# Structure Analysis of a Collagen Model Polypeptide, (Pro-Pro-Gly)$_{10}$

Kenji OKUYAMA, Nobuo TANAKA,* Tamaichi ASHIDA,** and Masao KAKUDO*

*Department of Applied Chemistry, Kyushu University, Fukuoka 812*
**Institute for Protein Research, Osaka University, Suita, Osaka 565*
***Department of Applied Chemistry, Nagoya University, Nagoya 464*

(Received February 19, 1976)

A detailed conformation study of the polypeptide (Pro–Pro–Gly)$_{10}$, synthesized as a collagen model, has been done using X-ray diffraction. Several structures having the observed helical parameters (unit height=8.61 Å, unit twist=51.4°) together with standard bond lengths and angles were generated and tested in terms of an existence of intermolecular van der Waals contacts and least squares calculations using 49 equatorial reflections. It was found from these studies that three polypeptide chains make a three-stranded rope of which the helical parameters are similar to but significantly different from those of the collagen models proposed so far. There exist two possible models for (Pro–Pro–Gly)$_{10}$ according to the type of hydrogen bonding. One is the model having a hydrogen bond so-called 'standard' structure type, and the other has a hydrogen bond of collagen II type. The two-dimensional electron density map, synthesized with the observed structure factors, agrees well with these models. In the least squares calculations, atomic scattering factors were modified by taking into account of effects of amorphous water between the three-stranded ropes.

The first three-stranded coiled coil model for collagen was that suggested by Ramachandran and Kartha.[1] Since then, this structural model has been widely accepted as correct at least in its general features. The structure was derived from a prototype structure, shown in polyglycine II and polyproline II structures, by the introduction of a right handed rope twist so that the individual chains had a coiled coil conformation. It is almost impossible, however, to determine the detailed molecular structure from only the poor X-ray fiber patterns of native collagen; so various models for the molecular conformation have been suggested from the X-ray data. In order to obtain more decisive information to determine the structure of collagen precisely, a number of polypeptide models have been synthesized.[2,3] Since the features of the chemical compositions of the collagen are the occurrence of glycyl residues at every third positions in the sequence and a relatively high contents of proline and hydroxyproline residues, many of these synthetic polypeptides have the sequence in the form of (Gly–Pro–X)$_n$, in which X is a variety of amino acids. A number of structure analyses of these synthetic polypeptides have been reported so far,[4] although X-ray diffraction patterns obtained from these were also poor fiber patterns similar to those from collagen.

In the case of (Pro–Pro–Gly)$_{10}$, single crystals available for X-ray diffraction study were obtained by a dialysis from 10% acetic acid.[5] This is the first one that could be crystallized among the polypeptides belonging to collagen group proteins. The crystal is orthorhombic, space group $P2_12_12_1$, with twelve molecules per unit cell of dimensions: $a$=26.93, $b$=26.42 and $c$=100.4 Å. The crystal has a subcell structure with the $c'$ dimension of one-fifth of the true period, $c'$=100.4/5=20.08 Å. Intensity data of 787 reflections (exclusive of reflections with $l \neq 5n$) were collected. It was given in the previous paper[6] that the crystal of (Pro–Pro–Gly)$_{10}$ consists of a three-stranded coiled coil structure, of which the helical parameters are similar to but significantly different from those of collagen models proposed so far. The long extended helix of the polypeptides lies parallel to the c axis. In this paper we will describe more detailed molecular structures found by a systematic evaluation of possible molecular conformations and crystal structures derived by a method of two-dimensional least squares refinement.

## Structure Analysis

At the beginning of this study, we tried to crystallize a heavy atom derivative of (Pro–Pro–Gly)$_{10}$, that is 4-bromoprolylprolylglycyl-(Pro–Pro–Gly)$_9$. Although crystals of this derivative were obtained by a dialysis from acetic acid, these had a disorder along the c axis, and they were not suitable for X-ray study. Therefore, we built molecular models systematically as in fibrous protein research. These models were tested for acceptable intra- and intermolecular van der Waals contacts and also hydrogen bonds which stabilize the three-stranded coiled coil structure. Finally the structures were refined by the least squares calculations.

*Relations among the Individual Chains in Triple Helix.* As mentioned earlier,[6] the helical parameters of (Pro–Pro–Gly)$_{10}$ are considerably different from those of collagen models proposed so far.[7–9] There are 21 amino acid residues in one pitch of (Pro–Pro–Gly)$_{10}$ (Fig. 1, left). The vector from the first amino acid residue to the

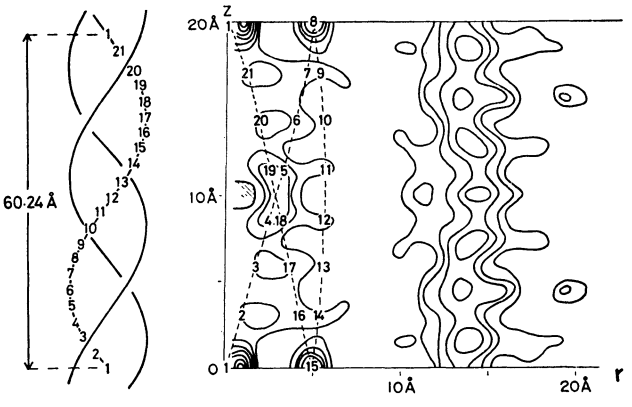


Fig. 1. Left: Three stranded helices. Each helix has 21 amino acid residues in one pitch. Right: Cylindrical Patterson map of (Pro–Pro–Gly)$_{10}$.

