Hydration structure of a collagen peptide

Jordi Bella[1], Barbara Brodsky[2] and Helen M Berman[1]*

[1]Department of Chemistry, Rutgers University, Piscataway, NJ 08855, USA and
[2]UMDNJ-Robert Wood Johnson Medical School, Piscataway, NJ 08854, USA

Background: The collagen triple helix is a unique protein motif defined by the supercoiling of three polypeptide chains in a polyproline II conformation. It is a major domain of all collagen proteins and is also reported to exist in proteins with host defense function and in several membrane proteins. The triple-helical domain has distinctive properties. Collagen requires a high proportion of the post-translationally modified imino acid 4-hydroxyproline and water to stabilize its conformation and assembly. The crystal structure of a collagen-like peptide determined to 1.85 Å showed that these two features may be related.
Results: A detailed analysis of the hydration structure of the collagen-like peptide is presented. The water molecules around the carbonyl and hydroxyprolyl groups show distinctive geometries. There are repetitive patterns of water bridges that link oxygen atoms within a single peptide chain, between different chains and between different triple helices. Overall, the water molecules are organized in a semi-clathrate-like structure that surrounds and interconnects triple helices in the crystal lattice. Hydroxyprolyl groups play a crucial role in the assembly.
Conclusions: The roles of hydroxyproline and hydration are strongly interrelated in the structure of the collagen triple helix. The specific, repetitive water bridges observed in this structure buttress the triple-helical conformation. The extensively ordered hydration structure offers a good model for the interpretation of the experimental results on collagen stability and assembly.

Structure 15 September 1995, **3**:893–906
Key words: collagen, hydration, hydroxyproline, triple helix, water bridges

Introduction

The triple helix is a protein motif found in fibril-forming collagens, a range of other extracellular matrix proteins, a series of host defense proteins, and at least three membrane proteins [1–5]. The basic triple-helical conformation consists of three close-packed supercoiled polyproline II helices, which requires the occurrence of glycine residues at every third position in the polypeptide sequence and the presence of a high content of imino acids. This results in an $(X–Y–Gly)_n$ repeating pattern in which the X and Y positions are frequently occupied by proline and 4-hydroxyproline (Hyp) residues, respectively. These strict requirements are landmarks of collagen-like domains in sequence analyses of proteins of unknown structure. The high content of sterically restricted imino acids stabilizes the extended nature of the individual chains, and the post-translational modification of proline in the Y position to Hyp has been shown to confer significant additional stability [1]. The highly extended nature of the triple helix, where every X and Y residue is substantially exposed to solvent, appears to make the triple-helical domain important for self-association and binding other molecules.

Triple-helical domains are found in a striking variety of supramolecular aggregates ranging from the characteristic collagen fibrils with a 67 nm axial repeat, to antiparallel forms of type VII collagen, to network forms seen in basement membranes, and to the parallel clusters of six or more triple helices seen in the complement component C1q or mannose-binding protein [1,2]. A large variety of ligands, including glycosaminoglycans, enzymes, and phospholipid vesicles, can bind to type I collagen with varying degrees of specificity. In addition, the triple-helical domain has been identified as the ligand-binding region of the macrophage scavenger receptor and the region which binds the C1q receptor [2,5,6]. Triple helices are intrinsically rod-like domains. However, kinks and other perturbations have been reported at points where the $(X–Y–Gly)_n$ repeating pattern is interrupted. Such interruptions, resulting from mutations in fibril-forming collagens or mannose-binding proteins, have been related to the development of pathological conditions [2,5,7].

Two key features of collagen are its high content of Hyp and its unique interaction with water. In spite of the accumulation of considerable amounts of experimental data, the structural rationale for these features has remained elusive, partly because of the lack of a high-resolution structure.

Triple-helical domains in collagenous and non-collagenous proteins contain significant amounts of the unusual imino acid Hyp. Studies have demonstrated that Hyp residues play a critical role in stabilizing the triple-helical conformation, both in collagen [1,8,9] and in synthetic peptides with collagen-like sequences [10,11]. A number of explanations have been proposed to account for its stabilizing effect. Even before the correct triple-helical models were proposed, Gustavson [12] suggested that Hyp residues stabilize collagen by participating in direct interchain hydrogen bonding. Once the collagen model was proposed from fiber diffraction data [13], it became apparent that steric considerations would not permit direct hydrogen bonding between the Hyp hydroxyl groups and peptide carbonyls in the same chain, between different chains in the same molecule, or between

*Corresponding author.

different molecules. Alternative models were proposed based on the formation of hydrogen bonds that involved the participation of water molecules as well as the Hyp hydroxyl group [9–11,14,15].

It was recognized early on that water plays a role in maintaining the conformation of the native collagen molecule (for a review, see [16]). A variety of techniques and measurements, including NMR, dielectric measurements, water sorption, longitudinal acoustic velocity, dynamic mechanical spectroscopy and heat capacity, indicate that the structure of water in collagen fibrils is very different from bulk water in terms of restricted mobility and asymmetric orientation [17–22]. These results have been interpreted in some cases to mean that water molecules are tightly bound to specific sites on collagen chains, while others have proposed the concept of water chains filling in the spaces between molecules. In the Rich and Crick II (RCII) triple helix structure [13], only one hydrogen bond is formed directly between amide groups (Gly:NH···O=C:X) for each tripeptide unit. This leaves two carbonyl groups and any amide groups present in the X or Y positions (when these residues are not imino acids) available for hydrogen bonding, in contrast to the complete hydrogen bonding of all backbone amide groups observed in α helices or β sheets. On the basis of model building, specific water bridges were proposed linking carbonyls on different chains within a molecule, or linking amide and carbonyl groups of different chains [14,15,23]. X-ray diffraction analysis of the polymer-like crystals of the peptide $(Pro–Pro–Gly)_{10}$ revealed possible hydration sites that are bound to carbonyl groups in the peptide and follow the helical conformation [24].

The inability of the Hyp hydroxyl to form direct hydrogen bonds with backbone carbonyl groups together with the data supporting specific binding of water molecules to collagen chains led to the idea that stabilization of the collagen triple helix by Hyp residues took place through intramolecular water bridges, involving those carbonyl and amide groups which were not participating in the interchain RCII hydrogen-bonding pattern [14,15]. The possible location of ordered water molecules was the subject of speculation, but without structural data an accurate atomic description was not possible.

We have reported the crystal structure of a triple-helical peptide with a collagen-like sequence containing a Gly→Ala substitution: $(Pro–Hyp–Gly)_4–Pro–Hyp–Ala–(Pro–Hyp–Gly)_5$ [25]. This crystal structure has provided the first precise experimental proof for the existence of water bridges in a collagen-like triple-helical molecule, and has allowed us to determine the locations of ordered water molecules. Triple helices in this crystal structure are surrounded by an extensive cylinder of hydration, and Hyp residues seem to act as 'keystones' connecting the water network to the peptide molecules. In addition, the interruption of the Pro–Hyp–Gly pattern by a Gly→Ala substitution results in a local substitution of the RCII hydrogen-bonding pattern by single water bridges that interconnect the peptide chains. These water molecules, bridging amide and carbonyl groups at the site of substitution, are called interstitial waters [25]. The solvent structure around a given triple helix goes beyond the immediate neighborhood and reaches adjacent triple helices in the crystal. This suggests a role for hydration in the process of molecular association leading to the formation of collagen assemblies. Long-range attractive forces between collagen triple helices have been measured in solution by Parsegian and co-workers [26], who have proposed that these forces may be responsible for spontaneous assembly of collagen.

In view of the important consequences of the hydration structure both in collagen triple-helical stability and molecular association, we have performed an extensive refinement of the solvent structure around the peptide in the crystal. Here we present a detailed analysis of repetitive hydration patterns and specific interactions. Several aspects are considered: selective hydration of candidate groups for hydrogen bonding in the peptide molecule; identification and characterization of repeating hydration motifs around the triple helix; characterization of specific hydration patterns at the zone surrounding the Gly→Ala substitution; and identification of water clusters and solvent networks interconnecting neighboring triple helices in the crystal. The possible implications of these patterns in collagen triple-helical stability, their generalization to other collagen sequences, and their transferability to collagen fibrillar assemblies are discussed.

Results

Global statistics for the water molecules

The current model for the Gly→Ala peptide at 1.85 Å resolution is divided into zones as described in [25]: termini zones, the collagen zone and the substitution zone. The 141 water molecules and 7 acetic acid molecules, which constitute the solvent zone, account for 27% of the total content of the unit cell. Without a definitive value for the density, it is impossible to know how much water is missing. All the solvent molecules that have been included are at suitable distances (<3.25 Å) and appropriately oriented to form hydrogen bonds to at least one immediate neighbor, and in fact, most of them have more that one hydrogen-bonded neighbor. Most waters have three (46%) or four (33%) neighbors.

Further classification of the water (W)···peptide hydrogen bonds yields 66 W···O=C hydrogen bonds, 50 W···Oδ–Cγ hydrogen bonds to the hydroxyl groups in Hyp residues, and 4 W···H–N hydrogen bonds (coming only from the interstitial waters). In addition, seven hydrogen bonds involving acetic acid molecules, which appear as a consequence of the crystallization conditions, have been identified, and seem to participate in the same kinds of hydrogen bonds as water molecules: five acetic···O=C and two acetic···Oδ–Cγ hydrogen bonds. In most of the discussion that follows they will be included among the water interactions.

Selective hydration around the carbonyl groups in the peptide

There are three distinctive carbonyl groups per tripeptide unit. Proline carbonyl groups only participate in the interchain N–H···O=C hydrogen bonding and do not exhibit hydration, except for those four carbonyl groups connected to interstitial waters in the Gly→Ala substitution zone (Pro46, Pro73, Pro76 and Pro13 following the nomenclature of [25]). Even without including those from prolyl residues, carbonyl groups are proportionally less hydrated than Hyp hydroxyl groups. On average, carbonyl groups from residues other than proline participate in 1.1 hydrogen bonds with solvent molecules — 0.9 from glycine and alanine residues and 1.3 from Hyp residues. Because carbonyl groups can only act as acceptors, there seems to be some tendency for Hyp carbonyl groups to receive hydrogen bonds from more than one solvent molecule. Figure 1 shows the observed distribution of solvent sites around the Hyp and glycine carbonyl groups. It is obvious that these two residues have differential hydration arising from their different positions in the basic Pro–Hyp–Gly repeat. Hyp carbonyl groups often show double hydrogen bonding (14 doubles, 11 singles and 5 zeros), whereas glycine carbonyl groups adopt almost exclusively single hydration.

