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

Giuseppe Melacini, Yangbo Feng, and Murray Goodman*

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

Received December 4, 1996; Revised Manuscript Received April 14, 1997⊗

ABSTRACT: Molecular modeling and ^{1}H-NMR were employed to study the structure and stability of collagen-like triple helices composed of Gly-Nleu-Pro repeats. The compounds studied include the acetyl analogs Ac-(Gly-Nleu-Pro)$_n$-NH_2 (where n = 1, 3, 6, and 10) and the KTA conjugates KTA-[Gly-(Gly-Nleu-Pro)$_n$-NH_2]$_3$ (where n = 3 and 6 and KTA denotes the Kemp triacid). The presence of collagen-like assembled structures is supported by a consistent set of experimental observations, which include the appearance of a distinct set of resonances, low hydrogen-exchange rates for Gly NH, cooperative melting transition, and observation of several interchain NOEs. Using ^{1}H-NMR, the triple helicity was monitored as a function of chain length, template, and temperature. These studies show that (Gly-Nleu-Pro)$_n$ sequences have a somewhat higher triple-helical propensity than (Gly-Pro-Nleu)$_n$ sequences. In addition, our investigations have shown that unlike the triple helices composed of Gly-Pro-Nleu repeats those composed of Gly-Nleu-Pro repeats can access conformations in which the Nleu side chains are arrayed between Pro residues belonging to different triple-helix cross sections. These structural features may serve as a basis for free energy computations and for the study of higher-order structures such as collagen-like fibrils containing peptoid moities.

One of the major goals of collagen research is to understand how its conformation is controlled by its typical Gly-X-Y repeating sequence (where X and Y represent any amino or imino acid residues). In particular, the difference in triple-helix stability between Gly-Pro-X and Gly-X-Pro collagen-like sequences is one of the central issues of collagen structural biology. Studies on synthetic model compounds have suggested the rule that Gly-Pro-X triple helices are generally more stable than those composed of Gly-X-Pro repeats [see Feng (1996) and references cited therein]. However, as shown in the preceding paper (Feng, 1996), when X is the peptoid *N*-isobutylglycine (denoted as Nleu;[1] Simon et al., 1992), the rule is inverted. This observation suggests that the Nleu-containing collagen-like sequences reveal novel structural determinants for triple helicity which account for the unusual sequence dependence of the triple-helix stability. In order to investigate these novel structural determinants of triple helicity, it is essential to obtain detailed knowledge of the structural features of both Gly-Nleu-Pro and Gly-Pro-Nleu triple helices.

The structure of Gly-Pro-Nleu triple helices has been previously characterized using an integrated biophysical approach which included CD, optical rotation, NMR, and molecular modeling (Goodman et al., 1996a; Feng et al., 1996; Melacini et al., 1996a). It was shown that Gly-Pro-Nleu triple helices adopt a poly(proline-II)-like backbone and can access conformations in which the hydrophobic Nleu side chains are in the proximity of the Pro residues belonging to the same triple-helix cross sections. As for Gly-Nleu-Pro triple helices, the poly(proline-II)-like main chain conformation has been elucidated (Feng, 1996), but no direct information is available on the side chains because CD spectra reflect only the average alignment of the backbone dipoles. In this paper, NMR and molecular modeling are therefore used to investigate both main and side chain conformations of Gly-Nleu-Pro triple helices, resulting in the first model for Gly-Nleu-Pro triple helices. Specifically, the (Gly-Nleu-Pro)$_n$ sequence was analyzed using compounds 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) (Kemp & Petrakis, 1981) (Table 1).

MATERIALS AND METHODS

Sample Preparation. All compounds were synthesized as described elsewhere [see the Supporting Information of Feng et al. (1997)]. The NMR samples were prepared in H_2O/D_2O (9:1) and in D_2O (purchased from Isotec, Inc.) solutions with a peptide concentration in the range of 1.3−3.8 mg/mL. After dissolution in water, all samples were kept at 5 °C for at least 4 months prior to any measurement of the triple-helical percentage, in order to allow for the equilibrium of the sample (Long et al., 1993). The pH was adjusted to 2.8 (direct pH meter reading without correction for isotope effects) for all samples.

