Collagen-Based Structures Containing the Peptoid Residue *N*-Isobutylglycine (Nleu). 6. Conformational Analysis of Gly-Pro-Nleu Sequences by ^{1}H NMR, CD, and Molecular Modeling

Giuseppe Melacini, Yangbo Feng, and Murray Goodman*

Contribution from the Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, California 92093-0343

Received May 20, 1996[⊗]

Abstract: Molecular modeling, ^{1}H NMR, and CD were employed to study the structure and stability of collagen-like triple helices composed of Gly-Pro-Nleu repeats. The compounds studied include the acetyl analogs Ac-(Gly-Pro-Nleu)$_n$-NH$_2$ (where $n = 1$, 6, 9) and the KTA conjugates KTA-[Gly-(Gly-Pro-Nleu)$_n$-NH$_2$]$_3$ (where $n = 1$, 3, 6, 9 and KTA denotes the Kemp triacid). The presence of collagen-like assembled structures was supported by a consistent set of experimental observations, including the appearance of a distinct set of resonances, low hydrogen exchange rates for Gly NH, KTA signal splitting, cooperative melting transition, and analysis of NOESY cross peaks. In this regard, the concept of *ensemble interchain NOEs* was introduced and used to establish the close packing of Gly, Pro, and Nleu residues in triple helices composed of Gly-Pro-Nleu repeats. In addition, the ensemble interchain NOEs gave insight into the puckering of the Pro ring and the conformations accessible to the Nleu side chain. The effect of the KTA template on triple helicity was studied and shown to consist in a net gain in the free energy of triple-helix formation, as also seen for Gly-Pro-Hyp sequences. This free energy gain led to the induction of an assembled collagen-like structure in the KTA conjugate containing six Gly-Pro-Nleu repeats per chain and to an increase in thermal stability of the compound containing nine Gly-Pro-Nleu repeats per chain.

Introduction

This paper is part of a series of investigations on assembled collagen-like triple helices composed of Gly-Pro-Nleu repeats (where Nleu represents the peptoid *N*-isobutylglycine[1]). The Gly-Pro-Nleu sequences studied are terminated either by an acetyl group or by KTA (*cis*, *cis*-1,3,5-trimethylcyclohexane-1,3,5-tricarboxylic acid, also known as the Kemp triacid[2]). These compounds were designed to mimic collagen-like triple helices which can be used in the development of novel biomaterials. In addition, the introduction of an unnatural imino acid (Nleu) can also lead to enhanced resistance to enzymatic degradation.[1]

In the first paper of the series[3] the combined use of circular dichroism (CD), optical rotation spectroscopy, and preliminary nuclear magnetic resonance (NMR) studies provided evidence of ordered assembled structures in solution with properties similar to those of collagen triple helices. Additional insight into collagen-like triple helices composed of Gly-Pro-Nleu repeats can be gained by further NMR and molecular modeling studies. In this paper we present the characterization of the compounds in Table 1 using 1D and 2D ^{1}H NMR to obtain sequence specific assignments, solvent shielding, triple-helix percentage, and NOE information. The NOEs are used to critically test modeled triple-helical structures and investigate the packing of Gly, Pro, and Nleu residues within the triple-helical assembly. The resulting studies provide also new insight on the multiple conformations accessible to the Nleu isobutyl side chain and to the Pro five-membered ring.

Table 1. Compounds Studied

compound	*n*	no.
Ac-(Gly-Pro-Nleu)$_n$-NH$_2$[a]	1	**I**
	6	**II**
	9	**III**
KTA-[Gly-(Gly-Pro-Nleu)$_n$-NH$_2$]$_3$[b]	1	**IV**
	3	**V**
	6	**VI**
	9	**VII**

[a] For $n = 1$, the C-terminal amide is replaced by an *N*-methyl amide in order to increase the structure similarity between this tripeptide and the longer-chain compounds. [b] KTA denotes the Kemp triacid.

Materials and Methods

Sample Preparation. All compounds were synthesized as described elsewhere.[4] The NMR samples were prepared in H_2O/D_2O (9:1) and in D_2O (purchased from Isotec, Inc.) with a peptide concentration in the range of 1.5−2.0 mg/mL. After dissolution in water, all samples were kept at 5 °C at least 49 days prior to any measurement of triple-helical percentage, in order to allow for a proper equilibration of the sample.[5] The pH was adjusted to 2.8 ± 0.1 (direct pH meter reading without correction for isotope effects) for all samples.

NMR and CD Spectroscopies. All NMR experiments were carried out on an AMX-500 Bruker spectrometer using the same data acquisition and processing scheme[6−10] previously used for the study of triple helices composed of -(Gly-Pro-Hyp)$_n$- sequences.[11] For the hydrogen exchange measurements, 1D spectra were collected at 5 °C over a period of 65 days. The NOESY spectra were obtained at mixing

[⊗] Abstract published in *Advance ACS Abstracts,* October 15, 1996.

(1) Simon, R. J.; Kania, R. S.; Zuckermann, R. N.; Huebner, V. D.; Jewell, D. A.; Banville, S.; Ng, S.; Wang, L.; Rosenberg, S.; Marlowe, C. K.; Spellmeyer, D. C.; Tan, R.; Frankel, A. D.; Santi, D. V.; Cohen, F. E.; Bartlett, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89* (20), 9367−9371.

(2) Kemp, D. S.; Petrakis, K. S. *J. Org. Chem.* **1981**, *46*, 5140−5143.

(3) Goodman, M.; Melacini, G.; Feng, Y. *J. Am. Chem. Soc.* **1996**, *118*, 10928−10929. This is paper 4 in our series on collagen-based structures.

(4) Feng, Y.; Melacini, G.; Taulane, J. P.; Goodman, M. *Biopolymers* In press. This is paper 5 in our series on collagen-based structures.

(5) Long, C. G.; Braswell, E.; Zhu, D.; Apigo, J.; Baum, J.; Brodsky, B. *Biochemistry* **1993**, *32*, 11688−11695.

(6) (a) Aue, W. P.; Bartholdi, E.; Ernst, R. R. *J. Chem. Phys.* **1976**, *64*, 2229−2246. (b) Bax, A.; Freeman, R. *J. Magn. Reson.* **1981**, *44*, 542−561. (c) Rance, M.; Sorensen, O. W.; Bodenhausen, G.; Wagner, G.; Ernst, R. R.; Wuthrich, K. *Biochem. Biophys. Res. Commun.* **1984**, *117*, 479−485. (d) Shaka, B.; Freeman, R. *J. Magn. Reson.* **1983**, *51*, 169−173.

