Polymers of Tripeptides as Collagen Models

I. X-Ray Studies of Poly (L-prolyl-glycyl-L-proline) and Related Polytripeptides

W. TRAUB AND A. YONATH

Department of X-Ray Crystallography
Weizmann Institute of Science, Rehovoth, Israel

X-Ray diffraction studies have been made of several polytripeptides related to collagen. The structure of poly(L-prolyl-glycyl-glycine) has been found to consist of helices which resemble the individual strands of the triple helix models for collagen. However, the helices are not coiled about each other as in the collagen models. Both poly(L-prolyl-L-alanyl-glycine) and poly(L-prolyl-glycyl-o-acetyl-L-hydroxyproline) give X-ray patterns which resemble that of collagen, including the characteristic 2·9 Å spacing, but are too diffuse for detailed analysis. Poly(L-prolyl-glycyl-L-proline), however, gives an X-ray pattern which has all the main features of the collagen pattern and is appreciably sharper in detail. As in collagen, water absorption leads to an increase in the equatorial, but not in the meridional, spacings. The X-ray pattern and the density indicate a helical structure for $(\text{Pro}.\text{Gly}.\text{Pro})_n$ with an axial translation of 2·85 Å and a rotation of approximately 108° per tripeptide. Only a structure consisting of three strands coiled about a common axis can be fitted satisfactorily to these helical parameters. Of the three-stranded models that have been proposed for collagen, that with two hydrogen bonds per tripeptide can be excluded on chemical grounds, whereas the collagen I model is incompatible with the observed unit cell. However, slightly modified versions of collagen II or the closely similar alternative Madras structure satisfy both criteria. The results show that neither hydroxyproline nor more than one interchain hydrogen bond per tripeptide is required for the formation of a collagen-like structure. In the light of recent findings concerning the sequence and other properties of collagen, it is suggested that much of the protein may have a structure very similar to that of $(\text{Pro}.\text{Gly}.\text{Pro})_n$.

1. Introduction

Collagen shows a distinctive X-ray diffraction pattern which differs markedly from those of almost all other fibrous proteins. Its main features include a strong equatorial reflection varying with humidity from 10·5 Å in dry collagen to about 15 Å in wet, a strong meridional arc at 2·9 Å on the tenth layer line, near-meridional reflections on the third and seventh layer lines and a diffuse distribution of intensity around 4·5 Å mainly on and near the equator (Ramachandran & Ambady, 1954; Cowan, North & Randall, 1955; Lakshmanan, Ramakrishnan, Sasisekharan & Thathachari, 1962). This diffraction pattern was interpreted by Ramachandran & Kartha (1955) in terms of three helical polypeptide chains, each having every third residue glycine, which are twisted about each other to form a three-stranded coiled coil. The individual chains have a conformation similar to that of poly-L-proline II (Cowan & McGavin, 1955;

Sasisekharan, 1959) and polyglycine II (Crick & Rich, 1955), and the whole structure is based on a non-integral screw axis which relates equivalent units by a translation of 2·9 Å and a rotation of approximately 108°.

Although the essential correctness of this type of structure has been widely accepted, three alternative modifications of it have been proposed. They differ in details of conformation and the mode of interchain hydrogen bonding. In particular, one modification (Ramachandran, 1963) has two systematic hydrogen bonds of the type NH...O for every three amino acid residues, whereas the other two, the so-called collagen I and collagen II structures (Rich & Crick, 1961; Cowan, McGavin & North, 1955), have only one hydrogen bond for three residues. The unequivocal elucidation of such details of the structure has proved difficult because of the limited detail of the X-ray pattern of collagen and the complexity of its amino acid sequence, much of which is still unknown.

In recent years interest has turned to the study of possible polypeptide models of collagen as an aid to the understanding of its structure and physicochemical properties. To this end several polytripeptides have been synthesized which have every third residue glycine as well as residues of one or both of the imino acids proline and hydroxyproline, in accordance with features of the composition of collagen which are believed to play an important role in determining its structure (Kitaoka, Sakakibara & Tani, 1958; Berger & Wolman, 1961; Debabov, Kozarenko & Shibnev, 1961; Engel, Kurtz, Traub, Berger & Katchalski, 1964).

Two structural forms have been reported for poly(glycyl-L-prolyl-L-hydroxyproline); a low molecular weight form in which groups of three adjacent chains are hydrogen-bonded to each other as in the collagen II structure, but are parallel rather than coiled about each other (Andreeva & Millionova, 1964), and a high molecular weight form which resembles collagen in its optical rotation, infrared spectrum and X-ray pattern (Andreeva, Millionova & Chirgadze, 1963; Rogulenkova, Millionova & Andreeva, 1964).