NMR Spectroscopy. All NMR experiments were carried out on an AMX-500 Bruker spectrometer using the same

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

⊗ Abstract published in *Advance ACS Abstracts,* June 1, 1997.

[1] Abbreviations: Ac, acetyl; CD, circular dichroism; CFF91, one of the force fields used by the simulation program Discover; DG, distance geometry; DQF-COSY, double-quantum-filtered correlation spectroscopy; Hyp, 4-*trans*-hydroxyproline; KTA, *cis*,*cis*-1,3,5-trimethylcyclohexane-1,3,5-tricarboxylic acid; Nleu, *N*-isobutylglycine; NOE, nuclear Overhauser enhancement; NOESY, two-dimensional nuclear Overhauser effect spectroscopy; NMR, nuclear magnetic resonance; rmsd, root mean square deviation; ROESY, rotating frame nuclear Overhauser spectroscopy; Sar, *N*-methylglycine; TOCSY, total correlation spectroscopy.

S0006-2960(96)02981-9 CCC: $14.00

Table 1: Compounds Studied and Their Triple Helicity

compound	no.	*n*	average no. of assembled Nleu residues per chain[a]
Ac-(Gly-Nleu-Pro)$_n$-NH$_2$[b]	**I**	1	–
	II	3	–
	III	6	1.2
	IV	10	5.9
KTA-[Gly-(Gly-Nleu-Pro)$_n$-NH$_2$]$_3$[c]	**V**	3	–
	VI	6	1.4

[a] Computed on the basis of the one-dimensional ^{1}H-NMR integrals of the assembled Nleu C$_\beta$H resonances measured in D$_2$O at 27 °C and pH 2.8. See Materials and Methods for the details of the normalization. [b] 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. [c] KTA denotes the Kemp triacid.

data acquisition and processing scheme (Redfield & Kuntz, 1975; Aue et al., 1976; Jeener et al., 1979; Kumar et al., 1980; Bax & Freeman, 1981; Braunsweiler & Ernst, 1983; Marion & Wuthrich, 1983; Shaka & Freeman, 1983; Bothner-By et al., 1984; Rance et al., 1984; Wider et al., 1984; Bax & Davis, 1985; Otting et al., 1986) previously used for the study of triple helices composed of (Gly-Pro-Hyp)$_n$ and of (Gly-Pro-Nleu)$_n$ sequences (Melacini et al., 1996a,b). For the hydrogen-exchange measurements, one-dimensional spectra were collected at 5 °C over a period of 1 month. A heating rate of 1 °C per 45 min was used for the melting of Ac-(Gly-Nleu-Pro)$_6$-NH$_2$ (**III**), and the temperature was increased in increments of 1 °C. The NOESY spectra were obtained using mixing times of 50, 75, 100, and 150 ms. ROESY experiments were carried out with a mixing time of 200 ms and a spin locking field of 2.5 kHz. The integrals were normalized using the peaks in the spectral region from 2.7 to 3.3 ppm which contains a known number of Nleu C$_\beta$H$_{r,s}$ protons.