S0002-7863(96)01719-2 CCC: $12.00

times of 50, 100, and 150 ms. ROESY experiments were carried out with a mixing time of 150 ms and a spin-locking field of 2.5 kHz. The integrals were normalized using the peaks in the spectral region from 0.7 to 1.1 ppm which contains a known number of protons from the Nleu $C_\delta H_{3r,s}$ groups. See Supporting Information for the experimental section on CD spectra.

Molecular Modeling. Structures were built using InsightII (Biosym Technologies, Inc.). Energy minimizations were carried out using the Discover program[12] and the force field CFF91. The CFF91 force field provided high-quality parameters[13] for all the Nleu internal coordinates. A distance dependent dielectric constant was used to approximate the solvent effects.[14] For the molecular modeling of triple helices composed of Gly-Pro-Nleu sequences, we focused on the model molecule Ac-(Gly-Pro-Nleu)$_4$-NH_2. The same approach was used by Scheraga *et al*. for the molecular modeling of other triple helical peptides composed of collagen-like sequences.[15−20]

The molecular modeling of Ac-(Gly-Pro-Nleu)$_4$-NH_2 triple helices includes two basic steps. First, the backbones of three Ac-(Gly-Pro-Nleu)$_4$-NH_2 chains have to be assembled into a triple-helical assembly. Second the conformational space accessible to the Nleu side chain needs to be searched for minimum energy structures. The first step concerns only the chain backbone and therefore can be accomplished using the model molecule Ac-(Gly-Pro-Sar)$_4$-NH_2 (where Sar denotes *N*-methylglycine) which does not contain any flexible side chain and simplifies the process of triple-helical assembly. The three Ac-(Gly-Pro-Sar)$_4$-NH_2 molecules were constrained to have the same backbone torsions as the structure proposed by Miller, Nemethy, and Scheraga[17] for (Gly-Pro-Hyp)$_n$. This conformation has been shown to be very stable with respect to substitutions of other amino acids in place of Pro and Hyp[21] and is therefore a good candidate to use as starting backbone conformation for minimizations. Since it was not possible to *a priori* discriminate among the two Pro puckerings (up and down),[16] two independent simulations were carried out, one with Pro puckered up and the other with Pro puckered down.

Three Ac-(Gly-Pro-Sar)$_4$-NH_2 chains were assembled into a triple-helical array using a protocol similar to that used for the modeling of -(Gly-Pro-Hyp)$_n$-:[11] interchain hydrogen bond[15−17] distance restraints were used as input for distance geometry runs[22] which resulted in 100 triple-helical structures. The average backbone pairwise root mean square deviation (RMSD) between the average minimized structure and the ensemble of the 100 distance geometry triple-helical structures was computed to be 0.51 ± 0.03 Å. The average minimized structure was therefore used to build triple-helical Ac-(Gly-Pro-Nleu)$_4$-NH_2 conformations by addition of isopropyl groups to the Sar methyls. The conformation of the Nleu side chain was described in terms of the χ_1 and χ_2 torsional angles, defined by the Nleu atoms C_α-N-C_β-C_γ and N-C_β-C_γ-$C_{\delta,r}$, respectively (see also Figure 1 for the atom naming). The χ_1 and χ_2 torsions of the 12 Nleu residues contained in the three Ac-(Gly-Pro-Nleu)$_4$-NH_2 strands of the triple helix were set to identical values according to the triple-helical ternary screw symmetry. A grid search[23] was then carried out in the (χ_1,χ_2) torsional space, and the resulting structures were minimized and clustered until convergence was achieved (see Supporting Information).

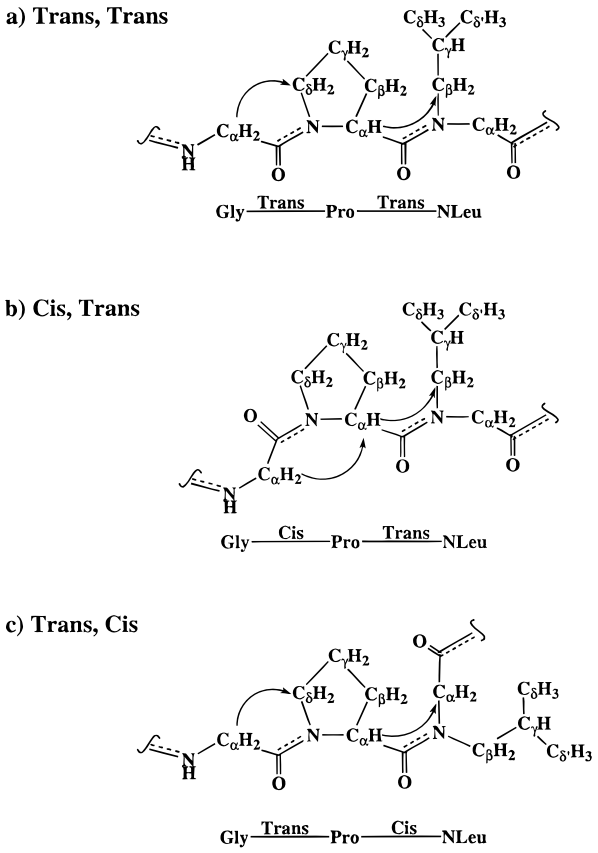

Figure 1. Schematic diagram of the structures corresponding to the three unassembled sets of resonances observed for sequences composed of Gly-Pro-Nleu repeats. The arrows indicate the critical NOEs used to assign the peptide bond structure (cis/trans) in each resonance set. The trans, trans set corresponds to 61% of the total intensity, while the cis, trans and trans, cis sets account for 6% and 33% of the total intensity, respectively (see also Supporting Information).

Results and Discussion

Assignments. For compounds **I**−**II** and **IV**−**V** (Table 1) at 27 °C, three sets of resonances can be detected using DQF-COSY, TOCSY, and ROESY experiments, and these sets correspond to different peptide bond structures (cis/trans), as explained in Figure 1 and in the Supporting Information section. For compounds **III**, **VI**, and **VII** (Table 1) at 27 °C an additional set of resonances can be observed which is absent in the other compounds under investigation. The sequential connectivities from Gly to Pro, from Pro to Nleu, and from Nleu to Gly of the new set of resonances are consistent with the NOESY spectra

(7) (a) Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. *J. Chem. Phys.* **1979**, *71*, 4546−4553. (b) Kumar, A.; Ernst, R. R.; Wuthrich, K. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1−6. (c) Otting, G.; Widmer, H.; Wagner, G.; Wuthrich, K. *J. Magn. Reson.* **1986**, *66*, 187−193. (d) Wider, G.; Macura, S.; Kumar, A.; Ernst, R. R.; Wuthrich, K. *J. Magn. Reson.* **1984**, *56*, 207−234.

