

Hydrogen bonding in the triple-helix

BARBARA BRODSKY

Department of Biochemistry, UMDNJ-Robert Wood Johnson Medical
School, Piscataway, NJ 08854, USA
email: brodsky@rwja.umdj.edu

Abstract. The increased level of detail obtained from high resolution crystal structures of triple-helical peptides and the residue specific properties that can be defined by NMR studies on ^{15}N -enriched peptides has given new information on hydrogen bonding in the triple-helix. In particular, it has given details about the principle originally posed by Ramachandran to try to maximize the hydrogen bonding in the collagen triple-helical structure in the face of the lack of direct favorable interactions. Now it can be seen that by the participation of water, and by the summation of many weak hydrogen bonding patterns, the structure is largely stabilized through hydrogen bonding, as indicated by calorimetry. The $\text{C}_\alpha\text{H}\dots\text{O}=\text{C}$ hydrogen bonds, the water mediated $\text{NH}\dots\text{CO}$ bonds, and the hydroxyproline–water hydrogen bonds proposed by Ramachandran have been visualized and now defined in the triple-helix.

Keywords. Collagen; triple-helix; hydrogen bonding.

1. Introduction

The years 1954–1955 represented a milestone in collagen structure. Ramachandran^{1–3} derived the concept of the correct type of triple-helical structure for collagen by applying the newly developed fiber diffraction theory⁴ to improved collagen fiber diffraction patterns with stereochemical considerations. It was Cohen and Bear⁵ who first recognized that the fiber diffraction pattern indicated collagen had a helical structure, but the elucidation of the correct structure required the additional insight that the unusually high Gly and imino acid contents of collagen must have a structural basis¹. The original proposal of three closely packed staggered helical chains, each with 3 residues per turn was modified by a supercoiled twist to fit well-oriented diffraction patterns^{1–3}. Independently, in 1955, two other groups proposed related triple-helical structures. Model building based on the structure of polyglycine II led Rich and Crick^{6,7} to a collagen triple-helical conformation built with more stringent stereochemical criteria. At King's College, the recording and analysis of highly oriented collagen X-ray patterns, together with the elucidation of the polyproline II structure, led to the proposal of a similar supercoiled structure⁸. These exciting years have been reviewed in depth^{9,10}, and a recent perspective has been published^{11–13}.

Hydrogen bonding in the triple-helix has been a source of controversy. In terms of solvent accessibility and interactions, the collagen triple-helix is analogous to the formation of secondary structure such as the α -helix and β -sheet conformations¹⁴. Pauling and Corey¹⁵ proposed that the formation of all possible peptide hydrogen bonds was an important factor in stabilizing helical conformations, and a similar maximization of

hydrogen bond formation might be expected for the triple-helix. But from the beginning^{1,3}, it was clear that no intrachain hydrogen bonds were possible in the triple-helix and that it would not be possible to have all backbone NH and CO groups involved in interchain hydrogen bonds. In particular, the NH group of the residue in the Y position points out of the triple-helix, into the surrounding solvent, with no possibility of interaction with other peptide groups. Anxious to maximize intramolecular hydrogen bonding as a stabilizing force in the triple-helix, Ramachandran and Kartha^{1,3} proposed that two interchain NH...O bonds could be formed for every three residues, one involving the Gly NH group (to the carbonyl of the X residue) and the second utilizing the NH group of the residue in the X position whenever this was not an imino acid. Rich and Crick^{7,16} disagreed with this proposal, and found through calculations that only one interchain hydrogen bond, Gly NH...O=C(X), could be formed and that disallowed short contact distances were present if the model was modified to allow a second interchain hydrogen bond. Later, Ramachandran and Chandrasekharan¹⁷ suggested that a second interchain hydrogen bond involving NH groups of the residue in the X position could be mediated by a water molecule. Ramachandran also suggested the possibility of interchain CH(Gly)...O=C bonds in polyglycine II and in collagen^{18,19} and was one of the researchers who proposed that the stabilizing effect of hydroxyproline was related to water-mediated hydrogen bonding²⁰⁻²².

