J. Biochem. 125, 310-318 (1999)

Structure Analysis of a Collagen-Model Peptide with a (Pro-Hyp-Gly) Sequence Repeat

V. Nagarajan,[1] Shigehiro Kamitori, and Kenji Okuyama[2]

Department of Biotechnology and Life Science, Faculty of Technology, Tokyo University of Agriculture and Technology, 2-24-16, Naka-cho, Koganei, Tokyo 184-8588

Received September 17, 1998; accepted October 26, 1998

The crystal structure of a triple helical peptide (Pro-Hyp-Gly)$_{10}$ has been determined at 1.9 Å resolution. Single crystals grown by the hanging drop method, diffracted to a resolution of 1.8 Å. The polymer-like structure of the triple helical repeat Pro-Hyp-Gly was in accordance with the 7/2 model proposed for collagen and very similar to the previously determined structure with a Pro-Pro-Gly sequence repeat. The solvent structure was also very similar to that previously observed, showing similar hydration patterns, but different crystal packing. The presence of hydroxyproline did not have any effect on the molecular structure or the hydration structure. This is in accordance with the recent finding that the inductive effect of the hydroxyl group attached to the C$^{\gamma}$ atom of hydroxyproline enhances collagen stability rather than the extensive water network.

Key words: collagen, hydration, hydroxyproline, model peptide, triple helix.

The triple helix is a unique structural motif found in fibril-forming collagen, a series of host defense proteins, and certain membrane proteins (*1*). The triple helix is also the functional element in certain receptors (*2*). The triple helical domains present in C1q and macrophage scavenger receptors act as the sites for binding interactions (*3-5*). The collagen molecule is a triple-helical coiled-coil formed by the super coiling of three polypeptide chains. Each strand is in a left-handed, extended polyproline II helical conformation. Three such strands coil about a common axis in a right-handed helical fashion. Each strand is staggered by one residue from its adjacent chain and the three strands are held together by inter-chain hydrogen bonding. The general amino acid sequence repeat of these polypeptide chains is of the form (Gly-X-Y)$_n$, in which X and Y can be any amino acid. The strict sequence constraint of Gly at every third position is required for the close packing of the three strands. The X and Y positions are frequently occupied by proline and hydroxyproline, respectively, comprising about 20% of all the residues in collagen. The high content of imino acids imposes a high degree of steric restriction which helps to stabilize the extended nature of the chains. Hydroxyproline results from the post-translational modification of proline by the enzyme prolyl hydroxylase, which adds a hydroxyl group to the C$^{\gamma}$ atom of proline. Hydroxyproline plays a very important role in the stability of the triple helix, as compared with proline (*6, 7*). This can be clearly observed by melting studies on (Pro-Pro-Gly)$_{10}$ and (Pro-Hyp-Gly)$_{10}$ (*7*); while the T_m for (Pro-Hyp-Gly)$_{10}$ is about 58°C in aqueous 10% acetic acid solution, the T_m for (Pro-Pro-Gly)$_{10}$ is only 24°C. Collagen lacking hydroxyproline is unstable (*8*), while the presence of Hyp at the Y-position has been found to enhance the stability of the assembly of collagen into microfibrils (*9*).

Initial models of collagen were proposed based on the fiber diffraction pattern of collagen and the diffraction pattern of synthetic polymers such as poly-Gly-II, poly L-Pro-II, *etc.* (*10, 11*). Rich and Crick (1961) proposed a model for collagen based on the high angle X-ray diffraction pattern of rat tail tendon (*12*). Later, in 1969, the fiber diffraction pattern of poly (Pro-Pro-Gly) was observed (*13*) to be very similar to the patterns obtained for native collagen, and detailed conformational analysis showed a similar type of structure as the Rich and Crick model. This model was further confirmed by fiber diffraction analysis on kangaroo tail tendon (*14*).

Meanwhile, synthetic oligopeptides such as (Pro-Pro-Gly)$_{10}$, (Pro-Hyp-Gly)$_{10}$, *etc.*, synthesized by solid-phase synthesis were found to be homologous and of defined molecular weight (*7, 15*). Single crystals of (Pro-Pro-Gly)$_{10}$ were grown by the dialysis method (*16*). Okuyama *et al.* (*17, 18*) reported the crystal structure of (Pro-Pro-Gly)$_{10}$ determined based on data collected from a single crystal and solved by the linked-atom least-squares method (LALS) (*19*). This structure shows fairly large differences in the helical parameters, leading to a new model for the triple helix (*20*). This model (Okuyama model) has a 7/1 helical symmetry with a 60 Å helical pitch for each strand of the triple helix, and the overall triple helix has a left-handed 7/2-helical symmetry. The observed result differs from the Rich and Crick model which has a 10/

[1] The author wishes to thank the Ministry of Culture, Science, Sports and Education of Japan, for the Monbusho fellowship.

[2] To whom correspondence should be addressed. Tel: +81-423-88-7028, Fax: +81-423-88-7209

Abbreviations: Gly→Ala peptide, structure determined by Bella *et al.* (1994); Hyp, 4-hydroxyproline; PPG-1, structure determined by Nagarajan *et al.* (1998); PPG-2, structure determined by Kramer *et al.* (1998) at 2.0 Å resolution; PPG-3, structure determined by Kramer *et al.* (1998) at 1.7 Å resolution.

1-helical symmetry with a helical pitch of 86 Å for each strand of the triple helix, and an overall structure with a left-handed 10/3-helical symmetry. The new 7/2 model could also explain the fiber diffraction pattern of native collagen (*20*).

Crystal structure analysis of a synthetic triple helical peptide, (Pro-Hyp-Gly)$_4$-(Pro-Hyp-Ala)-(Pro-Hyp-Gly)$_5$, hereafter called Gly→Ala peptide, revealed the structure to be in accordance with the Okuyama model except for regions around Ala (*21*). Recently, two crystal structure analyses (one by our group, hereafter called PPG-1, and the other by Kramer *et al.*, which contained two separate structural studies, hereafter referred to as PPG-2 and PPG-3) for the subcell structure of (Pro-Pro-Gly)$_{10}$ were solved (*22, 23*), which in both cases indicated structures similar to the model proposed by Okuyama *et al.* (*18*). The lowest energy models based on theoretical studies on poly (Pro-Pro-Gly) also agreed well with the experimental data of the Okuyama model (*24, 25*).

