A Template-Induced *Incipient* Collagen-Like Triple-Helical Structure

Murray Goodman,* Yangbo Feng, Giuseppe Melacini, and Joseph P. Taulane

Department of Chemistry & Biochemistry
University of California, San Diego
La Jolla, California 92093-0343

Received December 8, 1995

The folding of protein and peptide structures has been the focus of numerous recent investigations,[1] including three-[2] or four-helical bundles[3] and β-sheet structures.[4] Our study concerns the mimicry of the major domain found in all collagens: the triple helix. The triple helix consists of three extended polyproline-II-like chains intertwined around a common axis forming a superhelix.[5,6] The primary sequences of collagen triple helices are characterized by repeating triplets Gly-X-Y.[7−9] In particular, Gly-Pro-Hyp (Hyp represents 4-hydroxyproline) is a highly populated triplet sequence in naturally occurring collagens.[8,10−15]

In this communication, we report the use of a conformationally constrained organic template to induce collagen-like triple-helical structures composed of Gly-Pro-Hyp sequences. This template, *cis,cis*-1,3,5-trimethylcyclohexane-1,3,5-tricarboxylic acid (known as the Kemp triacid, KTA),[16] possesses three carboxyl groups which can be coupled to the N-termini of three peptide chains. We insert a Gly residue as a spacer between each peptide chain and each carboxyl group on KTA to compensate for the difference in diameters between the KTA and the collagen triple helix[17] and to facilitate the synthesis.[18] Similar template-assembling approaches have been reported previously, where the templates employed are either 1,2,3-propanetricarboxylic acid[19] or Lys-Lys dimer,[20−26] both of which are much more flexible than the KTA-based template.

On the basis of the above design rationale, we synthesized a template-assembled peptide composed of three repeating Gly-Pro-Hyp triplets per chain: KTA-[Gly-(Gly-Pro-Hyp)$_3$-NH$_2$]$_3$ **(KTAg-3,3)** (Figure 1). The corresponding analog with only one Gly-Pro-Hyp triplet per chain, KTA-[Gly-(Gly-Pro-Hyp)$_1$-NH$_2$]$_3$ **(KTAg-1,3)**, and the non-template-assembled analog Ac-(Gly-Pro-Hyp)$_3$-NH$_2$ **(Ac-3)** were also prepared as closely related parent structures. Our NMR data demonstrate that **KTAg-1,3** and **Ac-3** are incapable of forming triple-helical conformations.[17]

The ^{1}H-NMR data for **KTAg-3,3** establish a triple-helical structure. From the 2D ^{1}H-NMR TOCSY and NOESY experiments we obtained the residue specific assignments for KTA-[Gly-(Gly-Pro-Hyp)$_1$-NH$_2$]$_3$ **(KTAg-1,3)**, KTA-[Gly-(Gly-Pro-Hyp)$_3$-NH$_2$]$_3$ **(KTAg-3,3)**, and Ac-(Gly-Pro-Hyp)$_3$-NH$_2$ **(Ac-3)** in H$_2$O (5 °C, pH = 3.4 ± 0.1).[17] For **KTAg-3,3** an additional set of resonances was observed which is absent in **KTAg-1,3** and **Ac-3**. The additional set of resonances observed for **KTAg-3,3** is very similar to the triple-helical resonance set of (Pro-Hyp-Gly)$_{10}$ in H$_2$O at 10 °C (with a root-mean-square difference between the chemical shifts of the two sets of resonance of 0.07 ppm).[27] Among the resonances of the triple helical set, the Pro $C_\delta H_h$ hydrogen at 3.2 ppm is well resolved (Figure 2, left) and is not overlapped by any resonances of the non-triple-helical set. The resonance at 3.2 ppm is therefore useful for identification and quantification of triple-helical conformations.[28] Integration of the triple-helical Pro $C_\delta H_h$ resonance at 3.2 ppm shows that the average number of Pro residues per molecule in a triple-helix-like environment is 6 ± 1, out of a total of 9 Pro residues per molecule. This result is consistent with the typical triple-helix register shift which forces at least one Gly-Pro-Hyp triplet per chain to be only partially involved in the triple-helical array.