(8) (a) Braunsweiler, L.; Ernst, R. R. *J. Magn. Reson.* **1983**, *53*, 521−528. (b) Bax, A.; Davis, D. *J. Magn. Res.* **1985**, *65*, 355−360.

(9) Bothner-By, A. A.; Stephens, R. L.; Lee, J.; Warren, C. D.; Jeanloz, R. W. *J. Am. Chem. Soc.* **1984**, *106*, 811−813.

(10) (a) Redfield, A. G.; Kuntz, S. D. *J. Magn. Res.* **1975**, *19*, 250−254. (b) Marion, D.; Wuthrich, K. *Biochem. Biophys. Res. Commun.* **1983**, *113*, 967−974.

(11) Melacini, G.; Feng, Y.; Goodman, M. *J. Am. Chem. Soc.* **1996**, *118*, 10359−10364. This is paper 3 in our series on collagen-based structures.

(12) Hagler, A. T. *The Peptides*; Udenfriends, S., Meienhofer, J., Hruby, V. J., Eds.; Academic Press: Orlando, FL, 1985; Vol. 7, pp 214−296.

(13) Parameters assigned nonautomatically.

(14) McCammon, J. A.; Wolynes, P. G.; Karplus, M. *Biochemistry* **1979**, *18*, 927−942.

(15) Miller, M. H.; Scheraga, H. A. *J. Polym. Sci.* **1976**, Symposium No. 54, 171−200.

(16) Nemethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. *J. Phys. Chem.* **1992**, *96*, 6472−6484.

(17) Miller, M. H.; Nemethy, G.; Scheraga, H. A. *Macromolecules* **1980**, *13*, 470−478.

(18) Miller, M. H.; Memethy, G.; Scheraga, H. A. *Macromolecules* **1980**, *13*, 910−913.

(19) Nemethy, G.; Miller, M. H.; Scheraga, H. A. *Macromolecules* **1980**, *13*, 914−919.

(20) Zagari, A.; Nemethy, G.; Scheraga, H. A. *Biopolymers* **1990**, *30*, 967−974.

(21) Nemethy, G. *Biochimie* **1981**, *63*, 125−130.

(22) (a) Quantum Chemistry Program Exchange No. 590, Indiana University Department of Chemistry. (b) Havel, T. F.; Kuntz, I. D. *Bull. Math. Biol.* **1983**, *45*, 665. (c) Crippen, G. M. *Distance Goemetry and Conformational Calculations*; Bawden, D., Ed.; Research Studies Press (Wiley): New York, 1981. (d) Crippen, G. M.; Havel, T. F. *Distance Goemetry and Conformational Calculations*; Bawden, D., Ed.; Research Studies Press (Wiley): New York, 1988.

(23) Nemethy, G.; Scheraga, H. A. *Biopolymers* **1982**, *21*, 1535−155.

Table 2. Chemical Shifts of Assigned Proton Resonances in Water at 27 °C

		chemical shift (ppm)[a]					
resonance set	amino acid	NH	$C_\alpha H$	$C_\beta H$	$C_\gamma H$	$C_\delta H$	other
	Gly	7.68	3.86, 3.70				
	Pro		4.92	2.33, 1.97	2.07	3.58, 3.22	
assembled	Nleu		4.49, 3.82	3.45, 3.07	1.82	1.02, 0.91	
	Ac						∼2.05
	KTA						H_{ax} 1.26, 1.32, 1.27
							H_{eq} 2.83, 2.64, 2.51
							Me ∼1.23, 1.29, 1.26
	Gly	8.00[c]	4.11−4.19, 3.95−3.99				
	Pro		4.92	2.32, 1.94	2.04	3.63	
unassembled (trans, trans)[b]	Nleu		4.27−4.06, 4.02−3.98	3.39, 3.25[d]	2.00	0.97, 0.93	
	Ac						2.05
	KTA						H_{ax} 1.33
							H_{eq} 2.66
							Me 1.27
	Gly	7.98[c]	3.91−4.04, 3.60−3.69				
	Pro		5.07	2.47, 2.16	1.93, 1.83	3.58	
unassembled (cis, trans)[e]	Nleu		*f*	3.39, 3.23	2.00	0.97, 0.93	
	Ac						*f*
	KTA						*f*
	Gly	*f*	4.11−4.18, 3.95−4.06				
	Pro		4.67	2.23, 1.88	2.00, 2.08	3.63	
unassembled (trans, cis)	Nleu		4.37−4.52, 4.20−4.28	3.30−3.38, 3.08−3.12	1.92	0.86, 0.83	
	Ac						*f*
	KTA						*f*

[a] ^{1}H chemical shifts are reported relative to sodium 3-(trimethylsilyl)tetradeuteriopropionate (STP). [b] This set of resonances corresponds to structures with all trans peptide bonds. [c] A minor resonance can be observed also at 8.06−8.14 ppm as a result of differences between residues in the core and residues at the ends of the polypeptide chain. [d] In compound **I** these resonances appear at 3.43 and 3.30 ppm, respectively. In compound **IV** both these resonances appear at 3.35 ppm. [e] This resonance set corresponds to structures with a cis Gly-Pro peptide bond and a trans Pro-Nleu peptide bond. [f] Not accessible. [g] This resonance set corresponds to structures with a trans Gly-Pro peptide bond and a cis Pro-Nleu peptide bond.

leading to the definition of this new set of resonances as a distinct tripeptide unit.[3]

The set of resonances detected only for analogs **III** and **VI**−**VII** has proven to correspond to collagen-like triple-helical structures[3] and can therefore be denoted as the assembled set. Accordingly, the unassembled sets are those observed for compounds **I** and **IV**, which lack the structural requirements to form a triple helix and therefore assume less-ordered conformations. Table 2 shows the assignments for the assembled and unassembled sets of resonances. It is notable that among those resonances of the assembled set which are not significantly overlapped by any resonance belonging to the unassembled sets, the resonance of Nleu $C_\delta H_{3,l}$ at 1.02 ppm is the best resolved (the subscript *l* denotes the low-field resonance, while the subscript *h* denotes the high-field resonance). The resonance at 1.02 ppm can therefore be used to identify triple-helical conformations. We employed this resonance at 1.02 ppm to study the triple helicity as a function of chain length, of template, and of temperature, as discussed below.