Single crystals of the collagen-model peptide (Pro-Hyp-Gly)$_{10}$ were obtained. The cylindrical Patterson map deduced from the data collected from these single crystals showed a main chain conformation very similar to that of the homologous peptide (Pro-Pro-Gly)$_{10}$, further confirming the Okuyama model for collagen (*26*). For a better understanding of collagen structure, we are currently studying the crystal structure of collagen model peptides. In this report the structure of the triple helix forming peptide (Pro-Hyp-Gly)$_{10}$ is described.

MATERIALS AND METHODS

Crystallization—The peptide (Pro-Hyp-Gly)$_{10}$ was obtained from Peptide Institute, Osaka, and used for the crystallization experiments without further purification. Crystals suitable for crystallographic analyses were grown by the hanging drop method at 10°C. Peptide solutions consisted of (Pro-Hyp-Gly)$_{10}$ at a concentration of 4 to 5 mg/ml in 10% acetic acid. About 1 ml of 22% PEG 200 was used as the reservoir solute. A mixture of 3 μl of peptide solution and 3 μl of reservoir solution was used as the drop hanging over the reservoir. Crystals were obtained with two different morphologies, triangular and rectangular, in the same drop (Fig. 1). Triangular crystals grew to a size of $0.3\times0.2\times0.2$ mm^3 while rectangular crystals grew to $0.4\times0.3\times0.2$ mm^3. These crystals were used for preliminary analyses and data collection.

Data Collection—The crystals were sealed in a glass capillary along with the mother liquor and data were collected using a 4-circle diffractometer (AFC5R, CuKα radiation, generated by a Rigaku RU 200 rotating anode at 40 kV and 150 mA). Data collection was performed at 5°C due to the rapid decay of the crystal at room temperature. Both crystal types belonged to the monoclinic system in the space group $P2_1$ with cell parameters of $a=14.06(1)$, $b=26.67(1)$, $c=20.06(1)$ Å, $\beta=107.08(5)°$. The crystals diffracted to a resolution of 1.8 Å. A total of 1,434 reflections were collected up to a resolution of 1.82 Å of which

TABLE I. **Crystal data and data collection parameters of (Pro-Hyp-Gly)$_{10}$.**

Data collection device	AFC5R-Rigaku 4-circle diffractometer
Overall completeness(%)	100
Space group	$P2_1$
Cell dimensions	$a=14.06(1)$, $b=26.67(1)$, $c=20.06(1)$Å, $\alpha=\gamma=90°$, $\beta=107.08(5)°$
Volume	$V=7192(11)$Å^3
Crystal dimensions	$0.4\times0.3\times0.2$ mm^3
Scan mode	Omega
Scan speed	8°/min
Scan width	$(\Delta w)=(1.10+0.3\tan\theta)°$
No. of standard reflections	3
No. of reflections measured	1,434
No. of unique reflections	1,302
Range of h, k, l	$0<h<8$, $0<k<15$, $-12<l<12$
$2\theta_{max}$	50°
Data collection temp.	5°
High resolution limit	1.824 Å

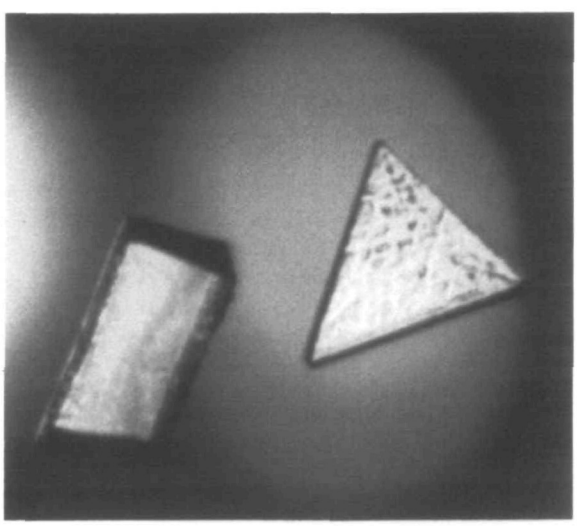

Fig. 1. **Two types of (Pro-Hyp-Gly)$_{10}$ single crystals found in the same drop.** The triangle shaped crystals grew to a size of $0.3\times0.2\times0.2$ mm^3 and the rectangular shaped crystals grew to a size of $0.4\times0.3\times0.2$ mm^3.

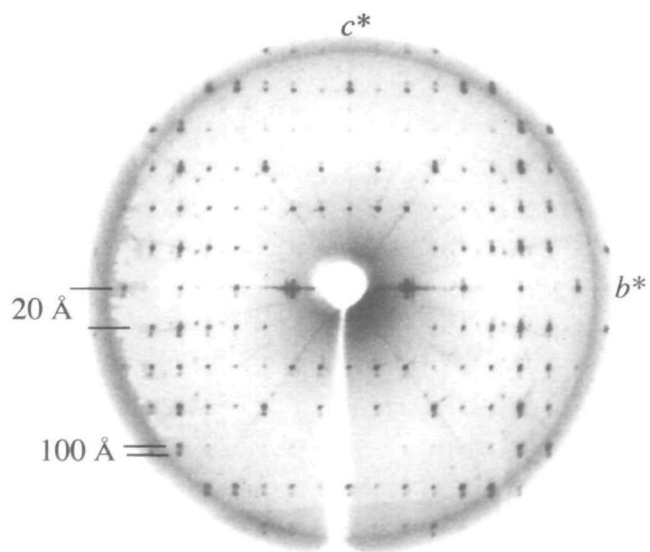

Fig. 2. **0*kl* precession photograph of (Pro-Hyp-Gly)$_{10}$, $\mu=15°$.**

1,302 reflections were found to be unique. Intensities were corrected for Lorentz-polarization and absorption effects with the teXsan software (*27*) from Molecular Structure. Decay corrections were also applied because 5% decay was observed at the end of the data collection process. The data collection parameters are shown in Table I.

Earlier, precession photographs were taken using a Rigaku ultra X18 rotating anode generator (40 kV, 200 mA) with CuKα radiation, $\mu = 15°$ (Fig. 2). Very bright spots corresponding to a spacing of 20 Å can be seen clearly and in addition to these spots, one or two spots (satellite spots) can also be seen on either side of the bright spots. The separation between the bright spots and the satellite spots was found to be around 100 Å. Similar satellite spots were also observed in the precession photograph of the homologous peptide $(Pro\text{-}Pro\text{-}Gly)_{10}$. Data were collected for this unit cell with $c' = 100.3$ Å and identical values for the a and b axes and angle β. Four different crystals were used for data collection (due to the rapid decay of the crystals) up a resolution of 1.8 Å at 5°C. Those diffraction spots corresponding to the cell with a 20 Å repeat (subcell) are the $l = 5n$ reflections, the intensities of which are quite large compared to those of the $l = 5n + i$ reflections (where $i = 1, 2, 3, 4$). With fewer than 50% of these $5n + i$ reflections being observed $[(F_o > \sigma F_o)]$ it was not possible to position the molecule along the c direction unambiguously [similar to the case of $(Pro\text{-}Pro\text{-}Gly)_{10}$ (*22, 23*)]. In $(Pro\text{-}Pro\text{-}Gly)_{10}$ and $(Pro\text{-}Hyp\text{-}Gly)_{10}$ crystals, the triple helical molecules may have a disordered alignment along the c-axis the details to be discussed elsewhere. For this paper, the average crystal structure corresponding to a subcell of $c = 20.06$ Å with an asymmetric unit consisting of 21 amino acids was solved and refined.