Calorimetric studies on collagen generated evidence that the major stabilizing force of the triple-helix is enthalpic, derived from a dominant contribution of hydrogen bonding²². Privalov's perceptive analysis of hydrogen bonding in the collagen triple-helix helped resolve several contentious issues by laying out the experimental evidence for water-mediated hydrogen bonding.

Recent advances in the understanding of collagen structure have been made through studies on triple-helical peptides of defined sequences, and the new information obtained from these studies concerning the basic collagen conformation, its hydrogen bonding, and its sequence-dependent features will be summarized here.

2. Peptide models of collagen

Studies on model peptides played a critical role in the original determination of the triple-helix conformation and in characterizing its features. The structures of both polyglycine II and polyproline II in the 1950's gave insight into the molecular conformation of individual chains, intermolecular hydrogen bonding, and molecular packing. Studies on poly(Gly-Pro-Hyp) and other repeating polytripeptides clarified requirements for triple-helices and their relative stabilities in solution²³⁻²⁵.

With the advent of solid-state peptide synthesis, the ability to create triple-helical peptides of defined length and sequence led to more quantitative studies and allowed modeling of specific collagen sequences, including cell-binding sites, heparin-binding sites, an epitope for a monoclonal antibody and the collagenase cleavage site²⁶⁻²⁸. The most stabilizing sequence is Gly-Pro-Hyp, and recent peptides have used repeating Gly-Pro-Hyp sequences to promote formation of the triple-helix structure. An additional approach to stabilization has been cross-linking of three chains, through dilysyl moieties, N-terminal links, or Kemp triacid²⁷⁻³⁰. One successful and striking example has been the synthesis of a cross-linked heterotrimer which models the type I collagenase cleavage site, and is a substrate for that enzyme³¹. Fortunately, some peptides have proved to have the high solubility needed for multi-dimensional NMR studies³², and several have formed crystals suitable for X-ray crystallography.

3. The first crystal structure: (Pro-Hyp-Gly)₁₀ peptide with a single Gly→Ala substitution

The first high-resolution triple-helical structure was obtained by X-ray crystallography for a peptide containing repeating Pro-Hyp-Gly motifs with a single Gly to Ala substitution near the center³³. Repeating Pro-Hyp-Gly sequences provided the first and most stable examples of defined triple-helical peptides³⁴, and the replacement of one Gly by an Ala decreased the thermal stability dramatically from 60°C for (Pro-Hyp-Gly)₁₀ to 30°C³⁵. Unperturbed Pro-Hyp-Gly regions at both ends of the peptide were similar to those predicted from fiber diffraction models, except for a small difference in symmetry (7/2 for the peptide vs. 10/3 for the collagen helix³⁶). The observed distances were consistent with (Gly)NH...CO(Pro) hydrogen bonds, except near the site of the Ala substitution, where a water mediated linkage is present. In addition, the crystal structure showed evidence for the existence of two repetitive patterns of C-H...O=C hydrogen bonds, with one set occurring between GlyC_α and C=O groups from Gly and Pro in the other chains and the second set connecting Hyp C_α atoms in one chain with Pro C=O groups in the neighboring chain³⁷. An extensive and ordered water network was observed, involving hydrogen bonds with all available carbonyl and Hyp hydroxyl groups^{33,38}. Such water-mediated bonding joins carbonyl and/or Hyp groups within one chain, between chains within one molecule, and in less regular arrangements, between molecules. The key participation of the Hyp OH groups in this ordered hydration network is likely to be the basis of the stabilizing role of Hyp in the triple-helix. The recent high resolution structure of (Pro-Pro-Gly)₁₀ again shows a highly ordered hydration network, involving the peptide carbonyl groups, surprisingly similar to that seen in the Pro-Hyp-Gly region of the Gly→Ala peptide, even though Hyp is absent³⁹. Comparison of the hydration patterns in Pro-Pro-Gly and Pro-Hyp-Gly peptides indicates 'that the existence of localized water surrounding the Y position imino acid is itself not dependent on hydroxyproline, but the presence of hydroxyproline induces the formation of additional water bridges, greater localization, and a more extensive hydration network'³⁹.