$n$-th amino acid residue is represented as a symbol $n$ in the cylindrical Patterson map (Fig. 1, right). These vectors are connected with dashed lines. Peaks in the range of $0 \leq r \leq 10$ Å in this map are well interpreted by these 21 vectors.

The structure of $(Pro\text{–}Pro\text{–}Gly)_{10}$ consists of three equivalent helical molecules, and each of them has the above mentioned conformation. If these three chains are arranged in an antiparallel fashion, the repeat distance along the chain axis coincides with a pitch, 60.24 Å, rather than a pseudoperiod of 20.08 Å (=60.24/3), which is clearly shown in the (0$kl$) precession photograph.[6] Therefore, these three are arranged in a parallel fashion. Electron microscopic study of this polypeptide has shown the same result.[10]

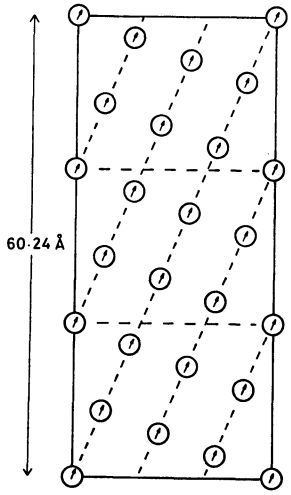


Fig. 2. Radial projection of $(Pro\text{–}Pro\text{–}Gly)_{10}$.

A radial projection of a right handed helix of $(Pro\text{–}Pro\text{–}Gly)_{10}$ is shown in Fig. 2, where an open circle represents a repeating unit, Pro–Pro–Gly, and an arrow in the circle represents a direction of the unit. It is easily shown that in order to have a pseudoperiod of 20.08 Å, the chains must be all parallel and also shifted by a distance of 2.87 Å or 5.74 Å which correspond to the heights of one-third or two-thirds of the unit along the chain axis, from the neighboring chains. Consequently, it is clear that the relation of the atomic coordinates of one chain and those of the corresponding in the other two chains are as follows;

$$(r,\ \phi,\ z)$$
$$(r,\ \phi-102.9°,\ z+2.87)$$
$$(r,\ \phi-205.8°,\ z+5.74)$$

If the polypeptide chain is a left handed helix, its radial projection is a mirror image of that of a right handed one and the relation of the atomic coordinates are as follows;

$$(r,\ \phi,\ z)$$
$$(r,\ \phi+102.9°,\ z+2.87)$$
$$(r,\ \phi+205.8°,\ z+5.74)$$

*Analytical Procedure.* Figure 3 (a) shows the structural formula of a tripeptide unit of $(Pro\text{–}Pro\text{–}Gly)_{10}$, together with its nine internal rotations. In

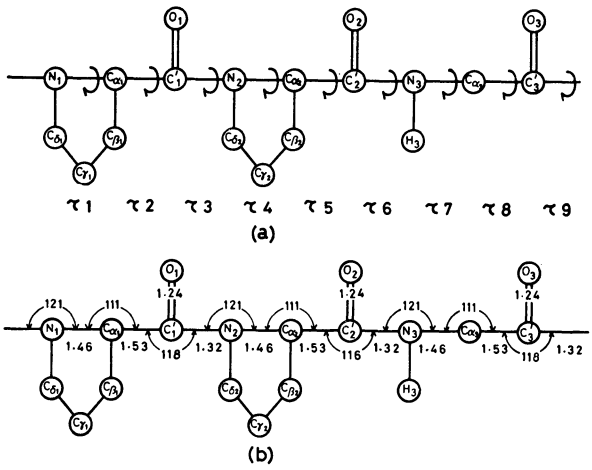


Fig. 3. (a) Structural formula and the notation of internal rotation angles in one tripeptide unit of $(Pro\text{–}Pro\text{–}Gly)_{10}$. (b) Standard stereochemistry of $(Pro\text{–}Pro\text{–}Gly)_{10}$.

