J. Mol. Biol. (1969) **43**, 461–477

Polymers of Tripeptides as Collagen Models

IV.† Structure Analysis of Poly(L-prolyl-glycyl-L-proline)

A. YONATH AND W. TRAUB

Department of Chemistry
Weizmann Institute of Science, Rehovot, Israel

(*Received 21 October 1968*)

Poly(L-prolyl-glycyl-L-proline) closely resembles collagen in its X-ray diffraction pattern, which indicates that it has a triple-helical structure with the same helical parameters as the protein. A detailed conformational analysis has been made of the polytripeptide. A computer was used systematically to generate structures having the observed helical parameters, as well as standard bond lengths and angles, and to test these structures in terms of acceptable intramolecular and intermolecular van der Waals distances and agreement between calculated and observed intensities. These computations narrowed down the possibilities to a unique conformation which resembles the collagen II model in its mode of interchain hydrogen bonding. Conformations of the collagen I type were found to make short intermolecular contacts, and those in which the glycyl NH is hydrogen bonded in the manner of the two-bonded model proved incompatible with the observed intensities of the equatorial reflections. It is suggested that the conformation found for the polytripeptide may be representative of the structure of collagen as a whole.

1. Introduction

The molecular structure proposed for collagen by Ramachandran & Kartha (1955) has been widely accepted as correct, at least in its general features. The 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°, and consists of three helical polypeptide chains, each having every third residue glycine, which are wound about each other to form a three-stranded coiled coil. However, an unequivocal elucidation of the precise molecular conformation has proved difficult, and three alternative modifications of the structure have been proposed. They differ in details of the conformation and the mode of interchain hydrogen bonding. In particular, one modification (Ramachandran & Sasisekharan, 1965; Ramachandran, 1967) has two systematic hydrogen bonds of the type NH . . . O (carbonyl) 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; Ramachandran & Chandrasekharan, 1968) have only one hydrogen bond for three residues.

As an aid to the understanding of the structure and properties of collagen and their dependence on amino acid sequence, a number of polypeptide models for the protein has been synthesized. Many of these resemble collagen in having every third residue

† Paper III of this series is Traub & Yonath, 1967.

glycine and containing appreciable amounts of the imino acids proline and hydroxyproline. Conformational studies of a number of such polytripeptide models for collagen have recently been reported by Andreeva, Esipova, Millionova, Rogulenkova & Shibnev (1967), Oriel & Blout (1966), Scatturin, Del Pra, Tamburro & Scoffone (1967) and in parts I, II and III of this series (Traub & Yonath, 1966; Engel, Kurtz, Katchalski & Berger, 1966; Traub & Yonath, 1967).

In Traub & Yonath (1966) we described a study of poly(L-prolyl-glycyl-L-proline), hereafter written $(\text{Pro-Gly-Pro})_n$, and showed it to have a well-defined X-ray pattern which has all the main features of that of collagen and a three-stranded structure with the same helical parameters as have been found for the protein. These results indicate a very close structural similarity between collagen and $(\text{Pro-Gly-Pro})_n$ and we have therefore sought to determine the detailed conformation of this polytripeptide. In this investigation we have had two distinct advantages not available to those who have attempted a detailed structure analysis of collagen itself. First, we have been able to test various possible conformations in terms of interatomic distances and X-ray intensities calculated for a known chemical formula rather than some approximation to the over-all amino acid composition of collagen. Second, presumably as a consequence of its simplified amino acid sequence, $(\text{Pro-Gly-Pro})_n$ molecules can pack more regularly than those of collagen and we have been able to determine unit-cell dimensions for the polytripeptide. We have therefore also been able to test possible conformations by their ability to pack into the available space.

In this paper we describe the systematic evaluation of possible molecular conformations for $(\text{Pro-Gly-Pro})_n$ in terms of the three criteria of acceptable intramolecular and intermolecular contacts and the agreement of calculated intensities with those of the observed X-ray pattern.

2. Experimental Procedure

In Traub & Yonath (1966) we described X-ray photographs of powders and oriented specimens of $(\text{Pro-Gly-Pro})_n$, taken at various degrees of hydration, and the resemblance of the X-ray pattern of the polytripeptide to that of collagen. We also reported how these data were used to determine the helical parameters of the molecule and the unit-cell dimensions and to establish that the structure has three polypeptide chains coiled about a common axis. Here, we shall confine ourselves to a description of more recent experimental work and some reassessment of the structural parameters of $(\text{Pro-Gly-Pro})_n$ which were used in the conformational analysis described in Structure Analysis below.

A 15,000 molecular weight preparation of $(\text{Pro-Gly-Pro})_n$, supplied by the Miles-Yeda Company of Rehovot, was used to prepare stroked films from aqueous solution. X-ray photographs of these specimens (Plate I) showed appreciably better orientation than was obtained from the 6000 molecular weight material described in Traub & Yonath (1966) and, unlike the latter, showed normal fibre, rather than reverse, orientation. That the orientation was, in fact, of the fibre and not of the uniplanar variety was shown by the similar appearance of photographs taken with the X-ray beam parallel and perpendicular to the film and the differences between the top and bottom halves of the photographs when the specimens were tilted towards the X-ray beam. Before describing how these photographs were used to measure the intensities of the various reflections, it is appropriate to consider the type and degree of crystallinity obtaining in $(\text{Pro-Gly-Pro})_n$.

Photographs of $(\text{Pro-Gly-Pro})_n$ containing up to about 8 water molecules per tripeptide indicate a hexagonal unit cell with the a-axis varying from 12·5 to 13·6 Å, depending on the degree of hydration (Traub & Yonath, 1966; Traub, Shmueli, Suwalsky & Yonath, 1967). However, the symmetry of the molecule, which has ten tripeptide units in three turns, is inconsistent with a truly crystalline hexagonal system. Either the unit cell must have a lower symmetry, but with pseudohexagonal dimensions, or the molecular packing

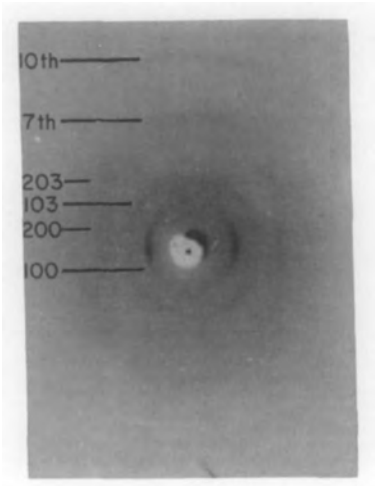

PLATE I. X-ray diffraction pattern of $(Pro\text{-}Gly\text{-}Pro)_n$. The fibre axis is vertical.

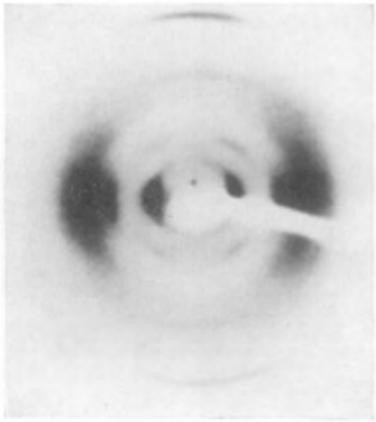

PLATE II. X-ray diffraction pattern of unstretched collagen. The fibre axis is vertical.