Triple Helicity As a Function of Chain Length and Template. Figure 2a−g shows the 1D spectral region containing the Nleu $C_\delta H_{3l,h}$ resonances of the compounds studied at 27 °C and sorted according to the number of Gly-Pro-Nleu repeats per chain. It can be seen that for the KTA conjugates six Gly-Pro-Nleu repeats are sufficient to observe the triple-helical Nleu $C_\delta H_{3,l}$ resonance at 1.02 ppm (Figure 2e), while for the acetyl analogs nine Gly-Pro-Nleu repeats are necessary to be able to observe the assembled resonance under our experimental conditions (Figure 2f). Integrations of the Nleu $C_\delta H_{3,l}$ resonance at 1.02 ppm indicate that for the KTA conjugate containing six Gly-Pro-Nleu repeats per chain (**VI**) the average number of triple-helical Nleu residues per chain is 3.9 at 27 °C, while for the KTA conjugate containing nine Gly-Pro-Nleu repeats per chain (**VII**) the average number of triple-helical Nleu residues per chain is 7.3 at the same temperature. This number of assembled Nleu residues per chain is higher than that measured for the corresponding acetyl analog (**III**) which is 3.1 even after an equilibration time of 4 months and 10 days.

Similar trends for the triple helicity as a function of chain length are observed by CD (see Supporting Information, Figure 1): only compounds **III** and **VI**−**VII** show the typical red shift (crossover and positive peak) associated with triple-helix formation.[3,4] These observations are fully consistent with previous studies on KTA conjugates composed of Gly-Pro-Hyp sequences which have shown that the KTA template reduces the entropy loss associated with triple-helical formation without any significant enthalpy variation.[11,24] Thus, the KTA template allows a net free energy gain which results in the induction of triple-helical structures even in short-chain molecules and increases the percentage of triple-helical structures of longer-chain compounds, as compared to the corresponding acetyl analogs.

Triple Helicity As a Function of Temperature. The melting of the triple-helical conformation of compound **VI**[25] was monitored using the normalized 1D integral of the assembled Nleu $C_\delta H_{3,l}$ resonance at 1.02 ppm (Figure 3). Figure 3 indicates that at 5 °C the average number of triple-helical Nleu residues per chain is close to 5, which is equal to the number of triplets per chain minus 1, as expected for the longest triple-helical structure allowed by the chain length. In this elongated triple helix, one triplet per chain is only partially packed in the triple-helical array as a consequence of the register shift (Figure 4a). Figure 3 shows also that the transition from assembled to unassembled structures of compound **VI** is cooperative and is centered at 33 °C. These results are fully

(24) Feng, Y.; Melacini, G.; Taulane, J. P.; Goodman, M. *J. Am. Chem. Soc.* **1996**, *118*, 10351−10358. This is paper 2 in our series on collagen-based structures.

(25) No reliable data could be collected for the melting of the longer-chain compounds **III** and **VII** because of the presence of aggregation phenomena at high temperatures.

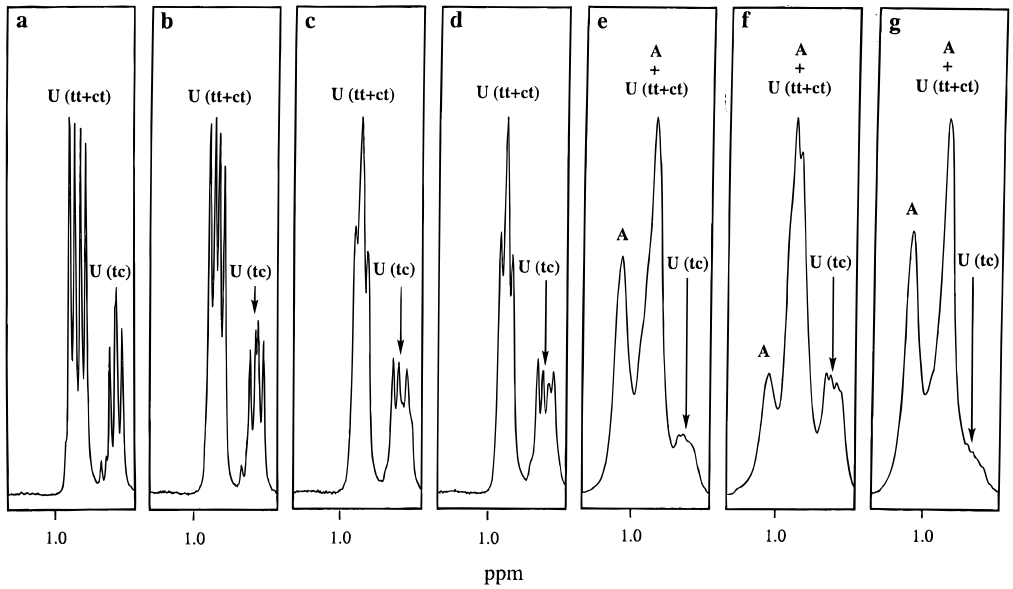

Figure 2. Expanded Nleu $C_\delta H_3$ region of 1D spectra at 27 °C of the compounds under investigation sorted according to their chain length: **I** (a), **IV** (b), **V** (c), **II** (d), **VI** (e), **III** (f), **VII** (g) (see Table 1). In spectra a−d only the resonances of the unassembled set can be observed, including (trans, trans) and (cis, trans) structures (labeled as **U(tt+ct)**) and (trans, cis) structures (labeled as **U(tc)**). In spectra e−g an additional set of resonances appears for the Nleu $C_\delta H_{3,l}$ and Nleu $C_\delta H_{3,h}$ protons, and this additional set is assigned to assembled triple-helical structures (labeled as **A**).