Structure Determination and Refinement—The probe model used for molecular replacement search, was based on the Okuyama model for collagen, and consisted of three strands, each strand consisting of 7 amino acid residues and containing as a whole 7 tripeptide units in a repeating period of 20 Å. With a space group of $P2_1$, the translational variable v becomes arbitrary. Since the helix axis can be aligned with the crystallographic c axis, orientation and translation searches had to be performed only for the azimuthal angle μ and variables u and w. All three variables were searched with reflections in the range 8 to 3.5 Å. The software package X-PLOR (*28*) was used for this molecular replacement (MR) search. A unique solution was obtained with the lowest R-factor and used for further refinement. The parameters for hydroxyproline were obtained from the "X-PLOR topology and parameter library for Hetero compounds"—a web site maintained by *Uppsala Universitet* (*29*).

Structure refinement was done using SHELX-97 (*30*). In all, 945 reflections with $(F_o > 2\sigma F_o)$ were used in the refinement process, of which 866 reflections were used for refinement while 79 reflections were used to monitor free R. Constraints for bond length and bond angles, as necessitated in refinements using SHELX-97, were applied, but no torsion angle restraints were applied. The parameters for Hyp (bond length and bond angle constraints) were calculated from the parameters obtained from the web site of *Uppsala Universitet* (*29*). Covalent bonding constraints were applied between the (translation) symmetry related molecules along the helix axis in order to hold the infinite triple helix. The model was refined for several cycles and corrections to the model were carried out based on the $2F_o - F_c$ electron density map with the molecular graphics program Xfit, part of the software package XtalView (*31*). The water molecules were then identified manually based on the $F_o - F_c$ electron density map. During the refinement cycles, both R and free R were monitored. Only 17 reasonable water molecules could be identified by this procedure. The final R factor converged to a value of 0.197 for the 866 reflections used in the refinement procedure, while free R was at 0.276 for 79 reflections. The R factor for the above model was checked using all 945 reflections and was found to be 0.199. The final refinement statistics are shown in Table II. The final structure has 21 amino acids and 17 water molecules and fits well with the $2F_o - F_c$ electron density map. The final structure, including the water molecules, was investigated by the omit map using the following procedure. The occupancy factors of one residue or five water molecules was set to zero, and the $(F_o - F_c)$ electron density map was calculated after 10 cycles of positional refinement. Electron density for the omitted region reappears showing that the omitted residue is correct. This procedure was repeated for all the 21 residues and water molecules in the asymmetric unit. This was done in order to avoid model bias.

TABLE II. **Final refinement statistics for the 20 Å model of $(Pro\text{-}Hyp\text{-}Gly)_{10}$.**

Resolution range	8.0-1.9 Å
Number of reflections used ($F > 2\sigma F$)	945
R factor	19.7%
Free R factor	27.6%
R_{all} factor	19.9%
Root mean square deviations	
Δbonds	0.018 Å
Δangles	3.72°
Average temperature factors (Å^2)	
All atoms	12.85
Peptide atoms	12.22
Solvent	18.22

R factor was based on 866 reflections. Free R factor was based on 79 reflections. R_{all} factor was based on all 945 reflections.

RESULTS AND DISCUSSION

The sub-cell structure of the collagen model peptide $(Pro\text{-}Hyp\text{-}Gly)_{10}$ along with 17 water molecules is shown in Fig. 3. The asymmetric unit consists of 21 residues arranged in three strands of equal length with 7 amino acids in each strand. Each strand has a left-handed polyproline II helical conformation. Three such strands align in parallel and wrap around a common helical axis in order to form the triple helix. Each strand is staggered by one amino acid residue along the helical axis with the average staggering being 2.86 Å. The triple helix has a right-handed helical-conformation following the Okuyama model, similar to that observed in the case of $(Pro\text{-}Pro\text{-}Gly)_{10}$.

The helical parameters (h, θ) for each tripeptide show a broad distribution from one tripeptide unit to another. The helical parameters were derived from the bond lengths, bond angles, and torsional angles by the method of Sugeta and Miyazawa (*32*). The average unit height (h) of a tripeptide unit is about 8.45 Å, and the average twist (θ)

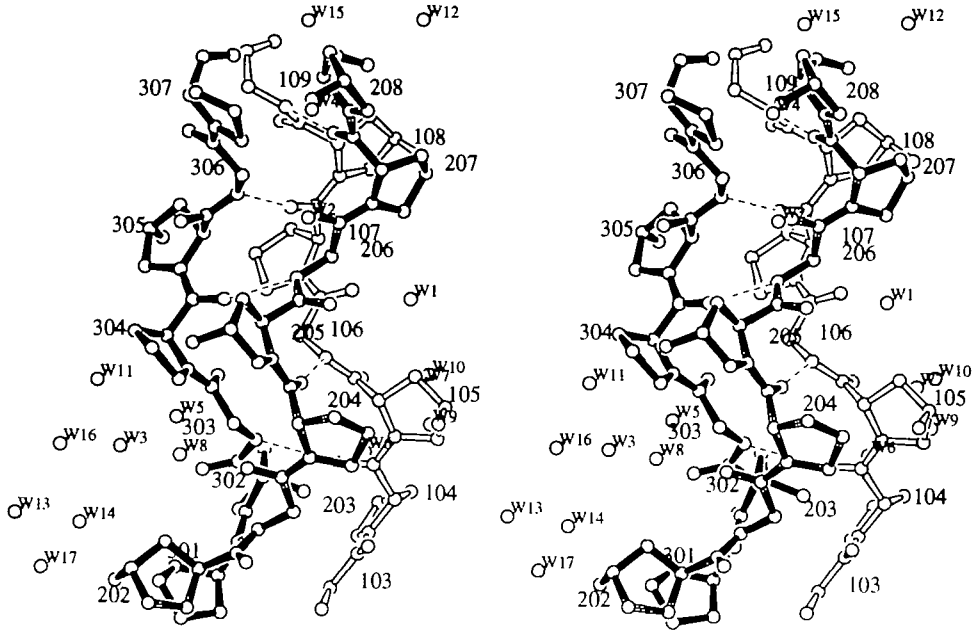

Fig. 3. **Stereo view of the sub-cell structure of the triple strand along with the 17 water molecules.** The inter-chain N···O hydrogen bonds are indicated by dotted lines.

