Crystal structure of the collagen triple helix model $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$

RITA BERISIO,[1] LUIGI VITAGLIANO,[1] LELIO MAZZARELLA,[1,2,4] AND ADRIANA ZAGARI[1,3,4]

[1]Centro di Studio di Biocristallografia, CNR, I-80134 Napoli, Italy
[2]Dipartimento di Chimica, Università degli Studi di Napoli "Federico II", I-80126, Napoli, Italy
[3]Dipartimento di Chimica Biologica, Università degli Studi di Napoli "Federico II", I-80134, Napoli, Italy
[4]CEINGE, Biotecnologie avanzate Scarl, Napoli, Italy.

(RECEIVED August 9, 2001; FINAL REVISION October 19, 2001; ACCEPTED October 31, 2001)

Abstract

The first report of the full-length structure of the collagen-like polypeptide $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$ is given. This structure was obtained from crystals grown in a microgravity environment, which diffracted up to 1.3 Å, using synchrotron radiation. The final model, which was refined to an R_{factor} of 0.18, is the highest-resolution description of a collagen triple helix reported to date. This structure provides clues regarding a series of aspects related to collagen triple helix structure and assembly. The strict dependence of proline puckering on the position inside the Pro-Pro-Gly triplets and the correlation between backbone and side chain dihedral angles support the propensity-based mechanism of triple helix stabilization/destabilization induced by hydroxyproline. Furthermore, the analysis of $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$ packing, which is governed by electrostatic interactions, suggests that charges may act as locking features in the axial organization of triple helices in the collagen fibrils.

Keywords: Collagen; protein structure stability; X-ray structure; triple helix; proline

In structural studies of collagen model compounds, the (Pro-Pro-Gly)n sequence is certainly the first and most widely investigated, since the X and Y positions of the collagen consensus X-Y-Gly sequence are frequently occupied by imino acids. A first structural analysis of the polypeptide poly(Pro-Pro-Gly) was reported by Yonath and Traub (1969). The poly(Pro-Pro-Gly) fiber diffraction pattern was clearly reminiscent of that observed for collagen, though showing a significantly higher definition. The model obtained exhibited the basic characteristics of the previously proposed collagen II model (Rich and Crick 1961), and for several years it was considered the best available model for the collagen triple helix. As for natural collagen, a model consisted of three left-handed polyproline II-like chains wrapped around a common axis. Gly residues are required at every third residue (X-Y-Gly) as the close packing of the chains near the central axis does not leave room for larger residues. Furthermore, backbone N atoms of Gly residues are involved in interchain hydrogen bonds with the carbonyl groups of residues located in the X positions.

Fiber diffraction studies were followed by a long series of theoretical works on both the Pro-Pro-Gly (Miller and Scheraga 1976; Némethy et al. 1992) and Pro-Hyp-Gly models (Miller et al. 1980). Additional structural information was provided by a low-resolution single-crystal X-ray study on the polypeptide model $(Pro\text{-}Pro\text{-}Gly)_{10}$ (Okuyama et al. 1981). In particular, this polypeptide exhibited significant differences in the triple helical parameters from the Rich and Crick model of real collagen (Rich and Crick 1961). These differences, which led to a 7_2 as opposed to 10_3 triple helical symmetry (Rich and Crick 1961), initiated a debate regarding the actual symmetry of natural collagen

Reprint requests to: Prof. Adriana Zagari, Centro di Studio di Biocristallografia, CNR, Via Mezzocannone 6, I-80134, Napoli, Italy; e-mail: zagari@chemistry.unina.it; fax: (39) 081-2536-603.

Abbreviations: PPG, average model of $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$ determined in the subcell approximation; Gly→Ala, polypeptide with sequence $(Pro\text{-}Hyp\text{-}Gly)_4$-Pro-Hyp-Ala-$(Pro\text{-}Hyp\text{-}Gly)_5$.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.32602.

(Okuyama et al. 1977; Kramer et al. 1999; Okuyama et al. 1999). Over the years, $(Pro\text{-}Pro\text{-}Gly)_{10}$ has been regarded as the reference structure for the study of the influence of pyrrolidine ring substituents on triple helix stability and structure. For example, it has been observed that hydroxylation of prolines in the Y position leads to a stabilization of the triple helix, whereas in the X position it strongly destabilizes (Fields and Prockop 1996; Inouye et al. 1982).

In the last decade, a variety of new polypeptide models have been synthesized and characterized, and they have shed light on important features related to collagen structure and stability. The use of specifically designed 'host–guest' peptides has provided a reliable scale of triple helical propensities of the various amino acids (Persikov et al. 2000). X-ray studies carried out on the host-guest polypeptide with sequence $(Pro\text{-}Hyp\text{-}Gly)_4$-Pro-Hyp-Ala-$(Pro\text{-}Hyp\text{-}Gly)_5$, named Gly→Ala, have shown the structural effects of a disease causative mutation (Bella et al. 1994, 1995). Furthermore, the study of fluorinated proline derivatives has indicated that inductive effects may play a role in collagen triple helix stability (Holmgren et al. 1998).

Although a large number of polypeptides have been studied to date, the interest in $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$ is still high, as demonstrated by several recent investigations (Kramer et al. 1998; Nagarajan et al. 1998; Vitagliano et al. 2001a). However, all of the structural models produced were obtained as approximate average structures, by using only a specific class of reflections which characterized a subcell of the $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$ crystals. Herein is the first report of the full-length structure of $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$, which was obtained using synchrotron radiation on crystals grown in microgravity conditions. Several conclusions have been extracted from the analysis of the structure, and their possible biological implications are discussed.

Results and Discussion

Overall description of the structure

The structure of $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$ was solved and refined at 1.3 Å by combining the use of improved quality crystals, grown in microgravity conditions (Berisio et al. 2000), and synchrotron radiation. The model obtained was refined to an R_{factor} of 0.181 (0.164 for $F > 4\sigma$) and an R_{free} of 0.297 (0.261 for $F > 4\sigma$). Unlike previous studies, which described only an average model (Kramer et al. 1998; Nagarajan et al. 1998; Vitagliano et al. 2001a), the first full-length structure of $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$ is here reported. A description of the relationship between the present structure and the average one is given in Figure 1A,B.

Electron density maps provide a very detailed picture of the whole asymmetric unit, which contains two triple helical molecules, related by a rotation of 51.5° around the shared triple helix axis. Each of the molecules has a rod-shaped structure about 90 Å long with a diameter of 10 Å, and is composed of three chains staggered by one residue. As shown in Figure 1A,C, the two molecules in the asymmetric unit are arranged in a head-to-tail fashion. An example of the quality of the electron density maps is given in Figure 1D, which shows an omit (Fo-Fc) map of a typical Pro-Pro-Gly triplet.