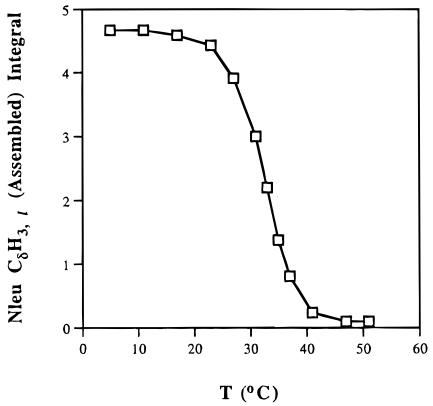

Figure 3. (a) Thermal melting curve monitored by ^{1}H NMR for KTA-[Gly-(Gly-Pro-Nleu)$_6$-NH_2]$_3$ (**VI**, Table 1). The vertical axis reports the normalized integral of the Nleu $C_\delta H_{3,l}$ signal in the assembled set (see Materials and Methods for the details of the normalization). This integral represents the average number of Nleu residues per chain in a triple-helical environment at a given temperature.

consistent with those independently obtained by optical rotation measurements (see Figure 3 of ref 3).

Conformational Characterization

Conceptual Basis. For a proper analysis of NOESY spectra it is essential to consider that in triple-helical molecules three polypeptide chains are symmetrically and closely packed and that these polypeptide chains have identical repetitive sequences. A given NOESY cross peak may therefore result from interactions between atoms of the same polypeptide chain and/or from interactions between atoms of two different polypeptide chains.[26] For instance the NOE between Nleu $C_\alpha H_r$ and Gly NH can result from either inter- or intrachain interactions. Because of these complications it is difficult to proceed to a direct structure determination of triple-helical compounds using NOE-based distance constraints. However, the inverse process can be accomplished: given a triple-helix model built on the basis of other structural techniques, such as molecular modeling, it is possible to predict which NOEs are expected to be observed (interproton distances < 4.5 Å) and check these expected NOEs against the NOESY spectra. In this regard, we took an initial structural approach in which we limited our comparison to the NOEs predicted to arise uniquely from interchain interactions, since these interchain NOEs provide a critical test of the proposed model.[26] The first step in the analysis of NOESY spectra is therefore the selection of a set of minimum energy, triple-helical structures.

The systematic conformational search for triple-helical structures composed of Gly-Pro-Nleu repeats resulted in 14 minimum energy conformational families (see Supporting Information). For the up Pro puckering, six conformational families were obtained (**u1**−**6**), while for the down Pro puckering, eight conformational families were found (**d1**−**8**). In the clusters **u1**−**3** and **d1**−**3**, the Nleu χ_1 torsion, which describes the relative orientation of the Nleu side chain and the chain backbone, is positive indicating that the Nleu side chain points toward the Pro ring. On the basis of this common feature, clusters **u1**−**3** and **d1**−**3** will be called *in* clusters (Figure 4b). In the remaining clusters Nleu χ_1 is negative, indicating that the Nleu side chain points far from the Pro ring. On the basis of this common feature, clusters **u4**−**6** and **d4**−**8** will be called *out* clusters (Figure 4b).

In the modeled Ac-(Gly-Pro-Nleu)$_4$-NH_2 triple helices (see Supporting Information), the differences in the average relative energies per Gly-Pro-Nleu repeat are not significant enough to rule out the possibility of conformational equilibria involving two or more of the 14 conformational clusters. In addition, the absent NOEs[27,28] did not allow to reliably exclude from the conformational equilibria any of the fourteen conformations. We therefore took the conservative approach of assuming that an equilibrium conformational ensemble is present in solution and that in this ensemble each of the 14 minimum energy triple-helical clusters may have a non-zero population.

(26) Li, M.-H.; Fan, P.; Brodsky, B.; Baum, J. *Biochemistry* **1993**, *32* (29), 7373−7387.

(27) Bruschweiler, R.; Blackledge, M. J.; Ernst, R. R. *J. Biomol. NMR* **1991**, *1* (1), 3−11.

(28) Blackledge, M. J.; Bruschweiler, R.; Griesinger, C.; Schmidt, J. M.; Xu, P.; Ernst, R. R. *Biochemistry* **1993**, *41* (32), 10960−10974.