TABLE III. **Helical parameters obtained for various collagen-model peptides based on single crystal analysis in comparison with two models for collagen.**

	Present study	PPG-1[a]	PPG-2[b]	PPG-3[c]	(Gly→Ala) peptide[f]	7/2 model[d]	10/3 model[e]
Unit height, h (Å)	8.45	8.79	8.65	8.75	8.4	8.61	8.58
Unit twist, θ(°)	46.5	50.7	51.4	51.4	60	51.4	36

[a]Nagarajan *et al.* (1998); [b]Kramer *et al.* (1998) structure at 2.0 Å resolution; [c]Kramer *et al.* (1998) structure at 1.7 Å resolution; [d]Okuyama *et al.* (1981); [e]Fraser *et al.* (1979); [f]Bella *et al.* (1994), values include the alanine substitution zone as well as the Pro-Hyp-Gly regions of the structure.

TABLE IV. **Average values of the main chain dihedral angles.** The values are compared with those of PPG-1, PPG-2, PPG-3, Gly→Ala peptide, and the 7/2 and 10/3 models for collagen.

Torsion angle	Present study	PPG-1[a]	PPG-2[b]	PPG-3[c]	7/2 model[d]	10/3 model[e]	(Gly→Ala) peptide[f]
ϕ Pro	−72.7(3.5)	−77.1(7.5)	−73.1(8.8)	−75.0(2.7)	−75.5	−72.1	−72.6(7.6)
ψ Pro	161.1(5.1)	163.8(4.7)	159.7(3.7)	161.4(3.1)	152.0	164.3	163.8(8.8)
ω Pro	179.6(2.1)	179.1(1.3)	178.6(0.6)	177.8(0.7)	−176.8	180.0	179.9(1.8)
ϕ (Y)	−58.4(4.8)	−62.4(7.9)	−58.7(8.1)	−61.2(1.1)	−62.6	−75.0	−59.6(7.3)
ψ (Y)	152.0(6.4)	154.0(8.8)	161.0(12.4)	153.3(2.2)	147.2	155.8	149.8(8.8)
ω (Y)	178.5(2.1)	177.6(3.9)	179.1(1.2)	176.7(2.1)	−172.8	180.0	178.5(1.5)
ϕ Gly	−74.8(5.6)	−76.7(6.6)	−83.7(11.2)	−75.8(2.0)	−70.2	−67.6	−71.9(9.6)
ψ Gly	172.8(3.0)	176.6(7.1)	179.8(5.8)	179.5(3.5)	175.4	151.4	174.1(11.9)
ω Gly	179.2(1.7)	179.0(1.8)	179.6(0.5)	−179.9(0.2)	178.2	180.0	177.3(3.1)

Y = proline in the case of PPG-1, PPG-2, PPG-3, and 7/2 model while Hyp in the case of POG-1, 10/3 model, and Gly→Ala. [a]Nagarajan *et al.* (1998); [b]Kramer *et al.* (1998) structure at 2.0 Å resolution; [c]Kramer *et al.* (1998) structure at 1.7 Å resolution; [d]Okuyama *et al.* (1981); [e]Fraser *et al.* (1979); [f]Bella *et al.* (1994).

per tripeptide unit is 46.5°. These values are closer to those of the 7/2 model than the 10/3 model for collagen (Table III).

Chain Conformation—The average ϕ, ψ, and ω torsion angles obtained for the Pro, Hyp, and Gly have been compared with those reported in previous studies (Table IV). It can be clearly seen that the average values obtained in the present study and those obtained in other model peptides are closer to the 7/2 model than to the 10/3 model.

The geometric values for the upward and downward puckering for the imino acids was given by Momany *et al.* (*33*). A general pattern of down puckering for the Pro at the X position and up puckering for the Hyp at the Y position has been reported for collagen (*14*) and in a two-dimensional NMR investigation of the solution structure of (Pro-Hyp-Gly)$_{10}$ (*34*). In the present study, the Pro at the X-position shows a downward puckering and Hyp at the Y-position shows an upward puckering, which complies well with the general rule. The puckering statistics for the proline and hydroxyproline residues are given in Table V.

Inter-Chain Hydrogen Bonding—All the Gly amide and Pro carbonyl groups take part in inter-chain (N-H···O=C) hydrogen bonds (Fig. 3) that run in the direction of strand 1 to strand 2 to strand 3 to strand 1, following the pattern of model II of Rich and Crick (*12*). In addition to this, the Pro carbonyl group is also at hydrogen bonding distance

TABLE V. **Puckering statistics for the proline and hydroxyproline residues at the X- and Y-positions.**

Proline at X-position	
Residue No.	Deviation of C^{γ} atom[a]
104	(+)0.4497
107	(+)0.3404
204	(+)0.4546
207	(+)0.4409
301	(+)0.5060
304	(+)0.4493
307	(+)0.5119
Hydroxyproline at Y-position	
Residue No.	Deviation of C^{γ} atom[a]
105	(−)0.5004
108	(−)0.4775
202	(−)0.5706
205	(−)0.4349
208	(−)0.4713
302	(−)0.5576
305	(−)0.3484

[a](+) indicates down conformation, (−) indicates up conformation.