[*facing p. 462*

must be disordered. In Traub & Yonath (1966) we indexed the reflections of (Pro-Gly-Pro)$_n$ in accordance with the first possibility. We now believe that the packing is, in fact, hexagonal but with screw disorder. This is supported by the regular expansion of the hexagonal cell dimensions with increasing water content and the relative sharpness of the 100, 110, 200 and 103 reflections compared with the rather more diffuse reflections on the seventh and tenth layer lines, indicating that the former, but not the latter, are subject to crystalline sampling.

Screw disorder occurs fairly commonly among helical molecules with pronounced grooves (Klug & Franklin, 1958; Marvin, Spencer, Wilkins & Hamilton, 1961) as it enables ridges and troughs of neighbouring molecules to interlock without their being in register along their lengths. Its effect on the X-ray pattern is that only those regions for which $m = 0$ in the selection rule for helical structures (Cochran, Crick & Vand, 1952) are sampled. For (Pro-Gly-Pro)$_n$ these would occur on layer lines with $l = 3n$, where n is an integer. Therefore we have treated the reflections on the equator and the third layer line as crystalline spots, but others, notably those on the seventh and tenth layer lines, as layer-line streaks.

In order to get the true meridional spacing, the streak on the tenth layer line was measured near the inner edge, as suggested by Lakshmanan, Ramakrishnan, Sasisekharan & Thathachari (1962) for the case of collagen, and we obtained a value of 2·87 Å. The cell dimensions of dry (Pro-Gly-Pro)$_n$ are therefore $a = 12{\cdot}5 \pm 0{\cdot}2$ Å and $c = 28{\cdot}7 \pm 0{\cdot}4$ Å. The spacing of the seventh layer line could not be measured accurately because the orientation was insufficient to resolve the near-meridional streaks. Therefore the spacing of the crystalline 103 reflections was used to calculate the third layer line spacing and hence the helical rotation per tripeptide unit. This was found to be $108° \pm 3°$, which in fact corresponds to ten units in three turns of the helix.

The measurements of cell dimensions and helical parameters were made on specimens containing some two or three water molecules per tripeptide and showing sharp 100, 110, 200 and 103 reflections. Specimens which were dried further, and these include the oriented films used for intensity measurements, have the same cell dimensions, but show evidence of additional disorder. Equatorial reflections are somewhat broadened, probably as a result of net disorder (Vainshtein, 1966), and the intensity of the 110, but not of the other observed reflections, is greatly diminished. This latter effect may be due to a tendency for sheets of molecules to slip past each other as has been suggested for collagen by Sasisekharan & Ramachandran (1957). In any case, it made it impossible to measure the intensity of the 110 reflection.

Several other difficulties attended the measurement of intensities. Because of incomplete orientation the various reflections are drawn out into arcs (Plate I). In the case of crystalline reflections and the fairly sharp streaks near the meridian on the seventh and tenth layer lines it was found possible to measure the intensities of the arcs from radial traces taken with a Joyce–Loebl recording microdensitometer (Langridge *et al.*, 1960). However, the broad non-meridional streaks which occur particularly on the poorly resolved layer lines near the equator proved too weak and diffuse for any meaningful measurement. Such streaks overlap the 200 and 203 reflections and for the intensity measurements it was necessary to estimate the streak contribution to the background from the shapes of the radial traces through the reflections. Corrections were also made for the general background, which falls off radially from the centre of the photographs, by superimposing equatorial, meridional and diagonal traces and drawing out the areas between them. In this way it proved possible to resolve and correct areas under the radial traces corresponding to the 100, 200, 103 and 203 reflections and the seventh and tenth layer line meridional streaks. These areas were measured by cutting out and weighing the paper on which they were drawn. The measured areas were multiplied by $\tan 2\theta$ (where θ is the Bragg angle) to allow for the reflections being drawn out into arcs of different lengths which, however, all subtend the same angle at the centre of the photograph. In addition, a special correction was applied to the area measured for the tenth layer line streak from photographs of untilted specimens. This streak would in fact not be observed if the specimen were perfectly oriented, and it appears on the meridian only because of the contribution of parts of the specimen which are misaligned by some 16°. The ratio of such misaligned intensities to those at the optimum orientation, as measured for the other

reflections, was estimated by comparing traces across the arc of the 100 reflection along the equator and at 16° to the equator. A ratio of 1·7 was obtained, and in fact the meridional intensity was found to increase by about the same amount in a photograph of a specimen which had been tilted by roughly 16°.

Additional correction factors were applied to obtain integrated intensities on the same scale from the measurements on the spots and streaks. The nature and derivation of these corrections are described in the Appendix, and the corrected integrated intensities are shown in Fig. 5. We estimate the maximum error for these integrated intensity values as around 30% with the value for 100 being more reliable and that for 203 rather less.

We have also taken photographs of unstretched collagen (Plate II). We used fibres of sheep submucosa collagen provided by the Ethicon Company. The combination of relatively sharp reflections on the equatorial and third layer lines with streaks in other parts of the diffraction pattern suggest the existence of screw disorder, as in the case of $(\text{Pro-Gly-Pro})_n$. Consequently, we have measured and corrected the intensities of the 100, 200, 103 and 203 reflections and the seventh and tenth layer line meridional streaks in the same way as we have described for the polytripeptide. The accuracy of the intensity values should be about the same for the two materials. Figure 5(e) shows the corrected integrated intensities which we obtained for collagen.

3. Structure Analysis

(a) *Analytical procedure*

The paucity of data in X-ray fibre patterns of polymers precludes structure analysis by standard crystallographic procedures. Consequently, their interpretation depends in general on a search for stereochemically acceptable molecular structures which are consistent with the X-ray pattern. In the case of $(\text{Pro-Gly-Pro})_n$ we have tried to systematize this procedure with the aid of the CDC 1604-A and Golem computers of the Weizmann Institute. The computers have been used to generate structures having the observed helical parameters, as well as standard bond lengths and angles, and to test these in terms of acceptable intramolecular and intermolecular van der Waals contacts and agreement between calculated and observed intensities.

Figure 1 shows the structural formula of one tripeptide unit of $(\text{Pro-Gly-Pro})_n$. In calculating atomic co-ordinates for the various possible conformations of the polymer, we have made several basic assumptions. For glycine we have taken the standard peptide dimensions of Corey & Pauling (1953), and we have also followed them in assuming the co-planarity of the six atoms of the peptide group in all three residues. We have taken the dimensions found in leucyl-prolyl-glycine (Leung & Marsh, 1958) for the peptide group of proline, but, as described below, we have allowed for a range of possible conformations of the proline ring.

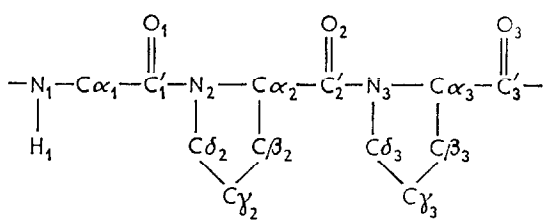

FIG. 1. Structural formula of one tripeptide unit of $(\text{Pro-Gly-Pro})_n$ indicating notation used in text.

We have also incorporated into the computations our findings that the molecule consists of three polypeptide chains and that equivalent units of the structure are related by a translation of 2·87 Å and a rotation of 108°. In addition, we have assumed that the NH of glycine is involved in interchain hydrogen bonding to a carbonyl group. In fact, only the three nearest carbonyl groups on each of the other two chains could possibly be near enough for this. It is difficult to conceive of a triple-chain structure without such hydrogen bonding, and, indeed, the observed properties of (Pro-Gly-Pro)$_n$ in solution (Engel *et al.*, 1966) provide experimental support for its existence.