building the various possible molecular conformations, we have made several stereochemically reasonable assumptions. Bond distances and angles of peptide backbone were fixed to the values shown in Fig. 3 (b). These are the mean values obtained from the results of structure analyses of oligopeptides. The value of $C_\alpha\text{–}C'\text{–}N$ angle is 118° in the peptide group followed by the prolyl residue, and 116° in that followed by the glycyl residue, because of the steric repulsion between the $C_\alpha$ atom of the preceding residue and the bulky $C_\delta$ methylene group of the pyrrolidine ring.[11] For the glycyl residue we have taken the values of 1.0 Å, 1.07 Å, 120° and 109.5° for the bond lengths of N–H and C–H and bond angles of $C_\alpha\text{–}N\text{–}H$ and $H\text{–}C_\alpha\text{–}H$, respectively. For the prolyl residue we have taken the dimension of *p*-bromocarbobenzoxy-glycyl-prolyl-leucyl-glycine.[12] It is known that *cis* as well as *trans* conformations are possible for the peptide bond of proline residues (*e.g.*, the structures of polyproline I and II). However, on the triple helix of $(Pro\text{–}Pro\text{–}Gly)_{10}$ at the lower temperatures, the study of proton magnetic spectra made it clear that the peptide bonds in Pro–Pro and Gly–Pro are all in *trans* conformations.[13] So, we also assumed that the six atoms of the peptide group in all three residues are coplaner, that is, $\tau 3 = \tau 6 = \tau 9 = 180°$. In this study, dihedral angles of the polypeptide chain are taken to be zero for a *cis* conformation and are taken to be positive for an anticlockwise rotation looking along the peptide backbone from the amino terminal to the carboxyl terminal. (This definition is different from that of a standard nomenclature[14] in the direction of rotation.)

Thus, the variable parameters reduced to only six dihedral angles, *i.e.*, $\tau 1$, $\tau 2$, $\tau 4$, $\tau 5$, $\tau 7$ and $\tau 8$. If we take into account of our finding that the structure of $(Pro\text{–}Pro\text{–}Gly)_{10}$ has 7 units (*i.e.*, 21 amino acid residues) in one turn of a single chain, there become not six but five independent parameters.[15] In our program the five parameters were varied systematically and a number of molecular models were built for each combination.[16] Intramolecular atomic coordinates for each model are

transformed into the cylindrical coordinates by using **T** and **L** matrices.[17] The unit height $h$ and one of the six dihedral angles, which was $\tau 2$ in our calculations, were obtained simultaneously. The molecular conformations, whose $h$ and $\tau 2$ are out of the ranges of criterion listed in Table 1, were eliminated in each stage. For the remaining conformations the atomic coordinates of the neighboring chains were given by the relations among the three-stranded chains. Further, the models which do not have a hydrogen bond, judged from a distance $b$(N, O) and an angle $\theta$(H, N, O), were excluded. Finally the remainders were tested against intra- and intermolecular van der Waals contacts. The minimum van der Waals contacts are based on the "outer limit" values cited by Ramakrishnan and Ramachandran.[18] The various stages of the analysis and the regions found to have the reasonable structures are described in Table 1.

The validities of the molecular structure were tested by the two-dimensional least squares refinements using 49 equatorial reflections. The parameters are the position of the molecular axis $(x, y)$, a rotation angle of a molecule about it, an over all temperature factor and a scale factor. In these calculations we have taken into account of the scattering by water molecules in the unit cell as described in the following section.

*Effects of Water Molecules.* The water content of (Pro-Pro-Gly)$_{10}$ crystals is 46% (by weight). Although several water molecules may be held at the definite positions by hydrogen bonds to the carbonyl oxygens of polypeptides, these positions may be clarified as the structure analysis proceeds. We took into account of the only water molecules which can be assumed to be uniformly distributed between the three-stranded ropes. For the first approximation we have assumed that all the atoms are spheres with van der Waals radii, 1.60, 1.35 and 1.40 Å for C, N and O, respectively. Each of the atomic scattering factors shown in Fig. 4 is calculated from the ordinary atomic scattering factors minus the scattering factor of water molecules which occupy uniformly the atomic volume with an electron

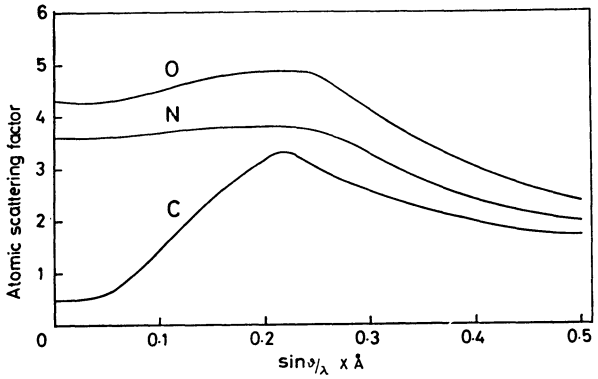


Fig. 4. Atomic scattering factors for atoms immersed in water.

density of 0.33 electrons/Å$^3$. This correction of atomic scattering factors decreased the $R$ index about 10% in the least squares calculations. The temperature factor was not decreased so immediately as the cases without a correction. A similar method of correcting for background has been used already in the structural studies of DNA.[19]