TABLE VI. **Inter-strand hydrogen bonding parameters observed in the triple helix.**

N-H···O hydrogen bonds involving N of Gly			
Hydrogen bonds	N···O[a]	H···O[a]	<N-H···O[b]
N(203)···O(301)	2.94	2.17	149.7
N(303)···O(104)	2.92	2.18	144.6
N(106)···O(204)	2.84	2.05	152.7
N(206)···O(304)	2.89	2.10	152.9
N(306)···O(107)	2.78	1.98	153.9
N(109)···O(207)	2.83	2.02	157.8
C^{α}-H···O hydrogen bonds involving C^{α} of Hyp			
Hydrogen bonds	C^{α}···O[a]	H···O[a]	<C^{α}-H···O[b]
C^{α}(202)···O(301)	3.48	2.66	142.2
C^{α}(302)···O(104)	3.28	2.44	142.9
C^{α}(105)···O(204)	3.44	2.62	141.3
C^{α}(205)···O(304)	3.22	2.37	144.7
C^{α}(305)···O(107)	3.14	2.34	138.1
C^{α}(108)···O(207)	3.17	2.36	140.1

[a]Distances in (Å). [b]Angles in (°).

from the C^{α} of Hyp (C^{α}-H···O=C) running in the same direction as in the case of N-H···O=C. This type of hydrogen bonding has been reported in the case of globular proteins (*35, 36*) as well as in the case of collagen model peptides (*22, 23, 37*). The hydrogen bonding parameters are tabulated in Table VI. The hydrogen atoms are fixed geometrically for the amide group of Gly and the C^{α} atom of Hyp, riding over the parent atom. These hydrogen atoms were not part of the refinement procedure.

Crystal Packing—The molecules are packed in a pseudo hexagonal manner as shown in Fig. 4a. This packing is slightly different from that of the pseudohexagonal but parallel packing (*38*) observed in the case of the semicrystalline rat tail tendon. The lateral spacing between the triple helices was found to be around 14 Å.

Role of Water Molecules—Only 17 reasonable water molecules could be included in the average structure of (Pro-Hyp-Gly)$_{10}$, while free R was taken into consideration. The hydrogen bonding interactions involving the triple helical carbonyl oxygens and water molecules are tabulated in Table VII. Of the 14 carbonyl oxygens (7 each in hydroxyproline and glycine) that can take part in hydrogen bonding interactions, only 8 (4 each in hydroxyproline and glycine) take part in hydrogen bonding interactions with water molecules. In addition to the carbonyl oxygens, the hydroxyl group of the Hyp are also in a position to form hydrogen bonding interactions with water molecules. Of the 7 hydroxyl groups, only 3 form hydrogen bonding interactions with water molecules. These water molecules form the first cylinder of hydration around the triple helix. These water molecules are also involved in hydrogen bonding interactions with other water molecules forming the second cylinder of hydration around the triple helix. There are 13 such interactions involving water molecules and these interactions are tabulated in Table VII. The water molecules form a network that gives rise to certain kinds of intra-chain and inter-chain water bridges. These water molecules mediate the interaction between the triple helices, since a direct interaction between the triple helices is not possible owing to the distance between them. The hydrogen bonding network, the intra-chain and inter-chain interactions mediated by the water molecules (water bridges), and the interaction between the triple helices are shown in Fig. 5.

A separate refinement of the structure without taking free R into consideration was attempted against all 945 reflections in the resolution range 8.0 to 1.9 Å. In all, 33 reasonable water molecules were identified and the R factor converged to a value of 0.163. The features of the triple helical peptide molecule remained the same, but there was a large variation in the number of water molecules and their positions. In this case, all the carbonyl oxygens were hydrogen bonded to water molecules, as is generally expected in the case of biological molecules. The water molecules also formed an extensive water bridge network, forming different kinds of intra-chain and inter-chain water bridges. Even though the hydrogen bonding network and the positions of the water molecules were in accordance with those of similar structures, the structure was not considered reasonable because of the method of structure refinement. When all 33 water molecules obtained by this method were assigned and refined while monitoring free R, there was a considerable decrease in the R factor, but free R increased to a very large value. This implies that the excess number of water molecules assigned only increased the number of parameters to fill the electron density, thereby decreasing the credibility of the structure refinement.

The smaller number of water molecules and water bridges reported in this structure may be against the prevailing general idea that the interaction of hydroxyproline with water molecules increases the stability of the triple helix (*22, 24, 39*). But recently it has been reported (*40*) that water bridges do not contribute significantly to collagen stability, but rather that collagen stability relies more on inductive effects. A highly electronegative element, fluorine, was introduced in place of the hydroxyl group of hydroxyproline forming 4(R)-fluoroproline (Flp). Flp was chosen because fluorine is the most electronegative element but it does not form hydrogen bonds (*40*). Thermal stability tests showed that the T_m values of the three triple helical peptides are in the order (Pro-Flp-Gly)$_{10}$ > (Pro-Hyp-Gly)$_{10}$ > (Pro-Pro-Gly)$_{10}$, and also that the thermal stability of (Pro-Flp-Gly)$_{10}$ far exceeds that of any known collagen model peptide of similar size. Since, there can be