Having fixed the dimensions of the peptide groups and the helical parameters, it is possible to describe the molecular conformation uniquely in terms of the eight parameters h_1, h_2, h_3, μ, ν, δ_1, δ_2 and δ_3, as is shown in Fig. 2. This way of describing collagen-like conformations was introduced by Ramachandran, Sasisekharan & Thathachari (1962).

In Fig. 2 the positions of only the α carbon atoms are shown, with $C\alpha_1$ and $C\alpha_4$ belonging to successive glycine residues along one chain. Both these atoms are at the same radial distance ρ from the helix axis. h_1, h_2 and h_3 are the vertical components of $C\alpha_1C\alpha_2$, $C\alpha_2C\alpha_3$ and $C\alpha_3C\alpha_4$, respectively, and δ_1, δ_2 and δ_3 define the orientations of the three peptide groups about these axes. Each δ is zero when the plane of the peptide group is parallel to the helix axis and CO points downwards; δ increases with clockwise rotation of the peptide group when viewed from $C\alpha_i$ to $C\alpha_{i+1}$. The supplement of the angle between the projections of $C\alpha_1C\alpha_2$ and $C\alpha_2C\alpha_3$ is μ and the supplement of that between $C\alpha_2C\alpha_3$ and $C\alpha_3C\alpha_4$ is ν.

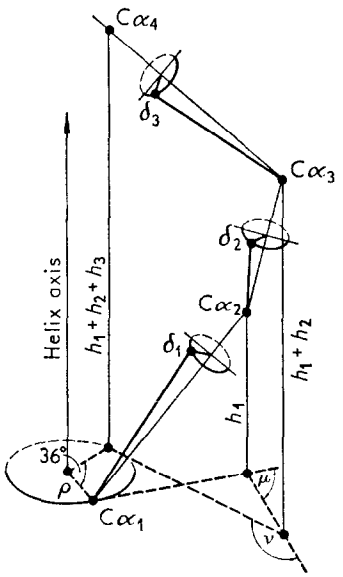

FIG. 2. Description of possible molecular conformations of (Pro-Gly-Pro)$_n$ in terms of inclinations and orientations of the three peptide groups. For further explanation see text.

In fact, there are not eight but five independent parameters. The sum of $h_1+h_2+h_3 =$ 8·6 Å, so any two of these vertical components specify the third. As the peptide dimensions fix the distance between successive Cα atoms to be 3·8 Å, h_1 and h_2 can only have values between 1·0 and 3·8 Å. Any set of parameters h_1, h_2, μ, δ_1 and δ_2 fixes the angle $N_2C\alpha_2C'_2$. Therefore, for each set of h_1, h_2, μ and δ_1, we have chosen five values covering the possible range for this angle (see below) and calculated the corresponding values of δ_2. At most there are two δ_2 values for any value of the angle $N_2C\alpha_2C'_2$; often there are none. Values of δ_3 corresponding to five different $N_3C\alpha_3C'_3$ angles were calculated in the same way, once h_1, h_2, μ, ν, δ_1, and a calculated value of δ_2 were fixed.

The five independent parameters were varied systematically in small steps, the order of increasing frequency of stepping being h_1, h_2, μ, ν and δ_1. For each combination of parameters, atomic co-ordinates were calculated and the structure tested in terms of the van der Waals contacts and other criteria described below. In the first stage of the analysis the complete ranges of possible values for the parameters were considered (1·0 to 3·8 Å for h_1 and h_2; 0° to 360° for μ, ν and δ_1), but the values were changed by relatively coarse increments (0·3 Å for h_1 and h_2; 10° for μ and ν; 15° for δ_1). This served to eliminate most parts of the five-parameter function, and in subsequent stages regions found to have structures which passed the tests applied were re-examined at finer intervals and with stiffer criteria. The various stages of the analysis with the corresponding ranges and increments of the five parameters and the regions found to have good structures are described in Table 1.

The minimum van der Waals contacts allowed in testing the structures were based on the minimum or "outer limit" values cited by Ramakrishnan & Ramachandran (1965) (see Table 2) and calculations of the maximum possible errors for the computed interatomic distances, due to the finite size of the steps in the parameters and to possible variations of the angles at the Cα atoms and conformations of the proline rings.

Table 3 lists the minimum contacts calculated for one set of increments in the five parameters. In fact, the maximum possible error was computed for each pair of atoms. For example for the intrachain N_2N_3 distance (of Fig. 1) the error should be less than

$$\frac{1}{2}\left[\Delta h\,\frac{(R_2+R_3)}{3{\cdot}8}+\Delta\delta(r_2+r_3)+\Delta\mu(\text{minimum of } R_2 \text{ and } R_3)\right],$$

where R_2 and R_3 are distances from $C\alpha_2$ of the projections of N_2 on $C\alpha_1C\alpha_2$ and of N_3 on $C\alpha_2C\alpha_3$, respectively, r_2 and r_3 are the distances of N_2 and N_3 from these axes, and Δh, $\Delta\delta$ and $\Delta\mu$ denote the increments in the parameters. Maximum possible errors in computed

TABLE 1

Details of computational cycles in structure analysis of (Pro-Gly-Pro)$_n$

	Parameters	Cycle 1	Cycle 2	Cycle 3	Cycle 4
Region tested	h_1	1·0–3·8 Å	†3·0–3·6 Å	†2·3–3·1 Å	2·25–3·25 Å
	h_2	1·0–3·8 Å	2·3–3·6 Å	2·3–3·6 Å	3·00–3·65 Å
	μ	0–360°	97–152°	97–152°	92–143°
	ν	0–360°	87–142°	87–142°	82–128°
	δ_1	0–360°	58–167°	58–167°	65–110°
	ρ	0–4·5 Å	†2·0–3·9 Å	†0·8–2·1 Å	1·3–2·1 Å
Increments					
	h_1	0·3 Å	0·1 Å	0·2 Å	0·1 Å
	h_2	0·3 Å	0·1 Å	0·2 Å	0·1 Å
	μ	10°	2°	5°	2°
	ν	10°	2°	5°	2°
	δ_1	15°	5°	10°	5°
			staggered	staggered	staggered
Region passed					
	h_1	2·50–3·55 Å	3·20–3·25 Å	2·4–3·1 Å	2·4–2·9 Å
	h_2	2·35–3·55 Å	2·95–3·10 Å	3·2–3·5 Å	3·2–3·5 Å
	μ	105–150°	136–142°	100–135°	116–136°
	ν	90–135°	86–102°	90–120°	104–126°
	δ_1	60–165°	135–145°	73–103°	75–105°
	ρ	0·95–3·71 Å	2·61–3·03 Å	1·36–2·00 Å	1·34–1·61 Å

† It was found that, of the structures passed by cycle 1, those with ρ greater than 2·0 Å have h_1 in the range 3·0 to 3·6 Å and those with ρ less than 2·0 Å have h_1 between 2·3 and 3·1 Å.

TABLE 2

Minimum interatomic contact distances

Contact	Normally allowed (Å)	Outer limit (Å)
C—C	3·20	3·00
C′—C′	3·00	2·90
C—O	2·80	2·70
C—N	2·90	2·80
C—H	2·40	2·20
O—O	2·70	2·60
O—N	2·70	2·60
O—H	2·40	2·20
N—N	2·70	2·60
N—H	2·40	2·20
H—H	2·00	1·90

This Table is taken from Ramakrishnan & Ramachandran (1965).