The synthesis, in this Institute, of poly(L-prolyl-glycyl-L-proline), hereafter written (Pro.Gly.Pro)ₙ, and the resemblance of its X-ray diffraction pattern and some of its properties in solution to those of collagen have already been reported (Engel *et al.*, 1964). A collagen-like X-ray pattern for this polytripeptide has recently also been reported by Shibnev, Rogulenkova & Andreeva (1965). More extensive investigations of (Pro.Gly.Pro)ₙ have now been made. Those concerning its behaviour in solution are described in Part II of this communication (Engel, Kurtz, Katchalski & Berger, 1966). This portion, Part I, is devoted to a description of X-ray structural investigations of (Pro.Gly.Pro)ₙ and some data on several related polytripeptides.

2. Materials and Methods

All polytripeptide samples used in this investigation were obtained from the Biophysics Department of the Weizmann Institute. The synthesis and fractionation of (Pro.Gly.Pro)ₙ are described in Part II. Specimens for X-ray study were prepared both from unfractionated material, average molecular weight 6000, and from various fractions. However, in experiments involving a quantitative estimate of the amount of water absorbed by the polymer, only samples from which low molecular weight material (less than 5000) had been removed were used. Samples of poly(L-prolyl-glycyl-glycine)-(Pro.Gly.Gly)ₙ, poly(L-prolyl-L-alanyl-glycine)-(Pro.Ala.Gly)ₙ and poly(L-prolyl-glycyl-o-acetyl-L-hydroxyproline)-(Pro.Gly.o-acetyl Hypro)ₙ of average molecular weight 3500, 2000 and 2700, respectively, were used (Berger & Wolman, 1961; Wolman, 1961).

Specimens of (Pro.Gly.Pro)$_n$ were photographed as powders and as oriented fibres or films grown from aqueous solution. In addition, specimens of various degrees of hydration were prepared by suspending the material in a sealed glass capillary connected to a reservoir of saturated salt solution of known relative humidity (Shmueli & Traub, 1965). Calcium chloride, sodium dichromate, sodium chloride, potassium chloride, sodium tartrate and potassium sulphate were used to obtain relative humidities of 0, 52, 76, 86, 92 and 98%, respectively. Partially oriented films of (Pro.Gly.Gly)$_A$ were grown from formic acid. We were not able to prepare specimens giving appreciably oriented X-ray photographs from either (Pro.Ala.Gly)$_n$ or (Pro.Gly.o-acetyl Hypro)$_n$.

Most powder photographs were taken with 114·6-mm and 57·3-mm diameter cylindrical powder cameras and standard X-ray units. Thin oriented specimens were photographed on a Norelco microcamera used with a Hilger microfocus X-ray tube. Some thicker oriented specimens and those enclosed in bulky cells were photographed on a flat-plate camera with a Philips fine-focus tube. Photographs were taken with nickel-filtered CuKα radiation.

Molecular models were built from rod components (5 cm = 1 Å) produced by Cambridge Repetition Engineers, and close-packing components (0·8 in = 1 Å) produced by Courtauld.

X-Ray photographs were measured with a travelling microscope as well as with a Joyce–Loebl recording microdensitometer.

The density of (Pro.Gly.Pro)$_n$ was measured by flotation in a toluene–carbon tetrachloride mixture.

3. X-Ray Diffraction Pattern of (Pro . Gly . Pro)$_n$

Powder photographs of dry (Pro.Gly.Pro)$_n$ show a remarkable resemblance in spacing and intensity distribution to those of unstretched collagen, Plate I. In fact the pattern of the polytripeptide is somewhat sharper than that of the protein and we have been able to measure some ten spacings, Table 1.

Oriented X-ray patterns resembling those of collagen fibres were obtained both from fibres and films grown from aqueous solutions of (Pro.Gly.Pro)$_n$, Plate II. Both

TABLE 1

Observed and calculated spacings for dry form of (Pro.Gly.Pro)$_n$

Orientation	I_0	d_0(Å)	hkl	d_a(Å)
Equatorial	vs	10·85	100	10·83
Meridional	w	7·21	103	7·16
Equatorial	vw	6·19	110	6·27
Equatorial	w	5·45	200	5·41
			203	4·71
Broad		4·8 to	114	4·71
Unoriented	mw	4·4	106	4·36
			204	4·32
Meridional	mw	3·85	107	3·82
Meridional	vw	3·35	117	3·42
			207	3·26
Meridional	mw	2·85	0, 0, 10	2·85
Meridional?	vvw	2·50	2, 0, 10	2·52
			219	2·51

Observed intensities (I_0) were estimated as very strong (vs), moderately weak (mw), weak (w), very weak (vw) and very very weak (vvw). Indices (hkl) were assigned to the reflections and their spacings calculated (d_a) on the basis of an hexagonal cell with a = 12·5 Å and c = 28·5 Å. Only values of hkl have been listed for which helical diffraction theory (Cochran *et al.*, 1952) indicates there could be appreciable intensity at the observed orientation.