The structure described here is the highest-resolution description of a collagen triple helix to date and, therefore, it provides a good opportunity to review the main architectural elements of collagen-like triple helices. The examination of φ, ψ, and ω backbone dihedral angles shows a regular trend of torsion angles along the peptide sequence with small variations at the triple helix terminations. These parameters, averaged over the two molecules in the asymmetric unit, are very similar to those previously reported using the subcell approximation (Table 1) and indicate a 7_2 triple helix symmetry. On the other hand, they are rather different from those derived for the models of natural collagen (Fraser et al. 1979) and poly(Pro-Pro-Gly) (Yonath and Traub 1969), which both exhibit a 10_3 superhelical symmetry.

Similar to the backbone dihedral angles, helical parameters of the individual peptide chains as well as superhelical parameters are somewhat constant along the molecule. As such, they do not reveal significant unfolding of the terminal ends, despite the higher mobility of the terminal ends indicated by the estimated uncertainties at the atomic positions. The values of the unit twist and unit height parameters are respectively centered around 51.5 (± 2.0)° and 8.53 (± 0.05) Å for the 7_1 single polypeptide chains, and around −102.5 (± 1.4)° and 2.70 (± 0.04) Å for the 7_2 triple helices.

Proline φ angles of $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$ are reported as a function of χ_1 in Figure 2. Together with the experimental data derived from the $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$ structure, the figure reports the results of a statistical analysis (Vitagliano et al. 2001b) carried out on all trans proline residues from nonredundant protein chains, belonging to structures refined at a resolution better than 2.0 Å and an R_{factor} lower than 0.23 (Fig. 2). In accordance with our previous results (Vitagliano et al. 2001a), proline residues from $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$ are allocated in triple helices without significant strain, since backbone φ angles correlate well with those intrinsically adopted by proline residues in globular proteins. Furthermore, proline rings in the Y positions systematically adopt an up puckering, whereas those in the X position adopt a down puckering. These findings provide statistical support for the 'propensity-based' mechanism for the stabilization of the collagen triple helix induced by the presence of Hyp in the Y position and its destabilization when Hyp is in the X position (Vitagliano et al. 2001a,b). Such a model, which is based on both the strict correlation between main and side chain dihedral angles in imino acids and on the observation that the hydroxyl group of Hyp

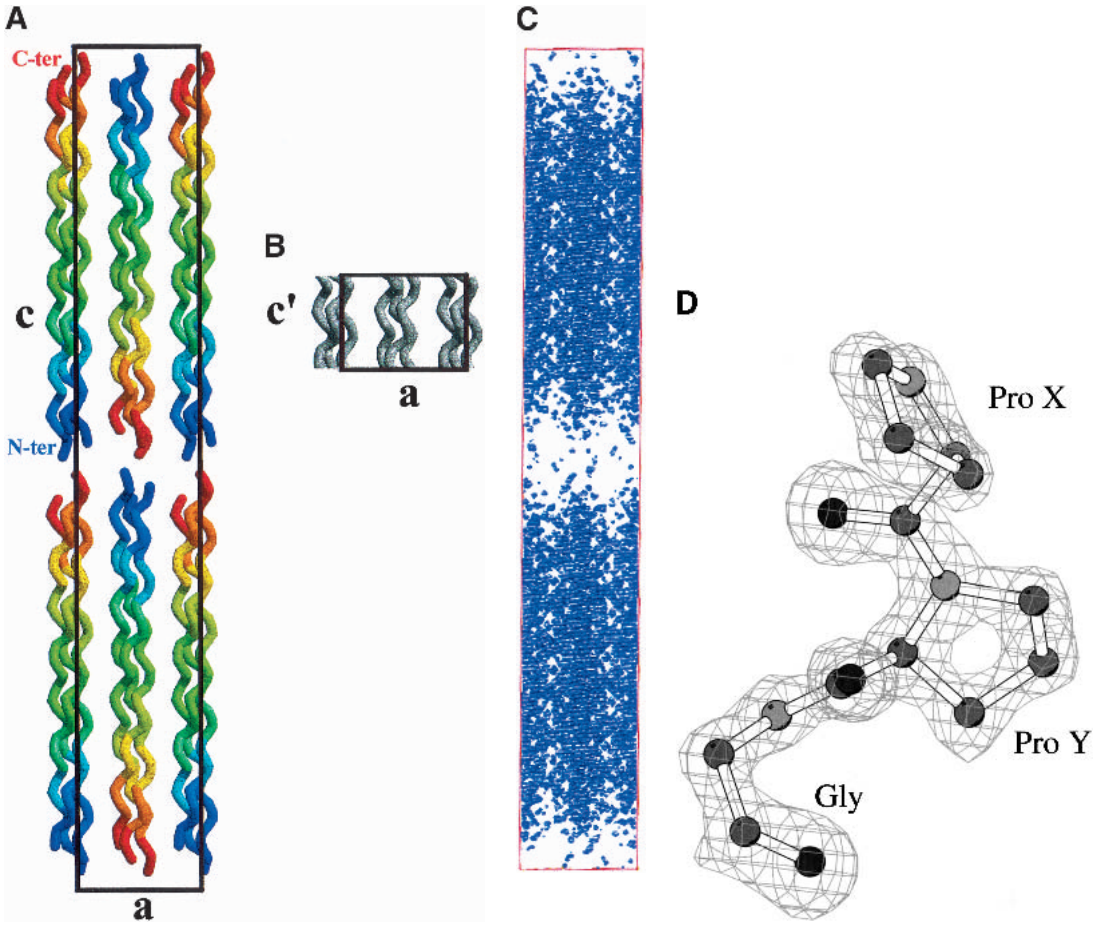

Fig. 1. (*A*) Organization of the $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$ triple helices in the *ac* plane. Chains are colored with a ramping code from blue (N-termini) to red (C-termini). The two molecules in the asymmetric unit are the top-center (molecule 1) and the bottom-center (molecule 2) molecules. (*B*) Average model obtained in the subcell approximation (Vitagliano et al. 2001a) in the *ac′* plane. The c′ corresponds to a ninth of the full-length *c* axis. (*C*) Electron density map (2Fo-Fc), extended to the whole unit cell, contoured at 2.0 σ. (*D*) Omit map (Fo-Fc) of a representative triplet contoured at 3.5 σ.

induces an up puckering, has also been very recently confirmed by theoretical calculations (Improta et al. 2001).

