crystals of poly-γ-methyl-L-glutamate,[6] in which the α helix, normally with approximately 3.69 residues per turn, is constrained to the value 18/5 = 3.60 in order to achieve a 6-fold screw axis. In the following discussion we describe a molecule with a 9-fold screw axis.

The translation of the operation that converts one structural element

into the following one in the chain is given for our model as 8.58 A, which corresponds to 2.86 A per residue, in exact agreement with the x-ray value.

The coordinates of atoms in a structural unit of the molecule are given in Table 1. The coordinates x, y, and z refer to cartesian axes centered at the α carbon atom C_2, and the coordinates ρ, θ, and z are cylindrical coordinates referred to the axis of the helix. These atomic positions correspond closely to the assumed bond angles, bond distances, and planarity of the

TABLE 1

ATOMIC PARAMETERS FOR THE COLLAGEN THREE-CHAIN HELIX

x, y, z, AND ρ IN A

ATOM	x	y	z	ρ	ϕ
C_1	−0.25	−0.25	2.83	3.59	0.0°
C_1'	1.16	−0.09	2.19	2.63	19.6°
O_1	2.18	−0.09	2.90	2.38	42.3°
N_1	1.17	0.00	0.88	2.55	18.9°
C_2	0.00	0.00	0.00	3.25	0.0°
C_2'	0.05	1.24	−0.90	2.47	−19.9°
O_2	0.22	2.38	−0.41	2.07	−47.5°
N_2	−0.04	1.06	−2.21	2.63	−17.2°
C_3	−0.25	−0.25	−2.83	3.59	0.0°
C_3'	1.01	−0.37	−3.73	2.94	19.3°
O_3	2.03	0.21	−3.41	2.10	37.9°
N_3	0.89	−1.09	−4.80	3.66	22.5°
C_1*	1.98	−1.32	−5.75	3.59	40.0°
Axis	2.29	2.29	...	0.00	...

TABLE 2

COMPARISON OF EQUATORIAL FEATURES OF COLLAGEN X-RAY PHOTOGRAPHS AND CALCULATED MAXIMA AND MINIMA OF THE FORM FACTOR

CALCULATED SPACINGS	OBSERVED SPACINGS
7.14 A, minimum intensity	7.8 A, center of light band
4.72 maximum	4.37 center of diffuse dark band
2.89 minimum	2.74 center of light band
2.25 maximum	2.18 center of dark band
1.64 minimum	

amide groups. Because of the complexity of the structure, we are not sure that a considerable displacement of the chains from the proposed positions might not exist, in such a way as to retain these structural features.

The nature of the inter-chain interactions is shown in figures 1 and 3. Two hydrogen bonds are formed per unit. The N_1—H···O_2 bond has the length 2.63 A. This is slightly shorter (by about 0.06 A) than any N—H··· O hydrogen bond so far reported, but it seems to us to be possible that the bond in this molecule is indeed this short. The N_2—H···O_3 hydrogen bond has the length 2.83 A. If our model were to be changed by a small addi-

tional rotation of the trans amide group the distance would be significantly shortened.

There is an interesting correlation between the chemical composition of the collagen-gelatin proteins and the existence in the structure of only two, rather than three, hydrogen bonds per element. It is found by chemical analysis that about one-third of the amino acids obtained by hydrolysis of collagen and gelatin are proline or hydroxyproline. In residues of these amino acids the nitrogen atom does not have an attached hydrogen atom and so does not enter into hydrogen-bond formation. The proposed structure may explain this aspect of the chemical composition of these proteins. It seems likely that if long polypeptide chains were to be synthesized with every third residue a proline or hydroxyproline residue and every third residue a glycine residue they would spontaneously aggregate into complexes of three, with the collagen structure, the other principal known structures for fibrous proteins, the α-helix structure, pleated-sheet structure, and γ-helix structure, being rendered unstable by the steric interference of the 5-membered rings that are held in a fixed orientation.