methods of preparation, however, gave rise to reverse orientation, i.e. (Pro.Gly.Pro)$_n$ reflections corresponding to near-meridional ones in collagen appeared near equatorial and *vice versa*. This phenomenon, in which the long molecular axes of linear polymers are oriented perpendicular to the length of a fibre or the plane of a film, is not an uncommon one, particularly in cases where the polymers are of relatively low molecular weight or have a strong tendency for side to side aggregation. It seems quite clear, both from the comparison with collagen and the indexing of the X-ray pattern described below, that reverse orientation in fact occurred in this case. Consequently, to avoid confusion, we will ignore this phenomenon in all further description and describe the X-ray pattern as if (Pro.Gly.Pro)$_n$ conformed to the more conventional mode of orientation exemplified by collagen.

All the reflections can be satisfactorily indexed, in accord with their observed spacings and orientations, in terms of hexagonal packing in a unit cell with $a = 12\cdot5$ Å and $c = 28\cdot5$ Å, see Table 1. On this basis, there is, as in collagen, a relatively strong meridional 2·85 Å reflection on the tenth layer line, indicating a helix with an axial translation of 2·85 Å per unit of structure and approximately 10 units in an integral number of turns. Also, as in collagen, there are relatively strong near-meridional reflections on the third and seventh layer lines, indicating, on the basis of helical diffraction theory (Cochran, Crick & Vand, 1952), that 10 units correspond in fact to three turns of the helix. Thus, (Pro.Gly.Pro)$_n$ appears to have, within experimental error, the same helical parameters as collagen. Also, as judged by the strong equatorial reflections at about 11 Å, they have similar lateral dimensions.

As a preliminary to testing the correctness of the above cell dimensions and helical parameters by a comparison of calculated and observed densities, the water content of (Pro.Gly.Pro)$_n$ at 0 and 52% relative humidity was determined. It was found that even after standing several days over phosphorus pentoxide *in vacuo* one molecule of water per tripeptide remained bound to the polymer. When equilibrated with a saturated salt solution at 52% relative humidity, (Pro.Gly.Pro)$_n$ was found to contain three molecules of water per tripeptide. As room humidity was in fact close to the latter value, the density determination was performed with material which had been equilibrated in a desiccator at 52% relative humidity. The observed density is 1·33 g/cm^3; that calculated on the basis of 10 tripeptide units in a cell of the above dimensions and three molecules of water per tripeptide is 1·31 g/cm^3.

Photographs of (Pro.Gly.Pro)$_n$ were taken at 0, 52, 76, 86, 92 and 98% relative humidity. Those at 0 and 52% R.H. are essentially identical, both showing the pattern described in Table 1. However, the X-ray patterns obtained at 76%, and higher relative humidities, though rather similar to that obtained at 0% R.H., are appreciably sharper and show some changes in spacing and intensity. The changes in spacing of the 100, 110 and 200 reflections in particular indicate a gradual expansion of the hexagonal unit cell with increasing humidity such that the a-axis increases from 12·5 Å at 52% R.H. to 13·6 Å at 98% R.H. without any appreciable change in the c-axis, Table 2. This expansion is reminiscent of the humidity-dependent changes in the collagen X-ray pattern. In addition, photographs at 98% R.H. and of very wet pastes of (Pro.Gly.Pro)$_n$ also show a new rich pattern with many sharp lines indicative of a major rearrangement in the packing of the polymer chains, Plate 1(e).

Especially in view of the indications that (Pro.Gly.Pro)$_n$ molecules may fold back on themselves in solution (see Part II), we have examined the X-ray photographs for any indications of an arrangement in which a single molecule by coiling back on itself

TABLE 2

Observed spacings in Ångstrom units of $(Pro.Gly.Pro)_n$ *axial reflections at various relative humidities*

hkl	Relative humidity (%)					
	0	52	76	86	92	98
100	10·85	10·90	11·23	11·58	11·74	12·00
110	6·19		6·45	6·60	6·67	6·89
200	5·45	5·47	5·64	5·71	5·84	5·93
0, 0, 10	2·85	2·83	2·86	†	†	2·87

† Photographs of specimens at 86 and 92% R.H. were calibrated assuming the spacing of the 0,0,10 reflection to be 2·85 Å.

might form a triple-stranded structure of two parallel and one antiparallel strands. We have not observed on any of the photographs meridional 8·6 Å or 4·3 Å reflections which might be expected to arise from such a structure (see section 4). However, apart from the possibility of a 4·3 Å reflection being hidden by the broad band in this region (see Table 1) and a very weak 8·6 Å reflection remaining unobserved, it should be noted that the photographs have a diffuse background indicating an appreciable proportion of amorphous material which may well have a different structure from the crystalline regions which give rise to the diffraction pattern.