Solvent organization and proline accessibility

The previous determinations of average PPG structures led to controversial findings regarding triple helix hydration (Kramer et al. 1998; Nagarajan et al. 1999; Vitagliano et al. 2001a). Although the role of hydration in triple helix stability has been undermined (Holmgren et al. 1998, 1999) we here show, in accordance with studies on other collagen-like polypeptides (Bella et al. 1995; Kramer et al. 1998; Melacini et al. 2000; Berisio et al. 2001; Vitagliano et al. 2001a), that there is an extensive participation of solvent in the triple helical structure. Indeed, a thick cylinder of hydration, composed of as many as 352 water molecules, surrounds the two triple helices in the asymmetric unit. These water sites were grouped in a first hydration shell, if directly hydrogen bonded to the polypeptide chain, and a second or higher hydration shell, if interacting only with other water molecules. The distribution of water sites as a function of the distance from the nearest polypeptide chain atom is shown in Figure 3. The very sharp peaks, centered on 2.7 and 3.6 Å correspond to the first and the second hydration shell, respectively (Fig. 3). In contrast to what is typically observed in high-resolution protein crystal structures (Esposito et al. 2000), the second and first hydration shells are nearly equally populated. This may be attributed to the close packing of the triple helices, which prevents the formation of large solvent channels. Less well defined hydration sites were found in the terminal regions as a consequence of the higher local disorder.

The high extent of first shell hydration in $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$ may be ascribed to the unusually high percentage of exposed unsaturated carbonyl groups, together with the triple helix peculiar rod-like shape. Indeed, in the Pro-Pro-

Table 1. *Average values of main chain dihedral angles*

	$[(PPG)_{10}]_3$[a]	PPG-0[b]	PPG-0″[c]	PPG-1[d]	PPG-2[d]	Collagen fiber[e]
ϕ Pro X(°)	−74.5 (2.9)	−75.5	−77.1 (7.5)	−73.1 (8.8)	−75.0 (2.7)	−72.1
ψ Pro X(°)	164.3 (4.1)	152.0	163.8 (4.7)	159.7 (3.7)	161.4 (3.1)	164.3
ω Pro X(°)	176.0 (2.5)	−178.2	179.1 (1.3)	178.6 (0.6)	177.8 (0.7)	180.0
ϕ Pro Y(°)	−60.1 (3.6)	−62.6	−62.4 (7.9)	−58.7 (8.1)	−61.2 (1.1)	−75.0
ψ Pro Y(°)	152.4 (2.6)	147.2	154.0 (8.8)	161.0 (12.4)	153.3 (2.2)	155.8
ω Pro Y(°)	175.4 (3.4)	−176.8	177.6 (3.9)	179.1 (1.2)	176.7 (2.1)	180.0
ϕ Gly(°)	−71.7 (3.7)	−70.2	−76.7 (6.6)	−83.7 (11.2)	−75.8 (2.0)	−67.6
ψ Gly(°)	175.9 (3.1)	175.4	176.6 (7.1)	179.8 (5.8)	179.5 (3.5)	151.4
ω Gly(°)	179.7 (2.0)	178.2	179.0 (1.8)	179.6 (0.5)	−179.9 (0.2)	180.0

[a] Present study.
[b] Okuyama et al. 1981.
[c] Nagarajan et al. 1998.
[d] Kramer et al. 1998.
[e] Fraser et al. 1979.

Gly triplets, the carbonyl groups of Gly and Pro residues in the Y positions are solvent exposed, and they are not involved in hydrogen bonding interactions with other groups of the triple helix. In accordance with the general trend observed in previous structure determinations (Bella et al. 1995; Kramer et al. 1998), the carbonyl groups of Gly residues are singly hydrated, whereas those of proline residues in the Y position are doubly hydrated. The water molecules bound to the carbonyl groups act as anchoring points for several water bridges, which surround the triple helices and mediate interactions between them in the crystal. The water sites also shield the triple helical apolar surface, which is peculiarly extended in $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$ compared to globular nonmembrane proteins. Indeed, 87.9% of the ex-

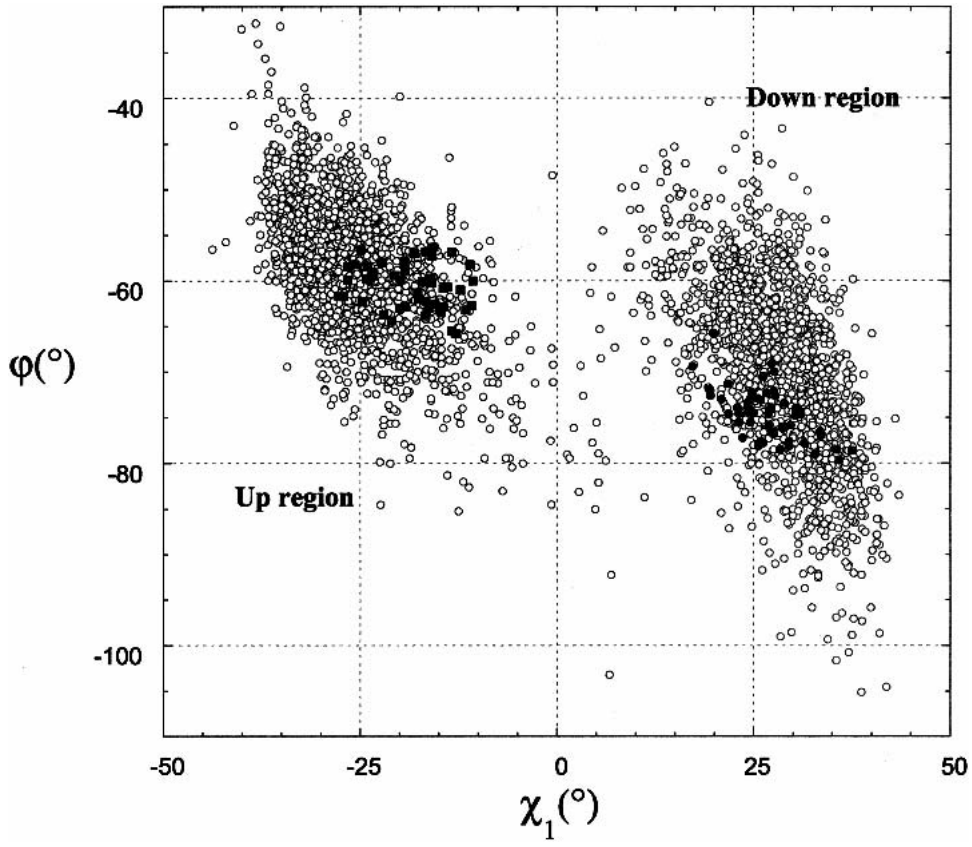

Fig. 2. Values of φ versus χ_1 for proline rings in X (●) and Y (■) positions in $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$ and derived from a statistical survey on trans proline rings in protein structures (○) (Vitagliano et al. 2001b).