TABLE 3

Minimum interatomic contacts allowed in cycle 1

Contact	*a* (Å)	*b* (Å)	*c* (Å)
Cα ... Cα	—	—	2·84
Cα ... Cβ	—	2·55	2·35
Cα ... Cγ	—	2·27	2·14
Cα ... Cδ	—	2·74	2·48
Cα ... C′	—	2·76	2·57
Cα ... N	—	2·61	2·42
Cα ... O	—	2·51	2·32
Cβ ... Cβ	2·27	2·25	2·00
Cβ ... Cγ	2·07	2·05	1·85
Cβ ... Cδ	2·52	2·50	2·25
Cβ ... C′	2·32*	2·29	2·13
Cβ ... N	2·41	2·39	2·23
Cβ ... O	2·28	2·25	2·05
Cγ ... Cγ	1·73	1·73	1·46
Cγ ... Cδ	2·27	2·25	2·00
Cγ ... C′	2·28	2·26	2·07
Cγ ... N	2·13	2·11	1·92
Cγ ... O	2·03	2·01	1·62
Cδ ... Cδ	2·76	2·74	2·48
Cδ ... C′	2·78	2·76	2·57
Cδ ... N	2·63	2·61	2·42
Cδ ... O	2·48	2·45	2·20
C′ ... C′	2·80	2·78	2·66
C′ ... N	2·31*	2·61	2·49
C′ ... O	2·50	2·48	2·31
N ... N	2·49	2·47	2·31
N ... O	2·43	2·41	2·22
O ... O	2·40	2·37	2·14

Columns (*a*) and (*b*) give intrachain distances for nearest neighbour and second nearest neighbour residues, respectively. Column (*c*) gives interchain distances. * denotes second nearest neighbour atoms.

interatomic distances for second nearest neighbour residues have an additional term involving $\Delta\nu$ and are therefore slightly larger than for neighbouring residues. Possible errors in interchain distances were estimated semi-empirically as approximately twice the corresponding intrachain distances.

Additional latitude was allowed for the $C\beta$ and $C\gamma$ atoms of the proline residues. The position of the latter was computed as being co-planar with the peptide group with $C\delta C\gamma = 1{\cdot}50$ Å and angle $NC\delta C\gamma = 103°$, but an additional 0·5 Å was subtracted from minimum van der Waals distances involving $C\gamma$. The $C\beta$ position was computed on the basis of a $C\alpha C\beta$ bond length of 1·53 Å and 110° for the angles $C'C\alpha C\beta$ and $NC\alpha C\beta$, but 0·25 Å was subtracted from minimum allowed contacts involving $C\beta$ to allow for possible variations in these angles.

For the calculation of δ_2 and δ_3 values from $NC\alpha C'$ angles the latter were considered to lie between 108° and 115°, but this range was extended to take account of the maximum possible error in NC′ distances. A similar extended range was allowed for the angle $N_1C\alpha_1C_1'$ which was used to test the structures.

Another stereochemical criterion used concerns the closure of the proline rings. It was found from molecular models that this could only be effected for $C_\beta C_\delta$ distances between 2·2 and 2·4 Å. In testing structures, this range was also extended to allow for errors in the computed $C_\beta C_\delta$ distances.

The test for a hydrogen bond involving the glycine NH and some CO group was based on the assumptions that the NH . . . O hydrogen bond length should be between 2·6 and 3·2 Å and the HNO angle less than 40°. In the final stages of the analysis, after the type of hydrogen bonding had been established, criteria of hydrogen bond length between 2·7 and 3·1 Å and hydrogen bond angle less than 25° were applied.

The requirement of NH . . . OC hydrogen bonding sets maximum limits on the possible distances between C_α atoms of glycine residues in adjacent polypeptide chains. For example, the type of hydrogen bond used in the standard two-bonded structure (Ramachandran & Sasisekharan, 1965) has the grouping $C_{\alpha 1}$—N_1—H_1 . . . O_3—C_3'—N_1—$C_{\alpha 1}$ (cf. Fig. 1). With the aid of molecular models, we found that in this case the distance between the two $C_{\alpha 1}$ atoms cannot be greater than 6·8 Å. Consequently, because of the 2·87 Å translation and 108° rotation of the helix, the radius of $C_{\alpha 1}$ (ρ in Fig. 2) cannot be greater than 3·8 Å. Similar measurements and calculations for the five other possible modes of interchain hydrogen bonding led to the conclusion that in no case can ρ be greater than 4·4 Å. Although not an independent stereochemical requirement, this proved a convenient test for structures which greatly shortened the computations.

The investigation of intermolecular short contacts provided a useful criterion for narrowing down the structural possibilities in a relatively small region of the five-parameter function. The structures in this region are all rather similar to collagen I (Rich & Crick, 1961) and we consequently knew, after preliminary examination, which pairs of atoms are likely to make short contacts and confined our tests to these.

For the calculations it was assumed that the molecules are 12·5 Å apart, at the same height and oriented in the same way, but rotated together about their centres to the orientation for which there are the best intermolecular contacts. We have shown previously, for a structure of this type, that allowing adjacent molecules to have different heights and orientations about their axes does not in fact enable them to pack more closely than in the crystalline type of arrangement (Traub & Yonath, 1966), for which the computations are far less formidable. Because of the hexagonal symmetry of intermolecular packing and the tenfold symmetry of the molecules, the same intermolecular contacts occur at every 12° as the molecules are rotated together about their helix axes.

Because of the screw disorder the molecules have all orientations about their helix axes, and the corrected observed intensities should therefore be proportional to the square of the cylindrically averaged Fourier transform. We calculated this by the method of Davies & Rich (1959) equating it with

$$\sum_n (A_n^2 + B_n^2)$$

$$\text{where } A_n = \sum_j f_j J_n(2\pi r_j \xi) \cos\left\{n\left(\frac{\pi}{2} + \phi_j\right) + \frac{2\pi l z_j}{c}\right\}$$

$$\text{and } B_n = \sum_j f_j J_n(2\pi r_j \xi) \sin\left\{n\left(\frac{\pi}{2} + \phi_j\right) + \frac{2\pi l z_j}{c}\right\}$$

and where l = layer line number,

ξ = distance of reciprocal lattice point from fibre axis,

f_j = scattering factor of jth atom,

c = length of c-axis of unit cell,

r_j, ϕ_j, z_j = cylindrical co-ordinates in real space of jth atom,

J_n = Bessel function of order n, which satisfies the helical symmetry (Cochran *et al.*, 1952). Bessel functions with the lowest three values of n were included in the calculations for each layer line.

Calculated intensities of this type were considered at various stages of the analysis. The most useful intensity criterion proved to be the ratio of the intensities of the 200 and 100 reflections and this was used extensively to test the structures. Both are low-order reflections and therefore only sensitive to fairly gross changes of structure, but by setting the lower limit for the 200/100 intensity ratio at 0·4 (half the observed value) we were able to discriminate effectively between structures with different modes of hydrogen bonding. Because both reflections are strong and the maximum possible contribution of the two water molecules in the unit cell could only account for a small proportion of their intensities, the 200/100 ratio provides a reliable test for the correctness of the polytripeptide structure without taking water into account. Near the end of the analysis, after the type of hydrogen bonding had been established, intensity calculations were considered for the other reflections as well in choosing the most acceptable structure and possible positions for the water molecules. These calculations were made for ξ values corresponding to the peaks of the measured intensities. However, in the final comparison of the observed intensities with those calculated for the refined structure and the various alternative structural types the intensities were computed at 0·001 intervals in ξ and plotted as a continuous function along all the layer lines up to and including the tenth (Fig. 5).