The formation of a triple-helix-like structure for **KTAg-3,3** in H$_2$O (5 °C, pH = 3.4 ± 1) is also supported by the NOESY spectra which were analyzed according to the procedure previously used for the NMR study of (Pro-Hyp-Gly)$_{10}$.[27] This approach relies on the distinction between intra- and interchain NOEs on the basis of the X-ray-derived triple-helical model proposed for sequences containing the Gly-Pro-Hyp repeat.[29] In particular eight interchain NOEs provide a critical test for the triple-helical array because they are anticipated to arise uniquely from interchain interactions on the basis of the X-ray model (distances smaller than 4.5 Å) and correspond to non-overlapped resonances.[27,29] These interchain NOEs are Pro $C_\gamma H_1$−Hyp $C_\beta H_1$, Pro $C_\gamma H_1$−Hyp $C_\beta H_2$, Pro $C_\gamma H_1$−Hyp $C_\gamma H$, Pro $C_\delta H_1$−Hyp $C_\beta H_1$, Pro $C_\delta H_1$−Hyp $C_\beta H_2$, Pro $C_\delta H_2$−Hyp $C_\beta H_2$, Pro $C_\delta H_1$−Hyp $C_\gamma H$, and Gly NH−Pro $C_\delta H_1$ (the hydrogens are named according to ref 30).[27] The first seven NOEs agree with the observed NOESY cross peaks (Figure 2, right), while the Gly NH−Pro $C_\delta H_1$ NOE could not be observed under our experimental conditions. This observation indicates that some structural distortions occur in **KTAg-3,3** as compared to the longer chain model compound (Pro-Hyp-Gly)$_{10}$. The differences between (Pro-Hyp-Gly)$_{10}$ and **KTAg-3,3** can be explained by the triple-helical end effects which are more significant in the short chain compound.

The circular dichroism (CD) spectrum of **KTAg-3,3** in H$_2$O at 20 °C is also consistent with the CD spectra (the peak, the

