THE STRUCTURE OF COLLAGEN

By Dr. ALEXANDER RICH* and Dr. F. H. C. CRICK

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge

VERY recently, Ramachandran and Kartha[1] have made an important contribution by proposing a coiled-coil structure for collagen. We believe this idea to be basically correct but the actual structure suggested by them to be wrong.

Their structure consists of three polypeptide chains, each having approximately a three-fold screw axis. In addition, the chains slowly wind around each other to form a coiled coil, thus reproducing the observed non-integer screw axis[2]. The major helix is right-handed, the minor one left-handed. Each chain is held to its neighbours by *two* sets of systematic hydrogen bonds.

The allowed sequence of residues is

—*G*—*R*—*P*—*G*—*R*—*P*—, etc.,

where *G* implies glycine only, *R* implies any residue except proline or hydroxyproline, and *P* implies any residue, but usually proline or hydroxyproline.

We believe this structure to be wrong for two reasons. (1) It is stereochemically unsatisfactory. In particular, there is a very short C_α—C_α contact of 3·3 A. (normally 3·6–4·0 A.) and an extremely short C_α—O contact of 2·6 A. (normally 3·2–3·5 A.). In addition, the hydrogen bond angles are on the outside limit of the values usually found. (2) It is not compatible with recent work[3] on the amino-acid sequence, which shows that

—gly—pro—hypro—

is a common sequence in collagen.

On the other hand, Dr. Pauline Cowan and her co-workers[4] have informed us that a preliminary optical diffraction pattern of this structure agrees qualitatively with the observed wide-angle X-ray pattern of collagen. They have also pointed out to us that the configuration of the backbone of the polypeptide chain is similar to that found by them for polyproline[5]. These facts suggest that the structure, *though incorrect, is on the right general lines.*

Our own work on collagen has sprung from our recent structure for polyglycine II[6]. We have taken a compact group of three adjacent chains out of the polyglycine lattice (space group $P\,3_1$) and twisted them, as in the Ramachandran–Kartha structure, to form a similar coiled coil. Such a group can be selected in two different ways, since the symmetry is *trigonal*, not hexagonal. We have called the two resulting arrangements structure I and structure II. Both have a right-handed major helix and a left-handed *minor* helix, arranged to fit the observed non-integer screw. Both have only one set of systematic hydrogen bonds linking neighbouring chains. In structure I the NH-groups point anticlockwise when viewed from the carboxyl ends of the chains; in structure II clockwise.

These two structures are thus similar to that of Ramachandran and Kartha in being three-chain coiled-coil structures; but they differ in having only one set of systematic hydrogen bonds instead of two. Moreover, they are both stereochemically completely satisfactory.

Both structure I and structure II will accommodate the sequence

—gly—pro—hypro—

but they do so in a different manner. In structure II the glycine position is near the axis, and every third residue must be glycine. The proline and hydroxyproline positions, on the other hand, are far removed from the axis. Either proline or hydroxyproline can go into either position—there seems to be no preference—and *in neither case can the OH of hydroxyproline make a hydrogen bond with a CO group of an adjacent chain within the same set of three chains.*

Structure I will also accommodate the above sequence; but when the hydroxyproline is in the expected position (that is, previous to glycine) its OH group can form a hydrogen bond to one of the adjacent chains within the same set. Careful structure building has shown that not all the possible positions for hydroxyproline can be occupied if the structure is to fit together comfortably. This is compatible with the amino-acid analyses of collagen, which show that the amount of hydroxyproline present in bovine collagen[7] would fill about one-third of these sites. The polypeptide backbones, being held together by only one set of hydrogen bonds, have a certain amount of flexibility. In the undeformed structure, only glycine can be accommodated in the glycine sites. However, if the structure is deformed somewhat, these sites can accommodate other residues, though only to a limited extent. This may explain certain minor features of the amino-acid sequence data.