Molecular Modeling. The Gly-Nleu-Pro triple helices were modeled following the protocol previously developed for Gly-Pro-Nleu triple helices (Biosyn Technologies, Inc.; Melacini et al., 1996a,b; Hagler, 1985; McCammon et al., 1979; Miller & Scheraga, 1976; Miller et al., 1980a,b; Nemethy & Scheraga, 1982; Nemethy et al., 1980, 1981, 1992; Zagari et al., 1990; see the Supporting Information). The model molecule Ac-(Gly-Nleu-Pro)$_4$-NH$_2$ was used, and the three poly(proline-II) backbones were assembled in a triple-helical array using Ac-(Gly-Sar-Pro)$_4$-NH$_2$ as input for distance geometry (DG) runs (Quantum Chemistry Program Exchange No. 590, Indiana University Department of Chemistry; Crippen, 1981; Havel & Kuntz, 1983; Crippen & Havel, 1988). The DG runs resulted in 100 triple-helical structures which could be clustered in two conformational families. One family, composed of 32 structures, showed the correct triple-helical hydrogen bond pattern. The average backbone pairwise root mean square deviation (rmsd) between the ensemble of the 32 distance geometry triple-helical structures with the correct hydrogen bonds and their minimized average structure was computed to be 0.77 ± 0.06 Å. The average minimized structure was therefore used to build triple-helical Ac-(Gly-Nleu-Pro)$_4$-NH$_2$ conformations by addition of isopropyl groups to the Sar methyl groups. The Nleu side chain torsional angles χ_1 and χ_2 were defined according to Melacini et al. (1996a).

RESULTS AND DISCUSSION

Assignments. The analysis of DQF-COSY, TOCSY, and ROESY spectra of compounds **I**, **II**, and **V** (Table 1) at 27 °C reveals four sets of resonances for each residue. As shown in Figure 1, these sets correspond to different combinations of peptide bond structures (*cis*/*trans*) (see also the Supporting Information). An additional set of resonances is detected for compounds **III**, **IV**, and **VI** (Table 1) at 27 °C (Figure 2). This new set of resonances arises from a distinct tripeptide unit as shown by the sequential connectivities from Gly to Nleu, from Nleu to Pro, and from Pro to Gly which are consistent with the NOESY spectra (see circled NOESY cross-peaks in Figure 3A).

