The Structure of Collagen Molecules and Fibrils.* By RICHARD S. BEAR. (*From the Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts.*)‡

Following the suggestion of triple-chain models for the helical molecules of collagen (Ramachandran and Kartha, 1955), improved similar structures were

* This work was supported in part by a grant-in-aid from the American Cancer Society, Inc., upon recommendation of the Committee on Growth of the National Research Council, and also in part by Research Grant A-901 from the National Institute of

derived *via* polyglycine II (Rich and Crick, 1955) and poly-L-proline (Cowan, McGavin, and North, 1955). Since, however, these models have not been evolved unambiguously, it seemed worthwhile to report briefly the results of a systematic derivation which does much to provide additional confidence in them.

The concept of the helix-net (Bear, 1955) furnishes a graphical representation of the dictates of the wide-angle x-ray diffraction data concerning the symmetry properties of the helical collagen molecule. Use of the helix-net greatly facilitates systematic consideration of all possibilities for helical molecular chains and also makes possible greater brevity in discussion of such structures. For these reasons liberal use of the helix-net and associated terminology is made in what follows.

The models under present consideration have asymmetric units, of three residues each, connecting every third node along the left-handed genetic screw of the helix-net. The appropriateness of three-residue asymmetric units throughout the collagen molecule has not yet been convincingly demonstrated. Earlier calculations regarding density were made on the assumption of pseudohexagonal packing of the molecules in the fibril. Then appreciably more than three (3.3) average residues per asymmetric unit were indicated (see Bear, 1955). A more accurate treatment of density has become possible since the determination of the basal cell section (a = 62 A, c = 76 A, β = 125°) for moist collagen fibrils by North, Cowan, and Randall (1954).

Whatever the nature of the asymmetric units or the manner of internodal connection, the helix-net shows that the main chains of the collagen molecule (but perhaps not all side-chains) project normal to its axis as a figure of tenfold rotational symmetry. As demonstrated in Fig. 1, molecular cross-sections of this symmetry pack satisfactorily into a basal cell

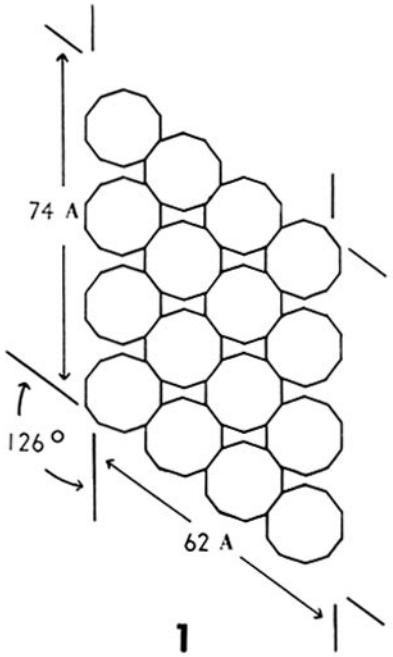
Arthritis and Metabolic Diseases of the National Institutes of Health, United States Public Health Service.

‡ *Received for publication, February 23, 1956.*

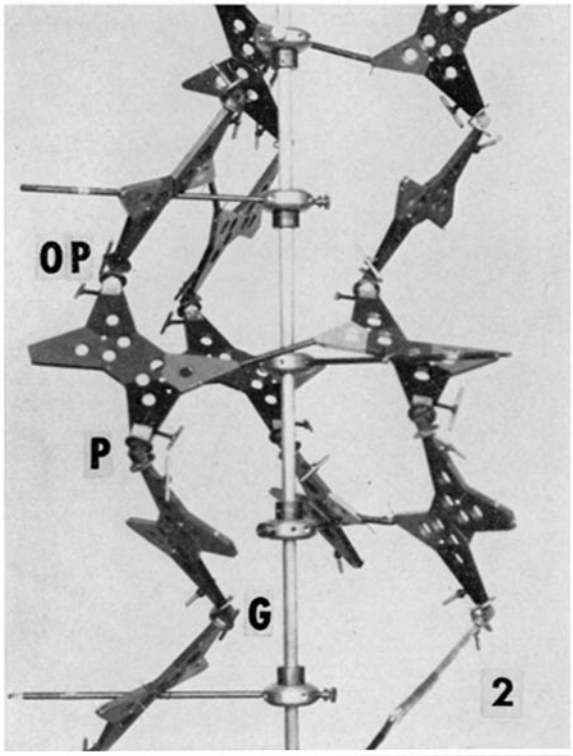
FIG. 1. Illustrating how decagonal figures, representing cross-sections of molecules having projections with tenfold rotational symmetry, pack into the basal cell section of the collagen fibril.

FIG. 2. A segment of the Rich-Crick model II for the collagen molecule, constructed of butterfly-shaped symbols, each representing amide or imide peptide links. Shading distinguishes bonds (not atoms) as follows: light areas correspond to N–H bonds, gray to C = O, dark gray to C-N and N-C_β, near black to N-C_α, and black to C_α-C. Springs joining N–H and O = C represent interchain hydrogen bridges. Several connectors are labelled *G*, *P*, and *OP* to mark typical C_α's which, respectively, *must* bear a glycine side-chain (hydrogen) or *can* carry proline or hydroxyproline side-rings. The rods supporting the structure are arranged according to the appropriate left-hand genetic screw (axial translation 2.86 A, rotation 108°). The model is complete only for the central three layers of links.