4. Alternative Models for $(Pro.Gly.Pro)_n$

The helical symmetry derived from the X-ray photographs implies that each unit cell contains 10 equivalent units of structure, and it is clear from the density that each unit in fact corresponds to one Pro.Gly.Pro tripeptide element of the polymer. This result is much easier to establish than the corresponding conclusion in the case of collagen, where, because of the complexity of the amino acid sequence and uncertainty regarding the unit-cell dimensions, the possibility of two or four amino acid residues per unit of structure requires serious consideration (Bear, 1956; Rich & Crick, 1961).

In theory the 10 tripeptide units in the unit cell may be joined chemically to form various numbers of polypeptide chains, each with an axial translation of $n(2\cdot85\,\text{Å})$ and a rotation of $n(108°)$ or $n(252°)$ per tripeptide, where n is the number of chains passing through each unit cell (Bear, 1955). Four or more chains would require a length of 11·4 Å or more per tripeptide; as a polypeptide chain cannot stretch so far, these cases are clearly impossible. Bear (1956) has reported a systematic investigation, by means of molecular models, of possible conformations of a polypeptide chain with a sequence corresponding in fact to $(Pro.Gly.Pro)_n$. He concluded that no single- or double-chain structure with systematic hydrogen bonding and conforming to the helical parameters of collagen is stereochemically possible. Pauling (1958) has reported a similar investigation leading to the conclusion that no single-chain structure is possible.

We are thus limited to a three-chain structure for $(Pro.Gly.Pro)_n$. To conform with the helical parameters, the three chains must be coiled about a common axis and each

must have an axial translation of 8·55 Å and a rotation of 36° or 324° per tripeptide. It can be easily shown with molecular models that, of the two, only 36° is stereochemically feasible.

We have considered whether the various models suggested for collagen are compatible with our results for (Pro.Gly.Pro)$_n$.

The standard structure of the Madras group (Ramachandran, 1963) has interchain hydrogen bonds linking the glycyl N_1H_1 to O_3 and N_2H_2 to O_2, following the notation of Fig. 1. This model can clearly be excluded on chemical grounds as each prolyl-glycyl-prolyl unit has only one NH group available for hydrogen-bond formation.
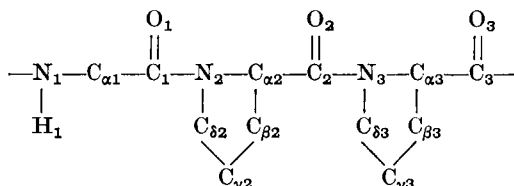

FIG. 1. Structural formula of one tripeptide unit of (Pro.Gly.Pro)$_n$ indicating notation used in text.

Models having only one hydrogen bond per tripeptide, and therefore compatible with the sequence (Pro.Gly.Pro)$_n$, include collagen I, which has a hydrogen bond between N_1H_1 and O_1, collagen II, which has an $N_1H_1 \ldots O_2$ hydrogen bond, and an alternative structure put forward by the Madras group which also has an $N_1H_1 \ldots O_2$ hydrogen bond. We have tested the possibility of fitting these structures into the observed unit cell. Unfortunately, there is some confusion in the literature concerning details of the various collagen models due to several errata and the fact that similar nomenclature has been used for different models as well as different nomenclature for essentially identical models. We have used co-ordinates for collagen I and collagen II given in Tables 5 and 4, respectively, by Rich & Crick (1961), omitting the O(H) of hydroxyproline. In Table 5 all the x and y co-ordinates should be transposed and the O of hydroxyproline should have $x = 3·4$ Å instead of 2·4 Å. The collagen II co-ordinates are close to those given by Burge, Cowan & McGavin (1958). For the alternative structure we have used co-ordinates given in Table VI by Ramachandran, Sasisekharan & Thathachari (1962) for the backbone atoms modified slightly so that the imide peptide groups are planar and conform with the dimensions found in L-leucyl-L-prolyl-glycine (Leung & Marsh, 1958); co-ordinates for the other atoms were derived by assuming the proline rings planar with bond lengths $C\alpha$—$C\beta$ and $C\gamma$—$C\delta$ equal to 1·52 Å and $C\beta$—$C\gamma$ 1·50 Å.

We have tested the packing of the three models in the following way. In each case, two drawings were made of the projection of the structure along the helix axis. The drawings were placed with their centres the equivalent of 12·5 Å apart and, while being kept parallel, rotated together about their centres. Short contacts between atoms of adjacent triple helices were noted for the whole range of possible orientations. In this way it was found that the best hexagonal packing of collagen II involves slightly short contacts between $C\gamma_3$ and the β, γ and δ carbon atoms of residue 2 (Fig. 1), the closest distance being 3·1 Å between the two γ carbon atoms. The Madras alternative structure proved slightly more compact, 3·3 Å between the γ carbon atoms being the only contact shorter than normal van der Waals distances in the best mode of hexagonal packing. Both structures, which are in fact similar, can be packed satisfactorily into the unit

cell if the pyrrolidine ring of proline 3 is somewhat distorted. In the case of collagen I, however, the best mode of packing still implies a 2·7 Å separation between γ carbon atoms as well as several other bad short contacts. We have considered the effect of lifting the restriction of a completely crystalline arrangement, and studied systematically by means of computations the best possible mode of packing collagen I with the triple helices parallel and in hexagonal array but randomly arranged along and about their helical axes. It turned out that even with these additional degrees of freedom no substantially better mode of packing is possible. It thus appears that $(Pro.Gly.Pro)_n$ cannot have the collagen I structure.