(b) *Results of analysis*

Table 1 gives details of the four successive computational cycles which served to determine the conformation of (Pro-Gly-Pro)$_n$. For these computations it was assumed that the 2·87 Å translation and 108° rotation between equivalent units follow a left-handed helix. A similar investigation of conformations of right-handed helices was completed in two cycles and served to eliminate this possiblity.

In the first cycle of Table 1 the structures generated by the computer were tested according to the criteria listed in Table 3. The 75 structures that passed these tests all lie in a fairly small region of the five-parameter function. This region, however, still includes the parameters of the two-bonded, collagen I and collagen II models for collagen, in which the NH of glycine makes interchain hydrogen bonds with O_3, O_1 and O_2 of Figure 1, respectively. The second hydrogen bond of the two-bonded structure is of the type $N_2H_2 \ldots O_2$ and implies an amino rather than an imino acid residue in the second position.

For cycles 2, 3 and 4 of Table 1, the various criteria used to test the structures took account of the maximum possible errors in interatomic distances as described above. These cycles were each made up of two computations with the same steps in the five parameters, but with structures calculated at two staggered sets of points. For such a

combination of staggered computations the maximum possible errors are half of those that could occur for either computation alone.

By fitting the criteria to the parameter intervals we avoided any possibility of missing good structures in between the points at which structures were calculated, so for cycles 2, 3 and 4 continuous searches were in fact made of the regions investigated. To have done this for cycle 1 would have required a prohibitive amount of computing or resulted in an enormous output of very bad structures. However, the fact that, outside the regions investigated in later cycles, not a single structure passed the quite lenient criteria of Table 3, makes it extremely unlikely that some stereochemically quite acceptable structure could have been missed.

The region passed in cycle 1 was divided into two parts which were investigated separately in cycles 2 and 3. Cycle 2 was concerned with a region for which h_1 is between 3·0 and 3·6 Å. This region includes collagen I and similar structures with an $N_1H_1 \ldots O_1$ hydrogen bond. The criteria for this cycle included a 200/100 intensity ratio of 0·4 as well as those concerning intramolecular van der Waals contacts, hydrogen bond length and angle and the conformations of the proline rings. In addition, for structures that passed these tests, short intermolecular contacts were printed out.

Only a very small region was in fact passed in cycle 2 and the structures in it all had rather large hydrogen bond lengths and angles, as well as short interchain or intermolecular contacts. As these generally involved C_γ or C_β atoms, they were further examined to see if possible variations in the proline ring conformations could lead to acceptable contacts. None of the structures proved capable of improvement in this way and the whole region was therefore eliminated.

The region investigated in cycle 3 has h_1 between 2·3 and 3·1 Å and includes conformations resembling the two-bonded structure and collagen II. The criterion that the 200/100 intensity ratio should exceed 0·4 proved a significant limitation in this cycle. It was found that all conformations with ρ less than 1·3 Å, and this includes all those with an $N_1H_1 \ldots O_3$ hydrogen bond, have intensity ratios appreciably less than 0·4. Thus after cycle 3 we remained only with conformations having a hydrogen bond of the collagen II type, i.e. $N_1H_1 \ldots O_2$ with the NH pointing clockwise when viewed from the carboxyl ends of the chains.

In cycle 4 the region of possible structures was narrowed down further by computing at finer intervals and applying stricter criteria, including the requirements that the 200/100 intensity ratio be greater than 0·6 and that the hydrogen bond have a length between 2·7 and 3·1 Å and an angle less than 25°.

Of the 214 structures that passed the tests of cycle 4 nearly half had intramolecular van der Waals contacts smaller than the "outer limit" values shown in Table 2. None had a hydrogen bond length less than 2·85 Å and, with the exception of the tenth layer line peak, values of calculated intensities varied rather little. Eleven structures which had good van der Waals contacts and small hydrogen bond angles were chosen from different parts of the region passed by cycle 4. These were further examined in terms of intra- and intermolecular contacts involving hydrogen atoms, and what appears to be the best structure by all criteria was picked out.

Some further adjustments were made in the atomic positions to improve van der Waals contacts and bond lengths and angles, particularly in the proline rings. The shifts from the computed positions were between 0·08 and 0·22 Å for C_β and C_γ and up to 0·06 Å for the other atoms. The final atomic co-ordinates are given in Table 4 and a projection of the structure is shown in Figure 3.

TABLE 4

Atomic co-ordinates of (Pro-Gly-Pro)$_n$

Atom	x(Å)	y(Å)	z(Å)	r(Å)	ϕ(°)
N_1	1·65	−1·11	−0·96	1·99	−33·8
H_1	0·86	−1·70	−1·08	1·91	−63·1
$C\alpha_1$	1·46	0·00	0·00	1·46	0·0
C'_1	2·59	0·16	1·03	2·59	3·5
O_1	3·23	−0·80	1·44	3·33	−14·0
N_2	2·76	1·38	1·56	3·09	26·6
$C\alpha_2$	3·79	1·64	2·60	4·13	23·4
$C\beta_2$	3·92	3·15	2·63	5·03	38·8
$C\gamma_2$	2·63	3·66	2·01	4·51	54·4
$C\delta_2$	2·01	2·57	1·14	3·26	51·9
C'_2	3·31	1·08	3·94	3·48	18·2
O_2	2·21	1·47	4·36	2·66	33·6
N_3	4·03	0·23	4·66	4·04	3·3
$C\alpha_3$	3·58	−0·35	5·94	3·60	−5·6
$C\beta_3$	4·83	−1·08	6·37	4·95	−12·6
$C\gamma_3$	5·72	−1·37	5·17	5·88	−13·5
$C\delta_3$	5·35	−0·28	4·16	5·36	−3·0
C'_3	2·97	0·63	6·92	3·04	12·0
O_3	3·37	1·80	7·02	3·82	28·2
$C\alpha_4$	1·18	0·86	8·61	1·46	36·0
$(H_2O)_1$	3·58	−1·46	9·09	3·87	−22·2
$(H_2O)_2$	2·86	−3·92	7·98	4·85	−53·9
$H_1(C\alpha_1)$	0·58	−0·10	0·46	0·59	−10·0
$H_2(C\alpha_1)$	1·33	0·85	−0·49	1·59	32·7
$H(C\alpha_2)$	4·65	1·20	2·34	4·80	14·5
$H_1(C\beta_2)$	4·32	3·74	3·38	5·72	39·5
$H_2(C\beta_2)$	4·57	3·27	1·88	5·64	35·1
$H_1(C\gamma_2)$	1·99	3·93	2·74	4·40	63·2
$H_2(C\gamma_2)$	2·82	4·48	1·34	5·29	57·8
$H_1(C\delta_2)$	2·19	2·75	0·21	3·52	51·4
$H_2(C\delta_2)$	1·04	2·46	1·47	2·67	67·1
$H(C\alpha_3)$	2·95	−1·11	5·74	3·15	−20·6
$H_1(C\beta_3)$	5·33	−0·53	7·14	5·36	−5·7
$H_2(C\beta_3)$	4·59	−1·94	6·69	4·98	−22·9
$H_1(C\gamma_3)$	6·68	−1·29	5·46	6·80	−10·9
$H_2(C\gamma_3)$	5·48	−2·27	4·76	5·93	−22·5
$H_1(C\delta_3)$	5·27	−0·67	4·20	5·31	7·2
$H_2(C\delta_3)$	6·04	0·46	3·24	6·06	4·4

The dihedral angles are $\phi = 129°$, 104° and 135° and $\psi = 333°$, 307° and 328° at $C\alpha_1$, $C\alpha_2$ and $C\alpha_3$, respectively, according to the notation of Edsall *et al.* (1965).