(1) Mutter, M.; Vuilleumier, S. *Angew. Chem., Int. Ed. Engl.* **1989**, *28*, 535−676.
(2) Ghadiri, M. R.; Case, M. A. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1594−1597.
(3) Choma, C. T.; Lear, J. D.; Nelson, M. J.; Dutton, P. L.; Roberson, D. E.; DeGrado, W. F. *J. Am. Chem. Soc.* **1994**, *116*, 856−865.
(4) Schneider, J. P.; Kelly, J. W. *J. Am. Chem. Soc.* **1995**, *117*, 2533−2546.
(5) Ramachandran, G. N. *Aspects of Protein Structure*; Ramachandran, G. N., Ed.; Academic Press: New York, 1963; pp 39.
(6) Bornstein, P.; Traub, W. *The Proteins*; Neurath, H., Hill, R. L., Eds.; Academic Press: New York, 1979; Vol. 2, pp 412−632.
(7) Traub, W.; Piez, K. *Adv. Protein Chem.* **1971**, *25*, 243−341.
(8) Ramachandran, G. N.; Ramakrishnan, C. *Biochemistry of Collagen*; Ramachandran, G. N., Reddi, A. H., Eds.; Plenum Press: New York, 1976; pp 45−84.
(9) Brodsky-Doyle, B.; Leonard, K. R.; Reid, K. B. M. *Biochem. J.* **1976**, *159*, 279−286.
(10) Bhatnagar, R. S.; Rapaka, R. S. *Biochemistry of Collagen*; Ramachandran, G. N., Reddi, A. H., Eds.; Plenum Press: New York, 1976; pp 479−521.
(11) Jones, E. Y.; Miller, A. *J. Mol. Biol.* **1991**, *218*, 209−219.
(12) Long, C. G.; Thomas, M.; Brodsky, B. *Biopolymers* **1995**, *35*, 621−628.
(13) Fietzek, P. P.; Kuhn, K. *Mol. Cell Biochem.* **1975**, *8*, 141−157.
(14) Fietzek, P. P.; Kuhn, K. *Int. Rev. Connect. Tissue Res.* **1976**, *7*, 1.
(15) Dolz, R.; Heidemann, E. *Biopolymers* **1986**, *25*, 1069−1080.
(16) Kemp, D. S.; Petrakis, K. S. *J. Org. Chem.* **1981**, *46*, 5140−5143.
(17) Melacini, G.; Feng, Y.; Goodman, M. Submitted for publication in *J. Am. Chem. Soc.*
(18) Feng, Y.; Melacini, G.; Taulane, J. P.; Goodman, M. Submitted for publication in *J. Am. Chem. Soc.*
(19) Greiche, Y.; Heidemann, E. *Biopolymers* **1979**, *18*, 2359−2361.
(20) Fields, G. B. *J. Theor. Biol.* **1991**, *153*, 585−602.
(21) Fields, C. G.; Lovdahl, C. M.; Miles, A. J.; Matthias Hagen, V. L.; Fields, G. B. *Biopolymers* **1993**, *33*, 1695−1707.
(22) Fields, C. G.; Mickelson, D. J.; Drake, S. L.; McCarthy, J. B.; Fields, G. B. *J. Biol. Chem.* **1993**, *268*, 14153−14160.
(23) Germann, H. P.; Heidemann, E. *Biopolymers* **1988**, *27*, 157−163.
(24) Roth, W.; Heidemann, E. *Biopolymers* **1980**, *19*, 1909−1917.
(25) Tanaka, T.; Wada, Y.; Nakamura, H.; Doi, T.; Imanishi, T.; Kodama, T. *FEBS Lett.* **1993**, *334*, 272−276.
(26) Thakur, S.; Vadolas, D.; Germann, H. P.; Heidemann, E. *Biopolymers* **1986**, *25*, 1081−1086.
(27) Li, M.; Fan, P.; Brodsky, B.; Baum, J. *Biochemistry* **1993**, *32*, 7377−7387.
(28) Brodsky, B.; Li, M.; Long, C. G.; Apigo, J.; Baum, J. *Biopolymers* **1992**, *32*, 447−451.
(29) Fraser, R. D. B.; MacRae, T. P.; Suzuki, E. *J. Mol. Biol.* **1979**, *129*, 463−481.
(30) Brown, F. R., III; Carver, J. P.; Blout, E. R. *J. Mol. Biol.* **1969**, *39*, 307−313.

S0002-7863(95)04132-1 CCC: $12.00

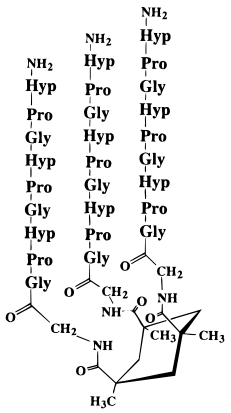

Figure 1. Structure of KTA-[Gly-(Gly-Pro-Hyp)$_3$-NH$_2$]$_3$, denoted as **KTAg-3,3**. In this simplified notation, the first number indicates the number of repeats of Gly-Pro-Hyp per chain, while the second number refers to the number of chains substituted on the Kemp triacid derivative.