*Results of Analysis.* Structures of the left handed helices as well as those of the right handed ones were examined. However, none of the structures of the left handed helices was acceptable in view of their intra- and intermolecular atomic contacts. As shown in Table 1, the ranges of five independent internal rotation angles became narrow at each cycle. In cycle 1, the dihedral angles, $\tau 1$ and $\tau 4$ in pyrrolidine rings were fixed to 60°, while the dihedral angles of glycine, $\tau 7$ and $\tau 8$, were searched over all ranges. From structure analyses of oligopeptides[11,20] and the energy calculations of prolyl oligomers,[21] the dihedral angles of $C_\alpha$–C′ in proline residue, $\tau 2$ and $\tau 5$, were assumed to be in the ranges of 170~260°. There are some structures in which these angles are in the range of −50~40°, but the range is for the α-helix type dihedral angles.[11] Increments of the three parameters, $\tau 5$, $\tau 7$ and $\tau 8$, were 10° at the first cycle. The structures in the range of

Table 1. Computational details in structure analysis of (Pro-Pro-Gly)$_{10}$

| | Parameters | Cycle 1 | Cycle 2 | Cycle 3 | Cycle 4 | |
|---|---|---|---|---|---|---|
| Region tested (Increments) | $\tau 1$ | 60° | 45—85° (5) | 50—85° (5) | 53—77° (3) | |
| | $\tau 4$ | 60° | 45—80° (5) | 50—70° (5) | 47—68° (3) | |
| | $\tau 5$ | 170—260° (10) | 170—240° (10) | 175—205° (5) | 182—203° (3) | |
| | $\tau 7$ | 0—350° (10) | 40—100° (10) | 75—105° (5) | 77—98° (3) | |
| | $\tau 8$ | 0—350° (10) | 180—240° (10) | 175—215° (5) | 183—210° (3) | |
| Criteria | $\tau 2$ | 170—270° | 170—270° | 170—270° | 170—270° | |
| | Unit height | 6.61—10.61Å | 8.44—8.78Å | 8.51—8.71Å | 8.58—8.64Å | |
| | H-bond | | | | | |
| | $b$(N, O) | — | 3.2—2.6Å | 3.2—2.6Å | 3.2—2.6Å | |
| | $\theta$(H, N, O) | — | — | 0—35° | 0—30° | |
| Region passed | | | | | Model 1 | Model 2 |
| | $\tau 1$ | 60° | 55—85° | 55—75° | 53—68° | 68—77° |
| | $\tau 2$ | 170—240° | 177—209° | 188—209° | 186—202° | 193—208° |
| | $\tau 4$ | 60° | 50—65° | 50—65° | 53—68° | 50—65° |
| | $\tau 5$ | 170—260° | 180—200° | 185—200° | 185—188° | 194—197° |
| | $\tau 7$ | −20—140° | 80—100° | 80—95° | 92—95° | 83—86° |
| | $\tau 8$ | 180—240° | 180—210° | 185—210° | 198—210° | 183—192° |

6.61～10.61Å of the unit height were selected in all combinations of these three parameters. After the first cycle it was shown that the angles of $\tau 7$ and $\tau 8$ of glycine residue were in the range of −20～140° and 180—240°, respectively. These ranges show that the glycyl residue has the left handed polyglycine II type structure, as in the collagen models. Therefore, taking into account of energy map of diglycine,[22] the range of $\tau 7$ was restricted to 40～100° at the second cycle.

In cycles 2, 3 and 4, structures without NH···O hydrogen bonds were eliminated. In the case of $(Pro–Pro–Gly)_{10}$, there were two possible types of hydrogen bonds, that is, $N_3(Gly)\cdots O_2(Pro)$ and $N_3(Gly)\cdots O_1(Pro)$. We call hereafter the structure having the former type hydrogen bond model 1 and that having the latter model 2. The ranges of parameters for model 1 and 2 at cycle 4 are shown in the last column of Table 1. Illustrations of a hydrogen bond between neighboring chains of each model are shown in Fig. 5.

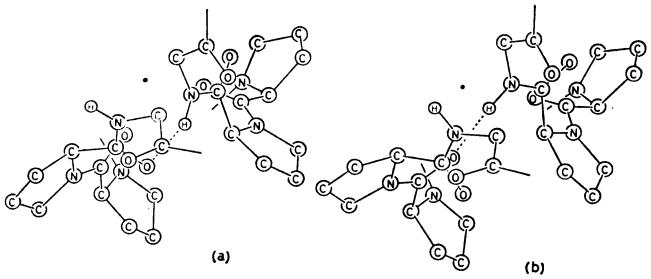


Fig. 5. An axial projection of segments of the two $(Pro–Pro–Gly)_{10}$ models. (a) Model 1, (b) Model 2. The filled circle represents the axis of the molecule for each model. The dashed line represents a hydrogen bond.

