J. Biochem. **138**, 135–144 (2005)
DOI: 10.1093/jb/mvi108

Repetitive Interactions Observed in the Crystal Structure of a Collagen-Model Peptide, $[(Pro\text{-}Pro\text{-}Gly)_9]_3$

Chizuru Hongo[1,*], Keiichi Noguchi[1], Kenji Okuyama[1,*,†], Yuji Tanaka[2] and Norikazu Nishino[2]

[1]Faculty of Technology, Tokyo University of Agriculture & Technology, Koganei, Tokyo 184-8588; and [2]Faculty of Engineering, Kyushu Institute of Technology, Kitakyushu, Fukuoka 804-8500

Received March 7, 2005; accepted April 30, 2005

The crystal structure of a collagen-model peptide $[(Pro\text{-}Pro\text{-}Gly)_9]_3$ has been determined at 1.33 Å resolution. Diffraction data were collected at 100 K using synchrotron radiation, which led to the first structural study of $[(Pro\text{-}Pro\text{-}Gly)_n]_3$ under cryogenic conditions. The crystals belong to the $P2_1$ space group with cell parameters of a = 25.95, b = 26.56, c = 80.14 Å and β = 90.0°. The overall molecular conformation was consistent with the left-handed 7/2-helical model with an axial repeat of 20 Å for native collagen. A total of 332 water molecules were found in an asymmetric unit. Proline residues in adjacent triple-helices exhibited three types of hydrophobic interactions. Furthermore, three types of hydrogen-bonding networks mediated by water molecules were observed between adjacent triple-helices. These hydrophobic interactions and hydrogen-bonding networks occurred at intervals of 20 Å along the c-axis based on the previous sub-cell structures $[(Pro\text{-}Pro\text{-}Gly)_n]_3$ (n = 9, 10), which were also seen in the full-cell structure of $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$. Five proline residues at the Y position in the X-Y-Gly triplet were found in a down-puckering conformation, this being inconsistent with the recently proposed propensity-based hypothesis. These proline residues were forced to adopt opposing puckering because of the prevailing hydrophobic interaction between triple-helices compared with the Pro:Pro stacking interaction within a triple-helix.

Key words: collagen, crystal structure, hydrogen-bonding network, hydrophobic interaction, triple-helix.

Abbreviations: Hyp, 4(*R*)-hydroxyproline; PPG, (Pro-Pro-Gly); POG, (Pro-Hyp-Gly).

Collagen is the major fibrous element responsible for the structural integrity of skin, bone, tendons and teeth, and is abundant in mammals. Furthermore, collagen functions as a scaffold for cell adhesion as extracelluar matrix of various organs. Its amino acid sequence is usually assumed to be $(Gly\text{-}X\text{-}Y)_n$, because of strict sequence constraints with a glycine residue (Gly) at every third position. Here, X and Y positions are frequently occupied by a proline residue (Pro) and a 4(*R*)-hydroxyproline (Hyp), respectively. Due to the above features of its amino acid sequence, the collagen molecule has a triple-helical structure in which the smallest amino acid residue, Gly, is located in the central region of the triple-helix, while the amino acids at the X and Y positions protrude from the helix. Triple-helical molecules further assemble and form a micro-fibril. When they assemble, each molecule is axially staggered by 670 Å along the fiber axis.

A structural study of native collagen at atomic resolution is extremely difficult because structural information from its fiber diffraction pattern is very limited. Two molecular models with different helical symmetries and axial repeats can explain the diffraction pattern of native collagen: the Rich and Crick model, in which three strands form a left-handed 10/3-helix with an axial repeat of 28.6 Å (*1*), was proposed based on the fiber diffraction pattern of native collagen; on the other hand, in the Okuyama model, three strands form a left-handed 7/2-helix with an axial repeat of 20.0 Å (*2*). This triple-helical structure was found in single crystals of $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$ (hereafter called PPG10) (*3–5*) and could also explain the fiber diffraction pattern of native collagen (*2*). In earlier studies, the structure of PPG10 was determined using the Linked-Atom Least-Squares method (*6*) for fibrous polymers, assuming the peptide molecule to be infinite. In this case, a helical asymmetric unit consists of one triplet, Pro-Pro-Gly. Nine dihedral angles in a main chain conformation were taken as variables, while bond lengths and angles were fixed at the standard values (*5*).

Structure analyses of PPG10 have been performed several times during the last decade by using the method for protein crystallography. Precession and oscillation photographs of PPG10 (*5*, *7*) clearly showed strong diffraction spots on the layer lines corresponding to a 20 Å axial repeat along the c-axis. Weak satellite diffraction spots were also observed on both sides of these strong spots. The axial repeat corresponding to the latter spots was assumed to be 100 Å for many years (*5*, *7*, *8*). Similarly,

*Present address: Department of Macromolecular Science, Graduate School of Science, Osaka University, 1-1 Machikaneyama, Toyonaka, Osaka 560-0043.
†To whom correspondence should be addressed at the present address. Tel: +81-6-6850-5455, Fax: +81-6-6850-5455, E-mail: okuyamak@chem.sci.osaka-u.ac.jp

strong spots with a 20 Å axial repeat and weak satellite spots with a longer axial repeat have also been observed in the diffraction patterns of $[(\text{Pro-Hyp-Gly})_{10}]_3$ (hereafter called POG10) peptide (*9–11*). In these structural analyses, only reflections on the layer lines corresponding to the 20 Å axial repeat were used, assuming that the molecule has an infinite helical conformation (*7–12*). Later, PPG10 crystals grown under microgravity in the space shuttle revealed that the true axial repeat along the *c*-axis is not 100 Å but 182 Å (*13*). Based on this axial repeat, the full-cell structure of PPG10 was determined at 1.3 Å resolution (*14*).

The torsion angles (ϕ) of Pro residues are closely related to the side chain conformation (χ_1) according to the results of statistical analysis of Pro residues (*15*), and the sub-cell (*12*) and full-cell (*14*) structures of PPG10. To explain the stabilization of the triple helical structure by Hyp at the Y position and destabilization at the X position, a propensity-based hypothesis was put forward by Zagari's group (*12*). According to this hypothesis, imino acids in the X-Y-Gly sequence must adopt "down" and "up" puckering at the X and Y positions, respectively. Here "up" and "down" puckering correspond to negative and positive values of χ_1 torsion angles, respectively. This hypothesis was based on the sub-cell structure of PPG10 at room temperature (RT). However, in the sub-cell structure of $[(\text{Pro-Pro-Gly})_9]_3$ (hereafter called PPG9) at RT, one opposing puckering was found at seven Y positions (*16*). Furthermore, for $[(\text{Pro-Hyp-Gly})_n]_3$ (hereafter called POGn), three opposing puckerings out of seven were also observed at the X position of POG10 and POG11 at 100 K. On the other hand, one opposing puckering was observed in the POG11 structure at RT (*11*). In order to obtain detailed information about proline ring puckering, it is necessary to determine the full-cell structure of PPG9 at high resolution.