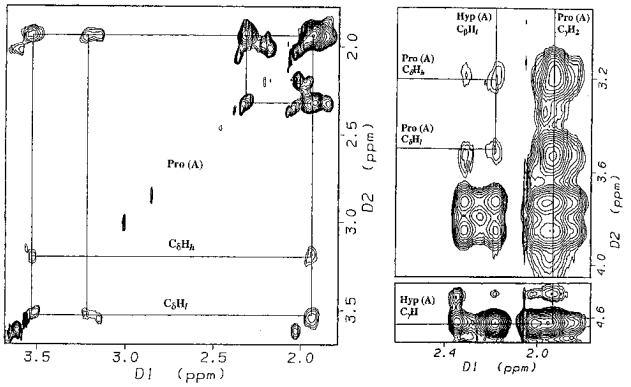

Figure 2. (Left) Expanded region of the 2D ^{1}H-NMR TOCSY spectrum (t_{mix} = 49 ms) of **KTAg-3,3** in H_2O, pH = 3.4 ± 0.1 at 5 °C. The solid lines indicate the side chain connectivity of a representative Pro residue in the triple-helical set of resonances denoted by (A). (Right) Expanded region of the 2D ^{1}H-NMR NOESY spectrum (t_{mix} = 300 ms) of **KTAg-3,3** in H_2O, pH = 3.4 ± 0.1 at 5 °C. The solid lines indicate cross peaks corresponding to representative interchain NOEs.

crossover, and the trough) of collagen[9,30] and (Gly-Pro-Hyp)$_{10}$[9,27,31−33] as reported in the literature. In addition, the presence of helical conformations in **KTAg-3,3** is clearly seen in melting curve measurements. Optical rotations were measured for all three analogs in H_2O (Figure 3). In Figure 3, **KTAg-1,3** and **Ac-3** do not show any transitions while **KTAg-3,3** exhibits a melting curve. The melting curve of **KTAg-3,3** with the midpoint of the broad transition at 30 °C indicates some cooperativity. This result is appropriate for the short chains present in **KTAg-3,3**.

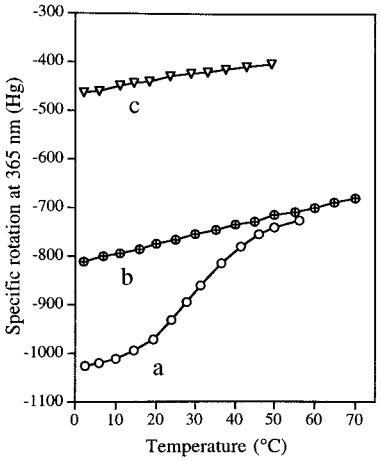

Figure 3. Thermal melting curves of **KTAg-3,3** (a), **Ac-3** (b), and **KTAg-1,3** (c) in H_2O (0.2 mg/mL) obtained by optical rotation measurements.

On the basis of the NMR and optical rotation results we can conclude that **KTAg-3,3** forms a triple-helical conformation in H_2O with a melting temperature of 30 °C. We believe this compound represents the shortest chain polypeptide able to form a triple helical structure at room temperature in H_2O reported to date. This compound will be useful as a model system for the study of the early stages of triple-helix folding. Therefore, we denote this borderline ordered structure as an *incipient* triple helix. In related papers, we report on a series of KTA-terminated sequences (KTA-[Gly-(Gly-Pro-Hyp)$_n$-NH$_2$]$_3$ with $n > 3$) where the triple-helical structures are more highly defined and more stable.[17,18]

Acknowledgment. This project is funded by grants from the National Science Foundation (NSFDMR-9201133) and Ciba-Vision Inc.

Supporting Information Available: The synthesis, HPLC profiles, 1-D ^{1}H-NMR spectra, and mass spectra of the collagen analogs (10 pages). Ordering information is given on any current masthead page.

JA954132H

(31) Venugopal, M. G.; Ramshaw, J. A. M.; Braswell, E.; Zhu, D.; Brodsky, B. *Biochemistry* **1994**, *33*, 7948−7956.
(32) Inouye, K.; Kobayashi, Y.; Kyogoku, Y.; Kishida, Y.; Sakakibara, S.; Prockop, D. J. *Arch. Biochem. Biophys.* **1982**, *219*, 198−203.
(33) Sakakibara, S.; Inouye, K.; Shudo, K.; Kishida, Y.; Kobayashi, Y.; Prockop, D. J. *Biochim. Biophys. Acta* **1973**, *303*, 198−202.