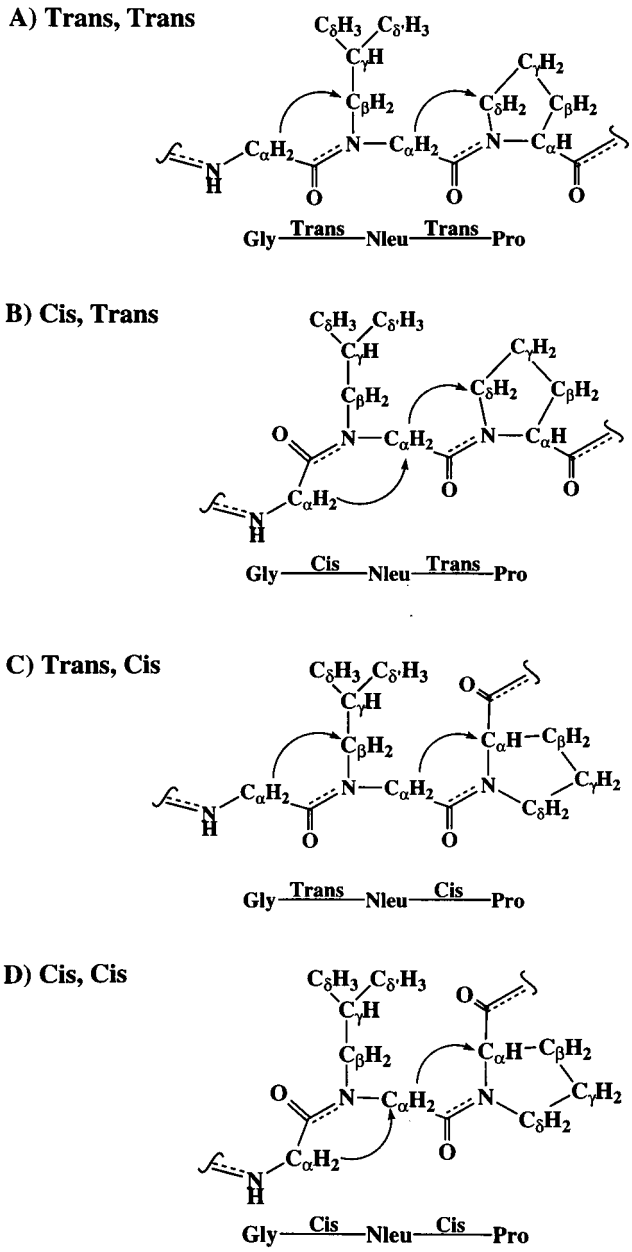

FIGURE 1: (A−D) Schematic diagram of the structures corresponding to the four unassembled sets of resonances observed for sequences composed of Gly-Nleu-Pro repeats. The arrows indicate the critical NOEs which can be used to assign the peptide bond structure (*cis*/*trans*) in each resonance set. In compound **I** (Table 1), the *trans*,*trans* and the *cis*,*trans* sets correspond to approximately 60 and 30% of the total intensity, respectively, while the *trans*,*cis* and *cis*,*cis* sets account together for about 10% of the total intensity.

The new set of resonances observed for compounds **III**, **IV**, and **VI** is assigned to residues in a collagen-like triple helix as indicated by several independent experimental observations. (1) For analog **IV** in D$_2$O (5 °C, pH 2.8), the Gly NH of this set of resonances exhibits a low hydrogen-

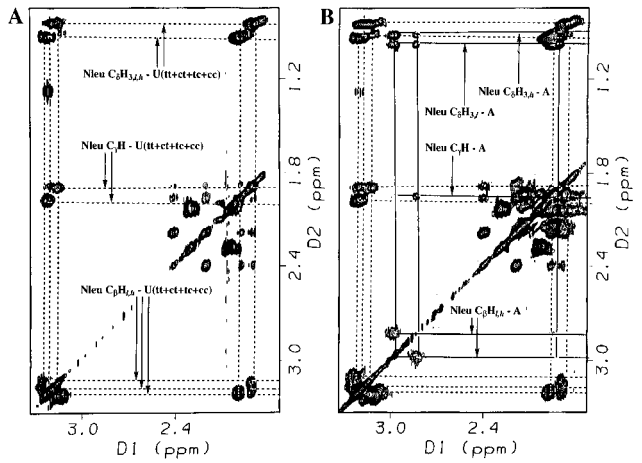

FIGURE 2: (A) Expanded region of the TOCSY (τ_{mix} = 39 ms) spectrum of Ac-(Gly-Nleu-Pro)$_3$-NH_2 (**II**) in D_2O (27 °C, pH 2.8) showing the Nleu side chain connectivities. Only non-triple-helical, unassembled (U) resonances are observed, including *trans*,*trans*, *cis*,*trans*, *trans*,*cis*, and *cis*,*cis* structures labeled as tt, ct, tc, and cc, respectively (see also Table 2). (B) Expanded region of the TOCSY (τ_{mix} = 39 ms) spectrum of Ac-(Gly-Nleu-Pro)$_{10}$-NH_2 (**IV**) measured in H_2O (27 °C, pH 2.8). An additional set of resonances (solid lines) is observed and assigned to assembled (A) triple-helical structures.