Carbonyl groups from Hyp residues have two hydration sites: W_A and W_N (Fig. 1a). Only one hydration site (W_N) appears for carbonyl groups of glycine (Fig. 1b). The potential W_A sites for glycine carbonyl groups are occupied by Cα atoms from a neighboring chain in the triple helix, and hence no water molecules can be placed there (the only exception occurs at the Gly→Ala substitution site, where the triple helix has a slightly untwisted conformation [25]).

Average values for W···O=C distances and angles are consistent with water···carbonyl hydrogen bonding (Table 1). They are similar to the values obtained for globular proteins [27], although overall they show slightly shorter W···O distances and more linear W···O=C angles. The positioning of the waters with respect to the plane of the carbonyl groups is represented by the torsion angle Cα–C–O–W for W_A waters, and the torsion angle N–C–O–W for W_N waters.

Selective hydration around the hydroxyl group from hydroxyproline residues

The hydroxyl group Cγ–Oδ from Hyp residues has both hydrogen-bond donor and acceptor properties; thus, double hydration would be expected. In fact, there are 2 hydroxyl groups with triple hydration, 19 doubles, 8 singles and 1 zero (due to its location at a terminal region of the peptide). On average, hydroxyl groups from Hyp residues are bonded to 1.7 solvent molecules. As shown in Figure 1c, there are two major clusters around the Cγ–Oδ bond: W_{D2} is in close proximity to Cδ, oriented approximately *gauche*$^+$ (*g*+) with respect to this atom in the hydroxyprolyl ring; W_B is closer to Cβ and is roughly in a *gauche*$^-$ (*g*$^-$) orientation from Cβ. There is an additional minor cluster, W_{D1}, approximately *cis* with respect to Cδ, too close to W_{D2} to allow simultaneous occupation of both water sites. This double preference for solvent molecules near Cδ is a consequence of two different water-bridging patterns that will be discussed later.

Average hydrogen-bonding angles for the Cγ–Oδ···W hydrogen bonds (Table 1) differ from the W···O=C values, consistent with the single bond character of Cγ–Oδ. Corresponding angles for serine and threonine in globular proteins peak around 118° [28]. The small differences between the various hydration sites come mostly from the water bridge geometry in which their waters are involved, and probably are not significant in terms of their possible donor or acceptor character.

Repeating hydration motifs: water bridges

The term water bridge refers to those associations of hydrogen bonded water molecules that link two different groups capable of hydrogen bonding in the triple helix. A generic water bridge can then be described with a

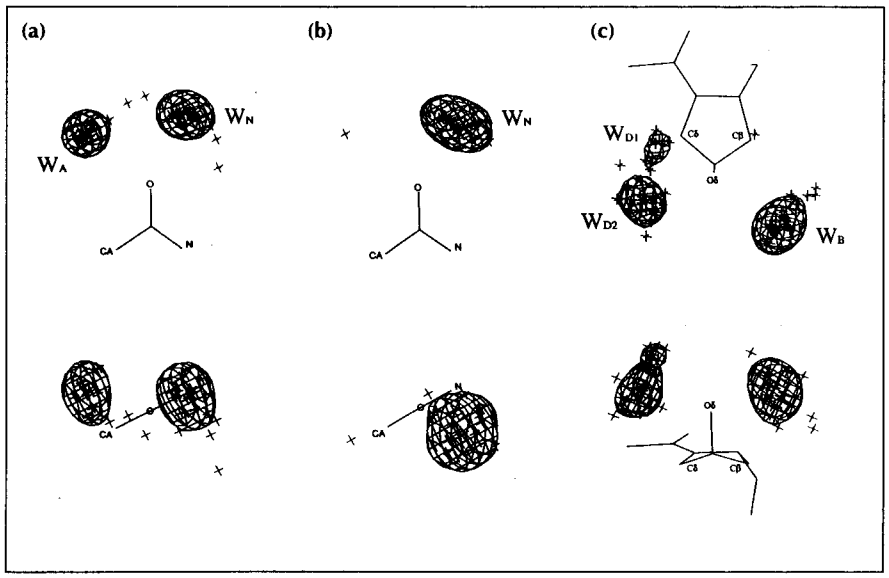


Fig. 1. Orthogonal views of the distribution of water molecules around the carbonyl groups of **(a)** Hyp and **(b)** glycine (or alanine) residues. Waters have been selected by their distance to the carbonyl oxygen atoms, applying a cut-off of 3.25 Å. A few oxygen atoms from acetic acid molecules have been included because they also fulfill this criterion. Three-dimensional contours have been calculated using the method of Schneider *et al.* [42], which is a modification of the method of Murray-Rust and Glusker [43]. Water sites are labeled according to their proximity to atoms in the N–C(=O)–Cα group: W_A site for the Cα atom and W_N site for the N atom. **(c)** Orthogonal views of the distribution of water molecules around the hydroxyl group from Hyp residues. These water sites are called W_B and W_D and are named for the Cβ and Cδ atoms in the hydroxyprolyl ring, respectively.

Table 1. Average hydrogen bonding for the individual hydrated groups in the crystal structure of the Gly→Ala peptide.

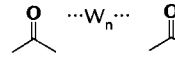
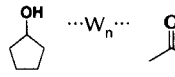
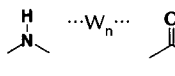
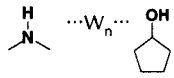
	W···O (Å)	W···O–C (°)	X–C–O···W dihedral angles (°) Value	Defining atoms	No. of cases
Gly C=O					
Overall	2.90±0.14	144±11			27
W_N	2.88±0.12	144±11	–57±23	N–C–O–W	24
Hyp C=O					
Overall	2.84±0.16	141±15			39
W_N	2.85±0.16	147±13	–12±30	N–C–O–W	23
W_A	2.84±0.16	133±13	–40±16	Cα–C–O–W	16
Hyp Cγ–Oδ					
Overall	2.86±0.14	114±15			52
W_B	2.88±0.12	112±13	–54±37	Cβ–Cγ–Oδ–W	26
W_D	2.83±0.15	116±18	42±32	Cδ–Cγ–Oδ–W	26

sequence X···W(···W)$_n$···Y, where X and Y are the hydrogen-bonding groups from the triple helix. These groups will be referred to collectively as anchoring groups, and include all carbonyl groups, hydroxyl groups from Hyp residues, and free amide groups from glycine and alanine residues (only those free amide groups from the interstitial zone are involved in any water bridge). Unless stated otherwise, the water bridges discussed below are characterized by distances between non-hydrogen atoms <3.25 Å, and angles X(W)–W–W(Y) >60°. Water bridges are categorized as follows: intrachain, connecting anchoring groups in the same peptide chain; interchain, connecting anchoring groups in different peptide chains from the same triple helix; and intermolecular, connecting triple-helical molecules related by the symmetry elements of the crystal lattice.

In order to describe bridges connecting different chains we have adopted the following convention. Interchain connectivity going in the same direction as the N–H···O=C hydrogen bonds will be referred to as the 1→2→3→1 direction, by analogy to the RCII hydrogen-bonding model; interchain connectivity going in the opposite direction will be referred to as the 1→3→2→1 direction.

Table 2 shows the different types of water bridges identified in the Gly→Ala crystal structure. The α and β bridges have no *a priori* sequence requirements, because they only involve carbonyl groups and could, in principle, appear connecting any two residues in a triple helix; γ and δ bridges require one of the groups to have a hydroxyl group (Hyp, Ser, Thr), and the other, a non-specific carbonyl group. The letters ϵ, ζ and η are used here to describe the interstitial water bridges connecting amide groups with others in the Gly→Ala substitution zone, but they could also be used to designate hypothetical water bridges connecting carbonyl groups with amide groups from non-imino acid residues in the X or Y position of the triple helix. Finally, the letter ω designates water bridges connecting neighboring triple helices in the crystal. These are less likely to participate in repeating patterns as they are strongly affected by the particular symmetry and dimensions of the crystal lattice.

Table 2. Nomenclature used for the different water bridges observed in the crystal structure of the Gly→Ala peptide.

Bridge*	Connectivity	Name*	Requirement for residue 1	residue 2
	Intrachain Interchain	αn βn	None	None
	Intrachain Interchain	γn δn	Hyp	None
	Intrachain Interchain	ϵn ζn	No Pro	None
	Intrachain	ηn	No Pro	Hyp
Any ···W$_n$··· Any	Between triple helices	ωn	None	None

*The 'n' index refers to the number of waters participating in a particular bridge.

A topological view of the most repetitive hydration motifs found in the crystal structure of the Gly→Ala peptide is shown in Figures 2 and 3. Only bridges of a given class are represented in each diagram. In fact, no water bridge is isolated and most of the water molecules act as nodes from which the water network extends.

Carbonyl–carbonyl water bridges: α and β types

Eleven α3 bridges and three α2 bridges have been identified. An example of an α3 bridge is shown in Figure 4a. Inspection of adjacent Hyp–Gly carbonyl pairs throughout the peptide seems to indicate that α3 bridges come from an optimal interaction of two of the hydration sites of these carbonyls connected through a capping water molecule. It seems, however, that only in cases where there is no steric interference from neighboring triple helices can the bridge be completed. Either hydration site at the Hyp carbonyl group seems to be able to participate in α3 bridges. Three additional α2 bridges can be thought of as α3 bridges for which the capping water is missing. They have been included in the list because their water–water distance is <3.25 Å and their hydrogens can be built with reasonable hydrogen-bonding geometry. However, the low number of cases observed indicates that they are not a dominant feature in this peptide. Figure 2a also includes two 'interrupted' bridges which contain non-water molecules. For example an α-like bridge between O_8 and O_9 is disrupted by an interaction with a symmetry-related triple helix which involves an ω1 bridge between O_8 and $O\delta_{41}$ and an ω2 bridge between O_9 and $O\delta_{41}$. The result is a pseudo-α4 bridge in which one of the corners is the hydroxyl group from Hyp41 in a neighboring triple helix. A similar situation is observed between O_{83} and O_{84} where the hydroxyl group from Hyp23 in a symmetry-related triple helix also introduces an interruption in the α pattern. The resulting geometry is comparable to that observed for O_8···O_9.