It is important to note that the crystallographic structure itself cannot directly give information on molecular stability. The structure results from molecular packing as well as intramolecular interactions, and there is no way to estimate the stabilization effect of visualized hydration networks or other interactions. However, the ordered hydration network involving Hyp visualized in these triple-helical peptides is strikingly similar to that predicted on thermodynamic grounds²², where the thermodynamic data does provide evidence for stabilization by such hydrogen bonding. Recent reports from Raines and his group suggest that the mechanism of Hyp stabilization is likely to be related to the inductive effect of an electron-withdrawing group on the pyrrolidine ring, rather than to hydration^{40,41}. This is based on the observation that the stabilization of the triple-helix is greater for fluoroproline >Hyp> Pro in the Y position of peptides and this stability order follows the observed trans:cis imide bond ratio. Since fluoroproline is unlikely to play an important role in a hydration network, the inductive effect is suggested as the principal means of stabilization. However, it is possible that the high degree of stability of (Pro-Fluoroproline-Gly)₁₀ derives from factors different from those that stabilize (Pro-Hyp-Gly)₁₀ and collagen, e.g. the unfavorable nature of a monomer with fluoroproline in aqueous solution compared with the trimer (J A M Ramshaw, personal communication).

Interestingly, the Gly→Ala peptide molecules were packed in a quasi-hexagonal arrangement in the crystal, with intermolecular distances comparable to those seen in

collagen fibrils in tendon or skin³³. No direct contacts were made with atoms in adjacent molecules; the water mediated bridges appeared to be the determining factor for intermolecular packing. Recent studies on collagen fibrils using osmotic pressure and Raman spectroscopy have emphasized the critical nature of water in the packing of molecules⁴².

4. The second hydrogen bond: Studies on peptide T3-785

Gly-Pro-Hyp tripeptides promote triple-helix formation and are highly stabilizing, but more varied sequences are needed in collagen to mediate biological functions and to modulate stability to just above physiological temperature. NMR studies and the recent crystallographic solution of a peptide including a 9 residue region from type III collagen (residues 785-793) just C-terminal to the unique collagenase site give a view into less stabilizing and more functional triple-helical regions. Peptide T3-785, a 30-mer of sequence Pro-Hyp-Gly-Pro-Hyp-Gly-Pro-Hyp-Gly-Ile-Thr-Gly-Ala-Arg-Gly-Leu-Ala-Gly-Pro-Hyp-Gly-Pro-Hyp-Gly-Pro-Hyp-Gly-Pro-Hyp-Gly, forms a stable triple-helix, with a T_m near 18°C, far below the 60°C T_m of the 30-mer (Pro-Hyp-Gly)₁₀. ¹⁵N-enriched residues were incorporated at specific sites in the peptide, and NMR studies showed that the entire peptide was in a stable triple-helix conformation, with a rigid chain as determined from relaxation parameters^{43,44}. Hydrogen exchange studies were carried out for each residue in the Gly-Leu-Ala sequence of this peptide⁴⁴. The Gly NH and Leu NH both exchanged slowly (protection factors of ~1000 and ~400, respectively) while the Ala NH exchanged very rapidly (protection factor of ~35), almost as fast as in the case of the monomer unfolded form. These studies confirm the interpretation of hydrogen exchange done on collagen²², indicating that the 1/3 very slowly exchanging hydrogens are from the Gly NH groups, and the slowly exchanging hydrogens come from the NH group of non-imino acids in the X position, while those from the NH group of the Y residue exchange very rapidly.

The high resolution structure of peptide T3-785 was determined recently by Rachel Kramer in Helen Berman's laboratory^{45,46}. The Gly NH groups form hydrogen bonds with the C=O groups of the X residue, as expected, with additional hydrogen bonding of the carbonyl to the Hyp C α H. The backbone NH of residues in the X position in the central region of the peptide (Ile, Ala, and Leu) all form a hydrogen bond through one water molecule to the C=O group of Gly, similar to the water mediated interchain hydrogen bonding scheme proposed by Ramachandran and Chandrasekharan¹⁷. Together, the NMR and X-ray data indicate that the NH group of Leu791 which has a water mediated hydrogen bond exchanges almost as slowly as the Gly790 with a direct interchain H bond. This suggests a stabilizing effect of this water mediated hydrogen bond, in accordance with the thermodynamic analysis of Privalov²². Studies on various peptides, including a well defined host-guest set, indicate that it is more favorable to have an imino acid in the X position than an amino acid, suggesting the enthalpic stabilization of a second hydrogen bond is less than the entropic stabilization of imino acids, but provides a way of increasing stability while allowing functional sequences.