The bond lengths and angles, which are shown in Figure 4, are all in good agreement with accepted values. All the van der Waals distances are greater than the normally allowed limits shown in Table 2. Only a few intrachain distances and some interchain distances between hydrogens are near these limits. The NH . . . O hydrogen bond distance of 2·96 Å and the HNO angle of 9° also conform with normal values. Both proline rings are puckered and the dihedral angles about the N—C_α bonds (104 and 135° for residues 2 and 3 of Fig. 1, respectively) are near the extreme possible values for proline.

We have presented atomic co-ordinates to 0·01 Å in order to show that the structure is compatible with accepted bond lengths and van der Waals distances, but we do not mean to imply that we have refined the structure to this degree of precision.

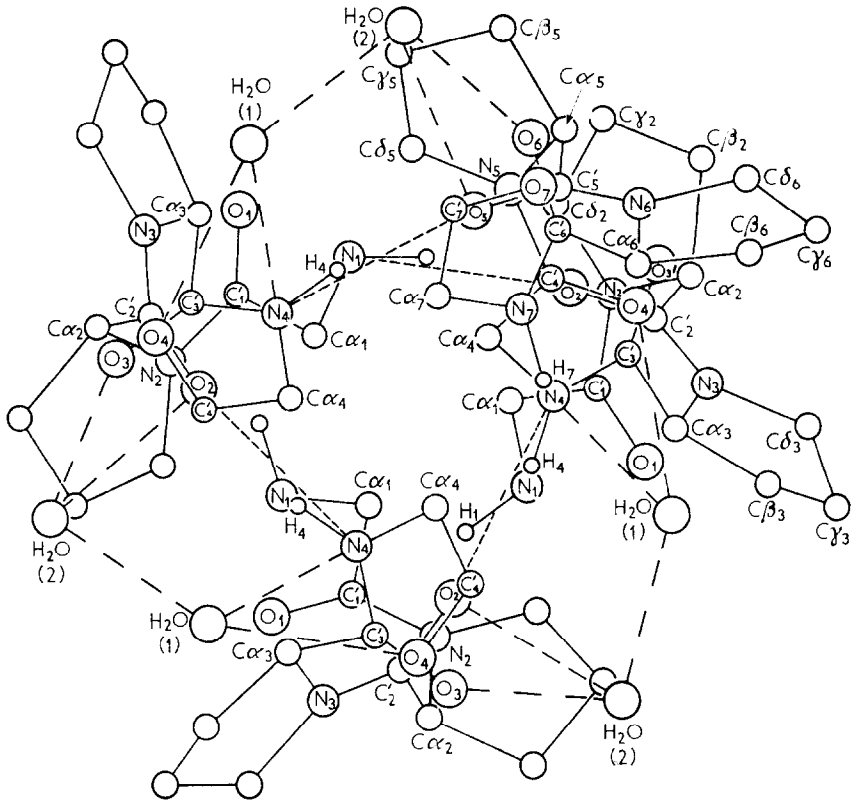

FIG. 3. Projection down the *c*-axis of the structure of (Pro-Gly-Pro)$_n$, including two water molecules per tripeptide. Dashed lines indicate hydrogen bonds.

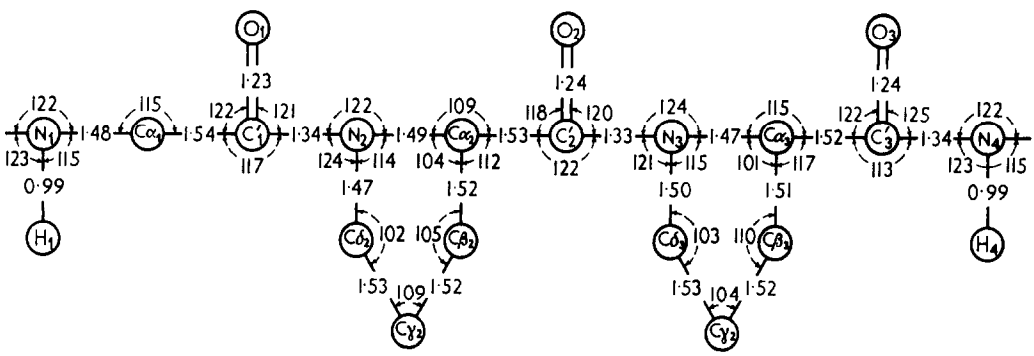

FIG. 4. Bond lengths and angles of one tripeptide unit of (Pro-Gly-Pro)$_n$ corresponding to the atomic co-ordinates shown in Table 4.

In our final comparison of calculated and observed intensities, we have tried to take account of the water in the structure. The photographs from which the intensities were measured were taken at fairly low humidity, and judging from our studies of the hydration of (Pro-Gly-Pro)$_n$ (Traub *et al.*, 1967), we estimate that the specimens must have contained about two molecules of water per tripeptide. We have considered various possible positions for water in the structure in terms of possible hydrogen bonding, packing in the unit cell and its contribution to the calculated intensities. There is a considerable range of possible ways of incorporating two water molecules in the unit cell with hydrogen bonding to the oxygen or nitrogen atoms of the polytripeptide. In all reasonable positions the water contribution raises the calculated intensity for the seventh layer line greatly and increases the 200/100 intensity ratio by a fairly small amount. The effect on the calculated intensities of 103, 203 and the tenth layer line (it can be particularly large for the latter) varies with the positions of the water molecules. In no position is the contribution of water to the strong 100 and 200 intensities large enough to invalidate our use of them to determine the structure of (Pro-Gly-Pro)$_n$.

In Table 4 and Figure 3 we show positions for two water molecules which meet the criteria of cell packing, hydrogen bonding and good calculated intensities rather well. One water molecule makes hydrogen bonds to O_1 and N_1, the other to O_2 and O_3, and they are also hydrogen bonded to each other, thus indirectly providing additional binding between the chains. These five hydrogen bond lengths are all between 2·64 and 2·90 Å. The effect of water molecules in these positions on the calculated intensities is shown in Figure 5(a).

In Figure 5 we also show, for comparison, intensities calculated for (Pro-Gly-Pro)$_n$ according to various other structural models. For Figure 5(b) and (c) we have used the atomic co-ordinates given by Rich & Crick (1961) for the collagen I and collagen II models respectively, and Figure 5(d) is calculated for a structure with an $N_1H_1 \ldots O_3$ interchain hydrogen bond. Figure 5(a), (b), (c) and (d) also show observed intensities for (Pro-Gly-Pro)$_n$.

In Figure 5(e) we have plotted our measured intensities for unstretched collagen together with the intensities calculated for the (Pro-Gly-Pro)$_n$ structure with and without water as in Figure 5(a). Collagen of course differs appreciably from (Pro-Gly-Pro)$_n$ in amino acid composition, and the specimens we photographed probably had a higher water content than we have assumed for the polytripeptide. Nevertheless, as can be seen from the Figure, the intensity distribution of collagen is quite similar to that of (Pro-Gly-Pro)$_n$ and, especially in view of the relatively high 200/100 intensity ratio, appears to be consistent with the conformation we have found for the polytripeptide. We have not measured the intensity of the diffuse near-equatorial blob around 4·5 Å. Though this is rather stronger in collagen than in (Pro-Gly-Pro)$_n$ it can probably be accounted for by the calculated peaks around this spacing on the first and second layer lines.