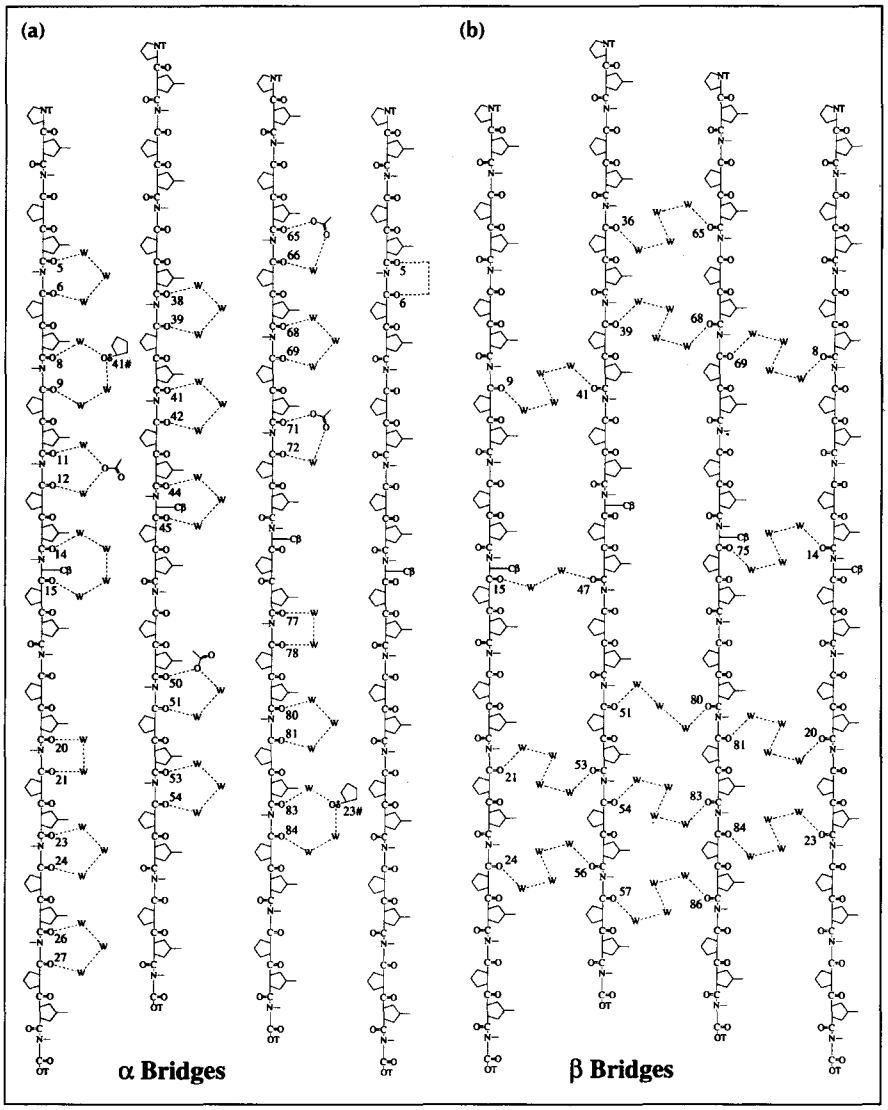


Fig. 2. Carbonyl–carbonyl water bridges in the collagen zones of the Gly→Ala peptide [25]. The triple helix is shown in open cylindrical projection with, from left to right, chains 1, 2, 3 and 1 repeated on the right to provide a clearer description of the chain 3→chain 1 interchain water bridges. The hash (#) symbol indicates a residue from a symmetry-related triple helix. **(a)** Intrachain water bridges connecting two carbonyl groups (α bridges). **(b)** Interchain water bridges connecting two carbonyl groups (β bridges). OT and NT are terminal atoms.

β bridges are especially interesting because they are interchain and do not involve Hyp residues. Most of the β bridges identified in this crystal structure are β4 (Fig. 2b), with a few exceptions: one β2 bridge connects O_{47} and O_{15}, although each one of its three oxygen–oxygen distances is quite long (3.15 Å, 3.21 Å and 3.11 Å); another β2 bridge between O_{15} and O_{46} (not shown in Fig. 2b) involves one of the interstitial waters and will be discussed later. Only one β3 bridge, between O_{51} and O_{80}, has been identified. It is shown in Figure 4a, where its fivefold-looking geometry is similar to that of an adjoining α3 bridge. The remaining cases show β4 connectivity. There is no clear correlation between the number of water molecules in a β bridge and the interatomic distance between its anchoring points. Measured carbonyl oxygen–carbonyl oxygen distances for the β bridges average 5.23±0.35 Å, in the range 4.84–6.10 Å. Paradoxically, the longest distance corresponds to a β2 bridge.

It is unclear whether or not β4 bridges are an intrinsic repetitive feature of hydrated triple helices or a consequence of the water network in this crystal structure. They seem to form interrupted filaments of water molecules that wrap around the triple helix (Fig. 4b), and, as a result, are exposed. Their geometry is also less well defined and two cases depicted in Figure 2b have one water–water distance longer than 3.5 Å. In spite of that, their repetitiveness may be indicative of a preferential hydration pattern.

Intrachain γ bridges

Intrachain bridges between Hyp hydroxyl and backbone carbonyl groups are one of the dominant features in the repetitive patterns of hydration in this crystal structure (Fig. 3a). γ bridges connect Hyp hydroxyl groups with the preceding glycine carbonyl groups in the same peptide chain. Glycine carbonyl groups participate with their only water site, W_N, whereas Hyp hydroxyl groups use their W_{D1} or W_{D2} sites, depending on the number of water molecules in the bridge. The occurrence of γ bridges seems to be rather independent of local distortions of the triple-helical geometry, consistent with their intrachain nature, and they are the only type of water bridge that appears relatively unaffected by the Gly→Ala substitution. Of 23 identified γ bridges, there are fourteen γ2, seven γ1 and only two γ3 bridges. The reasons

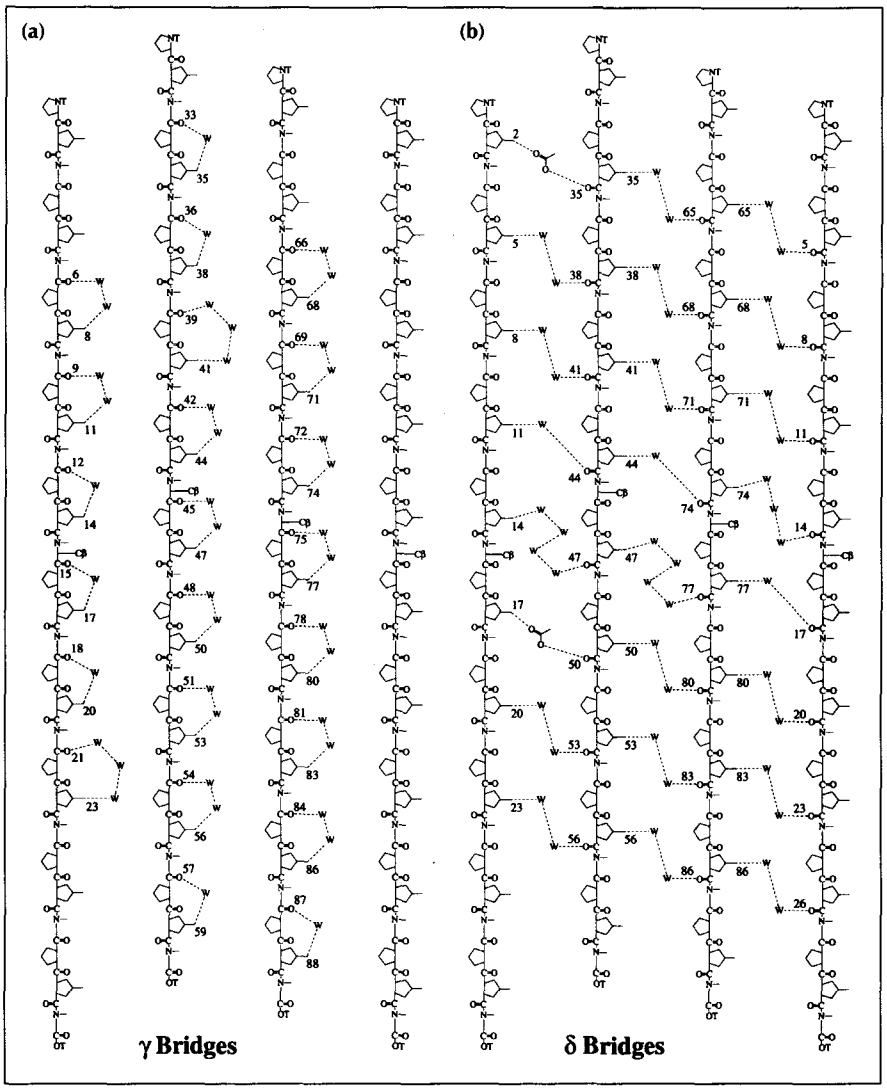


Fig. 3. Water bridges involving the hydroxyl group of Hyp and carbonyl groups. **(a)** Water bridges connecting carbonyl groups with hydroxyl groups from Hyp residues in the same chain (γ bridges). **(b)** Water bridges connecting carbonyl groups with hydroxyl groups from Hyp residues in the preceding chain, following a 1→2→3→1 connectivity (δ bridges). Longer δ3 and δ4 bridges occur in the Gly→Ala substitution zone and have been included for comparison.

for the particular number of waters observed in each case are not obvious, although there seems to be a certain preference for γ1 bridges at the chain termini, and at the Gly→Ala substitution site in chain 1.

Figure 4c shows a hydroxyl group simultaneously engaged in a γ2 bridge and a δ2 bridge. The geometry of the γ bridges is relatively uniform (Table 3). A set of dihedral angles can be used to characterize the pseudo-conformation of γ bridges. γ2 bridges can be divided into two groups, $s^-g^+g^+$ and g^-cs^+, using the notation commonly adopted for conformational angles in a polyatomic chain for the central τ dihedral angles (as defined in Table 3). γ1 bridges are more irregular, with an approximate g^-g^+ pseudo-conformation around the single water molecule. γ2 bridges (and also γ3 bridges) use the W_{D2} site from the Hyp hydroxyl group, whereas γ1 bridges use the W_{D1} site.