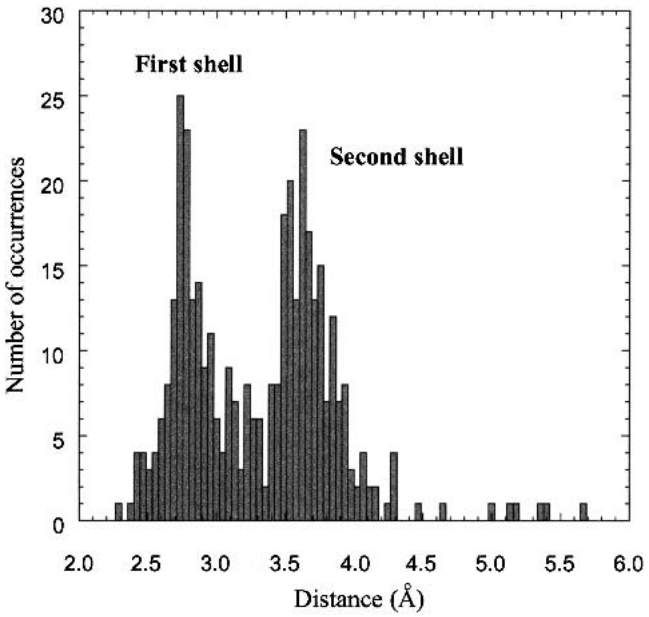

Fig. 3. Distribution of water molecules as a function of their distance from the nearest protein atom.

posed surface is apolar. Furthermore, the percentage of buried apolar surface is as low as 39.1%, in contrast with globular proteins, for which a higher percentage, about 60%–70% (Finney et al. 1980), is needed to compensate for the loss of main chain conformational entropy upon folding (Jones and Miller 1991). These findings may be ascribed to the more limited loss of conformational entropy of the $[(\text{Pro-Pro-Gly})_{10}]_3$ polypeptide chain upon folding, which is due to the high content of imino acids.

Solvent accessibility of Pro residues in the Pro-Pro-Gly triplets was evaluated by averaging side chain accessibility over all proline rings in the X and Y positions. Results show that the side chain accessibility per atom is similar in the X and Y positions, with average values of 22.1 ± 7.5 Å^2 and 25.2 ± 9.7 Å^2, respectively. However, each of the atom types of the proline side chains exhibits significantly different accessibility in the X and Y positions. In particular, the C^β atoms are much more exposed in the X positions. This finding is in accordance with the previous studies showing that the X position is generally more exposed for residues different than Pro (Jones and Miller 1991) and that mutations of nonimino acidic residues in the X position have relatively little effect on triple helix stability (Ramshaw et al. 1998). On the other hand, both C^γ and C^δ atoms are more accessible in Y positions (Fig. 4A,B). Indeed, C^γ and C^δ atoms of the down Pro rings in the X positions are partially buried by the up Pro rings in the Y positions belonging to the neighboring staggered single chain. It is worth noting that the occurrence of down and up puckerings, in X and Y positions respectively, maximizes the hydrophobic interactions between proline residues in adjacent chains.

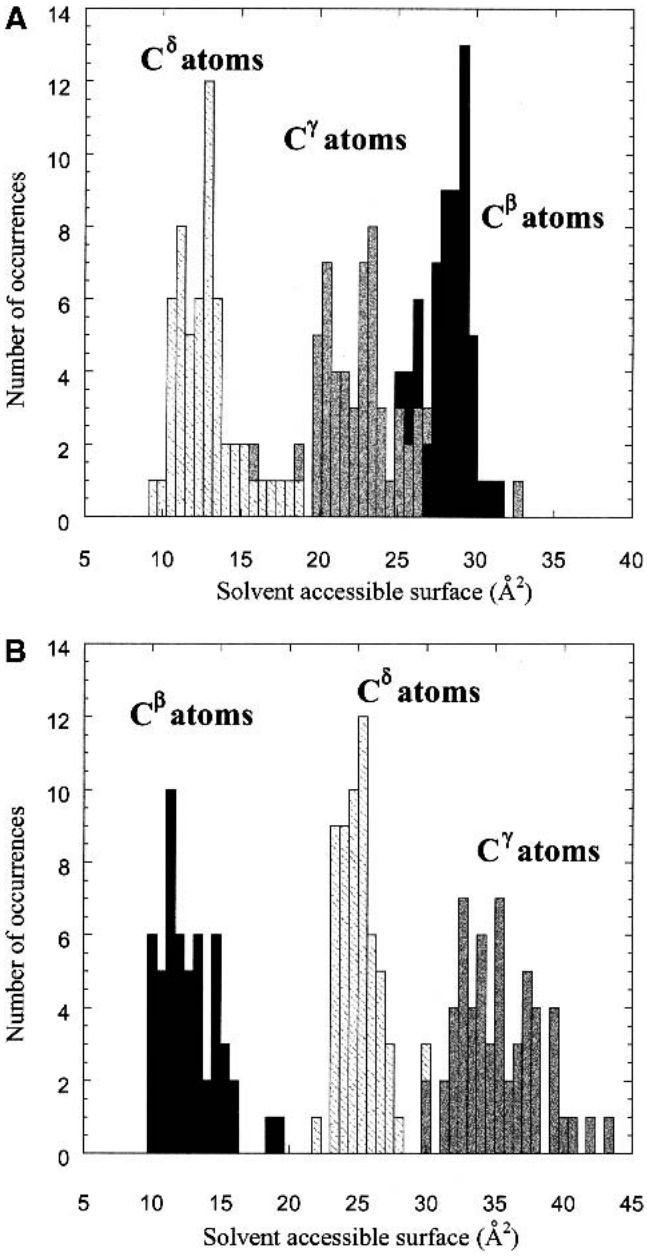

Fig. 4. Solvent accessibility, averaged over the two molecules in the asymmetric unit, of the Pro side chain atoms in X positions(*A*) and Y positions (*B*).