Although we could not obtain suitable single PPG10 crystals to analyze the full-cell structure, we succeeded in the crystallization of PPG9 suitable for full-cell analysis at high resolution. In this study, we analyzed the full-cell structure of PPG9 at 1.33 Å resolution using synchrotron radiation under cryogenic conditions. Based on the obtained structure, Pro ring puckering, hydration networks of water molecules and the hydrophobic interaction between triple helices will be discussed.

MATERIALS AND METHODS

Crystallization Experiments—Peptide PPG9 was synthesized by the solid phase method. The procedure was presented in detail elsewhere (*16*). Single crystals were grown within a week by the hanging drop-vapor diffusion method at 4°C. The peptide solution consisted of 4.5 mg/ml of PPG9, 5% (v/v) acetic acid and 0.5% (w/v) sodium azide. As a reservoir solution, 1 ml of 20% polyethylene glycol (PEG) 400 was used. A mixture of 3 μl of the peptide solution and 3 μl of the reservoir solution was used as one drop. PPG9 crystals were rectangular, measuring approximately $0.2 \times 0.2 \times 0.3$ mm^3.

Data Collection—X-ray diffraction experiments were performed using the synchrotron radiation source at the BL40B2 beam line of SPring-8, Hyogo, Japan. Diffraction data were obtained with an ADSC Quantum 4R CCD detector system with a total oscillation range of 180° and an oscillation angle of 1.0° at 100 K. The crystal was soaked for several seconds in a mixture of 20% (w/v) MPD and 10% (w/v) PEG 400 as a cryoprotectant solution, and then flash-cooled at 100 K under a vaporized nitrogen stream. The diffraction data included reflections with very strong intensity of mainly $l = 4n$ (n: integer) and with very weak intensity of mainly $l = 4n + m$ (m: 1, 2, 3). In order to obtain very strong reflections together with weak reflections, two series of data collection were performed under different conditions. The data collection parameters are presented in Table 1.

Structure Determination and Refinement—The initial examined structure was elucidated by the molecular replacement method. A probe model was built from a previously obtained structure (*16*). The previous analysis indicated four triple-helices in a unit cell. Since statistical analysis of the intensity data suggested the presence of *mmm* symmetry, structural analysis was performed assuming a space group of $P2_12_12_1$ based on the experience in the previous sub-cell structure (*16*). In these space groups, an asymmetric unit contains only one triple-helix. The appropriate position on the *a–b* plane was identified using the *R*-search method in X-PLOR (*17*), and very similar values to the previous ones were obtained (*16*). Using a similar strategy for the $P2_1$ space group cited below, 14 models were investigated thoroughly using SHELX-97 (*18*). However, none gave an *R*-factor of less than 28% even after the introduction of water molecules. As stated in the discussion section, these space groups were ruled out and $P2_1$ was used for the subsequent analysis.

In the $P2_1$ space group, one asymmetric unit contains two parallel molecules. Similar to those in the analysis of the sub-cell structure, the positions of the two molecules on the *a–b* plane were clearly determined. However, the position along the *c*-direction and the rotation angle around the molecular axis could not be determined individually. Therefore, 392 (=14 × 28) models were generated by considering the 7/2-helical symmetry with a 20 Å axial repeat. Here, 14 and 28 arose from the 7/2-helical symmetry of the molecule. Although we can determine the rotation angle around the helical axis, there are seven candidate positions within a helical asymmetric unit with a 20 Å axial repeat. Therefore, there are 28 candidate positions in a unit cell with an axial repeat of c = 80 Å. The space group $P2_1$ decreased the number of candidate positions to one half for one of the two molecules in an asymmetric unit. In addition, the direction of crystallographic 2_1-symmetry should be checked since there is no clear-cut evidence. We therefore analyzed both cases, the 2_1 axis parallel to the longer axis (26.56 Å) and the shorter axis (25.95 Å). Seven hundred eighty-four models were refined using X-PLOR and SHELX-97, and 24,205 unique reflections with $[F_o > \sigma((F_o)]$ in the resolution range of 8.0 to 1.33 Å were used for structure refinement, while 1,185 reflections were used for the R_{free} monitoring with SHELX-97. Restraining conditions for bond lengths and bond angles were applied during refinement, while no torsion angle restraint was applied. Water molecules were introduced using an automated water-diving program, SHELXWAT in the SHELX-97 package (*18*). Models with low *R*-factors had triple-helices arranged in

Table 1. **Data-collection parameters and refinement statistics.**

A. Data collection		
Data-collection device	ADSC Quantum 4R CCD at SPring-8 BL40 beam line	
Temperature (K)	100	
Camera length (mm)	85	135
Resolution limit (Å)	1.33	1.73
(Last shell)	1.38–1.33	1.79–1.73
No. of unique reflections	24,904	11,647
R_{merge} (last shell) (%)	6.8 (26.9)	6.6 (29.2)
Completeness (last shell) (%)	98.8 (99.8)	99.7 (99.9)
R_{merge} of two data sets (%)	6.7	
No. of reflections observed in both data sets [$F > \sigma(F)$]	10,402	
No. of reflections after merging [$F > \sigma(F)$]	24,336	
Space group	$P2_1$	
Unit cell dimensions		
a (Å)	25.95	
b (Å)	26.56	
c (Å)	80.14	
β (°)	90.0	
B. Refinement statistics		
Resolution	8.0–1.33	
No. of reflections	24,205	
R-factor	18.16	
R_{free}-factor	22.87	
No. of molecules in the asymmetric unit	2	
No. of peptide non-hydrogen atoms	958	
No. of water molecules	332	
Root-mean-squares deviation		
1–2 dist. (Å)	0.009	
1–3 dist. (Å)	0.023	
Flat planes (Å)	0.027	
Mean B-factors		
Peptide nonhydrogen atoms (Å^2)	16.97	
Water molecules (Å^2)	31.25	

a layer of approximately molecular-length thickness. This arrangement was consistent with the electron density map obtained using the infinite triple-helices located at appropriate positions on the a–b plane. Therefore, at this stage of the analysis, most models were ruled out and 24 models with the above molecular arrangement were refined further. The positions of water molecules obtained with SHELXWAT were modified based on $2F_o$–F_c and F_o–F_c electron density maps. Anisotropic treatment of the atomic displacement parameters was used for non-hydrogen atoms of peptides, while isotropic treatment was used for water oxygen atoms. Although a few models showed almost identical R-factors and R_{free}-factors, the final structure was determined in terms of $2F_o$–F_c and F_o–F_c electron density maps. Four Gly residues at the C-terminals, Gly127, Gly227, Gly427 and Gly527, could not be identified in the F_o–F_c electron density map. The atomic coordinates have been deposited in the Protein Data Bank, code 2CUO.