Measured distances for the 14 intrachain pairs O···Oδ connected by γ2 bridges average 5.32±0.21 Å, in the range 5.07–5.67 Å. The seven O···Oδ pairs linked by γ1 bridges are separated by 4.88±0.38 Å on average, and within the range 4.40–5.49 Å. The distance of 5.49 Å is unusually long for a single-water bridge, and indeed the two hydroxyl–water distances are long, being 3.20 Å and 3.17 Å. This particular bridge is found between O_{87} and $O\delta_{89}$, well into the C-terminal region of the peptide, and its geometry may be affected by the existing disorder in that zone. If this $O_{87}\cdots O\delta_{89}$ bridge is excluded, the average and range for the six remaining γ1 bridges become 4.78±0.38 Å and 4.40–5.11 Å, respectively.

The two O···Oδ pairs connected by γ3 bridges are separated by 5.36 Å and 5.24 Å, values which are typical of two-water bridges. From their location in the peptide there is no obvious reason for them having three waters, and probably the local hydration geometry will favor this instead of the more frequent γ2 connectivity.

Interchain δ bridges

The δ2 bridge is the dominant interchain feature in the collagen zones of this peptide. This bridge connects Hyp hydroxyl groups with Hyp carbonyl groups on the next chain following a 1→2→3→1 directionality. It involves the flanking residues of those already connected by direct interchain N–H···O=C hydrogen bonds: given a hydrogen-bonded pair $Gly_i\cdots Pro_j$, the δ2 bridge is formed between Hyp_{i-1} and Hyp_{j+1}. δ2 bridges are amazingly

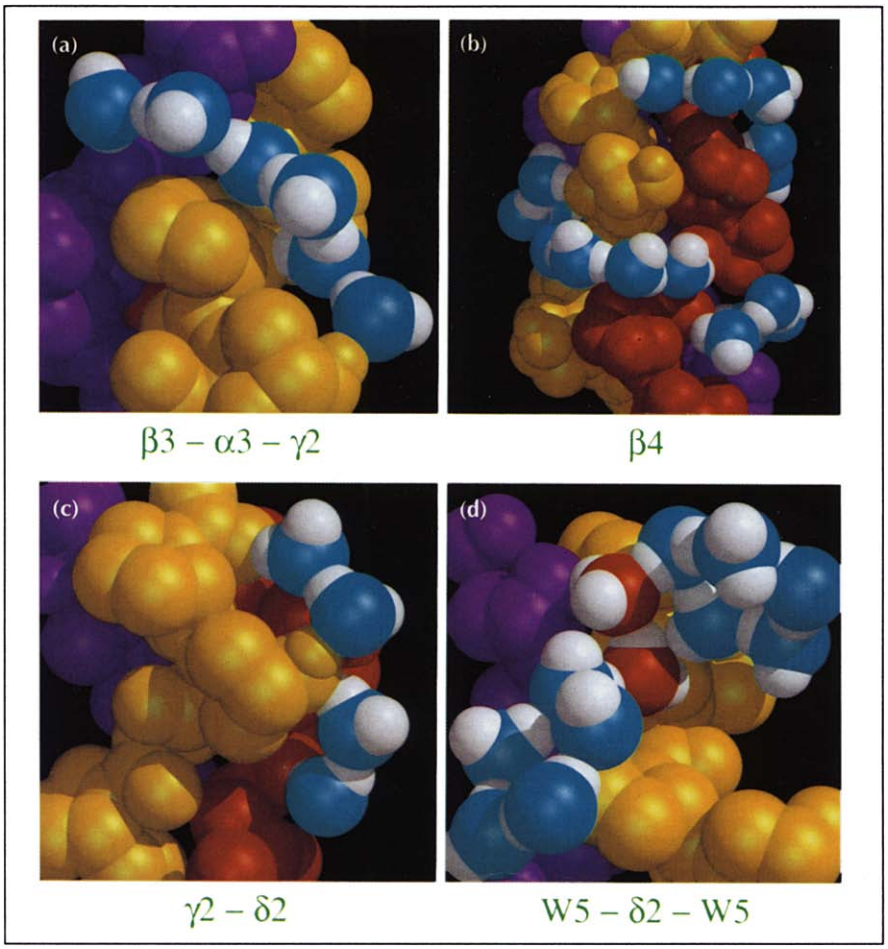


Fig. 4. Space-filling representations for several examples of water bridges observed in the crystal structure of the Gly→Ala peptide. The positions of hydrogen atoms have been built according to stereochemical criteria. **(a)** A central α3 bridge interconnected with other water bridges. **(b)** Long β4 bridges establish a network of water filaments that surround the triple helix. [The scale of this figure is smaller than panels (a),(c) and (d).] **(c)** A Hyp residue participating in both a γ2 bridge and a δ2 bridge. The incorporation of a single hydroxyl group in the central Hyp ring causes local organization of the solvent into two water bridges. Consequently, a seven-membered hydrogen-bonded chain is generated between the two carbonyl groups. **(d)** A δ2 bridge, shown by water molecules with red (oxygen) and white (hydrogen) atoms. The two waters act as nodes to which two pentagonal clusters of waters are hydrogen bonded (designated as W5). The peptide chains are colored magenta, red and yellow. Water molecules are shown in cyan (oxygen atoms) and white (hydrogen atoms), with the waters involved in the δ2 bridge in (d) shown in red (oxygen) and white (hydrogen).

preserved except at the central region, where the Gly→Ala substitution introduces a local untwisting in the triple-helical conformation [25] and dramatic changes in this interchain hydration pattern. Figure 3b includes, for comparison, three longer water paths connecting corresponding hydroxyl and carbonyl groups in the substitution zone. $O\delta_{14}$ needs a δ4 bridge to reach O_{47}; similarly, $O\delta_{47}$ connects to O_{77} through another δ4 bridge; and finally, $O\delta_{74}$ connects with O_{14} through a δ3 bridge. In three cases, the same kind of connection is achieved with only one water molecule: the δ1 bridge. δ1 bridges occur between pairs immediately before or after the major disruption zone.

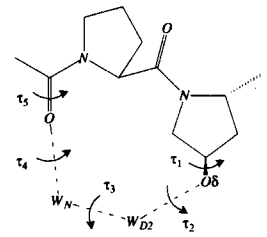
Finally, there are two cases in which an acetic acid molecule connects exactly the same groups, substituting for a δ water bridge. Interestingly, in both cases the acetic acid molecule bridges two groups in a zone where the predominant hydration patterns are likely to be distorted: one acetic acid connects the pair Hyp17···Hyp50 immediately after a δ4 bridge at the Gly15→Ala substitution; another acetic acid connects the pair Hyp2···Hyp35, adjacent to the N-terminal zone where the triple helix starts to unravel and the solvent structure is more subject to disorder.

Each of the two anchoring groups connected by δ bridges has two possible hydration sites. Nevertheless, δ bridges always occur between W_B from Hyp:Cγ–Oδ and W_N from Hyp:C=O (in δ1 bridges the same water occupies both sites). Furthermore, the geometry of the δ1 and δ2 bridges is relatively consistent along the crystal structure (Table 3), and the measured τ dihedral angles show a smaller dispersion than those for γ-type bridges. If we use the pseudo-conformational notation for the central τ dihedral angles, δ2 bridges are $g^+g^+s^-$, whereas δ1 bridges are g^+g^-.

Figure 4d shows an example of a δ2 bridge connected to two pentagonal clusters of water molecules. δ2 bridges themselves have a clear fivefold-like geometry [25]. This geometry is not so obvious for the δ1 bridges, and it seems that they can only form when two anchoring groups come close enough to allow a single water molecule to make the bridge. Measured distances between the 16 Oδ···O pairs connected by δ2 bridges average 5.31±0.20 Å, ranging from 4.93–5.63 Å. Corresponding distances for the three δ1 bridges average 4.88±0.12 Å, with a minimum of 4.74 Å and a maximum of 4.96 Å. These values suggest that δ1 bridges can only form when the Oδ···O distance is ≤5.0 Å, and that δ2 bridges can occur between Oδ···O groups separated by longer distances (up to 5.6 Å in this crystal structure). At the 4.9–5.0 Å interval, either kind of bridge seems to be possible and the choice will probably depend on the local hydration geometry. Similar values have been obtained

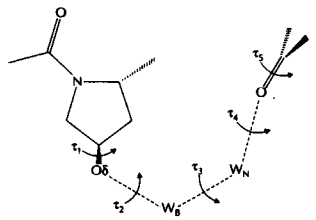
Table 3. Average hydrogen-bonding parameters for γ, δ and ζ bridges in the crystal structure of the Gly→Ala peptide.

γ2 bridges



Distances (Å)	Oδ···W_{D2}		W_{D2}···W_N		W_N···O
γ2, overall	2.79±0.12		2.82±0.13		2.85±0.10
Angles (°)	Cγ–Oδ–W_{D2}	Oδ–W_{D2}–W_N		W_{D2}–W_N–O	W_N–O=C
γ2, overall	117±16	108±20		118±17	145±10
Dihedrals (°)†	τ1	τ2	τ3	τ4	τ5
γ2, $s^-g^+g^+$	47±07	−120±13	63±17	71±22	−72±14
γ2, $g^-c\ s^+$	38±23	−72±25	2±19	109±16	−72±16

δ2 bridges



Distances (Å)	Oδ···W_B		W_B···W_N		W_N···O
δ2, overall	2.85±0.11		2.86±0.14		2.80±0.15
Angles (°)	Cγ–Oδ–W_B	Oδ–W_B–W_N		W_B–W_N–O	W_N–O=C
δ2, overall	112±12	99±26		108±11	150±07
Dihedrals (°)	τ1	τ2	τ3	τ4	τ5
δ2, overall	−70±06	42±15	65±13	−148±13	5±14

γ1 and δ1 bridges

Distances (Å)	Oδ···W		W···O	
γ1, overall	2.87±0.15		2.93±0.17	
δ1, overall	2.90±0.18		2.80±0.11	
Angles (°)	Cγ–Oδ···W	Oδ···W···O	W···O=C	
γ1, overall	117±13	115±11	146±12	
δ1, overall	109±21	118±04	146±09	
Dihedrals (°)	τ1	τ2	τ4	τ5
γ1, overall	3±19	−44±33	79±26	−39±16
δ1, overall	−43±5	83±19	−61±46	−43±25

ζ1 bridges*

Distances (Å)	H···W	N···W	W···O
ζ1, overall	1.87±0.10	2.82±0.11	2.75±0.10
Angles (°)	N–H···W	N···W···O	W···O=C
ζ1, overall	164±05	110±10	134±13

*Hydrogens for the amide groups in ζ1 bridges have been built with standard geometry. †g^+=~60°, g^-=~−60°, c=~0°, s=~120°, s^-=~−120°. This nomenclature refers to the torsion angles of τ2, τ3 and τ4.

for γ bridges, which indicates that there is some correlation between the number of waters in the bridge and the distance between anchoring groups, at least for the two types of Hyp hydroxyl centered bridges.