Triple helix packing

Analogous to the previously determined structure of the Gly→Ala polypeptide, $[(\text{Pro-Pro-Gly})_{10}]_3$ triple helices are arranged in layers, with a layer thickness of 91 Å (about half of the *c* axis), corresponding approximately to the length of the polypeptide (Fig. 1A,C). Crystal packing leads N- and C-termini of symmetry-related molecules close in space (Fig. 1A), though there are no directional intermolecular interactions between the terminal ends, which are charged in the crystallization medium (pH = 5.6). Therefore, the axial

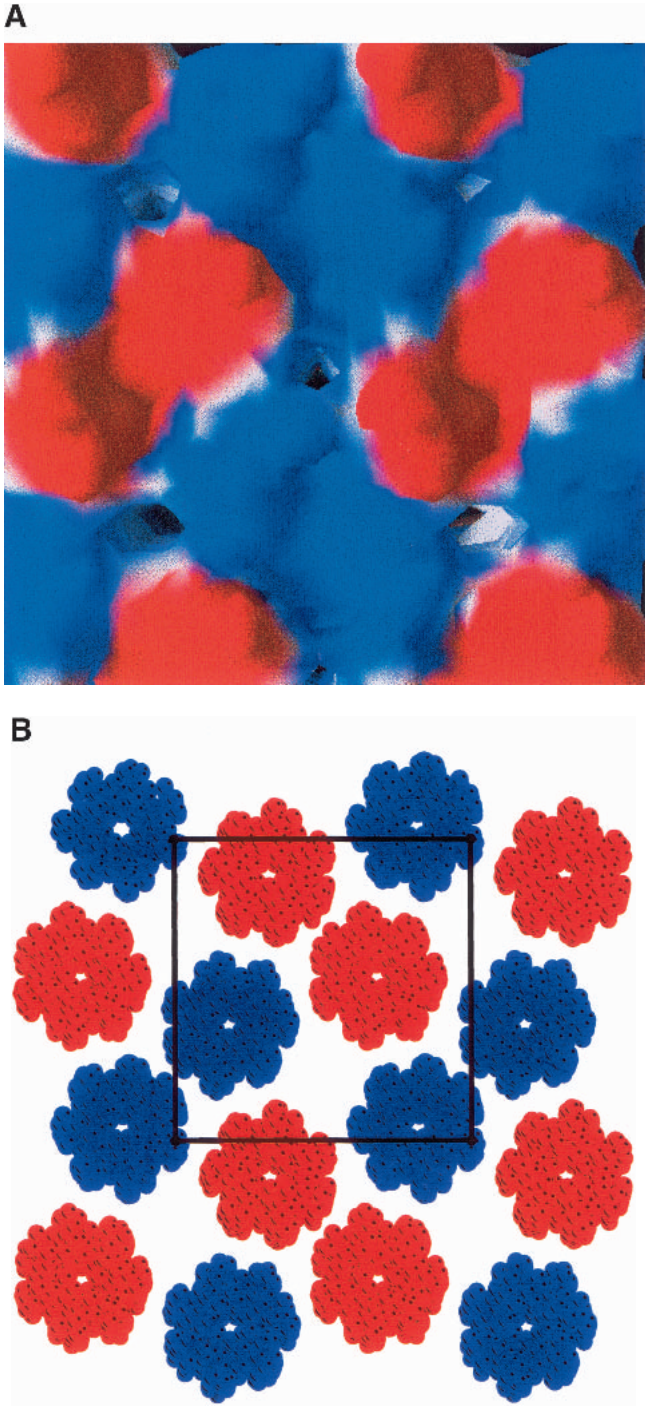

Fig. 5. Lateral packing: (*A*) Electrostatic potential surface of a charged layer. The color code is blue for positive and red for negative electrostatic potential. (*B*) Pseudotetragonal packing of the triple helices. The triple helices shown in red are directed upward from the paper, whereas those shown in blue are directed downward from the paper.

packing of [(Pro-Pro-Gly)$_{10}$]$_3$ is governed by the electrostatic interactions generated by the charged triple helix ends. The lack of requirement for a specific stereochemistry in electrostatic interactions explains the higher disorder of the terminal residues.

The charged layers through the crystal are characterized by a scattered electrostatic potential surface (Fig. 5A). Each of the triple helices is five-coordinated and surrounded by one parallel and four antiparallel helices. Alternatively, each couple of helices is surrounded by four couples with opposite-sign electrostatic potential. Packing of layers on top of each other generates a three-dimensional lattice of charges.

As shown in Figure 5B, the triple helices are organized in a pseudotetragonal packing, as previously reported for the average structures of PPG (Kramer et al. 1998; Nagarajan et al. 1998).

The structural findings found in the full-length [(Pro-Pro-Gly)$_{10}$]$_3$ structure led us to a series of considerations regarding triple helix assembly. The structure here presented, together with that of Gly→Ala (Bella et al. 1994), is charged at the terminal ends. We recently crystallized and solved the X-ray structure of the collagen model [(Pro-Hyp-Gly)$_{10}$]$_3$ (R. Berisio, L. Vitagliano, L. Mazzarella, and A. Zagari, unpubl.), which is also charged. It is worth noting that all of these three structures form double layers of charges, generated by the triple helix ends. Furthermore, in the case of the polypeptide with sequence (Pro-Hyp-Gly)$_4$-Glu-Lys-Gly-(Pro-Hyp-Gly)$_5$, which has also charged residues in the center, triple helices are not in the axial register (Kramer et al. 2000). However, their axial stagger occurs in correspondence with the charged Lys and Glu residues. Therefore, in all of these cases, charges act as locking features in the triple helix axial stagger.

Conclusions

[(Pro-Pro-Gly)$_{10}$]$_3$ has been the object of pioneering studies on collagen-like systems, as it was the first to be synthesized and crystallized. Since then, several X-ray determinations of its structure have been performed over the past nearly 30 years. However, only an approximate X-ray structure, obtained using only the reflections characterizing a subcell with a 20 Å repeat along the *c* axis, has thus far been reported. Indeed, most of the other reflections were very weak and could hardly be recorded. The difficulty in registering the weak reflections was attributed to a fiber-like disorder of the crystals, due to the high level of symmetry of the triple helix. Such difficulty led to an incorrect determination of the unit cell parameters and, consequently, did not allow the determination of the full-length structure. Therefore, the structure presented here is the first crystallographic view of a regular collagen-like model system.

Analysis of the proline puckering supports the 'propensity-based' model for collagen triple helix stabilization by Hyp (Vitagliano et al. 2001a). Indeed, such a model suggests that, due to inductive effects generated by the hydroxyl group (Holmgren et al. 1999; Vitagliano et al.

2001a), Hyp has main chain dihedral angles which are suitable only for the Y position in the triple helix. As such, imino acid propensities also allow the explanation of the destabilization induced by Hyp in the X position (Vitagliano et al. 2001a). Furthermore, they account for the observed shift of the *cis*/*trans* equilibrium observed in Hyp, compared to Pro (Bretscher et al. 2001; Vitagliano et al. 2001b). The present high-resolution structure confirms that triple helices are highly hydrated in the crystal state (Bella et al. 1995; Kramer et al. 1998; Berisio et al. 2001), although the role of water molecules in the stabilization of the collagen triple helix has been questioned (Holmgren et al. 1999; Nagarajan et al. 1999).