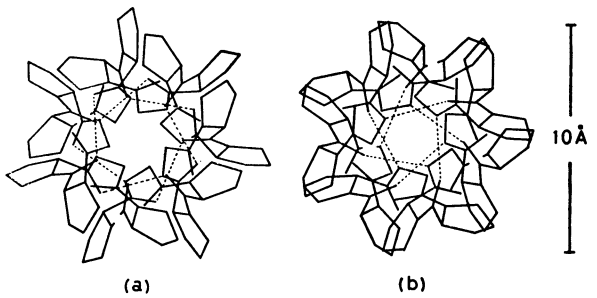


Fig. 6. The molecular structure viewed down the c axis. (a) Model 1, (b) Model 2.

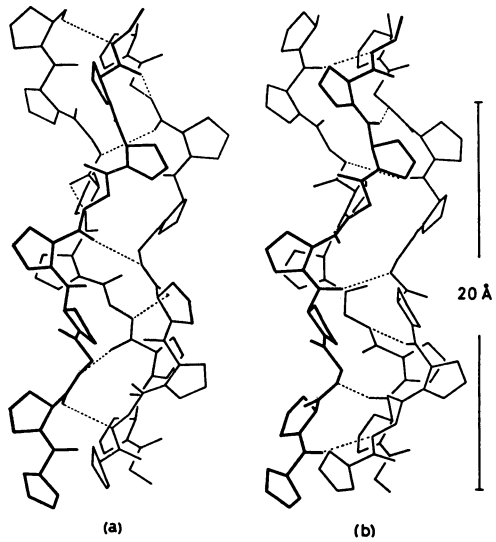


Fig. 7. The molecular structure viewed down the b axis. (a) Model 1, (b) Model 2.

Fifteen structures, which were chosen from different parts of regions for each model, have been tested by the two-dimensional least squares refinements. The $R$ index $(\Sigma||F_o|^2-|F_o|^2|/\Sigma|F_o|^2)$ was calculated after three cycles with the same initial setting parameters for each structure. The dihedral angles and the hydrogen bond parameters of all the fifteen structures are listed in Table 2, together with their $R$ indices.

As typical structures, the structure of No. 4 in Table 2 for model 1 and that of No. 13 for model 2 are shown in Figs. 6 and 7, since they had small $R$ indices and good hydrogen bond parameters. Atomic coordinates for the repeating unit are given in Table 3. The intermolecular short contacts (less than 3.5 Å) of the structures are as follows;

$O_1\cdots C\delta_1$ 3.27, $N_3\cdots O_1$ 3.14, $N_3\cdots O_2$ 3.09, $C\alpha_3\cdots C\alpha_3$ 3.49, $C'_3\cdots C\alpha_3$ 3.38, $O_3\cdots C\alpha_3$ 3.28, $O_3\cdots C\delta_1$ 2.90, $O_1\cdots C\alpha_3$ 2.99 and $O_1\cdots C\delta_1$ 2.27 Å in model 1 (No. 4). $C\alpha_2\cdots O_1$ 2.96, $C\beta_2\cdots O_1$ 3.42, $C_2'\cdots O_1$ 3.41, $N_3\cdots O_1$ 2.97 and $O_3\cdots C\alpha_3$ 2.99 Å in model 2 (No. 13).

Since the molecule has 7 units in one turn (7/1 helix), there is a 7-fold rotational symmetry in Fig. 6. At the molecular center of this figure, there exists a hole surrounded by glycyl residues. The dimension of this

TABLE 2. PARAMETERS OF FIFTEEN STRUCTURES CHOSEN FROM THE DIHEDRAL ANGLE REGIONS FOR EACH MODEL

| | Model 1 | | | | | | | Model 2 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| $\tau 1$/° | 54 | 56 | 58 | 60 | 62 | 64 | 66 | 67 | 69 | 71 | 73 | 73 | 75 | 75 | 77 |
| $\tau 2$/° | 193 | 195 | 192 | 193 | 197 | 199 | 199 | 199 | 198 | 202 | 202 | 202 | 205 | 199 | 207 |
| $\tau 4$/° | 62 | 60 | 62 | 60 | 56 | 54 | 54 | 60 | 60 | 58 | 54 | 58 | 52 | 58 | 50 |
| $\tau 5$/° | 185 | 185 | 188 | 186 | 186 | 186 | 188 | 194 | 197 | 196 | 194 | 197 | 194 | 198 | 194 |
| $\tau 7$/° | 95 | 95 | 91 | 94 | 94 | 94 | 92 | 86 | 82 | 83 | 86 | 82 | 86 | 82 | 86 |
| $\tau 8$/° | 209 | 207 | 207 | 205 | 203 | 201 | 199 | 192 | 192 | 188 | 188 | 186 | 186 | 186 | 184 |
| $b$(N, O)/Å | 2.95 | 2.95 | 2.93 | 3.09 | 3.03 | 3.02 | 3.14 | 3.08 | 2.74 | 2.99 | 2.94 | 3.07 | 2.97 | 3.12 | 3.01 |
| $\theta$(H, N, O)/° | 22 | 20 | 29 | 30 | 27 | 25 | 22 | 29 | 30 | 28 | 30 | 28 | 29 | 27 | 28 |
| $\theta'$(180-NOC)/° | 76 | 78 | 70 | 68 | 70 | 72 | 76 | 27 | 22 | 33 | 23 | 33 | 25 | 20 | 28 |
| $R$ index × 100 | 42 | 42 | 39 | 32 | 35 | 36 | 34 | 29 | 34 | 30 | 28 | 26 | 27 | 27 | 26 |