The indication that one water molecule per tripeptide is particularly strongly bound to $(Pro.Gly.Pro)_n$ has led us to consider possible modes of attachment of water to the polymer. As in each tripeptide unit there are two carbonyl groups which are not hydrogen-bonded to NH, one water molecule could only lead to a system of maximum possible hydrogen bonding by means of bridges of the type CO...HOH...OC. If the chains are linked as in collagen II, the conformation is rather unfavourable for such a water bridge linking two carbonyl groups of the same chain, but is particularly well suited to a water bridge between carbonyl groups on two different chains such that, looking from the C-terminal end, the bridge joins O_1 to O_3 of the next chain in a clockwise direction. A water molecule in this position would not affect the close packing of adjacent triple helices.

Starting from a model with the collagen II type hydrogen bonding between chains, we have found it possible to reverse the direction of one of the chains and reconnect the same number of hydrogen bonds to the other two chains. As pointed out by Rich & Crick (1961), the reversed chain would now be hydrogen-bonded in the manner of collagen I. This seems stereochemically a perfectly satisfactory structure and one that could arise from a single chain coiling back on itself. However, it should be noted that, whereas the parallel system of chains has an exact 2·85 Å periodicity, the shortest periodicity for any antiparallel arrangement would be 8·6 Å. Thus meridional reflections could occur at spacings of 8·6 Å and its higher orders; though, if the arrangement of scattering matter approximated closely to a 2·85 Å periodicity, the third order at this spacing would be expected to have much greater intensity than the first and second orders. In fact, an antiparallel structure with interchain hydrogen bonds as described above departs appreciably from a 2·85 Å periodicity. Also, because it is somewhat thicker than the parallel structure, it would be harder to fit into the unit cell. It seems to us unlikely that any alternative hydrogen-bonded antiparallel conformation would conform much better with the X-ray pattern in these two respects.

5. Studies of Other Polytripeptides

The structure of $(Pro.Gly.Gly)_n$ has been determined. As a detailed description of the work will be provided in a separate communication (Traub, manuscript in preparation), we will only describe the main structural features here.