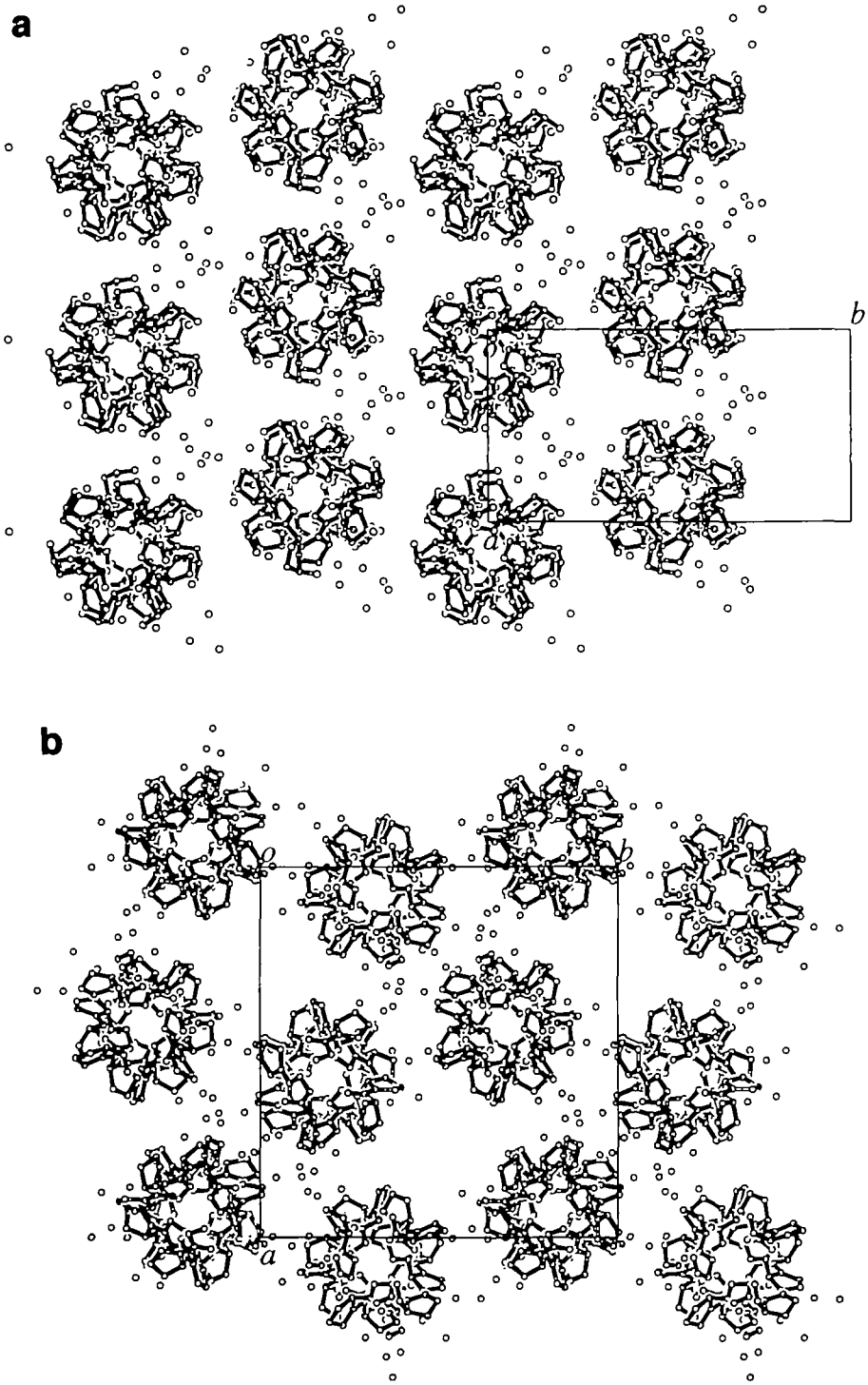

Fig. 4. **(a) Hexagonal packing of the (Pro-Hyp-Gly) sequence repeat viewed along the helical axis. (b) The packing of PPG-1, the packing can be considered to have two general regions, one triangular and the other square.** The triple helices shown by filled bonds are directed upward from the paper, while the open bonded triple helices are directed downward from the paper.

no water bridges, it can be concluded that the large inductive effect of the fluorine attached to the C^{γ} position enhances the collagen stability by favoring the *trans* conformation of the prolyl peptide bond. Based on this result, it was concluded that the inductive effects of the electronegative group attached to the C^{γ} atom of the proline rather than the water bridges enhance the stability of collagen. Therefore, it can be concluded that the triple helix is highly stable even in the absence of large number of water molecules attached to hydroxyproline, which is consistent with the observation about the water molecules around Hyp.

Comparison with Other Structural Studies on Collagen Model Peptides—The structures of the collagen model peptides, PPG-1, PPG-2, PPG-3 and the (Gly→Ala) peptide, reveal that three identical chains in the polyproline II conformation align in parallel and wrap around a common helical axis with a stagger of one residue between adjacent chains. The three chains are held together by hydrogen bonding interactions between the N-H groups of glycine

and the C=O groups of the proline residues in the X position of the neighbouring chain. Also, the ϕ and ψ values reported for these collagen model peptides (Table IV) are closer to the 7/2 model than to the 10/3 model for collagen.

The overall structures of all these collagen model peptides are very similar to one another, but there are small differences in some structural details. Table III gives the values of the helical parameters obtained in different cases. It can be seen that these values, even though closer to the 7/2 model, vary to a certain extent. The value of 60° for the unit twist obtained in the case of the (Gly→Ala) peptide includes the alanine substituted zone as well as the Pro-Hyp-Gly regions of the structure, and hence the large difference in the value compared to the 7/2 model.

A general pattern of Pro-down puckering (X-position) and Hyp-up puckering (Y-position) is reported for collagen (*14*). In the present study, PPG-3 and the (Gly→Ala) peptide the proline and hydroxyprolines comply with this generalization. In the case of PPG-1, the proline at the X position shows a greater tendency to down puckering, but the proline at the Y position shows no clear pattern of up puckering. In the case of PPG-2, the prolines at the X position show down puckering, while in the case of Y-position prolines, one shows a down conformation while all the others show an up conformation. From this observation, it can be stated that with Pro at the X-position and Hyp at the Y position, the puckering of the imino acids are in the down and up conformations, respectively. But with Pro at both the X-position and Y-position, there is considerable flexibility in the puckering of the proline rings at the Y-position.

The packing structure observed in the present study and in the (Gly→Ala) peptide are very similar to each other. Both structures show a pseudo hexagonal packing with similar lateral spacing of around 14 Å. This packing structure and the value of the lateral spacing are similar to those observed in the case of native collagen (*38*). The packing structure observed in the present study is entirely different from those found in the crystal structures of peptides with the (Pro-Pro-Gly) sequence, in which every triple helix is five coordinated and two different kinds of clusters, square and triangular cluster, appeared (Fig. 4b). A layer structure, with a layer thickness of around 86 Å, is observed in the (Gly→Ala) peptide, while no such layer structure was observed in the present study.

The number of water molecules assigned in the present study and in PPG-1 are nearly the same and fewer compared to the extensive water network seen in the case of PPG-2, PPG-3, and the (Gly→Ala) peptide. This is mainly due to the difference in the refinement procedure, monitoring free R is a common procedure in the structure analysis of macromolecules, and this method was used in the case of

TABLE VII. **Hydrogen bonds involving water molecules.**

Hydrogen bonds	O···Wat/Å	<C-O···Wat/°
Hyp_105 O···W(7)	2.54	132.7
Gly_109 O···W(15)	2.58	155.9
Gly_203 O···W(8)	2.71	148.7
Hyp_205 O···W(1)	2.85	150.1
Gly_206 O···W(2)	2.89	135.9
Hyp_208 O···W(12)	3.02	135.9
Hyp_302 O···W(3)	3.33	138.7
Gly_303 O···W(5)	2.89	115.9
Hyp_202 OD1···W(14)	2.75	106.4
Hyp_208 OD1···W(4)	3.11	118.7
Hyp_302 OD1···W(6)	2.93	122.8
W(6)···W(9)	2.59	
W(9)···W(7)	3.11	
W(9)···W(10)	2.72	
W(3)···W(14)	2.96	
W(3)···W(11)	2.84	
W(11)···W(16)	2.90	
W(16)···W(13)	3.24	
W(13)···W(14)	2.64	
W(13)···W(17)	2.97	
W(5)···W(11)	3.38	
W(9)···W(11)#1[a]	2.54	
W(12)···W(14)#1[a]	3.26	
W(13)···W(15)#1[a]	2.74	

[a]Symmetry related water molecule.