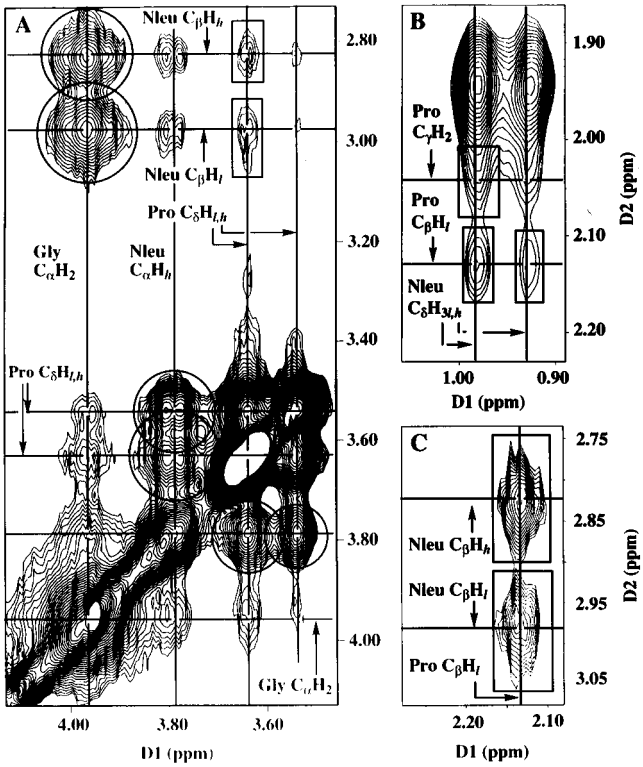

FIGURE 3: (A) Expanded region of the NOESY spectrum of Ac-(Gly-Nleu-Pro)$_{10}$-NH_2 (**IV**, Table 1) in D_2O at 27 °C with a mixing time of 150 ms. The circles indicate the sequential NOEs from Gly to Nleu and from Nleu to Pro. The rectangles indicate representative ensemble interchain NOEs between Pro and Nleu. The labels and arrows show the cross-peak assignments for the assembled set of resonances (see the text). (B and C) Other expanded regions of the same NOESY spectrum shown in panel A indicating additional ensemble interchain NOEs between the Pro and Nleu side chains of Gly-Nleu-Pro triple helices. The assignments for the assembled set of resonances are indicated by the labels and arrows in the figure.

exchange rate ($k \leq 10^{-5}$ min^{-1}) consistent with the triple-helix hydrogen bond network (Miller et al., 1980a). (2) The relative intensity of the new set of resonances decreases as the temperature increases. (3) The new resonances originate from conformations which exchange slowly on the NMR time scale. This is typically seen for collagen-like triple helices of other model peptides (Brodsky et al., 1992; Li et al., 1993; Goodman et al., 1996a,b). (4) Several interchain NOEs are observed in the NOESY spectra between the new resonances (Figure 3A−C), as explained below in the section concerning the conformational characterizations. We therefore denote the set of resonances detected only for analogs **III**, **IV**, and **VI** as the assembled set. Accordingly, the unassembled sets are those observed for compounds **I**, **II**, and **V**, which lack the structural requirements for forming a triple helix under our experimental conditions and therefore are assumed to adopt less ordered conformations.

The assignments for the assembled and unassembled sets of resonances (Table 2) show that the resonances of the assembled Nleu $C_\beta H_{l,h}$ (the subscript *l* denotes the low-field resonance, while the subscript *h* denotes the high-field resonance) at 2.98 and 2.83 ppm, respectively, are well-resolved and not significantly overlapped by any resonance belonging to the unassembled sets. Conversely, among the resonances of the unassembled set which are not significantly overlapped by any resonance belonging to the assembled set, the resonances of the unassembled Nleu $C_\beta H_{l,h}$ resonances in the range of 3.14−3.22 ppm are well-resolved. The Nleu $C_\beta H_{l,h}$ resonances can therefore be used for the identification and quantification of triple-helical conformations. We employed these resonances to study the triple helicity as a function of chain length, template, and temperature.

Triple Helicity as a Function of Chain Length and Template. Table 1 shows the measured average number of assembled Nleu residues per chain for the compounds studied at 27 °C. It can be seen that for both the acetyl-terminated analogs and the KTA conjugates six Gly-Nleu-Pro repeats are sufficient to observe the triple-helical Nleu $C_\beta H_{l,h}$ resonances. These observations show that (Gly-Nleu-Pro)$_n$ sequences have a higher triple-helical propensity than (Gly-Pro-Nleu)$_n$ sequences for which, under similar experimental conditions, six acetyl-terminated repeats are not sufficient to observe triple-helical conformations (Melacini et al., 1996a).