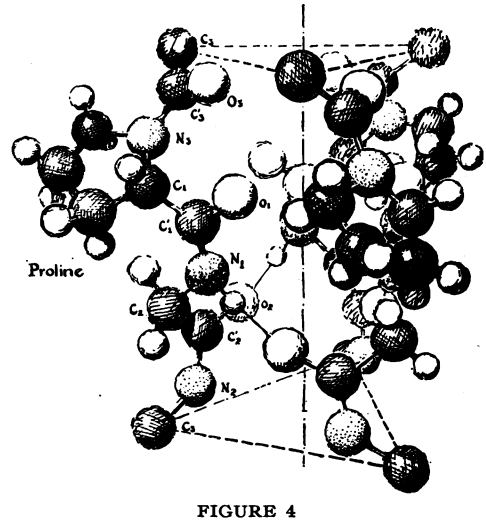

FIGURE 4

Drawing of a portion of the proposed structure of the collagen-gelatin molecule, showing the positions of the proline residues.

It is of interest that the sense of the collagen helix is uniquely related to the configuration of the amino acid residues. It is possible for only one of the two alternative positions of a β carbon atom on the α carbon atom C_1 of the proline residue to be connected with the nitrogen atom to form a 5-membered ring, for a given sense of the helix. The acceptable configuration is shown in figure 4.

The structure of the collagen-gelatin molecule can be represented diagrammatically as shown in figure 5. The three helical polypeptide chains are shown projected onto the surface of a cylinder. It is interesting to note that the structure of the molecule provides an immediate explanation of the principal mechanical property of collagen, its extensibility over only a limited range. The effective fiber-axis length of the trans residue in the molecule as shown is 2.93 A. If, on the application of force, this residue were to be twisted into a parallel orientation, its effective length would be

3.83 A, the C_3—C_1* distance. This maximum increase in length corresponds to a 10 per cent extension for the molecule. It is likely, however,

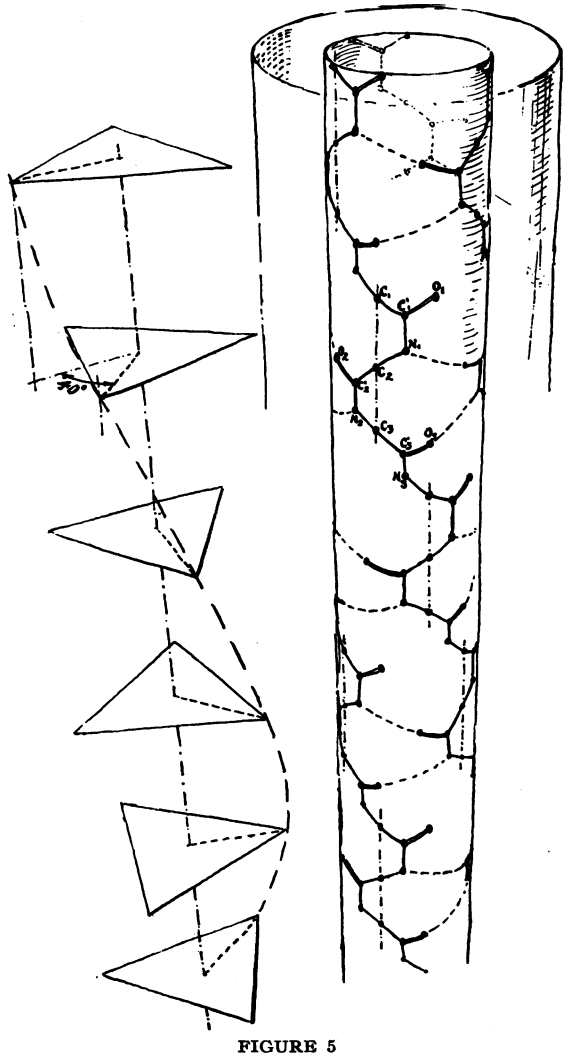

FIGURE 5

Diagrammatic representation of the collagen three-chain helix.

that the bond angles and planarity of amide groups would prevent complete parallel orientation, and that only a somewhat smaller extension could be achieved.