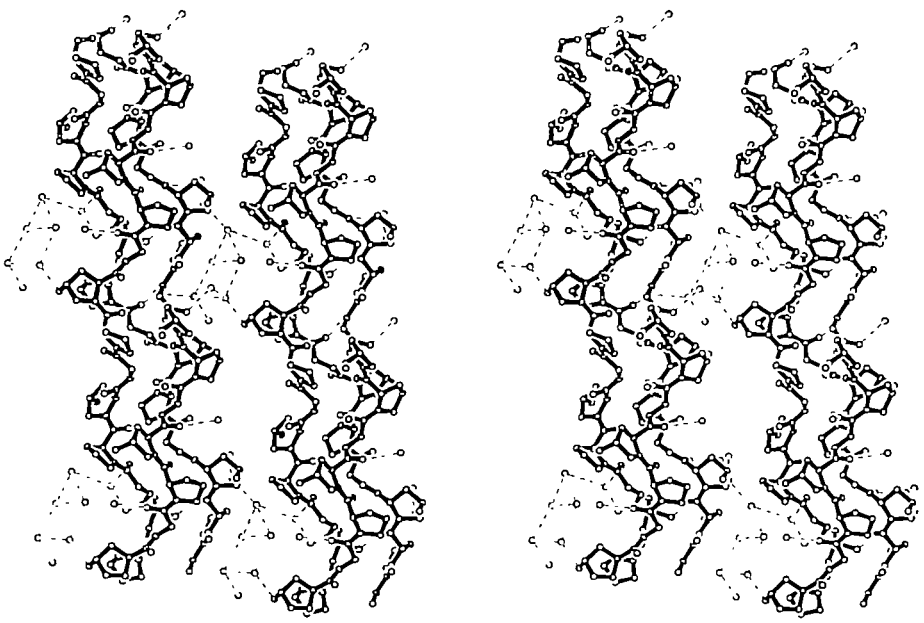

Fig. 5. **Stereo view of the hydrogen bonding network.** The open bonded part of the figure represents an asymmetric unit of the sub-cell structure, similar to that shown in Fig. 3. Two symmetry related triple helices and their hydrogen bonding interactions (dotted lines) are shown.

PPG-1 and the present study. However, free *R* was not monitored in the structure analyses of PPG-2, PPG-3, or the (Gly→Ala) peptide. Even in the present analysis we have seen that a similar kind of extensive network for water is observed when free *R* is not taken into account. Due to the differences in the refinement procedures, one must be careful when comparing the *R* factor of these analyses, the number of water molecules, and the hydrogen bonding schemes involving water molecules.

Comparing the present study with PPG-1, we find that nearly equal numbers of water molecules have been assigned (Ref. *17* in the present study and Ref. *15* in PPG-1). With the proline residue at the Y position being replaced by hydroxyproline, one expects a large number of water molecules to interact with the hydroxyl groups of the hydroxyprolines, thereby forming an extensive water bridge network. But as we have seen, only 3 out of the 7 hydroxyl groups take part in hydrogen bonding with water molecules. This result indicates that water molecules play a similar kind of role in sequences both with and without hydroxyproline, and it can be stated that the extra stability of the hydroxyproline-containing triple helices is due to inductive effects of hydroxyl group attached to the C^{γ} atom in hydroxyproline, rather than the water bridges.

Conclusions—The high resolution structure of a collagen model peptide with the sequence (Pro-Hyp-Gly) is described in detail. Since the molecules in the crystal form a polymer-like organization, the structure is a high resolution averaged model of the triple helical structure with a Pro-Hyp-Gly sequence.

The peptide structure is very similar to that obtained in a previous study involving the sequence (Pro-Pro-Gly), having a structure close to the 7/2 model for collagen (Okuyama model). The three chains are held together by hydrogen bonding interactions between the amide group of Gly and the carbonyl group of Pro. The puckering of the proline rings shows a clear pattern with the Pro-down and Hyp-up conformation, complying well with the general pattern for collagen. The packing of the triple helices in the crystal is pseudo-hexagonal and antiparallel, similar to that observed in the case of the (Gly→Ala) peptide, but entirely different from the packing observed in the case of PPG-1. The close similarity in the molecular structures of PPG-1 and the present structure confirms that the presence of hydroxyproline does not directly affect the molecular structure in an imino-acid rich region. Also, the nearly equal numbers of water molecules in structures with and without hydroxyproline clearly indicate that the extra-stability associated with sequences containing hydroxyproline is not due to the extensive water bridges, but may be due to inductive effects.