RESULTS AND DISCUSSION

Crystal Data—The diffraction data of PPG9 indicated *mmm* Laue symmetry with cell dimensions of a = 25.95, b = 26.56 and c = 80.14 Å. Based on the Laue symmetry and the sub-cell structure of this peptide (*16*), structural analysis was first performed assuming a space group of $P2_12_12_1$. As none of the refined structures gave R-factors less than 0.28, the crystal symmetry was lowered from orthorhombic to monoclinic $P2_1$. In earlier studies on PPG9 and PPG10, sub-cell structures were analyzed using only reflections on the layer lines corresponding to the axial repeat of 20 Å (*7*, *8*, *12*, *16*). In these analyses, molecules were assumed to be infinite, and hence there should have been no structural difference between PPG9 and PPG10. In fact, five sub-cell structures of PPG10 and PPG9 reported by different groups were essentially the same (*7*, *8*, *12*, *16*), that is, these structures had a space group of $P2_12_12_1$ with lattice dimensions of $a \approx 26.9$, $b \approx 26.4$, c (fiber period) ≈ 20.2 Å (Fig. 1a). Four triple-helices run through the unit cell and are related to each other by their crystallographic 2_1-symmetry. Although the number of water molecules introduced into cells differed depending on the quality of the diffraction data, the molecular conformations obtained in these analyses were essentially identical.

The full-cell structure of PPG10 with a $P2_12_12_1$ space group contains nine sub-cells in the c-direction (Fig. 1b). As the 2_1-symmetries of the full-cell structure (solid black lines) are not in conflict with those of the sub-cell structure (solid gray lines), the structural relationship between the adjacent triple-helices found in the sub-cell

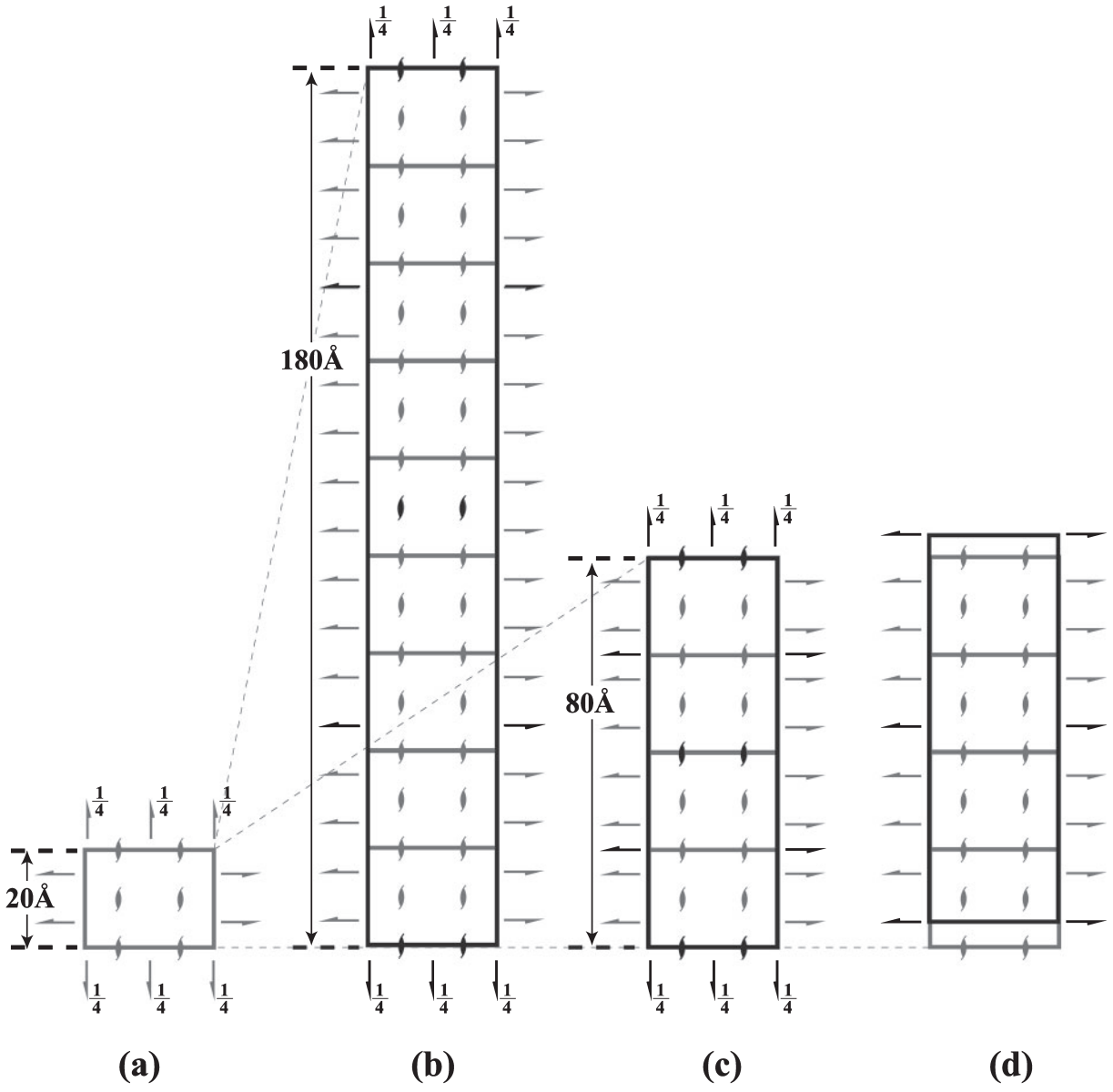

Fig. 1. **Space groups of PPGn crystals.** (a) The cell and symmetry elements of the sub-cell structure ($P2_12_12_1$) are shown as solid gray lines. (b) The cell and symmetry elements of the full-cell structure ($P2_12_12_1$) for PPG10 are shown as solid black lines, together with those of the sub-cell structure. (c) The cell and symmetry elements of the full-cell structure ($P2_12_12_1$) for PPG9 are shown as solid black lines, together with those of the sub-cell structure. The symmetry elements of the sub-cell structure (gray) are in conflict with those of the full-cell structure (black) in many places. (d) The cell and symmetry elements of the full-cell structure ($P2_1$) for PPG9 are shown as solid black lines, together with those of the sub-cell structure. The symmetry elements of the former do not conflict with those of the latter.

structure can be preserved in the full-cell structure. On the other hand, since the *c*-value of the full-cell structure of PPG9 is four times greater than that of the sub-cell, one of the 2_1-symmetries perpendicular to the *c*-axis in the full-cell (black) is in conflict with that of the sub-cell (gray) (Fig. 1c). Therefore, if the space group of the full-cell structure of PPG9 were $P2_12_12_1$, the structural relationship between triple-helices found in the sub-cell structure would be in conflict with that in the full-cell structure. If we adopt space group $P2_1$, however, there is no such conflict (Fig. 1d). Therefore, the space group of the full-cell structure of PPG9 must be $P2_1$ to preserve the relationship between adjacent triple helices found in the previous analyses (*7*, *8*, *12*, *14*, *16*). Furthermore, the electron density map obtained for $P2_1$ and the infinite triple-helical model clearly showed a peptide region and a gap region, while that for $P2_12_12_1$ did not (Fig. 2). $P2_1$ was therefore adopted as the space group for this crystal, and the refined structure gave reasonable *R*- and R_{free}-factors only for this space group.