Triple Helicity as a Function of Temperature. The normalized one-dimensional integral of the assembled Nleu $C_\beta H_{l,h}$ resonances at 2.98 and 2.83 ppm was used to monitor the melting of the triple-helical conformation of compound **III**. The transition from assembled to unassembled structures of compound **III** is cooperative and is centered at 26 °C. A melting transition is also observed for the KTA conjugate containing six Gly-Nleu-Pro repeats per chain (**VI**) for which the melting temperature measured on the basis of ^{1}H-NMR integrals is 36 °C. These results confirm previous studies on KTA conjugates composed of Gly-Pro-Hyp and Gly-Pro-Nleu sequences which have shown that the KTA template allows a net free energy gain and results in increased melting temperatures. No reliable data could be collected for the melting of the longer chain compound **IV** because of the presence of aggregation phenomena at high temperatures.

Sequence Effects on Triple Helicity. There is some difference in the extent of triple helicities for molecules containing the (Gly-Nleu-Pro)$_n$ sequence and those containing (Gly-Pro-Nleu)$_n$ and (Gly-Pro-Hyp)$_n$ sequences. Both (Gly-Pro-Nleu)$_n$ and (Gly-Pro-Hyp)$_n$ contain a chiral residue immediately following the glycines, while (Gly-Nleu-Pro)$_n$

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

		chemical shift (ppm)[a]					
resonance set	residue	NH	$C_\alpha H$	$C_\beta H$	$C_\gamma H$	$C_\delta H$	other
assembled[b]	Gly	7.53	3.96, 3.96				
	Nleu		4.61, 3.79	2.98, 2.83	1.95	0.97, 0.94	
	Pro		4.77	2.12, 1.93	2.04, 2.04	3.64, 3.54	
unassembled	Gly	8.27−8.32[d]	4.19, 4.19				
trans,*trans*[c]	Nleu		4.32,[e] 4.19	3.21, 3.21	1.97	0.94, 0.94	
	Pro		4.48[f]	2.27, 2.04[g]	2.04, 2.04[h]	3.67, 3.62	
	Ac						2.06
	KTA						H_{ax} 1.33
							H_{eq} 2.66
							Me 1.27
unassembled	Gly	8.27−8.31[j]	4.01,[k] 3.91				
cis,*trans*[b,i]	Nleu		4.36, 4.36	3.22, 3.14[l]	1.89	0.86, 0.86	
	Pro		4.48[f]	2.27, 2.04[g]	2.04, 2.04[h]	3.66, 3.61	
unassembled	Gly	8.23, 8.14[n]	4.05−4.11, 4.05−4.11[n]				
trans,*cis*	Nleu		4.39−4.31, 3.84[o]	—[p]	—[p]	—[p]	
and *cis*,*cis*[b,m]	Pro		4.59−4.63[q]	2.40, 2.19[r]	1.96, 1.89−1.86[s]	3.63−3.60, 3.55	

[a] ^{1}H chemical shifts are reported relative to sodium 3-(trimethylsilyl)tetradeuteriopropionate (STP). [b] In this set of resonances, the Ac and KTA assignments are not applicable. [c] This set of resonances corresponds to structures with *all-trans* peptide bonds. [d] A minor resonance can be observed also at 8.12 ppm as the result of differences between residues in the core and residues at the ends of the polypeptide chain (end effects). [e] In compound **I**, this resonance appears at 4.27 ppm. [f] A minor resonance can be observed also at 4.38 ppm as the result of the end effects. [g] Minor resonances can be observed also at 2.24 and 1.94 ppm as the result of the end effects. [h] A minor resonance can be observed also at 2.01 ppm as the result of the end effects. [i] This resonance set corresponds to structures with a *cis* Gly-Nleu peptide bond and a trans Nleu-Pro peptide bond.[j] A minor resonance can be observed also at 8.09 ppm as the result of the end effects. [k] A minor resonance can be observed also at 3.96 ppm as the result of the end effects. [l] In compound **I**, these resonances appear at 3.15 ppm. [m] These resonance sets correspond to structures with a *trans* Gly-Nleu peptide bond and a *cis* Nleu-Pro peptide bond and to structures with these peptide bonds both *cis*. [n] Data observed for compound **I**. [o] These resonances appear at 4.33−4.27 and 3.97−3.77 ppm in compound **I**. [p] The intensity of this set of resonances is too low to allow the observation of the NOEs necessary to assign the Nleu side chain protons. However, the TOCSY spectra indicate that the Nleu side chain resonances of this set are in the same region as those of the *trans*,*trans* and/or the *cis*,*trans* sets. [q] A minor resonance can be observed also at 4.52 ppm as the result of the end effects. [r] In compound **I**, these resonances appear at 2.35 and 2.13 ppm. [s] In compound **I**, this resonance appears at 1.85 ppm.