4. Discussion

The structure we have derived for (Pro-Gly-Pro)$_n$ resembles collagen II in its mode of hydrogen bonding, but there are appreciable differences between our atomic co-ordinates and those previously reported. Consequently, there are differences in the calculated intensities, particularly for the 100 reflection (cf. Fig. 5(a) and (c)) and in the lengths of the hydrogen bonds. Our value of 2·96 Å is considerably longer than the

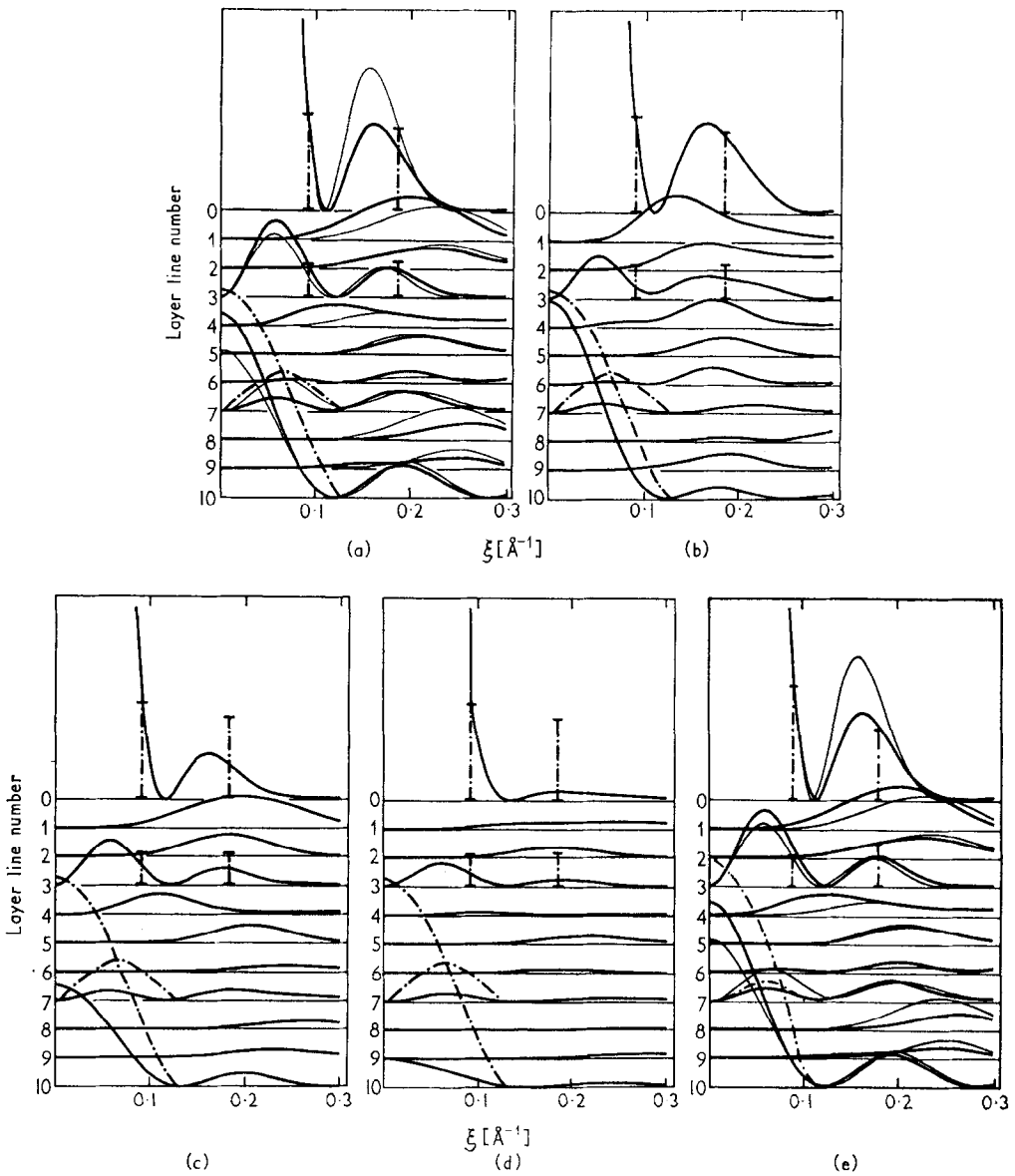

FIG. 5. Comparison of intensities observed for (Pro-Gly-Pro)$_n$ and collagen with those calculated for (Pro-Gly-Pro)$_n$ according to various structural models. Dashed lines show the observed intensities. Light and heavy continuous lines show calculated intensities with and without water contributions respectively. The various comparisons shown are (a) intensity observed for (Pro-Gly-Pro)$_n$ and calculated for the co-ordinates of Table 4; (b) intensity observed for (Pro-Gly-Pro)$_n$ and calculated for the collagen I model; (c) intensity observed for (Pro-Gly-Pro)$_n$ and calculated for the collagen II model; (d) intensity observed for (Pro-Gly-Pro)$_n$ and calculated for a model with $N_1H_1 \ldots O_3$ interchain hydrogen bonding; (e) intensity observed for collagen and calculated for the co-ordinates of Table 4. The calculated curves have been scaled in accordance with the observed intensity of 100.

2·85 and 2·70 Å reported by Rich & Crick (1961) and Burge, Cowan & McGavin (1958), respectively, but conforms better with commonly observed NH . . . O hydrogen bond lengths and the indication, from infrared data, of a long hydrogen bond in collagen (Ramachandran, 1967).

We have sought the stereochemically most plausible structure for (Pro-Gly-Pro)$_n$ with an $N_1H_1 \ldots O_3$ hydrogen bond, and using the analytical procedures described in the section on Structure Analysis we have derived structures with accepted values for bond lengths and angles and only a few interatomic distances slightly below the normal limits of Table 2. However, in all such structures the hydrogen bond angle is nearly 40° and the carbonyl group C_2O_2 points inwards towards the helix axis and is unavailable for hydrogen bonding even to water. These very unlikely structural features need not occur if there were an amino acid residue rather than proline in the second position, as is assumed in the standard two-bonded structure, but they would seem to preclude an $N_1H_1 \ldots O_3$ hydrogen bond for (Pro-Gly-Pro)$_n$ and probably for any collagen-like polytripeptide of the type (Gly-Pro-X)$_n$, where X can be any amino or imino acid residue. Finally, as we have described above and is illustrated in Figure 5(d), the high 200/100 intensity ratio observed in (Pro-Gly-Pro)$_n$ is incompatible with structures having a sufficiently small value of ρ to make interchain $N_1H_1 \ldots O_3$ hydrogen bonding possible.

We have already discussed (Traub & Yonath, 1966) some of the implications for collagen of the structure of (Pro-Gly-Pro)$_n$ and pointed out that extensive portions of the protein apparently have the sequence Gly-Pro-X on adjacent portions of all three chains and are consequently likely to have essentially the same structure as (Pro-Gly-Pro)$_n$.