The two acetic acid molecules substituting for equivalent δ bridges connect groups separated by 4.97 Å and 5.29 Å, which suggests that they can replace either δ2 or δ1 bridges. At the Gly→Ala substitution site, distances between the corresponding anchoring groups are far from the observed distances for δ2 bridges, and indeed the number of water molecules needed to connect them is higher. $O\delta_{74}$ and O_{14} are separated by 7.4 Å and they are connected by three water molecules. Distances between the other pairs, $O\delta_{14}\cdots O_{47}$ and $O\delta_{47}\cdots O_{77}$, are even longer, 9.1 Å and 7.7 Å respectively, and they need four water molecules each to connect.

Hydration patterns at the Gly→Ala substitution zone: interstitial waters

At the Gly→Ala substitution zone the alanyl methyl groups prevent the formation of direct interchain N–H···O=C hydrogen bonds. Instead, four water molecules 'sneak' in between the polypeptide chains and connect the equivalent groups allowing the RCII hydrogen-bonding scheme to resume after the disrupted zone [25]. These water bridges can be designated as ζ1 following the notation proposed in Table 2. The hydrogen-bonding geometry around the interstitial waters W_{IA} and W_{IB} (Fig. 5) suggests that these waters only connect one amide group with one carbonyl group. They are tetrahedrally coordinated to other water molecules and nucleate additional water bridges among the surrounding groups. W_{IC} is isolated from other water molecules and buried in a cavity created by the local untwisting of the three chains. It establishes hydrogen bonds with all three chains, acting as a donor to both O_{13} and O_{73} and as an acceptor from N_{45}. The last interstitial water, W_{ID}, acts as an acceptor from N_{75} and is within hydrogen-bonding distance of both O_{13} and O_{14}. If hydrogen atoms are placed in this water molecule keeping a tetrahedral disposition around the water oxygen atom, one of the hydrogen atoms lies perfectly on the plane $O_{13}\cdots W_{ID}\cdots O_{14}$, strongly suggesting the formation of a three-centered hydrogen bond. Hydrogen bonding-distances and angles for the ζ1 bridges are shown in Table 3.

The water molecules hydrogen bonded to the interstitial ones give rise to a local network in which other anchoring groups are involved. This allows three proline carbonyl groups to participate in the formation of several water bridges (Fig. 5). Their connectivity is similar to

that observed in the other zones of the peptide, but there are some new patterns. For example, two α3 bridges and one α2 bridge connect proline carbonyl groups with the next Hyp carbonyl group in a similar way to the previously described Hyp:C=O···O=C:Gly α bridges. Analogously, the probable three-centered hydrogen bond (O_{13}···W_{ID}···O_{14}) can be classified as the only example of an α1 bridge in this crystal structure.

Following the water paths in Figure 5, three new β bridges can be found involving proline carbonyl groups. A β2 bridge interconnects O_{15} with O_{46} through a water bridge that includes W_{IA}. These β bridges connect carbonyl groups placed at different relative vertical levels in the triple helix, thus differing from the previously described β bridges (Fig. 2b) between glycine and Hyp carbonyl groups which are at the same relative level.

Some new types of intrachain bridges arise from the hydrogen bonding of amide groups to water molecules (Fig. 5). Three ϵ bridges connect an amide group with a carbonyl group in the same chain: an ϵ2 bridge between NH_{15} and O_{15}; an ϵ3 between NH_{48} and O_{45}; and an ϵ4 between NH_{75} and O_{75}. Similarly, some η bridges can be described: an η3 between NH_{15} and $O\delta_{14}$; an η4 between NH_{15} and $O\delta_{17}$; an η2 between NH_{48} and $O\delta_{47}$; and an η3 between NH_{75} and $O\delta_{74}$.

All of these bridges are a consequence of the interstitial waters and, therefore, are not a repetitive feature in this crystal structure. However, they illustrate just how intricate hydrogen-bonding patterns can be in a triple-helical conformation when amino acids with free amide groups are incorporated into the sequence.

Cylinder of hydration around the Gly→Ala peptide: ω bridges

The last category of water bridges in Table 2 includes those bridges connecting anchoring points between neighboring triple helices, namely ω bridges. These bridges can be of three different types: Cγ–Oδ···W_n···Oδ–Cγ, Cγ–Oδ···W_n···O=C or C=O···W_n···O=C. They are much more dependent on the local environment than all other bridges discussed so far, and therefore no attempt has been made to further subclassify them on the basis of the identity of the two anchoring groups that they connect. Although ω bridges can involve any number of waters, the more interesting ones are those involving a minimum number of waters. Typically, ω bridges involving two or three waters can connect one group with a symmetry-related triple helix. There is a stronger preference for ω2 and ω1 bridges when the starting anchoring point is a hydroxyl group from a Hyp residue. This is consistent with the more peripheral position of the Hyp hydroxyl groups. For carbonyl groups, the minimal water path normally requires two or three waters, and rarely one or four.

The absence of direct contacts between triple helices suggests that triple helices are already coated by a cylinder of hydration before assembling into the crystal lattice. Figure 6 illustrates the progressive hydration of the Gly→Ala peptide. The grooves on the triple helix are filled with solvent molecules directly hydrogen bonded to the anchoring groups on the peptide, or connected to those waters in the first hydrogen-bonded shell. The third shell of waters completes the cylinder of hydration, which still shows some cavities left by the relatively open network of hydrogen-bonded water molecules. These cavities are closed by carbonyl and hydroxyl groups from neighboring triple helices, although they are deep enough to prevent direct contact between the peptide molecules.

Predominance of fivefold geometry in the water between triple helices

Most of the water bridges described so far show a pentagonal-like appearance as a consequence of the

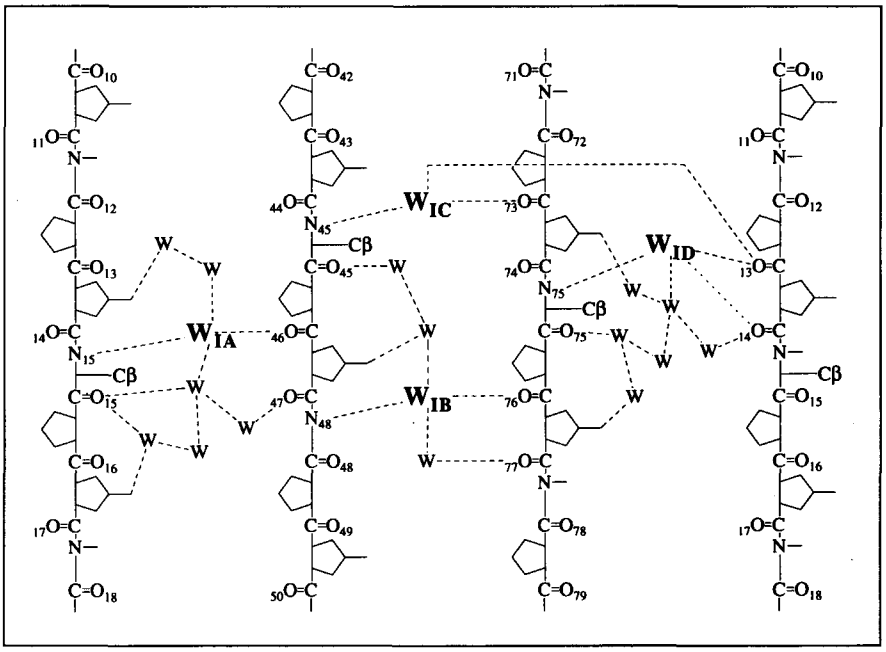


Fig. 5. Local hydrogen-bonding network around the four interstitial waters, which are labeled W_{IA}, W_{IB}, W_{IC} and W_{ID}, shown in larger type.

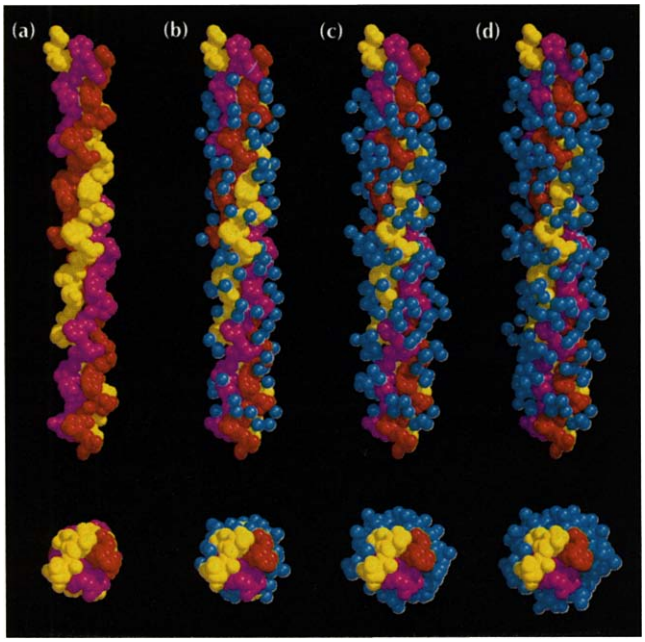


Fig. 6. Space-filling representation for the progressive hydration of the Gly→Ala peptide as seen in the crystal structure. The views in the top row are perpendicular to the molecular axis, whereas those in the bottom row are parallel to the molecular axis at the same hydration level. **(a)** A view of the naked triple helix; the three peptide chains are shown in different colors. **(b)** Incorporation of the first shell of water molecules, directly hydrogen bonded to carbonyl, hydroxyl or even amide groups on the peptide surface. **(c)** Incorporation of the second shell of water molecules, hydrogen bonded to the ones in the first shell; the filling of the superhelical groove by solvent molecules becomes more evident. **(d)** Third shell of water molecules.

predominance of the 110° angle between hydrogen-bonded waters. Only the single-water bridges may depart significantly from this tendency as their geometry is mostly dictated by the particular position of the two anchoring groups. This pentagonal tendency is reinforced when investigating the structure of the water molecules located between neighboring triple helices in this crystal structure. Several water clusters have been identified in this crystal structure with a predominance of the pentagonal motif, although a few cases of quadrilateral or hexagonal clusters have also been found. Water pentagons are sometimes fused with other pentagons in which one of the members is the oxygen from a carbonyl group or from a hydroxyl group on a Hyp residue (Fig. 7). Water pentagons are found in the solvent zone between two triple helices, and their water molecules participate in extensive hydrogen bonding to the anchoring groups in the peptide molecules or to other water molecules engaged in α, γ or δ bridges (Fig. 7).