The comparison of crystal packing of $[(\text{Pro-Pro-Gly})_{10}]_3$ with that of the other known collagen-like polypeptide structures has highlighted the importance of charges in the axial registration of collagen-like triple helices. This finding may be extended to natural collagen and supports the idea (Kramer et al. 2000) that charges may play a fundamental role in axial registration of triple helices in the collagen fibrils. It should also be mentioned that all Hyp-containing structures known to date form direct hydrogen bonding interactions between Hyp hydroxyl groups of adjacent triple helices (Kramer et al. 2000, 2001; Berisio et al. 2001), except for the structure of Gly→Ala (Bella et al. 1994). In this case, the lack of direct intermolecular hydrogen bonding interactions may be attributed to the loss of the axial triple helix repetition, which is due to the existence of a bulge in the center of the molecule (Bella et al. 1994). These findings suggest that Hyp–Hyp interactions may play a role in lateral assembly of triple helices in the fibrils. It may be concluded that charge–charge and Hyp–Hyp interactions can be regarded as two codes for three-dimensional aggregation of triple helices in collagen fibrils.

Table 2. *Statistics of data processing and refinement*

Data processing statistics	
Space group	$P2_12_12_1$
Unit cell (Å):	
a, b, c	26.91, 26.36, 182.50
Resolution range (Å)	15.0–1.3
No. of reflections	29431
Completeness (%)	88.3
R-merge (%)	9.0
Refinement statistics	
No. of molecules in the asymmetric unit	2
No. of protein atoms	1056
No. of water molecules	352
Rms deviations from ideal values:	
Bond distances (Å)	0.026 (0.04)
Angle distances (Å)	0.045 (0.08)
Planar groups (main chain peptides ($Å^3$))	0.056 (0.10)
Similar U_{ij} values for spatially close atoms ($Å^2$)	0.072 (0.10)
Atomic displacement parameters ($Å^2$):	
Main chain atoms	19.0
Side chain atoms	25.2

Materials and methods

Crystallization and data collection

Crystals of $[(\text{Pro-Pro-Gly})_{10}]_3$ were grown in microgravity conditions, using the dialysis technique, with a final polypeptide concentration of 5.0 mg mL^{-1} in 0.05 M acetic acid, 0.22 M sodium acetate, and 10%(v/v) PEG 400 (Berisio et al. 2000). Such crystals, which belong to the $P2_12_12_1$ space group, were significantly better diffracting than those grown on Earth (Berisio et al. 2000). Diffraction data were collected at 1.3 Å using synchrotron radiation at Elettra, Trieste, Italy. The diffraction pattern was characterized by very strong and very weak reflections. Indexing of only the strongest reflections was consistent with a subcell with axes $a' = 26.9$, $b' = 26.3$, $c' = 20.3$ Å, in accordance with previous results (Okuyama et al. 1981; Kramer et al. 1998; Nagarajan et al. 1998). Furthermore, the whole diffraction pattern was characterized by unit cell parameters a = 26.9, b = 26.4, c = 182.5 Å, with a much larger *c* axis than previously believed. Indeed, in all previous studies, conducted on crystals obtained under similar conditions, only a small fraction of weak reflections was measured. These reflections were tentatively indexed with a *c* axis of about 100 Å (Okuyama et al. 1981; Kramer et al. 1998; Nagarajan et al. 1998).

With the new cell axes, the strong reflections defining the subcell correspond to the reflections with indexes l = 9n, with n integer. As we showed previously (Berisio et al. 2000), reflections have markedly different intensities depending on the l index: besides those with l = 9n, reflections with l = 9n + 2 are the strongest, followed by those with l = 9n + 4 and l = 9n + 7. The presence of these sharp spots made the indexing unambiguous (Berisio et al. 2000). Statistics of data processing, carried out using the HKL package (Otwinowsky and Minor 1997), are shown in Table 2. A full report of the crystallization procedure and data quality has been published (Berisio et al. 2000).

Solution of the structure and refinement

From the analysis of the structure of PPG in the subcell approximation, much helpful information regarding structure determination was obtained. In particular, the noncrystallographic axis of the triple helical molecules is parallel to the *c* axis; in addition, the packing in the *ab* plane was determined (Kramer et al. 1998; Nagarajan et al. 1998; Vitagliano et al. 2001a). Furthermore, the length of the *c* axis suggested the existence of two molecules in the asymmetric unit arranged in a head-to-tail fashion. Molecular replacement was carried out with the program AMoRe (Navaza 1994), using an 81.2 Å-long search model which was generated from the structure refined in the subcell approximation (Vitagliano et al. 2001a). Probably due to the triple helix extensive noncrystallographic symmetry, molecular replacement was not of help in determining the axial positioning of the molecules in the asymmetric unit. The structure was solved by an unusual utilization of the *warpNtrace* approach, implemented in the ARP/wARP suite (Perrakis et al. 1999). The starting model in this strategy was an infinite triple helix parallel to the *c* axis. After ten rebuilding cycles, each followed by refinement cycles using the program REFMAC (Murshudov et al. 1999), triple helix terminations were clearly identified. Several randomly terminated models were also tested, and all of them converged to the same final model.

The model obtained from ARP/wARP contained only Ser and

Gly residues. As indicated by the electron density maps, Ser residues were mutated to Pro using the program O (Jones et al. 1991). The resulting model, which consisted of 153 residues, was refined using SHELX-L in a restrained mode (Sheldrick and Schneider 1997). R_{free}, calculated on 5% of the total number of reflections, was used throughout the refinement procedure. Potential water sites were identified using an automated water divining procedure, implemented in SHELX-L. Modeling of these sites was carried out after inspection of omit (Fo-Fc) electron density maps and was monitored using R_{free}. All of the missing terminal residues except the C-terminal glycines were built by alternating omit map calculations and manual building to SHELX-L refinement cycles. Although displaying higher atomic displacement factors, N-terminal Pro residues were substantially ordered. In contrast, for each of the two molecules in the asymmetric unit, C-terminal glycine residues display electron density only at the N atom of the two chains involved in interchain hydrogen bonds. On the other hand, no electron density is visible for the C-terminal glycines of the two most staggered chains (see inset in Fig. 6), for which interchain hydrogen bonds cannot be formed. Once the main features of the structure were modeled, an anisotropic treatment of the atomic displacement parameters was adopted. Introduction of anisotropy resulted in a decrease in R_{factor} and R_{free} of 0.05 and 0.03, respectively. During the anisotropic refinement, weak similarity restraints were applied to the corresponding U_{ij} components of atoms which were close in space (Table 2). The solvent water molecules were restrained to be approximately isotropic. All of the weights used for the restraints were monitored using R_{free}. The refinement converged with R_{factor}, calculated for all 21,970 positive reflections of 0.181 and R_{free} of 0.297. These values of R_{factor} and R_{free} are higher than normally observed at this resolution. This is most likely determined by the large proportion of weak reflections (41.6% of the whole set of reflections with $I/\sigma(I) < 2$), which is due to the extensive noncrystallographic symmetry. Since the diffraction pattern was characterized by very bright and very weak reflections, an R_{factor} for each of the nine reflection classes with a different l index was calculated (Table 3). The values of R_{factor} confirm the correctness of the final model. Indeed, among the reflections which were not used in the subcell approximation (reflection index $l \neq 9n$), the classes with $l = 9n + 2$ and $l = 9n + 4$ have an R_{factor} comparable with that having $l = 9n$. Namely, as expected for well refined structures, R_{factor} depends mainly on reflection intensities rather than on the specific class to which the reflections belong. Indeed, higher R_{factor} values correspond to reflections characterized by lower intensities (Table 3).