Using 24,205 unique reflections with [$F_o > \sigma(F_o)$] in the resolution range of 8.0 to 1.33 Å, the structure was refined to an *R*-factor of 0.182 [0.170 for $F_o > 4\sigma(F_o)$] and an R_{free}-factor of 0.229 [0.212 for $F_o > 4\sigma(F_o)$]. The final structure consisted of two triple-helical molecules and 332 water molecules in an asymmetric unit. This is the

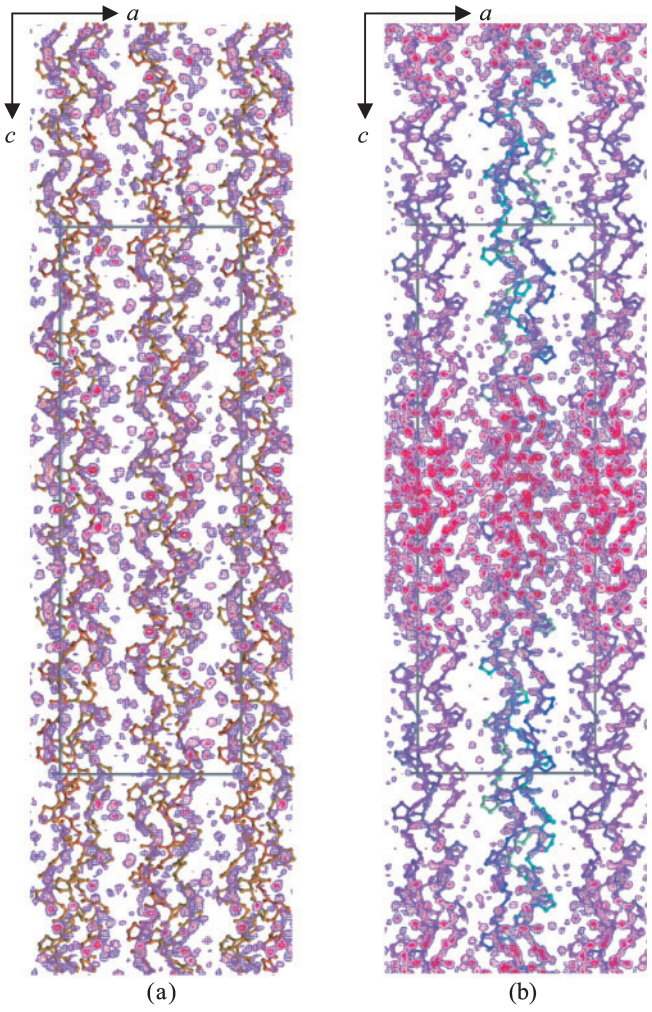

Fig. 2. **Electron density map ($2F_o$–F_c) obtained for infinite triple-helices with the space groups of (a) $P2_12_12_1$ and (b) $P2_1$.** The lowest contour level is 2σ.

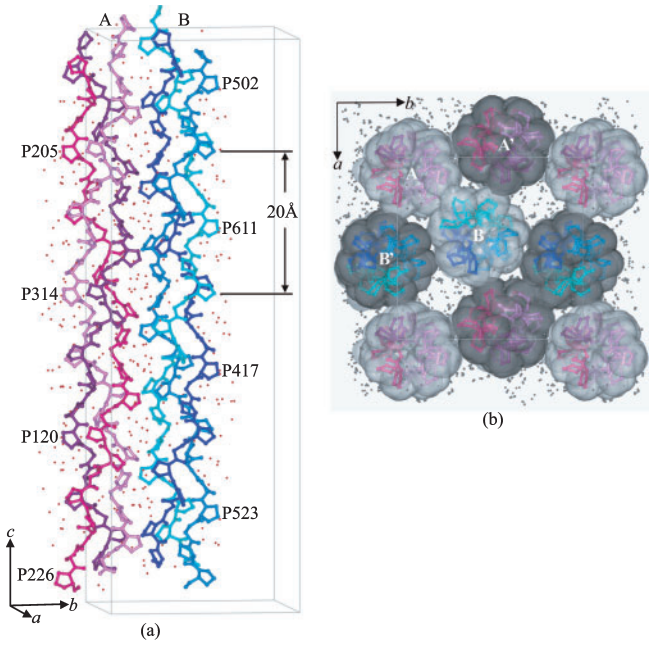

Fig. 3. **Crystal structure of PPG9.** (a) Two molecules (A and B) in an asymmetric unit together with water molecules (small red-filled circle). Residue numbers show the positions corresponding to the residues with opposing puckering in the previous sub-cell structure (*16*). (b) Lateral packing structure showing the pseudo-tetragonal packing of triple-helices, while water molecules between triple-helices are shown as small gray circles. Triple-helices in an asymmetric unit are shown as a ball-and-stick model. The asymmetric unit contains A and B molecules. Molecules A′ and B′ are related to A and B by crystallographic 2_1 symmetry, respectively. Those of parallel molecules are shown as the van der Waals surface in light gray and those of anti-parallel molecules are shown in dark gray.

second analysis of the full-cell structure and the first under cryogenic conditions of PPGn crystals. The details of the refinement statistics are presented in Table 1.

Overall Description of the Molecular Conformation—Two independent triple-helical molecules in an asymmetric unit were related approximately by non-crystallographic 2-fold symmetry and 10 Å translation along the *c*-axis, this being equivalent to the 2_1 symmetry of the sub-cell structure. Hereafter, these two molecules in an asymmetric unit are designated as the A and B molecules, and those related by crystallographic 2_1 symmetry as the A′and B′ molecules, respectively. The final structures of the A and B molecules are shown in Fig. 3a, and their structures looking along the fiber-axis can be seen in the lateral packing structure (Fig. 3b).