In the Y position, where no interchain H bond is possible, it is clearly more favorable to have Hyp, which is both entropically favorable and can participate in hydration hydrogen bonding network.

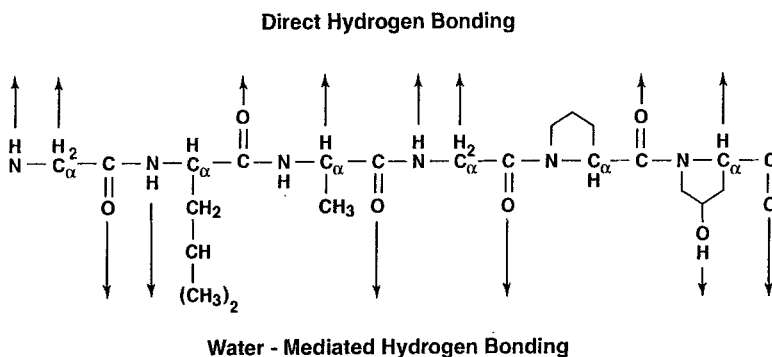


Figure 1. A schematic drawing of hydrogen bonding relationships for a collagen triple-helix sequence Gly-Leu-Ala-Gly-Pro-Hyp. The residues that have been shown to participate in direct hydrogen bonding in a crystal structure are indicated by arrows pointing up, while the residues which have been shown to participate in water-mediated hydrogen bonding are indicated by arrows pointing down^{33,37-39,45,46}.

5. Discussion

Application of modern biophysical techniques to peptides of defined sequences has opened the door to molecular characterization of the collagen triple-helical structure and dynamics. Stabilization of the triple-helix is seen to be the result of the entropic stabilization from imino acids, together with the summation of numerous sets of strong and weak hydrogen bonding patterns. Some hydrogen bonding sets are direct, while others link available groups on the peptide backbone and/or hydroxyproline through the mediation of one or more water molecules. An example of the kinds of hydrogen bonds which have been identified in the crystal structure is shown in figure 1. Of all available hydrogen bonding groups, only the NH in the Y position (when it is not an imino acid) bonds to water that does not further bind to a peptide chain^{45,46}.

Future challenges include the sequence specificity of the collagen features and the alterations in structure resulting from mutations in collagen. The major advances in recombinant DNA technology, together with progress on designed peptides, should provide an impetus to addressing these important challenges.

Acknowledgments

I thank Dr Jordi Bella for very helpful discussions. This work was supported by NIH AR 19626 and a grant from the Children's Brittle Bone Foundation.

References

1. Ramachandran G N and Kartha G 1954 *Nature (London)* **174** 269
2. Ramachandran G N and Ambady G K 1954 *Curr. Sci.* **23** 349
3. Ramachandran G N and Kartha G 1955 *Nature (London)* **176** 593
4. Cochran W, Crick F H C and Vand V 1952 *Acta Crystallogr.* **5** 581
5. Cohen C and Bear R S 1953 *J. Am. Chem. Soc.* **75** 2783
6. Crick F H C and Rich A 1955 *Nature (London)* **176** 780
7. Rich A and Crick F H C 1955 *Nature (London)* **176** 915
8. Cowan P M, McGavin S and North A C T 1955 *Nature (London)* **176** 1062