Water and water–peptide clusters with pentagonal shapes predominate in this crystal structure (Fig. 8). In addition, some of the most frequently observed water bridges can be considered as fragments of pentagons. The extensive interconnection of all these groups suggests a clathrate-like structure around this triple helical peptide.

Discussion

Synergy between the presence of hydroxyproline and water

The results obtained for this peptide shed some light on the long-debated issues of hydroxyproline stabilization of collagen triple helices and the structural role played by the water surrounding the triple helix. For the first time, it is directly demonstrated that hydroxyproline acts as a linchpin in a highly ordered water network, involving interchain and intrachain bridges. The presence of this highly ordered water network is consistent with the experimental results on hydration of collagenous tissues [17–22]. The solvent structure in this crystal is ordered to an extent that is usually found in crystal structures which diffract X-rays at much higher resolution [29–32]. The high degree of water structure must be attributed not only to the relatively small space left for solvent in between triple helices in this crystal, but also to some long-range organizing effect induced by the particular geometry of the collagen triple helix. The highly repetitive nature of this peptide sequence has proved beneficial in that it has been possible to identify highly repetitive solvent patterns. The persistence of most of these patterns through the different local environments of the peptide in the crystal strongly suggests that they are an intrinsic feature of collagen triple helices, at least those with a high content of Pro–Hyp–Gly triplets.

How compatible are these water bridges with other collagen sequences? The most repetitive water bridges in the Gly→Ala crystal structure are those involving the Cγ–Oδ group in Hyp residues: the γ and δ bridges. These bridges buttress the triple-helical conformation by bridging different chains cooperatively with the interchain N–H···O=C hydrogen bonds. Interestingly, only one γ bridge, one δ bridge and one N–H···O=C hydrogen bond are needed per tripeptide to ensure that all anchoring groups in a $(Pro–Hyp–Gly)_n$ triple helix participate in at least one hydrogen-bonding interaction (Fig. 3a,b). Formation of γ and δ bridges is only dependent on the presence of the hydroxyl group in the right place (Fig. 4c), as there will always be carbonyl groups at the other ends independent of the residues involved. It seems likely that, by prolyl hydroxylation, collagen triple helices acquire extra hydrogen-bonding capabilities that may reinforce their conformation. The X–Y–Gly triplets with Hyp in the Y position are a relatively common feature of collagens (about 30% of triplets in type I collagen, and 50% of triplets in type IV collagen).

α and β bridges, on the other hand, are totally independent of the amino acid sequence because they interconnect carbonyl groups. In this crystal structure, α and β bridges usually involve three and four waters respectively, which makes them, in principle, more susceptible to the local environment. In spite of this, they appear with certain regularity (Fig. 2a,b), possibly helped by the fact that several of the waters in α and β bridges participate simultaneously in shorter bridges of the γ or δ types.

A water bridge topologically equivalent to the γ2 bridge observed in this peptide was proposed on stereochemical

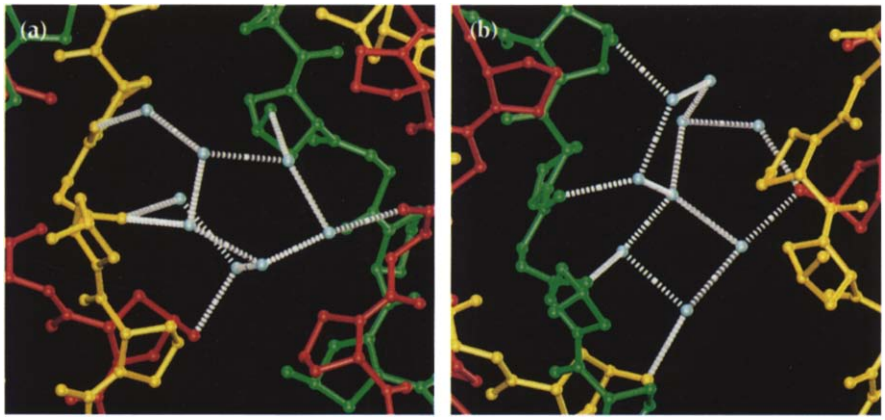


Fig. 7. Two examples of water cluster formation in the space between two neighboring triple helices in the crystal lattice. **(a)** The pentagonal cluster of waters at the center of the picture shares edges with three other water structures: its upper-left side is part of an α3 bridge; its right side is the central segment of a δ2 bridge in the neighboring triple helix; and its lower-left side is part of another pentagonal cluster comprising four water molecules and one oxygen atom from a carbonyl group. This second pentagon also contains another δ2 bridge. **(b)** The pentagonal cluster on the right contains four water molecules and one hydroxyl group. Two clusters of waters, one pentagonal and one quadrilateral, are fused with its upper and lower left sides, respectively. The pentagon of waters contains the central segment of a γ2 bridge, and also one side of an α3 bridge in the same chain. The quadrilateral cluster contains the central segment of a δ2 bridge.

grounds by Suzuki *et al.* [15], using the coordinates of their refined fiber diffraction model of kangaroo tail tendon collagen as a target [33]. Okuyama *et al.* [24] identified three water molecules per tripeptide in their analysis of the crystal structure of the peptide $(Pro–Pro–Gly)_{10}$, where only α or β bridges are possible. In our nomenclature, these are described as a β3 bridge and an α1 bridge.

Clathrate-like structures might be intrinsic to solvated triple helices

One of the most striking aspects of this crystal structure is the abundance of pentagonal water clusters in the cylinder of hydration surrounding the triple helix. Water around hydrophobic surfaces tends to form a more organized structure than the bulk water, and in a typical collagen triple helix, proline and Hyp rings are partially facing the solvent. In addition, collagen triple helices seem to provide good anchoring points to attach this hydrophobically enforced water structure to the peptide molecule (Fig. 2a). Remarkably, the frequently occurring α3, γ2 and δ2 bridges can be considered as truncated pentagons (Fig. 8d–f) that can fuse with clusters of waters into clathrate-like structures (Fig. 7). The observed preference for the pentagonal motif is a consequence of the favored hydrogen-bonding geometry introduced by the 105° H–O–H angle in the water molecule, and the 109° Cγ–Oδ–H angle in the Hyp hydroxyl group.

The formation of clathrate-like structures offers another way to understand, at least qualitatively, how Hyp residues can contribute to the stabilization of collagen triple helices. Water is likely to structure itself around a $(Pro–Pro–Gly)_n$ triple helix in which the hydrophobic prolyl rings are facing the solvent. Using the available carbonyl groups, these water structures will connect to the peptide molecules through α or β water bridges. Prolyl hydroxylation introduces extra anchoring points and extra water bridges, which participate in the water structure. The result is that the water immediately surrounding the peptide becomes highly ordered by extensive hydrogen bonding. This observation agrees with the hypothesis that the triple-helical conformation induces

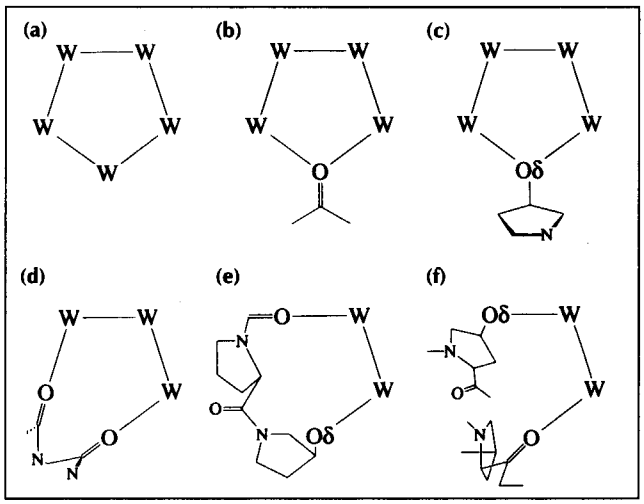


Fig. 8. Pentagonal arrays observed in the crystal structure of the Gly→Ala peptide. **(a)** The basic pentagon of waters. **(b)** A pentagon in which one of the vertices is the oxygen atom from a carbonyl group. **(c)** A pentagon in which one of the vertices is the oxygen atom from the hydroxyl group of a Hyp residue. **(d)** A typical α3 bridge in which two of the waters on the basic pentagon are substituted by oxygen atoms from two carbonyl groups. **(e)** A typical γ2 bridge in which two vertices of a pseudo-pentagon are a carbonyl oxygen and hydroxyl oxygen, and one edge is the polypeptide chain. **(f)** A typical δ2 bridge in which two of the waters on the basic pentagon are substituted by oxygen atoms — one from a carbonyl and one from a hydroxyl group — to form a pseudo-pentagon.

more order into the surrounding water that the denatured (coil) conformation [17,34]. According to this hypothesis, denaturing of collagen triple helices involves an entropy increase not only due to the macromolecule, but also to the water. This entropy increase upon denaturation will be higher for a $(Pro–Hyp–Gly)_n$ peptide than for a $(Pro–Pro–Gly)_n$ one. In agreement with this, careful thermodynamic experiments performed by Engel and co-workers [11] demonstrated that the entropy increase upon denaturation of $(Pro–Hyp–Gly)_{10}$ in water is much higher than that observed for $(Pro–Pro–Gly)_{10}$. However, because the enthalpy increase is even higher for the denaturation of $(Pro–Hyp–Gly)_{10}$, this peptide is still

much more stable in aqueous solution than its counterpart lacking Hyp residues.

Did ancestral collagens contain threonine residues in place of hydroxyproline?