The way in which the molecules are packed together in a fiber of tendon or connective tissue is shown in figure 6. The cylindrical molecules are arranged in hexagonal packing. As discussed above, they are all helical, with the same sense, either right-handed or left-handed. It cannot be predicted whether in tendon or other collagenous material the molecules would all be oriented similarly, or whether some would be oriented in one direction and others in the opposite direction. It seems likely that the distribution of side chains reflects the direction of the polypeptide chains in the molecules, in such a way that the packing would be better in case that the molecules are all similarly oriented than if they were to alternate in

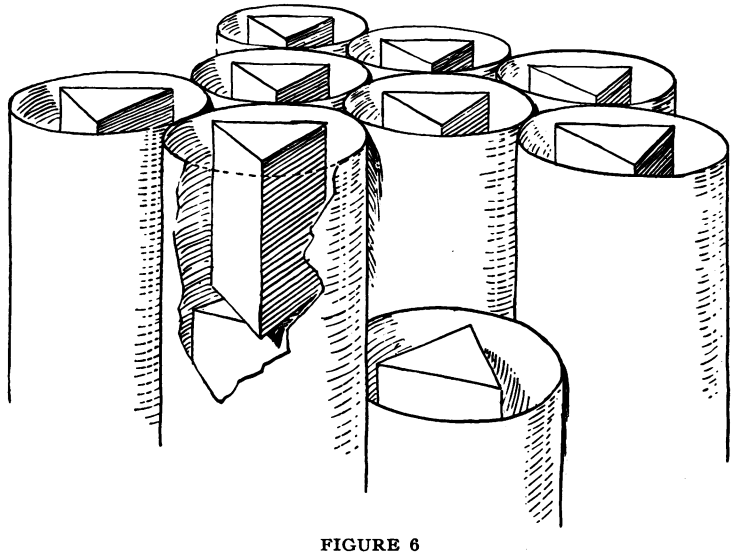

FIGURE 6

Diagrammatic representation of an aggregate of collagen molecules in a collagen fiber.

orientation. A preliminary study of the packing of side chains indicates that the glycine residues lie immediately below the proline and hydroxyproline residues (in the molecule oriented as shown in figure 2); that is, that the carbon atom C_2 is a methylene carbon atom of a glycine residue. It has been found by analysis that about one-third of the residues are glycine residues.

The proposed structure of the collagen-gelatin molecule accounts in a striking way for the principal features of the x-ray diffraction pattern of collagen and gelatin. A fiber consisting of these cylindrical molecules in hexagonal packing, with $a_0 = 12.50$ A (at normal humidity), would be predicted to produce equatorial reflections $\{10\cdot0\}$, with $d = 10.83$ A, $\{11\cdot0\}$,

with d = 6.25 A, {20·0}, with d = 5.42 A, and {22·0}, with d = 3.13 A. The lines reported by Corey and Wyckoff,[8] Astbury,[2] and other investigators are a strong reflection at 10.9 A and a medium reflection at 5.42 A. We may ask why these reflections appear, and {20·0} and {22·0} do not. The answer is given by the calculation of the form factor for the collagen molecule, with the parameters of table 1, and the assumption of cylindrical symmetry. The equation $F = \sum_i f_i J_0 (4\pi\rho_i \sin\theta/\lambda)$ leads to the function F shown in figure 7. This function has nodes and maxima as given in table 2, where comparison is made with the corresponding features as measured by us on the photographs of raw kangaroo tendon made by Corey and

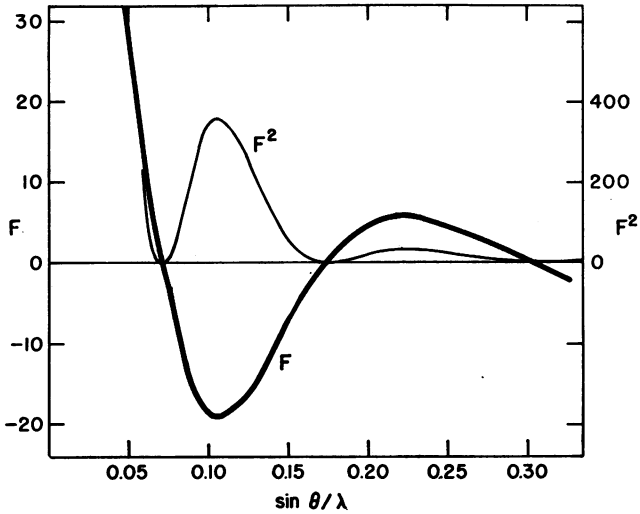

FIGURE 7

Calculated form factor and square of form factor for equatorial reflections of the collagen three-chain helix, calculated for cylindrical symmetry.