Table 3. Atomic coordinates of the repeating unit in the model structures of (Pro–Pro–Gly)$_{10}$

| Residue | Atom | Model 1 (No. 4) | | | Model 2 (No. 13) | | |
|---|---|---|---|---|---|---|---|
| | | $x$/Å | $y$/Å | $z$/Å | $x$/Å | $y$/Å | $z$/Å |
| Pro | $N_1$ | 2.94 | 0.00 | 0.00 | 3.40 | 0.00 | 0.00 |
| | $C\alpha_1$ | 3.88 | −0.27 | 1.08 | 4.08 | −0.52 | 1.18 |
| | $C\beta_1$ | 4.68 | 1.06 | 1.13 | 5.05 | 0.64 | 1.50 |
| | $C\gamma_1$ | 4.43 | 1.77 | −0.08 | 5.14 | 1.48 | 0.37 |
| | $C\delta_1$ | 3.17 | 1.27 | −0.68 | 3.93 | 1.27 | −0.49 |
| | $C_1'$ | 3.14 | −0.63 | 2.37 | 3.16 | −0.52 | 2.39 |
| | $O_1$ | 1.93 | −0.42 | 2.39 | 2.23 | 0.30 | 2.38 |
| Pro | $N_2$ | 3.85 | −1.15 | 3.36 | 3.40 | −1.38 | 3.36 |
| | $C\alpha_2$ | 3.21 | −1.52 | 4.62 | 2.56 | −1.44 | 4.55 |
| | $C\beta_2$ | 4.40 | −2.13 | 5.40 | 3.19 | −2.63 | 5.32 |
| | $C\gamma_2$ | 5.44 | −2.40 | 4.48 | 4.01 | −3.35 | 4.42 |
| | $C\delta_2$ | 5.23 | −1.58 | 3.25 | 4.37 | −2.47 | 3.27 |
| | $C_2'$ | 2.55 | −0.31 | 5.27 | 2.41 | −0.05 | 5.17 |
| | $O_2$ | 2.72 | 0.81 | 4.76 | 3.18 | 0.85 | 4.81 |
| Gly | $N_3$ | 1.83 | −0.55 | 6.36 | 1.46 | 0.08 | 6.08 |
| | $C\alpha_3$ | 1.16 | 0.53 | 7.07 | 1.22 | 1.36 | 6.74 |
| | $C_3'$ | 2.03 | 1.06 | 8.21 | 2.12 | 1.50 | 7.97 |
| | $O_3$ | 2.87 | 0.28 | 8.67 | 2.81 | 0.52 | 8.27 |

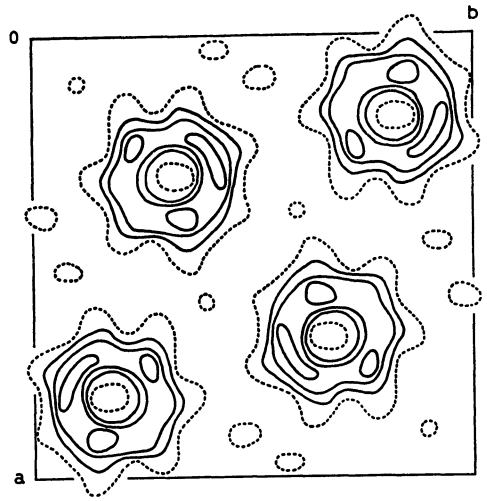


Fig. 8. Electron density map projected along the c axis.

hole is 2.0~2.5 Å for model 1 and 2.7~3.1 Å for model 2. The $R$ index of the least squares refinements decreases as the dimension of this hole increase.

The electron density map, calculated from the observed structure factors $|F_o|$ and phases calculated by the fifteen structures of Table 2, is shown in Fig. 8. Only 38 ($hk0$) reflections whose signs were the same in all the fifteen structures were used in this calculation. In this map, the center portion of the triple helix has low electron density, and high electron density parts are distributed cylindrically with about 5.7 Å in diameter. Further, there are seven projections of electron density from this cylinder.