We have made an exhaustive study of all possible structures (using topological enumeration, similar to that of Bragg, Kendrew and Perutz[8]) of this general type which are compatible with the observed screw axis—that is, with three parallel chains linked by at least one systematic set of hydrogen bonds—and we find that: (1) *no structure with two systematic sets* of hydrogen bonds is stereochemically possible; (2) no other structure with *one* systematic set of hydrogen bonds is stereochemically satisfactory except the two described above derived from polyglycine. Neither were we able to add occasional backbone hydrogen bonds to structures I and II in a convincing manner.

All the above findings are independent of arguments about the side-chain arrangements. We therefore conclude that structures I and II, though making few hydrogen bonds systematically, are the best that can be achieved along these lines. Note that, as in polyglycine II[6], it is not impossible that the direction of one of the three chains may be reversed.

Preliminary work on the optical transforms of these two structures (carried out with Dr. A. Elliott on his optical transform machine) show both structures to give rough agreement with the X-ray pattern, though structure II shows a discrepancy on the fourth layer-line. Both structures are also compatible with the infra-red[9] observations. Further work will therefore be required to decide between them. We are at the moment inclined to favour

* Permanent address: Physical Chemistry Section, National Institute of Mental Health, Bethesda, Maryland.

structure I, because it makes effective and systematic use of the hydroxyproline residues. Gustavson[10] has presented extensive evidence suggesting that this unusual amino-acid stabilizes the collagen structure. We feel that this is more likely to take place by formation of hydrogen bonds within a group of three chains than between different groups of chains[1]. In addition, structure I explains in a natural way the amino-acid sequence data. Thus while the peptide hypro—gly is common, gly—hypro is not found. Similarly, gly—pro is common, whereas pro—gly is rare. In structure I this preference is explained in terms of the hydroxyproline-hydrogen bond between the chains.

Although we cannot at the moment make a final decision between the two structures, we think it very probable that one of them is correct (or, less likely, both). The general agreement with the X-ray pattern, the close resemblance to polyproline and to polyglycine II, and the ability to explain the major features of the amino-acid sequence data all suggest that collagen is a structure of this type; that is, a three-chain, coiled-coil structure, with one set of systematic hydrogen bonds.

Full details of our findings together with co-ordinates of both structures will be presented elsewhere.

We should like to thank Dr. A. Elliott for his assistance in obtaining the optical transforms.

[Oct. 10

[1] Ramachandran, G. N., and Kartha, G., *Nature*, **176**, 593 (1955).
[2] Cohen, C., and Bear, R. S., *J. Amer. Chem. Soc.*, **75**, 2783 (1953). Cowan, P. M., North, A. C. T., and Randall, J. T., in "The Nature and Structure of Collagen" (Butterworth, London, 1953); and in "Fibrous Proteins and Their Biological Significance", Soc. Exp. Biol. Symp., **9** (Camb. Univ. Press, 1955). Bear, R. S., *ibid.*
[3] Schroeder, W. A., Kay, L. M., Le Gette, J., Hounen, L., and Green, F. C., *J. Amer. Chem. Soc.*, **76**, 3556 (1954). Kroner, T. D., Tabroff, W., and McGarr, J. J., *J. Amer. Chem. Soc.*, **77**, 3356 (1955).
[4] Personal communication. Also communicated at the 3rd International Congress of Biochemistry, Brussels, 1955.
[5] Cowan, P. M., and McGavin, S., *Nature*, **176**, 501 (1955).
[6] Crick, F. H. C., and Rich, A., *Nature*, **176**, 780 (1955).
[7] Bowes, J. H., and Kenton, R. H., *Biochem. J.*, **42**, 358 (1948).
[8] Bragg, W. L., Kendrew, J. C., and Perutz, M. F., *Proc. Roy. Soc.*, A, **203**, 321 (1950).
[9] Ambrose, E. J., and Elliott, A., *Proc. Roy. Soc.*, A, **206**, 206 (1951). Badger, R. M., and Pullin, A. D. E., *J. Chem. Phys.*, **22**, 1142 (1954). Sutherland, G. B. B. M., Tanner, K. N., and Wood, D. L., *J. Chem. Phys.*, **22**, 1621 (1954).
[10] Gustavson, K. H., *Nature*, **175**, 70 (1955).