The average ϕ, ψ, and ω torsion angles of the Pro at the X and Y positions [hereafter called Pro(X) and Pro(Y), respectively], and Gly are shown in Table 2, together with the full-cell structure of PPG10 (*14*), and 7/2- (*5*) and 10/3-helical models (*19*) of native collagen. The helical parameters calculated by Sugeta and Miyazawa's (*20*) method are also shown in Table 2. The helical parameters for both PPG9 and PPG10 strongly indicated the 7/2-helical structure rather than the 10/3 one. Three strands in a triple-helix were held together by consecutive hydrogen bonds between the NH of Gly in one strand and the O=C of Pro(X) in the adjacent strand. In addition, the carbonyl oxygens of Pro(X) and Gly were also within hydrogen-bonding distance of the Cα of Pro(Y) and Gly, respectively (Table 3). Similar values were observed for PPGn (n = 9, 10) (*7*, *8*, *12*, *14*, *16*) and (Pro-Hyp-Gly)$_n$ (n = 10, 11) (*9*–*11*), and two models for native collagen (*5*, *19*).

Proline Ring Puckering—Two molecules in an asymmetric unit contained 54 independent triplets. At the X position, 53 Pro residues exhibited down puckering, while one N-terminal residue (Pro201) exhibited up puckering with quite a high thermal factor. Since this up puckering was observed at the molecular terminal, it was not considered to be relevant to the hypothesis. Therefore, it can be said that proline ring puckering at the X position follows the propensity-based hypothesis.

On the other hand, six Pro(Y) residues (Pro302, Pro226, Pro502, Pro417, Pro611 and Pro523) of 54 exhibited down puckering instead of up puckering. Of these six, five residues, the exception being Pro302, were located at the same positions (Fig. 3a) as opposing puckering in the sub-cell structures of PPG9 (*16*) and PPG10 (*8*). Since the asymmetric unit of the full-cell structure consists of four sub-cells and two molecules, the total

Table 2. **Average torsion angles in the main chain of peptides with a Pro-Pro-Gly sequence.**

		PPG9	PPG10[a]	7/2-helix[b]	10/3-helix[c]
Torsion angles					
Pro(X)	ϕ	–74.7 (2.4)	–74.5 (2.9)	–75.5	–72.1
	ψ	164.8 (3.1)	164.3 (4.1)	152.0	164.3
	ω	174.7 (2.1)	176.0 (2.5)	–176.8	180.0
Pro(Y)	ϕ	–60.3 (2.0)	–60.1 (3.6)	–62.6	–75.0
	ψ	152.1 (2.0)	152.4 (2.6)	147.2	155.8
	ω	173.9 (2.3)	175.4 (3.4)	–172.8	180.0
Gly	ϕ	–70.1 (2.8)	–71.7 (3.7)	–70.2	–67.6
	ψ	176.8 (2.9)	175.9 (3.1)	175.4	151.4
	ω	179.6 (1.4)	179.7 (2.0)	178.2	180.0
Helical parameters					
h (Å)		8.49 (0.20)	8.53 (0.14)	8.61	8.58
θ (°)		52.4 (5.4)	51.4 (3.7)	51.4	36.0

[a]Berisio *et al.*, 2002; [b]Okuyama *et al.*, 1981; [c]Fraser *et al.*, 1979.

Table 3. **Average values of inter-strand hydrogen bonds.**

	H...A (Å)	D...A (Å)	∠(DHA) (°)
N(Gly)...O(Pro(X))	2.10 (0.06)	2.93 (0.06)	158.2 (2.9)
Cα(Pro(Y))...O(Pro(X))	2.43 (0.05)	3.29 (0.06)	143.5 (2.1)
Cα(Gly)...O(Gly)	2.59 (0.06)	3.09 (0.04)	111.5 (3.2)

number of positions corresponding to the opposing puckering should be eight, as shown in Fig. 3a. Of eight, three Pro(Y) residues (Pro205, Pro314 and Pro120) in the A molecule exhibited up puckering as expected. However, the absolute values of χ_1 angles were significantly smaller than for the others. Furthermore, the longest axes of the anisotropic thermal ellipsoid of Cγ atoms in these Pro residues were perpendicular to their ring planes. Similar tendencies were found in the sub-cell and full-cell structures of PPG10 (*7*, *8*, *12*, *14*), and the sub-cell structure of PPG9 (*16*). Since opposing puckering was observed at an identical position in the packing structures of PPG9 and PPG10, this opposing puckering could be attributed to the interaction with the adjacent triple-helices. The torsion angles (χ_1) of Pro(X) and Pro(Y) residues of the PPG9 structure are shown in Fig. 4 as a function of torsion angles (ϕ), together with those reported for PPGn (*7*, *8*, *12*, *14*, *16*). It is clearly shown that torsion angles (ϕ) of Pro residues were closely related to the side chain conformation (χ_1), as reported previously (*15*). Most Pro(X) and Pro(Y) residues exhibited down and up puckering, respectively, as expected from the propensity-based hypothesis (*12*). In these cases, the distance between the Cγ atom of Pro(X) in one strand and the Cγ atom of Pro(Y) in the adjacent strand was about 4.1 Å, and these Pro residues make a "Pro:Pro van der Waals interaction," which seems important for stabilization of the triple-helical structure (*21*). The ideal "Pro:Pro van der Waals interaction" occurs only when Pro(X) adopts the down and Pro(Y) adopts up puckering. On the other hand, when the puckering of Pro(Y) was down, the average distance between Cγ atoms was 4.8 Å, that is, significantly greater than the average for the former cases (4.1 Å). Therefore, there must be some compensatory mechanism for this energy loss, and this will be discussed in the next section.