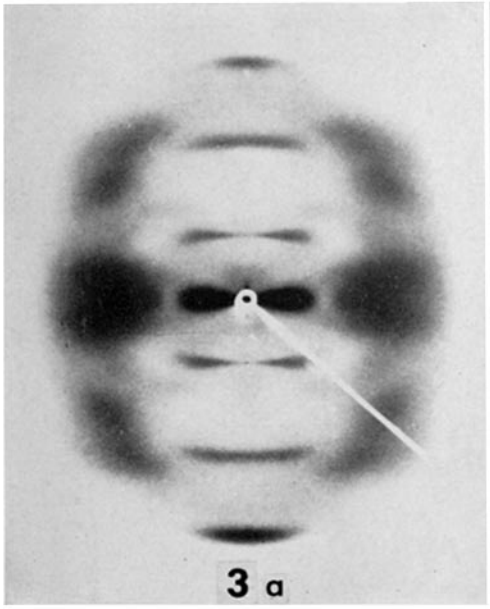
FIG. 3. A comparison of (*a*) the wide-angle x-ray pattern of moist collagen, with (*b*) an optical transform for model II, reproduced at the same magnification. The x-ray pattern shows, at higher layer lines, a drawing of intensity toward the meridian (because of the reflection sphere's curvature), which is effectively absent on the optical transform. The latter was made from a projection of the structure upon a plane through the helix axis, with use of annular "atoms" simulating the atomic structure factor's decline of amplitude with angle of scatter. All possible (excessive) pyrrolidine side chains were included.



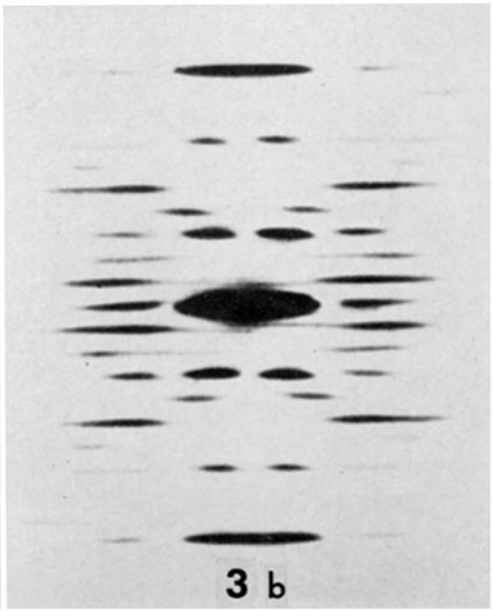
1



2



3 a



3 b

section differing only slightly from the reported shape. Indeed, 16 molecular sections apparently provide the correct content for the specific cell dimensions involved here, as can be shown by the following calculations.

Using three residues per node, each of average weight 93, one calculates 0.70 gm. per cc. for contribution of protein to the density of moist collagen fibrils. These residues, of average molal volume 66 cc., allow room for water furnishing 0.50 gm. per cc. of additional density. The total fibrillar density of 1.20 thus calculated may be compared to 1.16 estimated for the over-all density of a kangaroo tail tendon specimen at the state e described by Rougvie and Bear (1953). This sample had been minimally rewet to exhibit diffraction typical of moist collagen (110 gm. H_2O/100 gm. protein), and at least 56 per cent of the moisture entered the fibrils as shown by influence on the x-ray diffraction.

While drying introduces distortion sufficient to destroy most of the x-ray evidence for basal cell dimensions, symmetry considerations similar to those used above indicate that the cell section should not change shape. At constant cell shape, the dry fibrillar density of 1.41 is attained when the distance between molecular axes shrinks from 15.5 to 10.9 A, corresponding finally to 10.6 A between planes through axes and approximating the spacing derived from the prominent equatorial arc of the x-ray diagram of dry collagen.

Constancy of cell shape permits only the nearest alternative choices of 9 or 25 molecules per cell. One may also consider asymmetric units comprising 2 or 4 residues. All combinations of these alternatives are easily excluded. Indeed, 16 molecules per cell and 3 residues per asymmetric unit are shown above to furnish densities in agreement with experiment within less than 5 per cent, which is excellent for such cases.

Having thus determined the suitability of three-residue asymmetric units one can then show from the helix-net that models containing from one to three separate but coaxial polypeptide chains are formally possible (Bear, 1955). Before the unique suitability of the current triple-chain models can be concluded, the possibility of single- and double-chain models must be eliminated. The systematic procedure described below was developed for this purpose. Results of the intermediate stages are not considered worth detailed description, but the method is outlined because it provides a general procedure for systematically obtaining polypeptide chain configurations suitable for trial on any given helix-net.