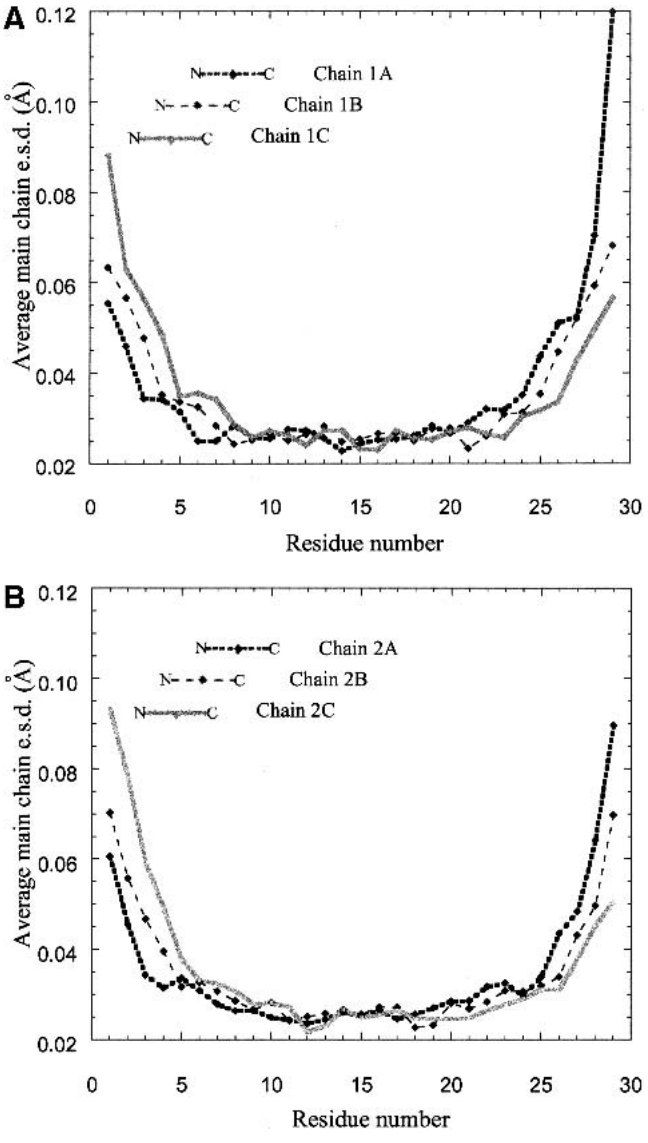

Fig. 6. Average main chain e.s.d. values versus residue number for the three chains of molecule 1 (*A*) and molecule 2 (*B*) in the asymmetric unit. The *insets* show the stagger of the three chains in each triple helix.

Table 3. *Distribution of <I/σ(I)> and R_{factor} in the nine reflection classes*

Reflection index 1	<I/σ(I)>	R_{factor}
9n	10.2	0.150
9n+1	2.2	0.281
9n+2	8.8	0.144
9n+3	1.3	0.344
9n+4	5.3	0.169
9n+5	2.0	0.228
9n+6	1.9	0.246
9n+7	2.4	0.197
9n+8	1.6	0.318

Atomic displacement parameters, averaged over all atoms, are 21.2 and 21.1 $Å^2$ for molecules 1 and 2 (Fig. 1A), respectively. Higher disorder characterizes the molecule terminal ends, as average B-factors of the N-terminal Pro-Pro-Gly triplets are 34.8 and 37.8 $Å^2$ for molecules 1 and 2, respectively. Similarly, B-factors averaged over the C-terminal Pro-Pro segments are 40.5 and 37.3 $Å^2$ for molecules 1 and 2, respectively. An analogous trend is followed by the estimated uncertainties in the atomic positions (Fig. 6), which were derived using block-matrix least-squares refinement implemented in SHELX-L (Sheldrick and Schneider 1997). For block-matrix refinement, the model was split into four overlapping blocks, each containing approximately the same number of atoms. Since a slight underestimation of the e.s.d. values may arise from the use of the restraints, most of them were removed during refinement, thus providing more reliable estimates of the errors. As shown in Figure 6, for both molecules in the asymmetric unit, average main chain e.s.d. values are close to 0.03 Å, whereas they rise significantly in the terminal regions. Furthermore, in all cases, highest e.s.d. values correspond to the most staggered single chains of the triple helices at both the N-terminus

(chains 1C and 2C in Fig. 6) and C-terminus (chains 1A and 2A in Fig. 6).

The pictures were generated using the programs MOLSCRIPT (Kraulis 1991), BOBSCRIPT (Esnouf 1999), O (Jones et al. 1991), GRASP (Nicholls et al. 1991), and RASTER3D (Merritt and Bacon 1997). Coordinates were deposited with the Protein Data Bank (Berman et al. 2000), code 1k6f.

Acknowledgments

We acknowledge EMBL/DESY, Sincrotrone Trieste Spa/Elettra, and CNR/Elettra for beam time and support. We thank Mr. Luca De Luca and Mr. Giosuè Sorrentino for technical assistance, Dr. L. Carotenuto (MARS) for discussions, and Dr. Antonello Mastrangelo for his help with the computations. This work was financially supported by the Agenzia Spaziale Italiana (ASI).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Bella, J., Brodsky, B., and Berman, H.M. 1995. Hydration structure of a collagen peptide. *Structure* **3:** 893–906.

Bella, J., Eaton, M., Brodsky, B., and Berman, H.M. 1994. Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. *Science* **266:** 75–81.

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.

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. D* **56:** 55–61.

Berman, H.M., Bhat, T.N., Bourne, P.E., Feng, Z., Gilliland, G., Weissig, H., and Westbrook, J. 2000. The Protein Data Bank and the challenge of structural genomics. *Nat. Struct. Biol.* **7:** 957–959.