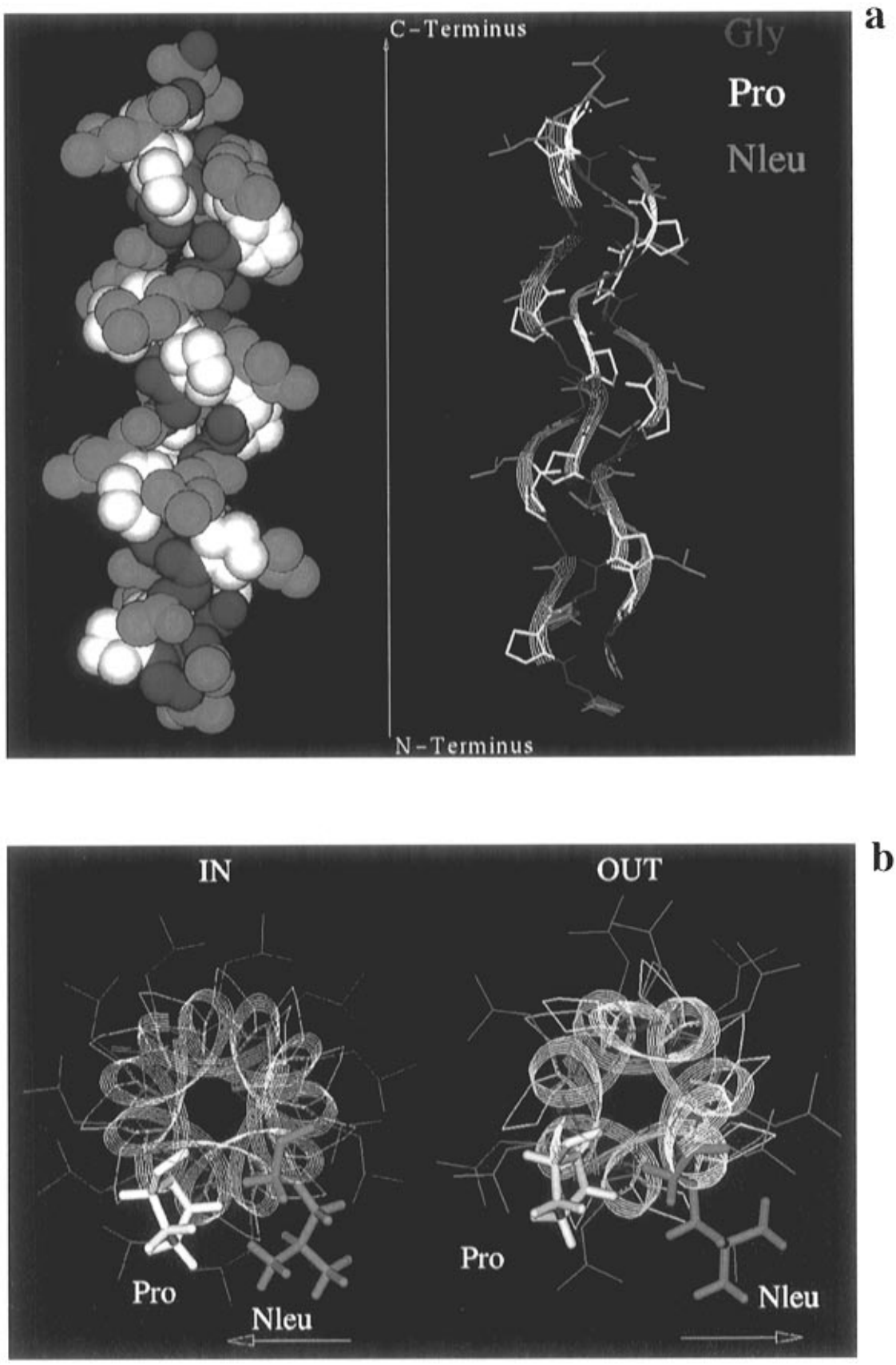

Figure 4. (a) Lowest energy triple-helical conformation of Ac-(Gly-Pro-Nleu)$_4$-NH$_2$ (cluster **u2** of Supporting Information Table 1) represented as a heavy atom CPK model (left) and as a stick model (right). For the stick model the backbone ribbon diagram is shown for each of the three chains, which are staggered by one residue. The Gly, Pro and Nleu residues are color coded with red, white, and green, respectively. (b) Top view of two representative triple-helical conformations of Ac-(Gly-Pro-Nleu)$_4$-NH$_2$ with different Nleu side chain orientations (clusters **u2** and **u5** of Supporting Information Table 1). The backbone ribbon is shown in white, while the Gly, Pro and Nleu residues are color coded as in Figure 4a. The Pro and Nleu residues of a central cross section are highlighted showing that the Nleu side chain points toward the Pro ring when Nleu $\chi_1 > 0$ (structures named *in*), while the Nleu side chain points far from the Pro ring when Nleu $\chi_1 < 0$ (structures named *out*).

Table 3. NOEs Anticipated to Arise Uniquely from *Interchain* Interactions Based on the Molecular Models of Triple Helices Composed of Gly-Pro-Nleu Repeats[a]

	pro up clusters						pro down clusters								
	in			*out*			*in*			*out*					
expected NOE	**u1**	**u2**	**u3**	**u4**	**u5**	**u6**	**d1**	**d2**	**d3**	**d4**	**d5**	**d6**	**d7**	**d8**	exptl NOE[b]
Gly NH-Pro $C_\beta H_r$	w	w	w	w	w	w									obsd
Gly NH-Pro $C_\gamma H_2$							w	w	w	w	w	w	w	w	obsd
Gly NH-Pro $C_\delta H_s$	w		w	w	w		w	w	w			w		w	obsd
Gly $C_\alpha H_s$-Pro $C_\alpha H$	w	w	w				w	w	w	w		w		w	?[c]
Gly $C_\alpha H_r$-Nleu $C_\alpha H_s$	w	w	w	w	w	w	w	w	w			w		w	obsd[c]
Gly $C_\alpha H_s$-Nleu $C_\alpha H_s$											w		w		
Gly $C_\alpha H_r$-Nleu $C_\beta H_s$											w		w		
Gly $C_\alpha H_\gamma$-Nleu $C_\beta H_r$											w		w		2 obsd[c]
Gly $C_\alpha H_s$-Nleu $C_\beta H_s$											w		w		2 ?[c]
Gly $C_\alpha H_s$-Nleu $C_\beta H_r$											w				
Gly $C_\alpha H_r$-Nleu $C_\delta H_{3,r}$							w								
Gly $C_\alpha H_r$-Nleu $C_\delta H_{3,s}$									w						2 obsd
Gly $C_\alpha H_s$-Nleu $C_\delta H_{3,r}$	w						w								2 ?[c]
Gly $C_\alpha H_s$-Nleu $C_\delta H_{3,s}$			w						w						
Pro $C_\beta H_r$-Nleu $C_\alpha H_r$	m	m	m	s	s	m	w	w	w	w		w		w	obsd[d]
Pro $C_\beta H_r$-Nleu $C_\alpha H_s$	w	w	w	m	m	m									obsd[d]
Pro $C_\beta H_s$-Nleu $C_\alpha H_r$	w		w	w	w	w									obsd[d]
Pro $C_\gamma H_2$-Nleu $C_\alpha H_r$	w	w	w	w	w	w	s	s	s	s	m	s	m	s	?[c]
Pro $C_\gamma N_2$-Nleu $C_\alpha H_s$				w	w	w	m	m	m	m	m	m	m	w	obsd[c]
Pro $C_\delta H_s$-Nleu $C_\alpha H_r$	s	s	s	s	s	s	m	m	m	s	m	m	m	s	?[c]
Pro $C_\delta H_s$-Nleu $C_\alpha H_s$	w	w	w	w	w	m	w	w	w	w	m	w	m	w	obsd
Pro $C_\delta H_r$-Nleu $C_\alpha H_r$	w	w	w	w	w	w	w	w	w	w	w	w	w	w	?[c]
Pro $C_\delta H_r$-Nleu $C_\alpha H_s$	w	w	w	w							w				obsd[d]
Pro $C_\delta H_s$-Nleu $C_\beta H_s$				m	m	m				m	w	w	m	w	
Pro $C_\delta H_s$-Nleu $C_\beta H_r$				m	m	m				m	s	m	m	m	2 obsd[d]
Pro $C_\delta H_r$-Nleu $C_\beta H_s$				w											2 ?[e]
Pro $C_\delta H_r$-Nleu $C_\beta H_r$					w	w					w		w	w	
Pro $C_\delta H_s$-Nleu $C_\gamma H$	m	m					w	m							obsd[d]
Pro $C_\delta H_r$-Nleu $C_\gamma H$		w						w							?[e]
Pro $C_\delta H_s$-Nleu $C_\delta H_{3,r}$	s	w	m				s		m		w			w	obsd[d]
Pro $C_\delta H_s$-Nleu $C_\delta H_{3,s}$	w	s	s					m	s				w		obsd[d]
Pro $C_\delta H_r$-Nleu $C_\delta H_{3,r}$	m		w				m		w					w	obsd[d]
Pro $C_\delta H_r$-Nleu $C_\delta H_{3,s}$		w	m			w		w	m						obsd[d]