Wyckoff.[8] We see that it is predicted that a minimum in intensities of reflections would occur at about d = 7.14 A and another minimum at d = 2.89 A; these values are close to those for {20·0} and {22·0}, 6.25 A and 3.13 A, respectively. Moreover, there is observable on the photographs, and on published photographs of collagen, such as those of Corey and Wyckoff and of Astbury, a general light band, at about the azimuthal angle predicted for the minimum at 7.14 A equatorial spacing, followed by a region of general darkening, for which Astbury has located the center at 4.4 A, and which we have also measured at 4.37 A. This band is followed by a lighter band, and then a faint darker ring, the center of which we measure at 2.18 A, in excellent agreement with the predicted maximum of inten-

sity, 2.25 A. It seems likely that the presence of the 3-chain collagen helix can be recognized in a fibrous material more easily from this general blackening, arising from disordered molecules, than from the crystallographic diffraction maxima.

The predicted unit of structure of crystalline collagen, except for perturbations due to the distribution of side chains, is a hexagonal unit with a_0 = 12.5 A (for collagen at normal humidity) and c_0 = 25.74 A, the fiber-axis distance for nine residues. We have noted that the wide-angle pattern of collagen can be completely or nearly completely indexed in terms of this unit. Except for the side chains, the suggested hexagonal structure involves only one undetermined parameter, the azimuthal angle fixing the orientation of the molecules relative to the crystal axes. The meridional small-angle reflections reported by Corey and Wyckoff,[8] Bear,[3] and other workers are also seen to be related to the unit. These reflections seem to represent orders from the large spacing of about 640 A (which appears not only in x-ray photographs but also in electron micrographs[9]), apparently enhanced when they bear a nearly rational relation to c_0.

(Added May 5, 1951.) An interesting paper on infrared spectra and structure of fibrous proteins, by E. J. Ambrose and A. Elliott, *Proc. Roy. Soc.*, **A 206**, 206 (1951), has just appeared. The results obtained agree well with our proposed structures. In particular, Ambrose and Elliott conclude that in collagen "the N—H bond must be nearly if not exactly normal to the chain axis, and this applies also to the C=O bond," in excellent substantiation of the prediction on the basis of our three-chain helical structure.

This work was aided by grants from The Rockefeller Foundation, The National Foundation for Infantile Paralysis, and The U. S. Public Health Service.

* Communication No. 1555.

[1] Herzog, R. O., and Jancke, W., *Ber.*, **53**, 2162 (1920).
[2] Astbury, W. T., *J. Int. Soc. Leather Trades' Chem.*, **24**, 69 (1940).
[3] Bear, R. S., *J. Am. Chem. Soc.*, **66**, 1297 (1944).
[4] Astbury, W. T., and Bell, F. O., *Nature*, **145**, 421 (1940).
[5] Pauling, L., Corey, R. B., and Branson, H. R., these PROCEEDINGS, **37**, 205 (1951).
[6] Pauling, L., and Corey, R. B., *Ibid.*, **37**, 241 (1951).
[7] Pauling, L., and Corey, R. B., *Ibid.*, **37**, 261 (1951).
[8] Corey, R. B., and Wyckoff, R. W. G., *J. Biol. Chem.*, **114**, 407 (1936).
[9] Schmitt, F. O., Hall, C. E., and Jakus, M. A., *J. Cell. Comp. Physiol.*, **20**, 11 (1942); *J. Am. Chem. Soc.*, **64**, 1234 (1942).