The persistence and structure-promoting effect of the γ and δ water bridges may be indicative of the importance of placing a sufficient number of hydroxyl groups in the periphery of the collagen triple helices. The presence of the γ and δ water bridges facilitates the linking of the polypeptide chains with the intrinsic structure adopted by their first cylinder of hydration. γ and δ bridges in this peptide involve the hydroxyl groups from Hyp residues, but in principle, similar water bridges could be achieved by serine or threonine residues having the appropriate conformation. For example, a high serine plus threonine content is frequently observed in invertebrate and lower vertebrate collagens and can be correlated with low Hyp content [35,36]. It is of interest that a cuticle collagen from worms collected at deep sea hydrothermal vents has an unusually high content of threonine that is not comparable to that found in any other collagen [37]. On the basis of this observation, Gaill *et al.* [37] proposed that threonine can substitute for Hyp residues whilst still providing similar hydrogen bonds. These authors also report a higher preference for threonine residues in the Y position of vertebrate collagens. It is tempting to speculate that ancestral collagens might have used hydroxylated amino acids such as threonine to promote a water structure around the triple helix similar to that observed for Hyp residues.

Triple-helical assemblies and hydration

The existence of a highly structured hydration shell around collagen triple helices is consistent with the very strong, distance-dependent forces between collagen triple helices observed under osmotic stress in solution [26]. Hydration structure is likely to play a role maintaining the macromolecular assemblies in which collagen molecules are involved, and water bridges equivalent to the ω ones found in this crystal structure are likely to appear in collagen fibrils [25,38]. Therefore, as Hyp hydroxyl groups are the key for the structuring of the water network around the triple helices, this suggests that Hyp residues not only play a crucial role in the stabilization of isolated triple helices but also participate in the formation of macromolecular assemblies by inducing and supporting the water-bridged network present between triple helices.

Triple helices from the Gly→Ala peptide pack in the crystal in such a way as to place the disrupted zones (where the substitution is located) next to one another. Thus, ω bridges, originating in the central zone of the triple helix, are most likely to connect with the central zones of neighboring triple helices. This contrasts with disorder phenomena observed in polymer-like crystals of the peptides $(Pro–Pro–Gly)_{10}$ [24] and $(Pro–Hyp–Gly)_{10}$ (J Bella, unpublished data), in which neighboring triple helices seem to aggregate longitudinally at several vertical levels. These peptides are so repetitive that they have no 'locking' features which can force the molecules to pack in an ordered, completely crystalline way. This difference points to a 'feature recognition' between the hydration shells of approaching Gly→Ala molecules when assembling into the crystal phase. Because there is a locking effect, and the peptide molecules most probably interact through their cylinders of hydration, some kind of hydration pattern recognition seems to be necessary (perhaps through some subtle differences in the hydration patterns which are not obvious in Fig. 8). This may be relevant for the sequence recognition that seems to play a role in the longitudinal aggregation of 'real' collagen molecules when forming fibrils. It is normally assumed that charged and hydrophobic groups interact specifically to produce the desired alignment. Alternatively, these charged groups may disrupt the local water structure in a way that offers 'recognizable' features to approaching hydrated triple helices.

Occurrence of ζ water bridges in non-Pro–Hyp–Gly segments

The water patterns observed around the interstitial waters provide clues about alternative hydrogen-bonding schemes. In this peptide, the RCII pattern is substituted in the central zone by a local network of water bridges in which the main feature is the formation of ζ1 bridges connecting the corresponding amide and carbonyl groups. An analogous situation is likely to occur wherever amino acids other than proline and Hyp occupy the X and Y positions in the collagen sequence. For example, ζ1 bridges between amide groups from amino acid residues in the X position and carbonyl groups from glycine residues in the preceding chain, that is in the 1→3→2→1 direction, form the basis for the modified one-bonded structure proposed by Ramachandran and co-workers [14,23]. Measured distances in this crystal structure between the corresponding Pro:N and Gly:O atoms are in the range 4.2–4.6 Å. These distances are identical to those observed between the groups connected by ζ1 bridges in the Gly→Ala substitution zone, although their relative orientation differs. If ζ water bridges $N–H\cdots W_n\cdots O=C$ indeed occur, additional water molecules attached to the ones involved in the ζ bridge will extend the local water network towards the rest of the anchoring groups placed nearby in the polypeptide chain, in a manner similar to the interstitial waters in this crystal structure (Fig. 5).

It is not easy to predict the exact number of waters in these bridges and the conformational changes which may result from water-mediating hydrogen bonds. Crystal structures of peptides containing non-imino acid triplets will answer these questions and will extend our knowledge of the range of possibilities available for water-bridge formation in collagen triple helices.

Biological implications

Collagens constitute the major structural proteins in the extracellular matrix, providing mechanical strength and structural integrity to the various connective tissues in the body. The most common collagens associate in a staggered array to form characteristic fibrils with an axial repeat of 67 nm,

while others form networks, microfibrils or supercoiled arrays, which may include antiparallel as well as parallel molecular association. In addition to self-association, collagens bind to different collagen types, other extracellular matrix molecules and cell-surface integrins, providing a framework for extracellular matrix organization. Mutations in specific collagen types have been shown to result in connective tissue hereditary diseases, including brittle bone disease (type I collagen), chondrodysplasias (type II collagen), Ehlers–Danlos Syndrome IV and familial aortic aneurysms (type III collagen), Alport syndrome (some chains of type IV collagen) and some cases of epidermolysis bullosa (type VII chains).

The defining feature of collagens is their large triple-helical domain. Distinctive sequence characteristics of this conformation include glycine at every third residue, (Gly–X–Y)$_n$, a high content of imino acids, and the post-translational modification of prolines in the Y position to hydroxyproline (Hyp). The Hyp content is highly correlated with collagen thermal stability. Water plays a critical role in maintaining the conformation of collagen molecules and the mechanical properties of collagen fibrils. Specific water binding to the peptide backbone as well as bridges of ordered water molecules between peptide chains have been proposed.

The determination of the first high-resolution crystal structure of a triple-helical molecule was accomplished for a peptide containing Pro–Hyp–Gly units, which models the imino acid rich regions of collagen. This peptide also contains one substitution of a glycine by an alanine, modeling the kind of mutations found in many connective tissue diseases. A highly ordered set of water molecules is present coating the triple helix. The pivotal feature of the regular water networks is the hydroxyl group of Hyp. After years of speculation and modeling, the precise role of this residue in collagen stabilization is finally revealed.

The regular arrangement of molecules in crystals, although key to the determination of the structure, does not normally have physiological relevance. In contrast, in this crystal the triple helices are organized such that the intermolecular 14 Å center to center distance is similar to that seen in collagen fibrils. Thus, the crystal packing, in which triple helices are bridged by ordered water molecules, appears to relate directly to collagen assembly in tissues.

Materials and methods

We have reported elsewhere the basic features of the crystal structure of the Gly→Ala peptide [25]. A more accurate modeling of its solvent structure required the incorporation of many additional solvent sites and a careful refinement of the solvent atomic positions. The resolution limits used were 12.5–1.85 Å. All reflections with F_{obs} greater than $1\sigma(F)$ were included in the calculations. The total number of reflections used was 4025, which represents the 73% of all unique data between 12.5 Å and 1.85 Å. The entire refinement was conducted using X-PLOR [39].

The final model shows good agreement between calculated and observed structure factors, with an R-factor of 0.172. The model also shows good agreement with the applied geometry restraints, with root mean square (rms) deviations of 0.007 Å for bond distances, 1.53° for bond angles and 1.53° for chirality and planarity constraints. The applied target values for glycine, proline and alanine residues are those determined by Engh and Huber [40] for use with X-PLOR. Some additional parameters were needed for the Hyp residue and we derived them from a statistical survey of X-ray structures from the Cambridge Structural Database [41], on compounds containing chemical fragments matching 4-hydroxyproline. Average bond lengths and angles are shown in Table 4. Ramachandran plots show that most of the residues have main-chain torsion angles strongly clustered around the polyproline II conformation. Only the two *cis*-Hyp residues, following N-terminal prolines in chains 2 and 3, have different conformations. The six ends of the chains (residues 1–3, 31–33, 61–63 in the N terminus zone; and 28–30, 58–60, 88–90 in the C terminus zone) were difficult to refine because they appear to be affected by increased conformational disorder. Electron-density maps showed some evidence for alternative conformations, mainly at the N terminus of chain 2 (residues 31–33), and the C terminus of chains 1 and 3. Only the dominant conformation of these terminal groups has been included in the final model, which contains all 90 residues, and partial occupancies have been allowed for the 18 residues involved in the first and last triplet on every chain. The analysis of solvent structure and interactions has specifically excluded the poorest defined residues.

Table 4. Parameters for the imino acid residue 4-hydroxyproline as obtained from the Cambridge Structural Database (CSD).

	d_{ave}/τ_{ave}*	σ	X-PLOR types†	d_{app}/τ_{app}‡	Weight
Bond					
C–N	1.340	0.017	C–N	1.341	
N–CA	1.472	0.016	N–CH1E	1.466	
CA–CB	1.532	0.015	CH1E–CH2E	1.530	
CB–CG	1.523	0.016	CH2E–CH1P	1.523	2312.5
CG–OG	1.422	0.025	CH1P–OH1	1.422	947.2
CG–CD	1.522	0.016	CH1P–CH2P	1.522	2312.5
CD–N	1.474	0.013	CH2P–N	1.473	
Angle					
C–N–CA	120.1	4.5	C–N–CH1E	122.6	
N–CA–C	112.7	3.8	N–CH1E–C	111.8	
N–CA–CB	102.9	0.8	N–CH1E–CH2E	103.0	
C–CA–CB	113.4	2.6	C–CH1E–CH2E	110.1	
CA–CB–CG	104.4	1.5	CH1E–CH2E–CH1P	104.4	863.7
CB–CG–OG	110.5	2.6	CH2E–CH1P–OH1	110.5	310.9
CB–CG–CD	103.3	1.9	CH2E–CH1P–CH2P	103.3	538.3
OG–CG–CD	110.4	2.2	OH1–CH1P–CH2P	110.4	401.5
CG–CD–N	104.6	2.2	CH1P–CH2P–N	104.6	401.5
CD–N–C	125.4	3.7	CH2P–N–C	125.0	
CD–N–CA	111.3	1.7	CH2P–N–CH1E	112.0	

*d_{ave} and τ_{ave} are average bond lengths and angles, respectively, obtained from a search for hydroxyprolyl fragments on the CSD [41]. †X-PLOR atom types follow the nomenclature adopted from reference [35]. ‡d_{app} and τ_{app} are the target values used during X-PLOR refinement; those with no specified weight were taken directly from the reference [35] parameter set since they did not differ significantly from the obtained average values. The only new parameters are related with a new atom type, CH1P, needed for Cγ of Hyp.