Mechanical models for —CO—NH— (herein termed for convenience "amide" or A) and for —CO—NC_δ- ("imide" or I) peptide links, useful in forming polypeptide main chains containing both amino and imino acid residues, were constructed at a scale 2 inches = 1 A. Currently approved interatomic distances and angles, involving "planar-trans" configuration of both amide and imide groups were adopted. The planar-trans nature of I links was suggested by the results of Leung and Marsh (1955) on the structure of L-leucyl-L-prolylglycine, and was also found satisfactory for poly-L-proline by Cowan and McGavin (1955). The links were joined by C_α connectors bearing dials about which N—C_α and C_α—C rotations could be controllably varied. The N—C_α rotations of the imides were fixed to agree with the crystallographic data on L-proline (Donahue and Trueblood, 1952).

Chemical evidence suggests the frequent occurrence of glycylprolylhydroxy-

prolyl sequences in collagen (Kroner, Tabroff, and McGarr, 1955). Two such three residue sequences were simulated as follows:—

$$- C_\alpha - A - C_\alpha - I - C_\alpha - I - C_\alpha - A - C_\alpha - I - C_\alpha - I - C_\alpha -$$

$$|\text{———————}\ l_1\ \text{———————}|$$

$$|\text{———————————————}\ l_2\ \text{———————————————}|$$

Distances l_1 and l_2, between C_α atoms as indicated, were measured for all allowed rotational variations which did not cause intrachain atom crowding and which kept corresponding rotations in the two AII segments identical. Configurations which offer trial structures for the collagen helix must have l_1 and l_2 as consecutive cylindrical chords interconnecting nodes of the helix-net at equal radii, r. This geometrical condition requires that

$$r = \frac{l_1^2 - p}{2(1 - q)} = \frac{l_2^2 - 4p}{4(1 - q^2)}.$$

Here $p = b_h^2/M^2$, $q = \cos(2\pi T/M)$, in which b_h is the axial period of any one of the possible individual helices consistent with the helix-net (*not* the period of the total combination), while M is the number of asymmetric units and T the number of turns in one period of the same individual helix. For example, with collagen b_h, M and T are, respectively b, 10 and 3 for single-chain models; b, 5 and 2 for double-chain structures; and 3 b, 10 and 1 for triple-chain models. The axial period, b, of the total configuration is 28.6 to about 30 A for collagen under varying degrees of tension.

The following numbers of configuration were found to satisfy the geometrical conditions for collagen: none suitable for single-chain models, 5 for double-chain structures, and 2 for triple-chain cables. However, all but one of the seven trial configurations could not be placed on either the right- or left-hand genetic screws of collagen to yield satisfactory intra- or interchain hydrogen bridging and/or lack of crowding of main-chain or side-chain atoms of separate chains.

The successful configuration is essentially that of the main-chain atoms in single chains of poly-L-proline (Cowan and McGavin, 1955) and of polyglycine II (Crick and Rich, 1955). The several ways in which this simple, left-handed (minor) coil can be oriented, axially and rotationally, in placing three coils on the left-handed genetic screw of collagen, to form, however, a right handed major twist of the resulting cable, have already been discussed (Rich and Crick, 1955). Fig. 2 illustrates the Rich-Crick model II constructed of the mechanical models described above.

Statements of Rich and Crick that the new models provide optical transforms in good agreement with x-ray diffraction have been confirmed with atom coordinates kindly supplied by them (see Fig. 3). Possibly their collagen II is better in this respect than collagen I. Experience with optical transforms of other models devised for collagen suggests, however, that the optical examinations provide what may be termed a "necessary" rather than a "sufficient" test, because of lack of detail in the x-ray diffraction and incomplete knowledge of side-chain distribution. The present independent and systematic means of arriving

at these models contributes an added confidence in their sufficiency by greatly reducing the likelihood that other very different ones need be considered.

A number of considerations, not to be developed here, suggest that eventually there will be required minimal but significant variations from the basic molecular structure and from the intrafibrillar packing indicated here. Thus it is likely that Fig. 1 is to be interpreted at present only in the limited sense of indicating (*a*) the number of molecules which density requires across the basal cell section, and (*b*) the relation of the cell section's shape to the approximate tenfold symmetry of the molecules.

Bibliography

Bear, R. S., *Symp. Soc. Exp. Biol. (Great Britain)*, 1955, No. 9, 97.

Cowan, P. M., and McGavin, S., *Nature*, 1955, **176**, 501.

Cowan, P. M., McGavin, S., and North, A. C. T., *Nature*, 1955, **176**, 1062.

Crick, F. H. C., and Rich, A., *Nature*, 1955, **176**, 780.

Donahue, J., and Trueblood, K. N., *Acta Cryst.*, 1952, **5**, 414, 419.

Kroner, T. C., Tabroff, W., and McGarr, J. J., *J. Am. Chem. Soc.*, 1955, **77**, 3356.

Leung, Y. C., and Marsh, R. E., American Crystallographic Association Meeting, Pasadena, California, 1955, Paper 75.

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

Ramachandran, G. N., and Kartha, G., *Nature*, 1955, **176**, 593; *Proc. Indian Acad. Sc., Sect. A*, 1955, **42**, 215.

Rich, A., and Crick, F. H. C., *Nature*, 1955, **176**, 915.

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