$(Pro.Gly.Gly)_n$ forms left-handed helices, with each tripeptide corresponding to an axial translation of 9·3 Å and a rotation of 360°, approximately equally divided between the three amino acid residues. The conformation of the chains is therefore essentially the same as has been found for polyproline II (Cowan & McGavin, 1955; Sasisekharan, 1959) and polyglycine II (Crick & Rich, 1955). The chains are held together by two NH...O hydrogen bonds per tripeptide to form double-layered
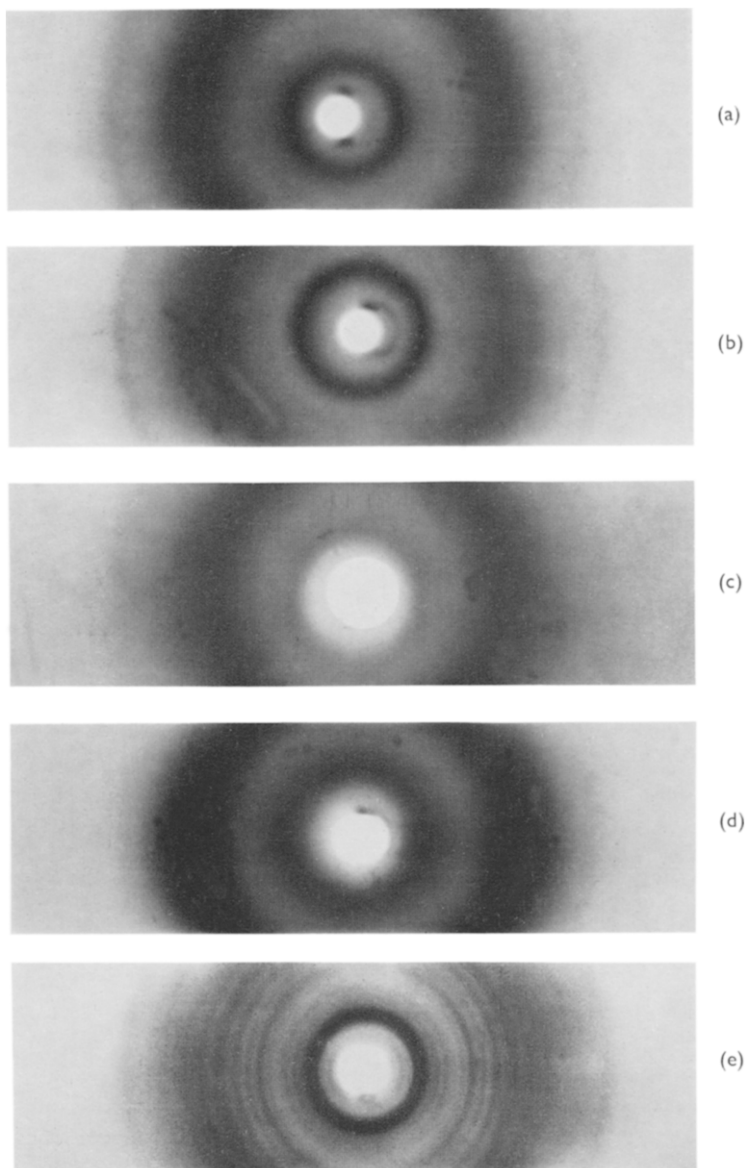
PLATE I. X-Ray diffraction powder photographs of (a) collagen, (b) $(Pro.Gly.Pro)_n$, (c) $(Pro.Gly.o\text{-acetyl Hypro})_n$, (d) $(Pro.Ala.Gly)_n$, (e) wet paste of $(Pro.Gly.Pro)_n$.
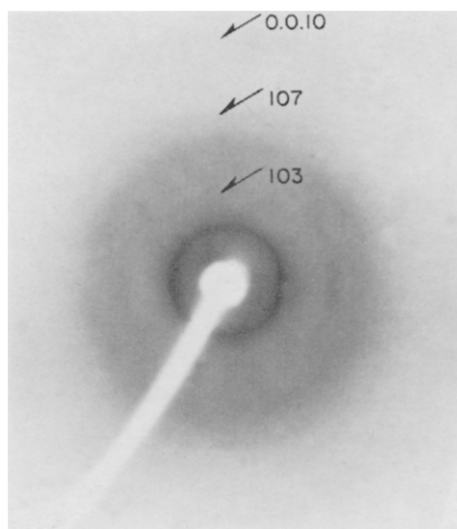
PLATE II. X-Ray diffraction photograph of oriented film of (Pro . Gly . Pro)$_n$; X-ray beam parallel to plane of the film.

sheets, with the pyrrolidine rings of proline residues on the outsides of the sheets. The chains are not coiled about each other and each has a lateral separation of 4·9 Å from each of its four nearest neighbours. It should be noted that such close packing of chains within the sheets is possible only because two out of every three residues, being glycyl, have no β carbon atoms.

It thus appears that all the members of the series of polypeptides (Gly.Gly.Gly)$_n$, (Pro.Gly.Gly)$_n$, (Pro.Pro.Gly)$_n$ and (Pro.Pro.Pro)$_n$ form chains with the same conformation, though the mode of association of the chains differs in the various members of the series.

Only three features were observed in powder photographs of (Pro.Gly.o-acetyl Hypro)$_n$, Plate I(c); relatively strong diffuse bands in the 11 to 12 Å and 4 to 5 Å regions and a weak but fairly sharp reflection at 2·85 Å. Though these data are clearly inadequate for an unambiguous determination of the structure, they are suggestive of its being similar to that of collagen. An outstanding reflection at 2·9 Å is a characteristic feature of the collagen pattern, which is generally believed to correspond to the axial repeat distance that results from the coiling of polyproline II-like chains about each other to form a triple helix (Ramachandran, 1963). Apart from the 2·9 Å reflection, the collagen powder pattern, like that of (Pro.Gly.o-acetyl Hypro)$_n$, consists mainly of two regions of strong intensity (Plate I(a)) in the neighbourhood of 11 Å and 5 Å which are believed to arise, respectively, from inter-triple-helix and intra-triple-helix vectors.

Powder photographs of (Pro.Ala.Gly)$_n$, Plate I(d), have a similar appearance, with strong bands around 10 Å and in the 4 to 5 Å region and a weak reflection at 2·9 Å. In the most crystalline specimens we were able to prepare, the former band was found to be appreciably sharper, with a spacing of 10·0 Å, and several additional lines were observed.

In a collagen-like triple helix, any tripeptide with the sequence Gly.X.Y, where X and Y may be any amino or imino acid, has the glycine residue near the axis of the triple helix whereas residue Y lies at the greatest radius with its NH group pointing outwards in a way that excludes its participation in an intra-triple-helix hydrogen bond. These are common features of the various triple helix models that have been proposed for collagen; the controversy regarding the amount of hydrogen-bonding in collagen concerns the possible participation of the NH group of residue X (Rich & Crick, 1961; Ramachandran, 1963). It follows that if (Pro.Ala.Gly)$_n$—that is (Gly.Pro.Ala)$_n$—has a triple-helical structure, as is suggested by the powder photographs, the alanine residues must be on the outside and there can be only one intra-triple-helix hydrogen bond per tripeptide.