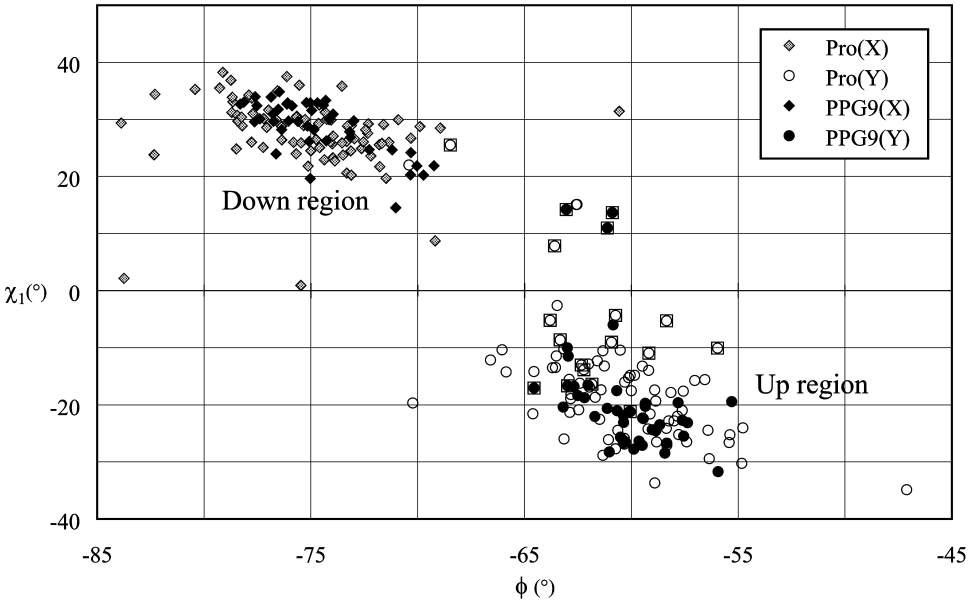

Fig. 4. **χ_1 *versus* ϕ plot for Pro residues at the X (solid diamonds) and Y (solid circles) positions observed in this study, together with those for the X (gray diamonds) and Y (open circles) positions in PPGn (*7*, *8*, *12*, *14*, *16*).** Points enclosed by squares show Pro(Y) corresponding to the position with opposing puckering in the sub-cell structure of PPG9.

The Packing Structure and Hydrophobic Interaction between Triple-Helices—PPG9 molecules were packed laterally in a pseudo-tetragonal manner, each triple helix being surrounded by one parallel and four anti-parallel helices with hydrophobic interactions between them (Fig. 3b). This was essentially the same as in the cases of the previous structures (*5*, *7*, *8*, *12*, *14*, *16*).

Since the molecular surface of PPG9 consists of hydrophobic Pro rings, every interaction between adjacent triple-helices is hydrophobic. These interactions can be classified into three types: first, an interaction between the A and B molecules in an asymmetric unit, as shown in Fig. 5a. In this interaction, the Cγ atom at the Pro(Y) in one molecule interacts with the Cδ atom at the Pro(X) and the Cβ atom at the Pro(Y) of the same strand in another molecule with a distance shorter than 4 Å. As there is a 2_1-symmetry along the *c*′-axis of the sub-cell structure with a 20 Å axial repeat, this interaction is repeated every 10 Å along the *c*-axis. Second, there is an interaction between anti-parallel A and B′ molecules along the *a*-axis, as shown in Fig. 5b. In this case, Cβ and Cγ at the Pro(X) in one molecule interact with Cγ at the Pro(Y) in the other molecule. In addition, Cγ and Cδ at the Pro(Y) in one molecule also interact with Cβ and Cγ at the Pro(X) in the other molecule with a distance shorter than 4 Å. These two interactions are repeated every 20 Å along the *c*-axis and form a sheet structure parallel to the *a*–*c* plane. Third, there is an interaction between anti-parallel A and A′, or B and B′ molecules along the *b*-axis. These two molecules are related by the crystallographic 2_1-symmetry and form a sheet structure parallel to the *b*–*c* plane, as shown in Fig. 5c. Each sheet consists of an A or B molecule, which accumulates alternately in the *a*-direction. These two sheets are very similar, and are related by the pseudo-2_1 symmetry along the *a*-axis. However, there are significant differences: in both cases, Cδ at the Pro(Y) in one molecule is close to Cγ at the Pro(Y) in the adjacent molecule, and Cγ at the Pro(Y) in one molecule is also close to Cγ at the Pro(X) in the other molecule. Although these interactions are longer than 4 Å in the case of the A and A′ molecules (for example, Fig. 5d), the distances between the B and B′ molecules are rather small. This strong interaction between Cδ and Cγ atoms in the adjacent chains was reflected by the opposing down puckering of the Pro(Y) residues in the B molecule. As mentioned in the previous section, because of this opposing puckering, the distance between the Cγ atoms of Pro(X) and Pro(Y) residues in the adjacent strands of the same molecule becomes greater (for example, 5.05 Å in Fig. 5e), and the Pro:Pro stacking interaction becomes weaker. On the other hand, in the case of the A molecule, the puckering of Pro(Y) retains the up conformation but with smaller absolute values of χ_1. Therefore, the interaction between adjacent A and A′ molecules is rather weak but the Pro:Pro stacking interaction remains strong, considering the distance between the Cγ atoms. These experimental results suggested that the proline rings easily adapt to both types of puckering depending on the surrounding circumstances.

Hydration of Water Molecules—A total of 332 water molecules were found in an asymmetric unit and were classified into two groups: water molecules in the first shell, which form hydrogen bonds directly with peptide atoms, and those in the second shell, which form hydrogen bonds with only water molecules. The numbers of water molecules in these hydration shells were 130 and 202, respectively. In the case of PPG10, the total number of water molecules is 352 and that in the first shell is 143 (*14*). Therefore, the number of water molecules per one triplet is essentially the same (6.2 for PPG9 and 5.9 for PPG10) even though the data collection temperatures were different (100 K for PPG9 and RT for PPG10). Similar numbers of water molecules were also found in the cases of POG10 at RT (6.0), and POG10 and POG11 at 100 K (7.0) (*11*). This evidence showed that the number of water molecules does not depend on the temperature, the existence of a hydroxyl group of Hyp or a difference in the lateral packing arrangement.

Water molecules in the first shell form typical hydrogen bonds with peptide atoms, which were observed in many single crystal structures of collagen-model peptides (*7*, *8*, *12*, *14*, *16*), that is, the carbonyl oxygen at the Y position has two hydration sites, while that of a Gly residue has a single site since the other site is sterically hindered by the adjacent chain. The carbonyl oxygen at the X position cannot form a hydrogen bond with water molecules since it faces the center of the helix and forms a hydrogen bond with NH of Gly in the adjacent chain. The locations of water molecules in the first hydration shell at 100 K were essentially the same as those found in PPGn (n = 9, 10) (*7*, *8*, *12*, *14*, *16*) at RT. The average distance between water molecules in the first hydration shell and carbonyl oxygen atoms is 2.76 Å in PPG9 at 100 K, and 2.78 Å in PPG10 at RT (*14*), which are similar to those of POG11 (2.77 Å at 100 K and 2.79 Å at RT). On the other hand, the average distances between water molecules in the second hydration shell and the closest peptide atom (3.54 Å for PPG9 at 100 K and 3.61 Å for PPG10 at RT) are quite similar, although that of POG11 at RT (3.83 Å) is greater than that at 100 K (3.54 Å) by about 0.3 Å (*11*).

Hydrogen-Bonding Networks Mediated by Water Molecules—Similar to hydrophobic interactions, three types of hydrogen-bonding networks mediated by water molecules were observed between adjacent triple-helices: the first is formed between parallel triple-helices in an asymmetric unit (Fig. 5a), and connects to the carbonyl oxygen of Pro(Y) in one triple-helix and that of Pro(Y) in the other helix mediated by two water molecules, shown by broken green lines. This network is repeated every 10 Å along the *c*-axis, which gives seven such hydrogen-bonding networks in an asymmetric unit since the geometry of this network is disturbed in the terminal region.