Bretscher, L.E., Jenkins, C.L., Taylor, K.M., DeRider, M.L., and Raines, R.T. 2001. Conformational stability of collagen relies on a stereoelectronic effect. *J. Am. Chem. Soc.* **123:** 777–778.

Esnouf, R.M. 1999. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D* **55:** 938–940.

Esposito, L., Vitagliano, L., Sica, F., Sorrentino, G., Zagari, A., and Mazzarella, L. 2000. The ultrahigh resolution crystal structure of ribonuclease A containing an isoaspartyl residue: Hydration and sterochemical analysis. *J. Mol. Biol.* **297:** 713–732.

Fields, G.B. and Prockop, D.J. 1996. Perspectives on the synthesis and application of triple-helical, collagen-model peptides. *Biopolymers* **40:** 345–357.

Finney, J.L., Gellatly, B.J., Golton, I.C., and Goodfellow, J. 1980. Solvent effects and polar interactions in the structural stability and dynamics of globular proteins. *Biophys. J.* **32:** 17–33.

Fraser, R.D.B., MacRae, T.P., and Suzuki, E. 1979. Chain conformation in the collagen molecule. *J. Mol. Biol.* **129:** 463–481.

Holmgren, S.K., Bretscher, L.E., Taylor, K.M., and Raines, R.T. 1999. A hyperstable collagen mimic. *Chem. Biol.* **6:** 63–70.

Holmgren, S.K., Taylor, K.M., Bretscher, L.E., and Raines, R.T. 1998. Code for collagen's stability deciphered. *Nature* **392:** 666–667.

Improta, R., Benzi, C., and Barone, V. 2001. Understanding the role of stereoelectronic effects in determining collagen stability. I. A quantum mechanical study of proline, hydroxyproline and fluoroproline dipeptide analogues in aqueous solution. *J. Am. Chem. Soc.* (in press).

Inouye, K., Kobayashi, Y., Kyogoku, Y., Kishida, Y., Sakakibara, S., and Prockop, D.J. 1982. Synthesis and physical properties of (hydroxyproline-proline-glycine)10: Hydroxyproline in the X-position decreases the melting temperature of the collagen triple helix. *Arch. Biochem. Biophys.* **219:** 198–203.

Jones, E.Y. and Miller, A. 1991. Analysis of structural design features in collagen. *J. Mol. Biol.* **218:** 209–219.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. 1991. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47:** 110–119.

Kramer, R.Z., Bella, J., Mayville, P., Brodsky, B., and Berman, H.M. 1999. Sequence dependent conformational variations of collagen triple-helical structure. *Nat. Struct. Biol.* **6:** 454–457.

Kramer, R.Z., Venugopal, M.G., Bella, J., Mayville, P., Brodsky, B., and Berman, H.M. 2000. Staggered molecular packing in crystals of a collagen-like peptide with a single charged pair. *J. Mol. Biol.* **301:** 1191–205.

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.

Kramer, R.Z., Bella, J., Brodsky, B., and Berman, H.M. 2001. The crystal and molecular structure of a collagen-like peptide with a biologically relevant sequence. *J. Mol. Biol.* **311:** 131–147.

Kraulis, P.J. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24:** 946–950.

Melacini, G., Bonvin, A.M., Goodman, M., Boelens, R., and Kaptein, R. 2000. Hydration dynamics of the collagen triple helix by NMR. *J. Mol. Biol.* **300:** 1041–1049.

Merritt, E.A. and Bacon, D.J. 1997. Raster3D: Photorealistic molecular graphics. *Methods Enzymol.* **277:** 505–524.

Miller, M.H., Nemethy, G., and Scheraga, H. 1980. Calculation of the structures of collagen models. Role of interchain interactions in determining the triple-helical coiled-coil conformation. 2. Poly(glycyl-prolyl-hydroxyprolyl). *Macromolecules* **13:** 470–478.

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.* **54:** 171–200.

Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S., and Dodson, E.J. 1999. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. D* **55:** 247–255.

Nagarajan, V., Kamitori, S., and Okuyama, K. 1998. Crystal structure analysis of collagen model peptide (Pro-Pro-Gly)10. *J. Biochem.* **124:** 1117–1123.

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.

Navaza, J. 1994. AMoRe, an automated package for molecular replacement. *Acta Crystallogr. A* **50:** 157–163.

Némethy, G., Gibson, K.D., Palmer, K.A., Yoon, C.N., Paterlini, G., Zagari, A., Rumsey, and S., Scheraga, H.A. 1992. Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in ECEPP/3 algoritm, with application to proline-containing peptides. *J. Phys. Chem.* **96:** 6472–6484.

Nicholls, A., Sharp, K.A., and Honig, B. 1991. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11:** 281–296.

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.

Okuyama, K., Nagarajan, V., and Kamitori, S. 1999. 7/2-helical model for collagen—Evidence from model peptides. *Proc. Indian Acad. Sci., Chem. Sci.* **111:** 19–34.

Okuyama, K., Takayanagi, M., Ashida, T., and Kakudo, M. 1977. A new structural model for collagen. *Polym. J.* **9:** 341–343.

Otwinowsky, Z. and Minor, W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276:** 307–326.

Perrakis, A., Morris, R., and Lamzin, V.S. 1999. Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6:** 458–463.

Persikov, A.V., Ramshaw, J.A., Kirkpatrick, A., and Brodsky, B. 2000. Amino acid propensities for the collagen triple-helix. *Biochemistry* **39:** 14960–14967.

Ramshaw, J.A., Shah, N.K., and Brodsky, B. 1998. Gly-X-Y tripeptide frequencies in collagen: A context for host-guest triple-helical peptides. *J. Struct. Biol.* **122:** 86–91.

Rich, A. and Crick, F.H. 1961. The molecular structure of collagen. *J. Mol. Biol.* **3:** 483–506.

Sheldrick, G.M. and Schneider, T.R. 1997. SHELXL: High-resolution refinement. *Methods Enzymol.* **277:** 319–343.

Vitagliano, L., Berisio, R., Mazzarella, L., and Zagari, A. 2001a. Structural bases of collagen stabilization induced by proline hydroxylation. *Biopolymers* **36:** 459–464.

Vitagliano, L., Berisio, R., Mastrangelo, A., Mazzarella, L., and Zagari, A. 2001b. Preferred proline puckerings in cis and trans peptide groups: Implications for collagen stability. *Protein Sci.* (in press).

Yonath, A. and Traub, W. 1969. Polymers of tripeptides as collagen models. IV. Structure analysis of poly(L-proly-glycyl-L-proline). *J. Mol. Biol.* **43:** 461–477.