The long spacings observed in X-ray patterns of polytripeptides with sequences (Gly.Pro.Hypro)$_n$, (Gly.Pro.Pro)$_n$ and (Gly.Pro.Ala)$_n$ are 11·9 (Rogulenkova et $al.$, 1964), 10·9 and 10·0 Å, respectively. This is consistent with their all having closely similar triple-helical structures with different diameters (13·7, 12·5 and 11·5 Å, respectively, assuming hexagonal packing) corresponding to the different sizes of the hydroxyproline, proline and alanine residues on the outsides of the helices.

6. Discussion

The X-ray pattern of (Pro.Gly.Pro)$_n$ resembles that of collagen so closely that it seems extremely probable that they have the same type of structure. The investigation of (Pro.Gly.Pro)$_n$ strongly supports the postulate that this type of structure is

indeed a three-stranded coiled coil with the helical parameters that have been assigned to collagen.

Our results also indicate that the —OH group of hydroxyproline is not required for the formation of a collagen-like structure. This bears out a similar conclusion derived from the fact that the stability against thermal melting of collagen from different species is proportional to the sum of the proline and hydroxyproline contents (Josse & Harrington, 1964). These facts, together with the position of hydroxyproline on the outside of the triple helix indicated by the common collagen sequence Gly.X.Hypro (Greenberg, Fishman & Levy, 1964), point to an external role for hydroxyproline through interaction with adjacent triple helices or other chemical components of connective tissue.

As pointed out in section 5, in a collagen-like triple helix any tripeptide with the sequence Gly.Pro.X or Gly.Hypro.X, where X is any amino or imino acid, has only one NH group in a position to make a hydrogen bond with chains of the same triple helix. The studies of the four polytripeptides $(Gly.Pro.Hypro)_n$, $(Pro.Gly.Pro)_n$, $(Pro.Ala.Gly)_n$ and $(Pro.Gly.o\text{-}acetyl\ Hypro)_n$ therefore all indicate that no more than one hydrogen bond per tripeptide is required for a triple helix. Sequence studies of collagen show 30 to 40% of the protein to be composed of tripeptide segments with the sequence Gly.Pro.X. These are concentrated in the apolar regions of collagen and account for at least two-thirds of the total imino acid composition (Grassmann, Nordwig & Hörmann, 1961; Grassmann, Hannig & Nordwig, 1963; Greenberg et al., 1964). The similar amino acid composition of the three chains of which the collagen molecule is composed (Piez, 1964) and the recent finding that an electron micrograph pattern similar to the segment long-spacing pattern of native collagen can be obtained from α1 chains alone (Kühn, Tkocz, Zimmermann & Beier, 1965) indicate a similar distribution of polar and apolar regions in the three chains of collagen, with like regions occurring on adjacent portions of all three chains. These facts indicate that much of the apolar regions of collagen can have only one hydrogen-bond per tripeptide and may have a structure very similar to what is probably the common structural conformation of $(Pro.Gly.Pro)_n$, $(Pro.Ala.Gly)_n$ and $(Gly.Pro.Hypro)_n$. It must of course be remembered that 60% of the collagen molecule does not have the sequence Gly.Pro.X. Our results are obviously much less relevant to such regions, which could conceivably have two hydrogen bonds per tripeptide.

Where there is only one NH...O hydrogen bond per tripeptide, additional stabilization of the structure could be provided by hydrogen-bonded water forming bridges between different chains of a triple helix in the way indicated by our model studies. There are several lines of evidence which suggest that water takes a role in stabilizing the collagen structure, and various ways in which water may be bound, including a bridge of the type we have proposed for $(Pro.Gly.Pro)_n$, have been considered (Harrington & von Hippel, 1961; Burge et al., 1958).

As regards the detailed conformation of $(Pro.Gly.Pro)_n$ our results are as yet only of a preliminary nature. The unit-cell dimensions of $(Pro.Gly.Pro)_n$ are incompatible with the collagen I structure of Rich & Crick (1961), but compatible with modified versions of their collagen II structure or the alternative structure proposed by the Madras group (Ramachandran et al., 1962). However, until we have completed a much more exhaustive analysis, we cannot rule out the possibility that many substantially different conformations may be consistent with the X-ray pattern. We are continuing investigation of this aspect of the problem.