The second hydrogen bonding network is formed between anti-parallel triple-helices along the *a*-axis, and connects the carbonyl oxygen of Pro(Y) in one molecule and the carbonyl oxygen of Gly in the other molecule mediated by two water molecules (Fig. 5b). In addition, one more hydrogen bond always occurs between the carbonyl oxygen of Gly in one molecule and that of Pro(Y) in the other molecule mediated by two water molecules. Since two water molecules directly linked to the carbonyl oxygen atom of Pro(Y) are also linked to each other, four carbonyl oxygen atoms of Pro(Y) and Gly in two adjacent molecules along the *a*-axis are involved in the "H"-shaped hydrogen-bonding network together with four water molecules. This

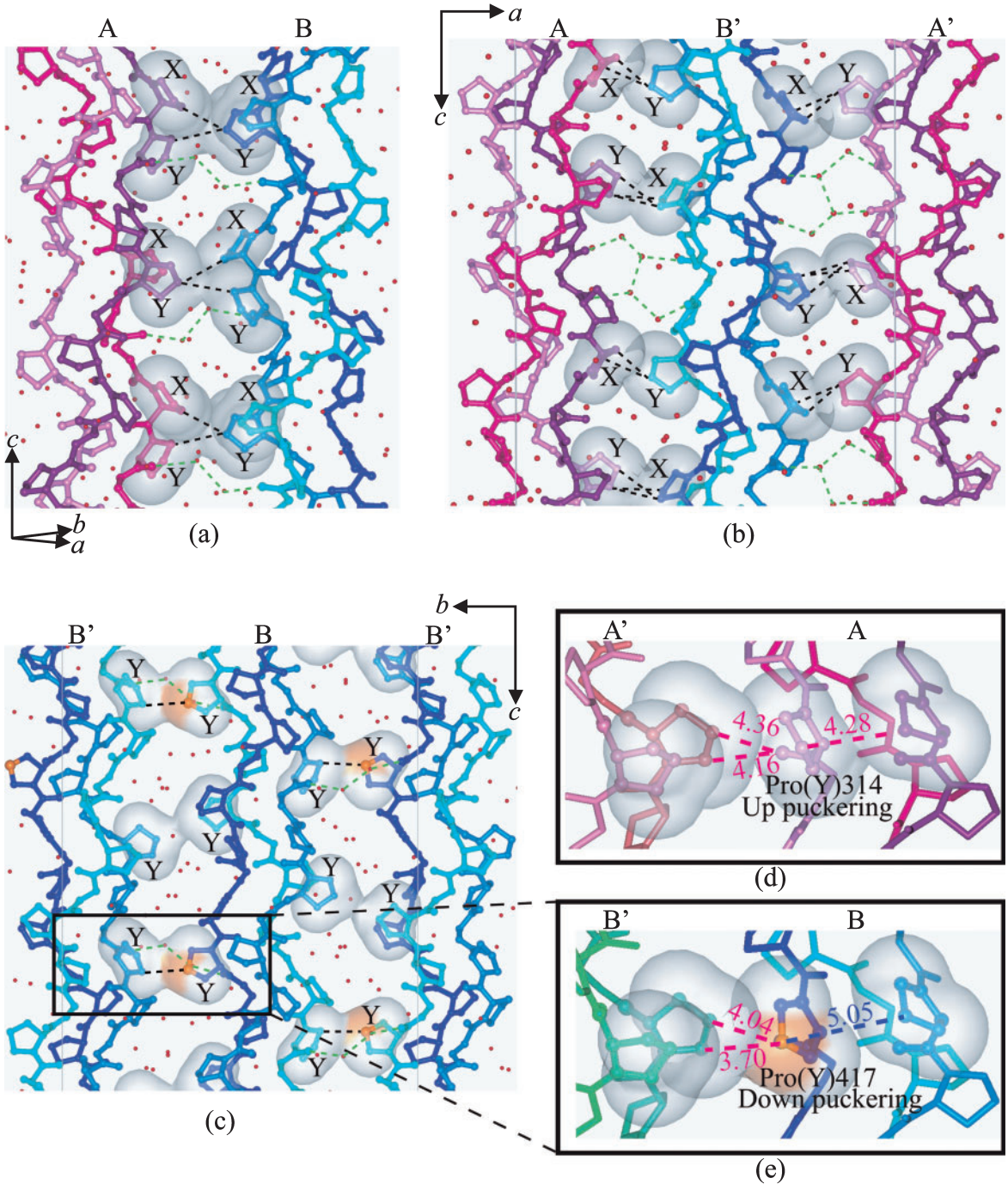

Fig. 5. **Three types of hydrophobic interactions and hydrogen bonding networks between triple-helices near the central region.** The side chains of related Pro residues are shown as the van der Waals surface in gray. Distances between carbon atoms of smaller than 4 Å are shown as broken black lines, while broken green lines denote hydrogen-bonding networks mediated by water molecules in the first shell between triple-helices. (a) Interaction between the parallel triple-helices (A and B) in an asymmetric unit is repeated every 10 Å. (b) Interaction between anti-parallel triple-helices (A and B′) along the *a*-axis occurs in a sheet structure parallel to the *a*–*c* plane. (c) Interaction between anti-parallel triple-helices (B and B′) along the *b*-axis is found in a sheet structure parallel to the *b*–*c* plane. There is a similar interaction between the A and A′ molecules that is found in the other sheet. Although these two sheets are similar, Pro(Y) residues in B molecules along the *b*-axis exhibit puckering opposing the hypothesis. The Cγ atoms of these Pro(Y) residues are shown as a van der Waals surface in orange. (d) Enlargement of the hydrophobic interaction around Pro(Y)314. The shortest distance of carbon atoms between triple-helices is 4.16 Å, and that of carbon atoms between adjacent strands is 4.28 Å. (e) Enlargement of the hydrophobic interaction around Pro(Y)417. The shortest distance of carbon atoms between triple-helices is 3.70 Å, and the distance of Cγ atoms between adjacent strands is increased (5.05 Å).

network is repeated every 20 Å along the *c*-axis, which forms a sheet structure parallel to the *a*–*c* plane.

The third hydrogen-bonding network is formed between anti-parallel triple-helices related by the crystallographic 2_1-symmetry along the *b*-axis (Fig. 5c), and connects the carbonyl oxygen of Pro(Y) in one molecule and that of Pro(Y) in the other molecule mediated by two water molecules. This network is repeated every 20 Å along the *c*-axis and forms a sheet structure parallel to the *b*–*c* plane. There are two such sheets comprising only A or B molecules.