[a] NOEs corresponding to distances equal to or smaller than 2.5 Å are classified as strong (s), NOEs corresponding to distances bigger than 2.5 Å and equal to or smaller than 3.5 Å are classified as medium (m), NOEs corresponding to distances bigger than 3.5 Å and equal to or smaller than 4.5 Å are classified as weak (w), NOEs corresponding to distances bigger than 4.5 Å are considered absent (blank space); pseudoatom corrections are applied for the Pro $C_\gamma H_2$, Nleu $C_\delta H_{3,r}$, and Nleu $C_\delta H_{3,s}$ protons. [b] The "obsd" symbol indicates an observed NOE, while the "?" symbol indicates an unresolved NOE cross peak. [c] Gly $C_\alpha H_l$ and Nleu $C_\alpha H_h$ resonances overlap. [d] See ref 29. [e] Overlap with the (cis, trans) set of resonances.

The assumption of a conformational equilibrium possibly involving each of the structures **u1**−**6** and **d1**−**8** led to the redifinition of the concept of interchain NOE: the same NOE cross peak can be anticipated to arise uniquely from interchain interactions on the basis of some triple-helical structures, while on the basis of other triple-helical structures it can be expected to arise from a combination of interchain and intrachain interactions or uniquely from intrachain interactions. For instance, the NOE between Pro $C_\beta H_r$ and Nleu $C_\gamma H$ is predicted to be interchain in the triple-helical conformation **u1**, but it is predicted to be intrachain in the triple-helical conformation **u4**. Thus, the definition of interchain NOE can be structure dependent, and therefore, in order to critically test the whole ensemble of triple-helical conformations, it is necessary to select the NOEs which can arise only from interchain interactions in every structure of the entire conformational ensemble. These NOEs are those which are expected to be interchain in at least one structural cluster and are either absent or interchain in all the other structural clusters of the ensemble. We define such NOEs as *ensemble interchain NOEs*. For the ensemble of the 14 triple-helical conformations **u1**−**6** and **d1**−**8**, 33 ensemble interchain NOEs were found and are listed in Table 3. The stereospecific assignments of the non degenerate methylene resonances will be published elsewhere.

Ensemble Interchain NOE Analysis. As shown in Table 3 and Figure 5 the expected ensemble interchain NOEs are consistent with the NOESY spectra giving further evidence that the assembled set of resonances refers to collagen-like triple-helical structures. It is interesting to notice that the unambiguously observed ensemble interchain NOEs listed in Table 3 can be sorted into three main classes: three NOEs between Gly and Pro, five NOEs between Gly and Nleu, and 13 NOEs between Pro and Nleu. This observation indicates that at each level of the triple helix Gly, Pro, and Nleu are closely packed (Figure 4a, left) and also supports the register shift which is typically found in collagen-like triple helices (Figure 4a, right).[3] Further evidence for the triple-helical register shift comes from considerations on the KTA symmetry (see below). Table 3 shows also that the NOEs Gly HN to Pro $C_\beta H_r$, Pro $C_\beta H_r$ to Nleu $C_\alpha H_s$, and Pro $C_\beta H_s$ to Nleu $C_\alpha H_r$ can arise uniquely from conformations with an up Pro puckering, while the NOE Gly HN to Pro $C_\gamma H$ and the four NOEs Gly $C_\alpha H_{r,s}$ to Nleu $C_\beta H_{r,s}$ can arise uniquely from conformations with a down Pro puckering. The simultaneous observation of these NOEs suggests that the Pro five-membered ring can adopt both up and down puckerings.

Table 3 indicates also that the four NOEs Gly $C_\alpha H_{r,s}$ to Nleu $C_\delta H_{3r,s}$ and the two NOEs Pro $C_\delta H_{r,s}$ to Nleu $C_\gamma H$ can arise uniquely from structures in which the Nleu χ_1 torsion is positive (*in* conformations of Table 3), while the four NOEs Gly $C_\alpha H_{r,s}$

(29) The four NOEs between Gly $C_\alpha H_{r,s}$ and Pro $C_\beta H_{r,s}$ are all predicted to be absent, and therefore, the overlap between the Gly $C_\alpha H_l$ and the Nleu $C_\alpha H_h$ resonances does not interfere with the measurement of NOESY cross peaks between the Nleu $C_\alpha H_h$ and the Pro $C_\beta H_{l,h}$ protons.