REFERENCES

1. Van der rest, M. and Garrone, R. (1991) Collagen family of proteins. *FASEB J.* **5**, 2814-2823
2. Fields, G.B. and Prockop, D.J. (1996) Perspectives on the synthesis and applications of triple-helical, collagen-model peptides. *Biopolymers* **40**, 345-357
3. Acton, S., Resnick, D., Freeman, M., Ekkel, Y., Ashkenas, J., and Krieger, M. (1993) The collagenous domains of macrophage scavenger receptors and complement component C1q mediate their similar, but not identical, binding specificities for polyanionic ligands. *J. Biol. Chem.* **268**, 3530-3537
4. Doi, T., Higashino, K., Kurihara, Y., Wada, Y., Miyazaki, T., Nakamura, H., Uesugi, S., Imanishi, T., Kawabe, Y., and Itakura, H. (1993) Charged collagen structure mediate the recognition of negatively charged macromolecules by macrophage scavenger receptors. *J. Biol. Chem.* **268**, 2126-2133
5. Hoppe, H.J. and Reid, K.B.M. (1994) Collectins: soluble proteins containing collagenous regions and lectin domains and their roles in innate immunity. *Protein Sci.* **3**, 1143-1158
6. Engel, J., Chen, H., Prockop, D.J., and Klump, H. (1977) The triple helix-coil conversion in aqueous and nonaqueous solvents. Comparison of the thermodynamic parameters and the binding of water to (L-Pro-L-Pro-Gly)$_n$ and (L-Pro-L-Hyp-Gly)$_n$. *Biopolymers* **16**, 601-622
7. Sakakibara, S., Inouye, K., Shudo, K., Kishida, Y., Kobayashi, Y., and Prockop, D.J. (1973) Synthesis of (Pro-Hyp-Gly)$_n$ of defined molecular weights. Evidence for the stabilization of collagen triple helix by hydroxyproline. *Biochim. Biophys. Acta* **303**, 198-202
8. Rosenbloom, J., Harsch, M., and Jimenez, S. (1973) Hydroxyproline content determines the denaturation temperature of chick tendon collagen. *Arch. Biochem. Biophys.* **158**, 478-484
9. Nemethy, G. and Scheraga, H.A. (1986) Stabilization of collagen fibrils by hydroxyproline. *Biochemistry* **25**, 3184-3188
10. Rich, A. and Crick, F.H.C. (1955) The structure of collagen. *Nature* **176**, 915-916
11. Ramachandran, G.N. and Kartha, G. (1955) Structure of collagen. *Nature* **176**, 593-595
12. Rich, A. and Crick, F.H.C. (1961) The molecular structure of collagen. *J. Mol. Biol.* **3**, 483-506
13. Yonath, A. and Traub, W. (1969) Polymers of tripeptides as collagen models. *J. Mol. Biol.* **43**, 461-477
14. Fraser, R.D.B., MacRae, T.P., and Suzuki, E. (1979) Chain conformation in the collagen molecule. *J. Mol. Biol.* **129**, 463-481
15. Sakakibara, S., Kishida, Y., Kikucha, Y., Sakai, R., and Kakiuchi, K. (1968) Synthesis of poly-(L-prolyl-L-prolylglycyl) of defined molecular weights. *Bull. Chem. Soc. Jpn.* **41**, 1273
16. Sakakibara, S., Kishida, Y., Okuyama, K., Ashida, T., and Kakudo, M. (1972) Single crystals of (Pro-Pro-Gly)$_{10}$, a synthetic polypeptide model of collagen. *J. Mol. Biol.* **65**, 371-373
17. Okuyama, K., Tanaka, N., Ashida, T., and Kakudo, M. (1976) Structure analysis of a collagen model polypeptide, (Pro-Pro-Gly)$_{10}$. *Bull. Chem. Soc. Jpn.* **49**, 1805-1810
18. Okuyama, K., Okuyama, K., Arnott, S., Takayanagi, M., and Kakudo, M. (1981) Crystal and molecular structure of a collagen-like polypeptide (Pro-Pro-Gly)$_{10}$. *J. Mol. Biol.* **152**, 427-443
19. Smith, P.J.C. and Arnott, S. (1978) LALS: A linked-atom least-squares reciprocal-space refinement system incorporating stereochemical restraints to supplement sparse diffraction data. *Acta Cryst.* **A34**, 3-11
20. Okuyama, K., Tanaka, N., Ashida, T., and Kakudo, M. (1977) A new structural model of collagen. *Polym. J.* **9**, 341-343
21. Bella, J., Eaton, M., Brodsky, B., and Berman, H.M. (1994) Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science* **266**, 75-81
22. Nagarajan, V., Kamitori, S., and Okuyama, K. (1998) Crystal structure analysis of collagen model peptide (Pro-Pro-Gly)$_{10}$. *J. Biochem.* **124**, 1117-1123
23. Kramer, R.Z., Vitagliano, L., Bella, J., Berisio, R., Mazzarella, L., Brodsky, B., Zagari, A., and Berman, H.M. (1998) X-ray crystallographic determination of a collagen-like peptide with the repeating sequence (Pro-Pro-Gly). *J. Mol. Biol.* **280**, 623-638
24. Miller, M.H. and Scheraga, H.A. (1976) Calculation of the structures of collagen models. Role of interchain interactions in determining the triple-helical coiled-coil conformation. I. Poly(glycyl-prolyl-prolyl). *J. Polym. Sci. Symp.* **54**, 171-200
25. Nemethy, G., Gibson, K.D., Palmer, K.A., Yoon, C.N., Paerlini, G., Zagari, A., Rumsey, S., and Scheraga, H.A. (1992) Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with applications to proline-containing peptides. *J. Phys. Chem.* **96**, 6472

26. Okuyama, K., Nagarajan, V., Kamitori, S., and Noguchi, K. (1998) Crystallization of (Pro-Hyp-Gly)$_{10}$ and its triple-helical structure deduced from cylindrical Patterson map. *Chem. Lett.* **5**, 385-386
27. teXsan (1993) *Single Crystal Structure Analysis Software Version 1.6*, Molecular Structure Corporation, The Woodlands, TX, 773
28. Brunger, A.T. (1992) A system for X-ray crystallography and NMR, *X-PLOR Version 3.1*. Yale University Press, New Haven and London
29. Kleywegt, G.J. (1997) HIC-UP: dealing with hetero compounds. *Protein Data Bank Newslett.* **82**, 6-7
30. Sheldrick, G.M. and Schneider, T.R. (1997) SHELXL: High resolution refinement in *Methods in Enzymology* (Sweet, R.M. and Cater Jr, C.W., eds.) Vol. 277, pp. 319-343, Academic Press, Orlando, Florida
31. Duncan, E.M. (1993) *Practical Protein Crystallography*, Academic Press, London
32. Sugeta, H. and Miyazawa, T. (1967) General method for calculating helical parameters of polymer chains from bond lengths, bond angles and internal-rotation angles. *Biopolymers* **5**, 673-678
33. Momany, F.A., McGuire, R.R., Burgess, A.W., and Scheraga, H.A. (1975) Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions and intrinsic torsional potentials for the naturally occurring amino acids. *J. Phys. Chem.* **79**, 2361-2381
34. Li, M-H., Fan, P., Brodsky, B., and Baum, J. (1993) Two-dimensional NMR assignments and conformation of (Pro-Hyp-Gly)$_{10}$ and a designed collagen triple-helical peptide. *Biochemistry* **32**, 7377-7387
35. Derewenda, Z.S., Lee, L., and Derewenda, V. (1995) The occurrence of C-H···O hydrogen bond in proteins. *J. Mol. Biol.* **252**, 248-262
36. Felcy Fabiola, G., Krishnaswamy, S., Nagarajan, V., and Pattabhi, V. (1997) C-H···O hydrogen bonds in β-sheets. *Acta Cryst.* **D53**, 316-320
37. Bella, J. and Berman, H.M. (1996) Crystallographic evidence for C$^{\alpha}$-H···O=C hydrogen bonds in a collagen triple helix. *J. Mol. Biol.* **264**, 732-742
38. Fraser, R.D.B., MacRae, T.P., Miller, A., and Suzuki, E. (1983) Molecular conformation and packing in collagen fibrils. *J. Mol. Biol.* **167**, 497-521
39. Bella, J., Brodsky, B., and Berman, H.M. (1995) Hydration structure of a collagen peptide. *Structure* **3**, 893-906
40. Holmgren, S.K., Taylor, K.M., Bretscher, L.E., and Raines, R.T. (1998) Code of collagen's stability deciphered. *Nature* **392**, 666-667