If other regions of collagen have the standard two-bonded structure, with six NH . . . O hydrogen bonds between three adjacent tripeptides, there would also have to be many border regions where the continuity of conformation is broken. At these borders there would be a change from $N_1H_1 \ldots O_2$ to $N_1H_1 \ldots O_3$ hydrogen bonding and small but general shifts of atomic positions, including about 0·35 Å in ρ. Regions with four or five NH . . . O hydrogen bonds, because of a Gly-Pro-X sequence on two or one of the three chains, would have inward pointing C_2O_2 carbonyl groups, unavailable for hydrogen bonding, and $H_1N_1O_3$ hydrogen bond angles greater than the 30° of the standard two-bonded structure (Ramachandran & Sasisekharan, 1965).

In fact, we now have evidence that sequences of the form Gly-Ala-Pro and Gly-Ala-Ala can be incorporated into a collagen-like structure with only one NH . . . O hydrogen bond per tripeptide (Traub, Yonath & Segal, 1969; Segal, Traub & Yonath, 1969; Segal, 1969), and we therefore believe that the structure we have derived for (Pro-Gly-Pro)$_n$ is representative of the structure of collagen as a whole.

We are grateful to Dr F. L. Hirshfeld and Dr J. Yonath for helpful discussions during the course of this work. This investigation was supported by research grant GM 08608 from the National Institutes of Health, U.S. Public Health Service.

APPENDIX

Derivation of Integrated Intensities for Streaks and Spots

It is convenient to consider X-ray reflections as arising from the intersection of a sphere of reflection, associated with the X-ray beam, and the Fourier transform of the molecular structure, associated with the specimen. An oriented fibre contains parallel crystallites with all possible orientations about the fibre axis, so we may consider the X-ray pattern as due to a superposition of all intersections of the sphere and transform when the latter is rotated about the fibre axis. For crystalline or sampled reflections the diffraction potentialities of the transform are concentrated at isolated points whose intersections with the sphere of reflection give rise to spots in the X-ray pattern. These points lie on the so-called reciprocal lattice and each corresponds to a cell of the transform with basal area A^* and height c^*. The volume of transform corresponding to a spot on a fibre photograph is therefore A^*c^*n, where n is the number of reciprocal lattice points contributing to the spot. For a hexagonal unit cell, as in $(\text{Pro-Gly-Pro})_n$, n is 12 in general, but 6 if $h = k$ or either h or k is zero.

Streaks in the X-ray pattern arise from regions where the diffraction potentialities of the transform vary continuously over a plane perpendicular to the fibre axis and separated by a distance c^* from the next plane. If a streak is due to regions in the plane which lie between $\xi - \frac{1}{2}\Delta\xi$ and $\xi + \frac{1}{2}\Delta\xi$ from the fibre axis, the volume of transform corresponding to it is $2\pi\xi\Delta\xi c^*$. For convenience of calculation, we have chosen a narrow streak at a relatively large distance from the meridian, but, in fact, intensities are recorded on the X-ray pattern on the same scale for all streaks for which the specimen is suitably aligned.

The probability of a volume element of the transform intersecting the sphere of reflection is inversely proportional to ξ the distance of the element from the fibre axis. ξ thus corresponds to the Lorentz factor of X-ray crystallography. We have noted in Experimental Procedure that the intensity recorded for a spot or narrow streak is proportional to $B \tan 2\theta$, where B is the area under a radial densitometer trace.

We can now calculate the observed integrated intensity scaled to the volume of transform to which it corresponds. This quantity is in fact F^2 observed, using crystallographic nomenclature, and should correspond to the calculated square of the Fourier transform for the correct molecular structure.

For spots,

$$F^2 \text{ obs} = \frac{\xi B \tan 2\theta}{A^*c^*n} \tag{1}$$

and for streaks, on the same scale,

$$F^2 \text{ obs} = \frac{\xi B \tan 2\theta}{2\pi\xi\Delta\xi c^*} = \frac{B \tan 2\theta}{2\pi\Delta\xi c^*}. \tag{2}$$

A^* is in fact $1/A$, where A is the c-axis projection of the unit cell in Å^2, ξ and $\Delta\xi$ are expressed in Å^{-1}, and c^* is a common factor for streaks and spots and can therefore be excluded from the calculation.

Equation (2) gives F^2 obs averaged over the whole length of the streak corresponding to $\Delta\xi$. The profiles of the streaks on the seventh and tenth layer lines were determined from densitometer traces along the layer lines and the total streak intensities, equal to $B \tan 2\theta/2\pi$, were distributed accordingly.

A rather similar method of scaling the intensities of spots and streaks on the same X-ray pattern has been used by Marvin *et al.* (1961).

REFERENCES

Andreeva, N. S., Esipova, N. G., Millionova, M. I., Rogulenkova, V. N. & Shibnev, V. A. (1967). In *Conformation of Biopolymers*, ed. by G. N. Ramachandran, vol. 2, p. 469. New York: Academic 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.

Corey, R. B. & Pauling, L. (1953). *Proc. Roy. Soc.* B, **141**, 10.

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

Davies, D. R. & Rich, A. (1959). *Acta Cryst.* **12**, 97.

Edsall, J. T., Flory, P. J., Kendrew, J. C., Liquori, A. M., Némethy, G., Ramachandran, G. N. & Scheraga, H. A. (1965). *J. Mol. Biol.* **15**, 399.

Engel, J., Kurtz, J., Katchalski, E. & Berger, A. (1966). *J. Mol. Biol.* **17**, 255.

Klug, A. & Franklin, R. E. (1958). *Disc. Faraday Soc.* **25**, 104.

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

Langridge, R., Seeds, W. E., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F. & Hamilton, L. D. (1960). *J. Mol. Biol.* **2**, 19.

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

Marvin, D. A., Spencer, M., Wilkins, M. H. F. & Hamilton, L. D. (1961). *J. Mol. Biol.* **3**, 547.

Oriel, P. J. & Blout, E. R. (1966). *J. Amer. Chem. Soc.* **88**, 2041.

Ramachandran, G. N. (1967). In *Treatise on Collagen*, ed. by G. N. Ramachandran, vol. 1, p. 103. London: Academic Press.

Ramachandran, G. N. & Chandrasekharan, R. (1968). *Biopolymers*, **6**, 1649.

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

Ramachandran, G. N. & Sasisekharan, V. (1965). *Biochim. biophys. Acta*, **109**, 314.

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

Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909.

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

Sasisekharan, V. & Ramachandran, G. N. (1957). *Proc. Indian Acad. Sci.* **45**, 363.

Scatturin, A., Del Pra, A., Tamburro, A. M. & Scoffone, E. (1967). *Chem. Ind. M.* **49**, 970.

Segal, D. M. (1969). *J. Mol. Biol.* **43**, 497.

Segal, D. M., Traub, W. & Yonath, A. (1969). *J. Mol. Biol.* **43**, 519.

Traub, W., Shmueli, U., Suwalsky, M. & Yonath, A. (1967). In *Conformation of Biopolymers*, ed. by G. N. Ramachandran, vol. 2, p. 449. London: Academic Press.

Traub, W. & Yonath, A. (1966). *J. Mol. Biol.* **16**, 404.

Traub, W. & Yonath, A. (1967). *J. Mol. Biol.* **25**, 351.

Traub, W., Yonath, A. & Segal, D. M. (1969). *Nature*, **221**, 914.

Vainshtein, B. K. (1966). *Diffraction of X-rays by Chain Molecules*. Amsterdam: Elsevier.