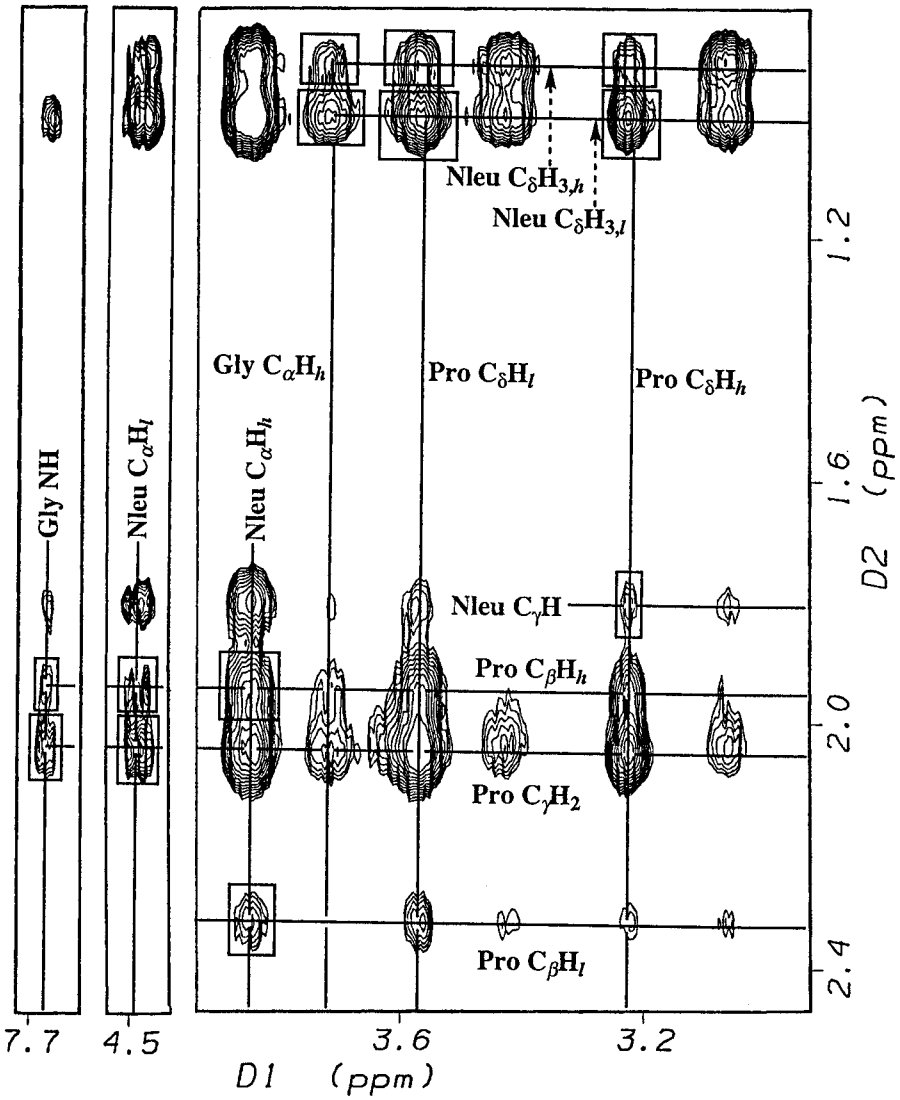

Figure 5. Expanded aliphatic and amide regions of the NOESY spectrum of KTA-[Gly-(Gly-Pro-Nleu)$_9$-NH_2]$_3$ (**VII**, Table 1) in D_2O at 27 °C, with a mixing time of 150 ms. Representative ensemble interchain NOE cross peaks are boxed, and the labels show their assignments (see text).

to Nleu $C_\beta H_{r,s}$ and the four NOEs Pro $C_\delta H_{r,s}$ to Nleu $C_\beta H_{r,s}$ can arise uniquely from structures in which the Nleu χ_1 torsion is negative (*out* conformations of Table 3). The simultaneous observation of these NOEs indicates therefore that the Nleu side chain can adopt both conformations in which it points toward the Pro ring (*in*) and conformations in which it points in the opposite direction (*out*, Figure 4b). It is difficult to obtain a reliable estimate of the relative populations of *in* and *out* conformations on the basis of NOE intensities, but previous studies on the comparison between Gly-Pro-Sar and Gly-Pro-Nleu sequences[4] indicate that the replacement of Sar with Nleu increases the stability of interchain interactions suggesting a high population of *in* structures in which the Nleu side chain hydrophobically interacts with the Pro ring.

Triple-Helix Register and KTA Symmetry. The screw symmetry of the triple-helix-coiled coils leads to a one-residue register shift along the triple-helix axis. The triple-helical register shift breaks the ternary rotational symmetry of the KTA template transfering to it the chirality of the triple helices. A positive proof of the loss of KTA symmetry as a result of triple-helix formation comes from the TOCSY spectrum of KTA-[Gly-(Gly-Pro-Nleu)$_6$-NH_2]$_3$ (**VI**, Table 1) at 17 °C. At this temperature the triple helix is not denatured, and three distinct cross peaks can be observed between the equatorial and the axial protons of the KTA methylenes (Table 2 and Supporting Information). However, when the triple helix is absent either because it has been denatured or because the chain is too short as for KTA-[Gly-Gly-Pro-Nleu-NH_2]$_3$ (**IV**, Table 1), then the KTA signal splitting disappears (Supporting Information) indicating that the ternary rotational symmetry of KTA is maintained. Similar results on the KTA splitting were obtained for KTA conjugates composed of Gly-Pro-Hyp repeats.[11] On the basis of these observations, we conclude that the KTA signal splitting can be considered as an additional support of triple helicity for the KTA terminated analogs.

The KTA Conformation. Further insight into the conformation of the KTA ring can be obtained from the differences in methylene chemical shifts, which are known to be very sensitive to the conformational preferences of KTA analogs.[2,30] For compounds **IV** and **V** which are not triple helical and do not show the KTA signal splitting, the difference in the methylene chemical shift is 1.33 ppm and the equatorial[31] methylene proton is found at 2.66 ppm, indicating that the KTA template adopts a chairlike conformation in which the three methyls are equatorial and the three carbonyls axial. For the

(30) Jackman, L. M.; Sternhell, S. *Applications of Nuclear Magnetic Resonance Spectroscopy in Organic Chemistry*, 2nd ed.; Pergamon: Oxford, 1969; pp 238−240.

(31) Morelle, N.; Gharbi-Benarons, J.; Acher, F.; Valle, G.; Crisma, M.; Toniolo, C.; Azerad, R.; Girault, J.-P. *J. Chem. Soc., Perkin Trans.* **1993**, *2*, 525−533.

other KTA conjugates **VI** and **VII**, which are triple helical and show the KTA signal splitting, the difference in methylene chemical shift is 1.24−1.57 ppm and the equatorial[31] methylene proton is found at 2.51−2.83 ppm, suggesting that the KTA template adopts a conformation similar to that proposed for compounds **IV** and **V**. These conclusions are fully consistent with previous investigations on the Kemp triacid which was assigned a chair conformation for its protonated, mono-, and dianionic forms with the three methyls equatorial.[2]

Conclusions

Molecular modeling, ^{1}H NMR, and CD provided new insight as to the incorporation of peptoid residues into collagen-like triple helices. It was shown that -(Gly-Pro-Nleu)$_n$- sequences can be assembled into collagen-like triple helices with a typical poliproline-II-like backbone structure. The -(Gly-Pro-Nleu)$_n$- triple-helical array can be further stabilized by favorable interactions between the Pro five-membered ring and the hydrophobic Nleu side chain. Even further stabilization can be obtained using a KTA-based template to assemble the three polypeptide chains. The structural details highlighted in this study offer a basis for the rational design of novel collagen-like peptidomimetics including peptoid residues different from Nleu. In addition, the concept of ensemble NOEs was introduced for the analysis of the multiconformational dynamics of triple helices, providing a useful tool for the study of other collagen-like triple-helical peptides containing residues with flexible side chains.

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

Supporting Information Available: CD spectra, molecular modeling, assignments, and TOCSY spectra for the KTA resonance splitting (7 pages). See any current masthead page for ordering and Internet access instructions.

JA9617193