Solvent molecules were located by using difference electron density maps. In order to minimize any bias from the X-PLOR energy function on the water geometries, the water molecules were refined first against the X-ray term only. Those that remained in electron density and maintained good hydrogen bonding to the peptide and other water molecules were refined further using the complete X-PLOR energy term. Hydrogen bonding in X-PLOR can be modeled either with an explicit term, or as the combination of a van der Waals repulsive term with an electrostatic attractive term. Both approaches have been tested. The explicit description models satisfactorily those water bridges in which solvent molecules from symmetry-related triple helices are not involved. However, because so many waters are involved in complex bridges interconnecting neighboring triple helices, we found that the non-explicit description of the hydrogen-bonding interaction was more appropriate. The main criterion for the acceptance of a particular water position was its agreement with the electron density. Omit maps were calculated routinely in order to confirm the positioning of solvent molecules, water bridges, or special hydrogen-bonded topologies. Solvent molecules appearing at electron density levels below 1σ in the $2F_o-F_c$ maps have been considered dubious, and have been removed whenever their omit maps have failed to bring them back at levels of at least 2.5σ in the F_o-F_c map.

Coordinates of the refined structure are being deposited in the Brookhaven Protein Structure Databank (entry code 1CGD).

Acknowledgements: This work was supported by grants from the National Institutes of Health to HMB (GM21589) and BB (AR19626) as well as a fellowship from the government of Spain to JB. We thank Christine Zardecki for her help in preparing the manuscript. X-PLOR refinements were carried out on the Pittsburgh Supercomputer under a grant from the National Science Foundation.

References

1. Kielty, C.M., Hopkinson, I. & Grant, M.E. (1993). Collagen: The Collagen Family: Structure, Assembly and Organization in the Extracellular Matrix. In *Connective Tissue and Its Hereditable Disorders.* (Royce, P.M. & Steinmann, B., eds), pp. 103–147, Wiley-Liss, Inc., NY.
2. Hoppe, H.-J. & Reid, K.B.M. (1994). Collectins — soluble proteins containing collagenous domains in animal lectins — and their role in innate immunity. *Protein Sci.* **3**, 1143–1158.
3. Kodama, T., Freeman, M., Rohrer, L., Zabrecky, J., Matsudaira, P. & Krieger, M. (1990). Type I macrophage scavenger receptor contains α-helical and collagen-like coiled coils. *Nature* **343**, 531–535.
4. Elomaa, O., *et al.*, & Tryggvason, K. (1995). Cloning of a novel bacteria-binding receptor structurally related to scavenger receptors and expressed in a subset of macrophages. *Cell* **80**, 603–609.
5. Kadler, K. (1994). Extracellular matrix fibril forming collagens. *Protein Profile* **1**, 519–638.
6. Acton, S., Resnick, D., Freeman, M., Ekkel, Y., Ashkenas, J. & 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.
7. Kuivaniemi, H., Tromp, G. & Prockop, D.J. (1991). Mutations in collagen genes: causes of rare and sometimes common diseases in humans. *FASEB J.* **5**, 2052–2060.
8. Rosenbloom, J., Harsh, M. & Jimenez, S. (1973). Hydroxyproline content determines the denaturation temperature of chick tendon collagen. *Arch. Biochem. Biophys.* **158**, 478–484.
9. Burjanadze, T.V. (1992). Thermodynamic substantiation of water-bridged collagen structure. *Biopolymers* **32**, 941–949.
10. Sakakibara, S., Inouye, K., Shudo, K., Kishida, Y., Kobayashi, Y. & Prockop, D.J. (1973). Synthesis of $(Pro\text{–}Hyp\text{–}Gly)_n$ of defined molecular weights. Evidence for the stabilization of collagen triple helix by hydroxyproline. *Biochim. Biophys. Acta* **303**, 198–202.
11. Engel, J., Chen, H.-T., Prockop, D.J. & Klump, H. (1977). The triple helix–coil conversion of collagen-like polytripeptides in aqueous and nonaqueous solvents. Comparison of the thermodynamic parameters and the binding of water to $(L\text{–}Pro\text{–}L\text{–}Pro\text{–}Gly)_n$ and $(L\text{–}Pro\text{–}L\text{–}Hyp\text{–}Gly)_n$. *Biopolymers* **16**, 601–622.
12. Gustavson, K.H. (1955). The function of hydroxyproline in collagens. *Nature* **175**, 70–74.
13. Rich, A. & Crick, F.H.C. (1961). The molecular structure of collagen. *J. Mol. Biol.* **3**, 483–506.
14. Ramachandran, G.N., Bansal, M. & Bhatnagar, R.S. (1973). A hypothesis on the role of hydroxyproline in stabilizing collagen structure. *Biochim. Biophys. Acta* **322**, 166–171.
15. Suzuki, E., Fraser, R.D.B. & MacRae, T.P. (1980). Role of hydroxyproline in the stabilization of the collagen molecule via water molecules. *Int. J. Biol. Macromol.* **2**, 54–56.
16. Fraser, R.D.B. & MacRae, T.P. (1973). *Conformation in Fibrous Proteins.* Academic Press, New York.
17. Berendsen, H.J.C. & Migchelsen, C. (1965). Hydration structure of fibrous macromolecules. *Ann. N. Y. Acad. Sci.* **125**, 365–379.
18. Grigera, J.R. & Berendsen, H.J.C. (1979). The molecular details of collagen hydration. *Biopolymers* **18**, 47–57.
19. Nomura, S., Hiltner, A., Lando, J.B. & Baer, E. (1977). Interaction of water with native collagen. *Biopolymers* **16**, 231–246.
20. Cusack, S. & Lees, S. (1984). Variation of longitudinal acoustic velocity at gigahertz frequencies with water content in rat tail tendon fibers. *Biopolymers* **23**, 337–351.
21. Peto, S., Gillis, P. & Henri, V.P. (1990). Structure and dynamics of water in tendon from NMR relaxation measurements. *Biophys. J.* **57**, 71–84.
22. Hoeve, C.A.J. & Tata, A.S. (1978). The structure of water absorbed in collagen. *J. Phys. Chem.* **82**, 1660–1663.
23. Ramachandran, G.N. & Chandrasekharan, R. (1968). Interchain hydrogen bonds via bound water molecules in the collagen triple helix. *Biopolymers* **6**, 1649–1658.
24. Okuyama, K., Okuyama, K., Arnott, S., Takayanagi, M. & Kakudo, M. (1981). Crystal and molecular structure of a collagen-like polypeptide $(Pro\text{–}Pro\text{–}Gly)_{10}$. *J. Mol. Biol.* **152**, 427–443.
25. Bella, J., Eaton, M., Brodsky, B. & Berman, H.M. (1994). Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science* **266**, 75–81.
26. Leikin, S., Rau, D.C. & Parsegian, V.A. (1994). Direct measurement of forces between self-assembled proteins. Temperature-dependent exponential forces between collagen triple helices. *Proc. Natl. Acad. Sci. USA* **91**, 276–280.
27. Baker, E.N. & Hubbard, R.E. (1984). Hydrogen bonding in globular proteins. *Prog. Biophys. Molec. Biol.* **44**, 97–179.
28. Jeffrey, G.A. & Saenger, W. (1991). Hydration of proteins. In *Hydrogen Bonding in Biological Structures.* pp. 459–486, Springer-Verlag, NY.
29. Karle, I.L. (1980). Water structure in $[Phe^4\ Ala^4]$ antamanide. $12H_2O$ crystallized from dioxane. *Int. J. Pept. Protein Res.* **28**, 6–12.
30. Neidle, S., Berman, H. & Shieh, H.-S. (1980). Highly structured water network in crystals of a deoxydinucleoside–drug complex. *Nature* **288**, 129–133.
31. Shieh, H.-S., Berman, H.M., Dabrow, M. & Neidle, S. (1980). The structure of drug–deoxydinucleoside phosphate complex: generalized conformational behavior of intercalation complexes with RNA and DNA fragments. *Nucleic Acids Res.* **8**, 85–97.
32. Teeter, M.M. (1984). Water structure of a hydrophobic protein at atomic resolution: pentagon rings of water molecules in crystals of crambin. *Proc. Natl. Acad. Sci. USA* **81**, 6014–6018.
33. Fraser, R.D.B., MacRae, T.P. & Suzuki, E. (1979). Chain conformation in the collagen molecule. *J. Mol. Biol.* **129**, 463–481.
34. Privalov, P.L. (1982). Stability of proteins. *Adv. Protein Chem.* **35**, 1–103.
35. Pikkarainen, J., Rantanen, J., Vastamäki, M., Lampiaho, K., Kari, A. & Kulonen, E. (1968). On collagen of invertebrates with special reference to *Mytilus edulis. Eur. J. Biochem.* **4**, 555–560.
36. Pikkarainen, J. & Kulonen, E. (1969). Comparative chemistry of collagen. *Nature* **223**, 839–841.
37. Gaill, F., Wiedemann, H., Mann, K., Kühn, K., Timpl, R. & Engel, J. (1991). Molecular characterization of cuticle and interstitial collagens from worms collected at deep sea hydrothermal vents. *J. Mol. Biol.* **220**, 209–223.
38. Ramachandran, G.N. (1967). Structure of collagen at the molecular level. In *Treatise on Collagen.* (Ramachandran, G.N., ed), pp. 103–183, Academic Press, New York.
39. Brünger, A.T. (1992). *X-PLOR Manual, Version 3.1.* Yale University Press, New Haven, CT.
40. Engh, R.A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst. A* **47**, 392–400.
41. Allen, F.H., Kennard, O. & Taylor, R. (1983). Systematic analysis of structural data as a research technique in organic chemistry. *Accounts Chem. Res.* **16**, 146–153.
42. Schneider, B., Cohen, D.M., Schleifer, L., Srinivasan, A.R., Olson, W.K. & Berman, H.M. (1993). A systematic method for studying the spatial distribution of water molecules around nucleic acid bases. *Biophys. J.* **65**, 2291–2303.
43. Murray-Rust, P. & Glusker, J.P. (1984). Directional hydrogen bonding to sp^2- and sp^3-hybridized oxygen atoms and its relevance to ligand–macromolecule interactions. *J. Am. Chem. Soc.* **106**, 1018–1025.

Received: **9 Jun 1995**; revisions requested: **7 Jul 1995**; revisions received: **20 Jul 1995**. Accepted: **25 Jul 1995**.