These hydrogen-bonding networks together with the interactions between adjacent molecules were also observed in the structures of PPGn (n = 9, 10) (*7*, *8*, *12*, *14*, *16*). This suggests that the crystal structures of PPGn have essentially the same basic sub-cell structure. Therefore, the local structure is independent of the chain length, space group and temperature, and the specific interaction between triple-helices is preserved. In addition, the $P2_1$ space group for the PPG9 crystal is strongly expected to maintain these specific interactions between triple-helices.

Conclusions—The crystal structure of a collagen-model peptide, [(Pro-Pro-Gly)$_9$]$_3$, was determined at 1.33 Å resolution at 100 K using synchrotron radiation. This is the second analysis for the full-cell structure of [(Pro-Pro-Gly)$_n$]$_3$ crystals and the first analysis under cryogenic condition.

The main chain conformation of [(Pro-Pro-Gly)$_9$]$_3$ was consistent with the left-handed 7/2-helical model with an axial repeat of 20 Å for native collagen. Based on the high-resolution structure, intra- and inter-molecular interactions were discussed. Intra-molecular Pro:Pro stacking interactions were observed between Pro(X) in one strand and Pro(Y) in the other strand of the same molecule. This attractive interaction seems most effective when Pro(X) and Pro(Y) adopt down and up puckering, respectively, which is also expected from the propensity-based hypothesis. In other words, the propensity-based hypothesis indicates Pro:Pro stacking interactions between imino acid residues in a triple-helical molecule. Three types of hydrophobic interactions between neighboring molecules are repeated every 20 Å so that the relationship between adjacent molecules in the sub-cell structure of the [(Pro-Pro-Gly)$_n$]$_3$ is tightly maintained. In order to maintain these interactions, the space group of [(Pro-Pro-Gly)$_9$]$_3$ crystal must be $P2_1$. The five opposing ring puckerings of Pro residues at the Y position found in this study were explained in terms of the above two competing attraction forces, that is, the Pro:Pro stacking interaction and hydrophobic interaction between neighboring molecules.

Since collagen molecules act as a fibrous material and an extracellar matrix after assembling together, the average molecular structure affects the structure of the assembly and the mode of assembly. For this reason, it is very important to know the average molecular structure of collagen. The present high resolution analysis clearly suggested that a peptide molecule takes the left-handed 7/2-helical conformation with a 20 Å axial repeat, which has been proposed as an average structure for collagen by our group. This 7/2-helical model has also been supported by many other single crystal analyses of model peptides. On the other hand, there is no such support for the Rich and Crick model.

We wish to acknowledge the use of the BL40B2 of SPring-8, with gratitude to the Japan Synchrotron Radiation Research Institute (JASRI) for the beam time and support. This study was supported in part by a Grant-in-Aid for Scientific Research (No. 16550107) from the Japan Society for the Promotion of Science (JSPS).

REFERENCES

1. Rich, A. and Crick, F.H. (1961) The molecular structure of collagen. *J. Mol. Biol.* **3**, 483–506
2. Okuyama, K., Takayanagi, M., Ashida, T., and Kakudo, M. (1977) A new structural model for collagen. *Polym. J.* **9**, 341–343
3. Okuyama, K., Tanaka N., Ashida T., Kakudo M., and Sakakibara S. (1972) An x-ray study of the synthetic polypeptide (Pro-Pro-Gly)$_{10}$. *J. Mol. Biol.* **72**, 571–576
4. 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
5. 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
6. 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 Crystallogr.* **A34**, 3–11
7. Nagarajan, V., Kamitori, S., and Okuyama, K. (1998) Crystal structure analysis of collagen model peptide (Pro-Pro-Gly)$_{10}$. *J. Biochem.* **124**, 1117–1123
8. 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
9. Nagarajan, V., Kamitori, S., and Okuyama, K. (1999) Structure analysis of a collagen-model peptide with a (Pro-Hyp-Gly) sequence repeat. *J. Biochem.* **125**, 310–318
10. Berisio, R., Vitagliano, L., Mazzarella, L., and Zagari, A. (2001) Crystal structure determination of the collagen-like polypeptide with repeating sequence Pro-Hyp-Gly: Implications for hydration. *Biopolymers* **56**, 8–13
11. Okuyama, K., Hongo, C., Fukushima, R., Wu, G., Narita, H., Noguchi, K., Tanaka, Y., and Nishino, N. (2004) Crystal structures of collagen model peptides with Pro-Hyp-Gly repeating sequence at 1.26 Å resolution: Implications for proline ring puckering. *Biopolymers* **76**, 367–377
12. Vitagliano, L., Berisio, R., Mazzarella, L., and Zagari, A. (2001) Structural bases of collagen stabilization induced by proline hydroxylation. *Biopolymers* **58**, 459–464
13. Berisio, R., Vitagliano, L., Sorrentino, G., Carotenuto, L., Piccolo, C., Mazzarella, L., and Zagari, A. (2000) Effects of microgravity on the crystal quality of a collagen-like polypeptide. *Acta Crystallogr.* **D56**, 55–61
14. Berisio, R., Vitagliano, L., Mazzarella, L., and Zagari, A. (2002) Crystal structure of the collagen triple helix model [(Pro-Pro-Gly)$_{10}$]$_3$. *Protein Sci.* **11**, 262–270
15. Vitagliano, L., Berisio, R., Mastrangelo, A., Mazzarella, L., and Zagari, A. (2001) Preferred proline puckerings in *cis* and *trans* peptide groups: Implications for collagen stability. *Protein Sci.* **10**, 2627–2632
16. Hongo, C., Nagarajan, V., Noguchi, K., Kamitori, S. Okuyama, K., Tanaka, Y., and Nishino, N. (2001) Average crystal structure of (Pro-Pro-Gly)$_9$ at 1.0 Å resolution. *Polymer J.* **33**, 812–818
17. Brünger, A.T. (1992) *X-PLOR Version3.1 A System for X-Ray Crystallography and NMR*, Yale University Press, New Haven and London

18. Sheldrick, G.M. and Schneider, T.R. (1997) SHELXL: High-resolution refinement. *Methods Enzymol.* **277**, 319–343
19. Fraser, R.D.B., MacRae, T.P., and Suzuki, E. (1979) Chain conformation in the collagen molecule. *J. Mol. Biol.* **129**, 463–481
20. 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**, 673–679
21. Bhatnagar, R.S., Pattabiraman, N., Sorensen, K.R., Langridge, R., MacElroy, R.D., and Renugopalakrishnan, V. (1988) Interchain proline:proline contacts contribute to the stability of triple-helical conformation. *J. Biomol. Struct. Dynanics* **6**, 223–233