The model studies indicate that an antiparallel triple-helical structure with one hydrogen bond per tripeptide is stereochemically possible and could presumably be formed by a single chain coiling back on itself. Such a situation would appear to be consistent with some of the results of the studies of (Pro.Gly.Pro)$_n$ in solution described in Part II. However, from packing considerations and the absence of a meridional 8·6 Å reflection it appears unlikely that such a structure is responsible for the X-ray diffraction pattern. Perhaps the two types of observation can be better reconciled if it is assumed that whereas antiparallel structures of single chains may occur in solution, greater concentration of molecules when they come out of solution favours parallel aggregation of different chains to form triple helices. In the solid state, the more regular parallel structure might tend to crystallize and thus dominate the X-ray pattern, whereas the antiparallel component would form amorphous regions.

REFERENCES

Andreeva, N. S. & Millionova, M. I. (1964). *Soviet Physics—Crystallography* (Eng. trans.), 8, 464.

Andreeva, N. S., Millionova, M. I. & Chirgadze, Yu. N. (1963). In *Aspects of Protein Structure*, ed. by G. N. Ramachandran, p. 137. London: Academic Press.

Bear, R. S. (1955). *Symp. Soc. Exp. Biol.* 9, 97.

Bear, R. S. (1956). *J. Biophys. Biochem. Cytol.* 2, 363.

Berger, A. & Wolman, Y. (1961). *Proc. 5th Int. Congr. Biochem.*, vol. 9, p. 82. London: Pergamon Press.

Burge, R. E., Cowan, P. M. & McGavin, S. (1958). In *Recent Advances in Gelatin and Glue Research*, ed. by G. Stainsby, p. 25. London: Pergamon Press.

Cochran, W., Crick, F. H. C. & Vand, V. (1952). *Acta Cryst.* 5, 581.

Cowan, P. M. & McGavin, S. (1955). *Nature*, 176, 501.

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

Cowan, P. M., North, A. C. T. & Randall, J. T. (1955). *Symp. Soc. Exp. Biol.* 9, 115.

Crick, F. H. C. & Rich, A. (1955). *Nature*, 176, 780.

Debabov, V. G., Kozarenko, T. D. & Shibnev, V. A. (1961). *Proc. 5th Int. Congr. Biochem.*, vol. 9, p. 63. London: Pergamon Press.

Engel, J., Kurtz, J., Katchalski, E. & Berger, A. (1966). *J. Mol. Biol.* in the press.

Engel, J., Kurtz, J., Traub, W., Berger, A. & Katchalski, E. (1964). *Structure and Function of Connective and Skeletal Tissue*, ed. by S. F. Jackson, p. 241. London: Butterworths.

Grassmann, W., Hannig, K. & Nordwig, A. (1963). *Hoppe-Seyler's Z.* 333, 154.

Grassmann, W., Nordwig, A. & Hörmann, H. (1961). *Hoppe-Seyler's Z.* 323, 48.

Greenberg, J., Fishman, L. & Levy, M. (1964). *Biochemistry*, 3, 1826.

Harrington, W. F. & von Hippel, P. H. (1961). *Advanc. Protein Chem.* 16, 91.

Josse, J. & Harrington, W. F. (1964). *J. Mol. Biol.* 9, 269.

Kitaoka, H., Sakakibara, S. & Tani, H. (1958). *Bull. Chem. Soc. Japan*, 31, 802.

Kühn, K., Tkocz, C., Zimmermann, B. & Beier, G. (1965). *Nature*, 208, 685.

Lakshmanan, B. R., Ramakrishnan, C., Sasisekharan, V. & Thathachari, Y. T. (1962). In *Collagen*, ed. by N. Ramanathan, p. 117. New York: Interscience.

Leung, Y. C. & Marsh, R. E. (1958). *Acta Cryst.* 11, 17.

Pauling, L. (1958). In *Recent Advances in Gelatin and Glue Research*, ed. by G. Stainsby, p. 11. London: Pergamon Press.

Piez, K. A. (1964). *J. Biol. Chem.* 239, 4315.

Ramachandran, G. N. (1963). In *Aspects of Protein Structure*, ed. by G. N. Ramachandran, p. 39. London: Academic Press.

Ramachandran, G. N. & Ambady, G. K. (1954). *Current Science*, 23, 349.

Ramachandran, G. N. & Kartha, G. (1955). *Nature*, **176**, 593.

Ramachandran, G. N., Sasisekharan, V. & Thathachari, Y. T. (1962). In *Collagen*, ed. by N. Ramanathan, p. 81. New York: Interscience.

Rich, A. & Crick, F. H. C. (1961). *J. Mol. Biol.* **3**, 483.

Rogulenkova, V. N., Millionova, M. I. & Andreeva, N. S. (1964). *J. Mol. Biol.* **9**, 253.

Sasisekharan, V. (1959). *Acta Cryst.* **12**, 897.

Shibnev, V. A., Rogulenkova, V. N. & Andreeva, N. S. (1965). *Biofisika*, **10**, 164.

Shmueli, U. & Traub, W. (1965). *J. Mol. Biol.* **12**, 205.

Wolman, Y. (1961). Ph.D. thesis, Hebrew University, Jerusalem.