## Discussion

It can be seen from the systematic study of molecular conformations that there exist two possible structures distinguished by hydrogen bonding. The structure of model 1 has the so called "standard" structure type hydrogen bond and that of model 2 has the collagen 11 type hydrogen bond. The difference between these two structures is illustrated in Fig. 5 and the values of hydrogen bond parameters are listed in Table 2. The angle $\theta$(H, N, O), between N–H and H···O, is 20~30° in model 1 and 27~30° in model 2. The angle $\theta'$(180−NOC'), between C=O and O···N, is about 70~80° in model 1 and 20~30° in model 2. These values as well as the hydrogen bond lengths are not extreme cases of hydrogen bonding parameters.[23]

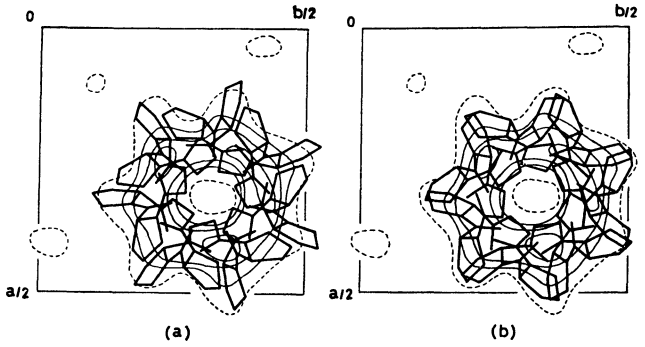


Fig. 9. Comparison of the electron density in an asymmetric unit with the molecular structures of (a) model 1 and (b) model 2 for (Pro–Pro–Gly)$_{10}$.

The two structures are supported by the electron density map as seen from Fig. 9. There is a remarkable coincidence, seven knobs of pyrrolidine rings being well fitted to those of the electron density distribution. The position of the molecular axis of the triple helix obtained from least squares refinements is $x=8.57\pm0.06$ Å and $y=8.42\pm0.06$ Å. Then, the distance between nearest neighbours of the triple helix is 13.7~12.9 Å, which agrees well with the peaks lined up at about $r=14$ Å in the cylindrical Patterson map (Fig. 1). The electron density map calculated from all the 49 ($hk0$) reflections for each model was somewhat different from the map in Fig. 8 if examined in detail. This may be due to the neglect of the water molecules which are at the definite positions in the unit cell. At this stage it cannot be decided which is more plausible for (Pro–Pro–Gly)$_{10}$, model 1 or 2.

A comparison of the hydrogen bond type and the helical parameters of (Pro–Pro–Gly)$_{10}$ with those of the several collagen models[7,9] and of a synthetic polypeptide (Pro–Gly–Pro)$_n$[4] is shown in Table 4. The amino acid sequence of (Pro–Gly–Pro)$_n$ is the same as that of (Pro–Pro–Gly)$_{10}$. The molecular weight is about 15000 for the former and 2528 for the latter. Our sample was practically monodispersed,[24] and single crystals available for X-ray study were obtained. On the other hand, molecular weights of (Pro–Gly–Pro)$_n$ seem to be polydispersed and only the fiber patterns similar to those of collagen could be obtained. Collagen 1 and "standard" structures can not be considered as good models for collagen because imino acids could not be incorporated in position $R_3$ or $R_2$, respectively,[25] while it is known that the amino acid sequence of collagen

Table 4. Comparisons of features for several collagen models and collagen model synthetic polypeptides

| | Amino acid sequence $(Gly_1\text{–}R_2\text{–}R_3)_n$ | Types of hydrogen bond | Helical parameters |
|---|---|---|---|
| Collagen I[7] | $R_2$ any residue<br>$R_3$ Gly only | $N(Gly_1)\cdots O(Gly_1)$ | 10/1 helix, $h=2.86\times3$Å, $t=36°$ |
| Collagen II[7] | $R_2$ any residue<br>$R_3$ any residue | $N(Gly_1)\cdots O(R_2)$ | 10/1 helix, $h=2.86\times3$Å, $t=36°$ |
| 'Standard' structure[9] | $R_2$ Pro or Hypro imposible<br>$R_3$ any residue | $N(Gly_1)\cdots O(R_3)$<br>$N(R_2)\cdots O(R_2)$ | 12/1 helix, $h=2.91\times3$Å, $t=30°$ |
| $(Pro\text{–}Gly\text{–}Pro)_n$[4] | $R_2$ Pro<br>$R_3$ Pro | $N(Gly_1)\cdots O(R_2)$ | 10/1 helix, $h=2.87\times3$Å, $t=36°$ |
| $(Pro\text{–}Pro\text{–}Gly)_{10}$ | $R_2$ Pro<br>$R_3$ Pro | $N(Gly_1)\cdots O(R_2)$<br>or $N(Gly_1)\cdots O(R_3)$ | 7/1 helix, $h=2.87\times3$Å, $t=51.4°$ |

proteins obtained from several sources has many tripeptide sequences of Gly–Pro–X. Structures of collagen II and $(Pro\text{–}Gly\text{–}Pro)_n$ have the same type hydrogen bond and the same helical parameters with each other. And the structure of collagen II is considered to be the most favorable model for collagen so far.[25] In the model I of $(Pro\text{–}Pro\text{–}Gly)_{10}$, the hydrogen bond is the same type as that in "standard" structure and the other (model 2) is the same as that of collagen II structure, although the helical parameters of this polypeptide are significantly different from those of collagen models. The major difference of these structures is in the pseudoperiod along the fiber axis. The pseudo period of $(Pro\text{–}Pro\text{–}Gly)_{10}$ is 20.08 Å, which is clearly shown in the (0*kl*) precession photograph.[6] On the other hand, that of collagen II structure was considered to be 28.6 Å, since the meridional reflection with 2.86 Å spacing was considered to be 10 th order reflection. Bear[26] thought, however, that the pseudoperiod of kangaroo tail tendon collagen was 20.0 Å. He ascribed the meridional reflection of 2.86 Å as the 7-th, although the molecular structure suggested was incorrect. If this is true, the helical parameters of $(Pro\text{–}Pro\text{–}Gly)_{10}$ are the same as those of collagen.