9. Ramachandran G N (ed.) 1967 In *Treatise on Collagen* (New York: Academic Press) pp. 103–183.
10. Fraser R D B and MacRae T P 1973 In *Conformation in fibrous proteins* (New York: Academic Press)
11. History 1998 *Nature Struct. Biol.* **5** 675
12. History 1998 *Nature Struct. Biol.* **5** 760
13. History 1998 *Nature Struct. Biol.* **5** 858
14. Jones E Y and Miller A 1991 *J. Mol. Biol.* **218** 209
15. Pauling L, Corey R B and Branson H R 1951 *Proc. Natl. Acad. Sci.* **37** 205
16. Rich A and Crick F H C 1961 *J. Mol. Biol.* **3** 483
17. Ramachandran G N and Chandrasekharan R 1968 *Biopolymers* **6** 1649
18. Ramachandran G N, Sasisekharan V and Ramakrishnan C 1966 *Biochim. Biophys. Acta* **112** 168
19. Ramachandran G N and Sasisekharan V 1965 *Biochim. Biophys. Acta* **109** 314
20. Ramachandran G N, Bansal M and Bhatnagar R S 1973 *Biochim. Biophys. Acta* **322** 166
21. Suzuki E, Fraser R D B and MacRae T P 1980 *Int. J. Biol. Macromol.* **2** 54
22. Privalov P L 1982 *Adv. Protein Chem.* **35** 1
23. Andreeva N S, Esipova N G, Millionova M I, Rogulenkova V N and Shibnev V A 1967 In *Conformation of Biopolymers-2* (ed.) G N Ramachandran (London: Academic Press) pp. 469–477.
24. Traub W, Yonath A and Segal D M 1969 *Nature (London)* **221** 914
25. Doyle B B, Traub W, Lorenzi G P and Blout E R 1971 *J. Am. Chem. Soc.* **10** 3052
26. Brodsky B and Ramshaw J A M 1997 *Matrix Biol.* **15** 545
27. Fields G B and Prockop D J 1996 *Biopolymers* **40** 345
28. Heidemann E and Roth W 1982 *Adv. Polymer Sci.* **43** 143
29. Tanaka T, Nishikawa A, Tanaka Y, Nakamura H, Kodama T, Imanishi T and Doi T 1996 *Protein Eng.* **9** 307
30. Goodman M, Bhūmralkar M, Jefferson E A, Kwak J and Locardi E 1998 *Biopolymers* **47** 127
31. Ottl J, Battistuta R, Pieper M, Tschesche H, Bode W, Kuhn K and Moroder L 1996 *FEBS Lett.* **398** 31
32. Baum J and Brodsky B 1997 *Fold Des.* **2** R53
33. Bella J, Eaton M, Brodsky B and Berman H A M 1994 *Science* **266** 75
34. Sakikabara S, Inouye K, Shudo K, Kishida Y, Kobayashi Y and Prockop D J 1973 *Biochim. Biophys. Acta* **303** 198
35. Long C G, Braswell E, Zhu D, Apigo J, Baum J and Brodsky B 1993 *Biochemistry* **32** 11688
36. Fraser R D B, MacRae T P and Suzuki E 1979 *J. Mol. Biol.* **129** 463
37. Bella J and Berman H M 1996 *J. Mol. Biol.* **264** 734
38. Bella J, Brodsky B and Berman H A M 1995 *Structure* **3** 893
39. Kramer R Z, Vitagliano L, Bella J, Beriso R, Mazzarella L, Brodsky B, Zagari A and Berman H A M 1998 *J. Mol. Biol.* **280** 623
40. Eberhardt E S, Panasik N J and Raines R T 1996 *J. Am Chem. Soc.* **118** 12261
41. Holmgren S K, Taylor K M, Bretscher L E and Raines R T 1998 *Nature (London)* **392** 666
42. Leikin S, Parsegian V A, Yang W-H and Walrafen G E 1997 *Proc. Natl. Acad. Sci.* **94** 11312
43. Li M-H, Fan P, Brodsky B and Baum J 1993 *Biochemistry* **32** 7377
44. Fan P, Li M-H, Brodsky B and Baum J 1993 *Biochemistry* **32** 13299
45. Kramer R Z 1998 *Crystallographic structural characterization of three collagen-like peptides*, Ph D thesis, Rutgers University, New Jersey
46. Kramer R Z and Berman H A M *J. Biomol. Struct. Dyn.* (in press)