During this study we have assumed that the molecule is an infinite chain, but actually the dimension of the molecules is 86.1 Å ($2.87\times30$) and the number of turns in this peptide chain is only 1.4. Therefore, it should be taken into account of this molecular terminal effects in the least squares refinements. However, the results at this stage are the same, whether the molecule is assumed to be finite or infinite.

We are grateful to Prof. M. Takayanagi of Kyushu University for his encouragement throughout this work. We are indebted to Drs. S. Sakakibara and Y. Kishida of Protein Research Foundation for suppling the valuable samples. All the numerical computations were done on a FACOM 230-75 of the Computer Center of Kyushu University.

## References

1) G. N. Ramachandran and G. Kartha, *Nature*, **176**, 593 (1955).
2) N. S. Andreeva, N. G. Esipova, M. I. Millionova, V. N. Rogulenkova, and V. A. Shibnev, "Conformation of Biopolymers," Vol. 2, Academic, New York (1967), p. 469.
3) D. M. Segal and W. Traub, *J. Mol. Biol.*, **43**, 487 (1969).
4) A. Yonath and W. Traub, *J. Mol. Biol.*, **43**, 461 (1969).
5) S. Sakakibara, Y. Kishida, K. Okuyama, N. Tanaka, T. Ashida, and M. Kakudo, *J. Mol. Biol.*, **65**, 371 (1972).
6) K. Okuyama, N. Tanaka, T. Ashida, M. Kakudo, S. Sakakibara, and Y. Kishida, *J. Mol. Biol.*, **72**, 571 (1972).
7) A. Rich and F. H. C. Crick, *Nature*, **176**, 915 (1955).
8) P. M. Cowan, S. McGavin, and A. C. T. North, *Nature*, **176**, 1062 (1955).
9) G. N. Ramachandran, "Treatise on Collagen," Vol. 1, ed. by G. N. Ramachandran, Academic, London (1967), p. 103.
10) B. R. Olsen, R. A. Berg, S. Sakakibara, Y. Kishida, and D. J. Prockop, *J. Mol. Biol.*, **75**, 589 (1971).
11) T. Ashida and M. Kakudo, *Bull. Chem. Soc. Jpn.*, **47**, 1129 (1974).
12) T. Ueki, T. Ashida, M. Kakudo, Y. Sasada, and Y. Katsube, *Acta Crystallogr.*, **B25**, 1840 (1969).
13) Y. Kobayashi and Y. Kyogoku, *J. Mol. Biol.*, **81**, 337 (1973).
14) IUPAC-IUB Commission on Biochemical Nomenclature, *J. Mol. Biol.*, **52**, 1 (1970).
15) M. Yokouchi, H. Tadokoro, and Y. Chatani, *Macromolecules*, **7**, 769 (1974).
16) S. Arnott and A. J. Wonacott, *Polymer*, **7**, 157 (1966).
17) H. Sugeta and T. Miyazawa, *Biopolymers*, **5**, 673 (1967).
18) C. Ramakrishnan and G. N. Ramachandran, *Biophys. J.*, **5**, 909 (1965).
19) R. Langridge, D. A. Marvin, W. E. Seeds, H. R. Wilson, C. W. Hooper, and M. H. F. Wilkins, *J. Mol. Biol.*, **2**, 38 (1960).
20) R. Balasubramanian, A. V. Lakshminarayanan, M. N. Sabesan, G. Tegoni, K. Venkatatesan, and G. N. Ramachandran, *Int. J. Protein Res.*, **3**, 25 (1971).
21) A. J. Hopfinger and A. G. Walton, *J. Macromol. Sci. Phys.*, **B3**, 171 (1969).
22) A. J. Hopfinger, Ph. D. thesis, Case Western Reserve University, Cleveland, Ohio (1970).
23) C. Ramakrishnan and N. Prasad, *Int. J. Protein Res.*, **3**, 209 (1971).
24) S. Sakakibara, Y. Kishida, Y. Kikuchi, R. Sakai, and K. Kakiuchi, *Bull. Chem. Soc. Jpn.*, **41**, 1273 (1968).
25) D. B. Fraser and T. P. MacRae, "Conformation in Fibrous Proteins," Academic, New York and London (1973), p. 371.
26) R. S. Bear, *Advan. Protein Chem.*, **7**, 69 (1952).