Table 3: NOEs Anticipated To Arise Uniquely from Interchain Interactions on the Basis of the Molecular Models of Triple Helices Composed of Gly-Nleu-Pro Repeats[a]

	Pro down clusters				Pro up clusters					
expected NOEs	**d1**	**d2**	**d3**	**d4**	**u1**	**u2**	**u3**	**u4**	**u5**	observed NOEs[b]
Gly NH−Nleu $C_\gamma H$	−	−	−	−	−	w	−	−	−	?[c]
Nleu $C_\alpha H_s$−Pro $C_\alpha H$	w	w	w	w	−	w	w	w	−	?[d]
Nleu $C_\alpha H_s$−Pro $C_\beta H_s$	−	w	w	w	−	−	−	−	−	?[e]
Nleu $C_\beta H_s$−Pro $C_\alpha H$	−	m	m	m	−	m	m	s	m	X,[f] ?[g]
Nleu $C_\beta H_r$−Pro $C_\alpha H$	−	m	m	m	−	m	m	m	−	
Nleu $C_\beta H_s$−Pro $C_\beta H_r$	m	−	w	w	m	−	−	−	−	?[c]
Nleu $C_\beta H_r$−Pro $C_\beta H_r$	w	w	w	w	w	−	−	−	w	?[c]
Nleu $C_\beta H_s$−Pro $C_\beta H_s$	w	m	m	m	m	w	m	w	−	X
Nleu $C_\beta H_r$−Pro $C_\beta H_s$	−	w	m	w	w	−	−	−	m	X
Nleu $C_\beta H_s$−Pro $C_\gamma H_2$	w	w	w	w	w	m	m	m	w	X, ?[e]
Nleu $C_\beta H_r$−Pro $C_\gamma H_2$	w	w	w	w	w	w	w	w	w	
Nleu $C_\beta H_s$−Pro $C_\delta H_r$	−	w	w	w	−	w	w	w	w	X
Nleu $C_\beta H_r$−Pro $C_\delta H_r$	−	w	w	w	−	w	w	w	−	X
Nleu $C_\gamma H$−Pro $C_\alpha H$	m	−	w	−	m	−	−	−	−	?[c]
Nleu $C_\gamma H$−Pro $C_\beta H_r$	w	m	w	−	−	m	w	−	−	?[c]
Nleu $C_\gamma H$−Pro $C_\beta H_s$	w	w	w	−	−	m	−	−	−	?[c]
Nleu $C_\gamma H$−Pro $C_\gamma H_2$	w	w	w	−	w	w	w	w	−	?[c]
Nleu $C_\gamma H$−Pro $C_\delta H_r$	m	−	w	−	w	−	w	−	−	?[c]
Nleu $C_\delta H_{3,r}$−Pro $C_\alpha H$	s	w	w	w	s	w	−	w	w	X[f]
Nleu $C_\delta H_{3,s}$−Pro $C_\alpha H$	w	−	w	w	w	−	w	w	s	X[f]
Nleu $C_\delta H_{3,r}$−Pro $C_\beta H_r$	w	m	−	s	w	m	−	s	w	?[c]
Nleu $C_\delta H_{3,r}$−Pro $C_\beta H_s$	s	w	w	m	m	m	−	s	w	X
Nleu $C_\delta H_{3,s}$−Pro $C_\beta H_s$	w	w	m	w	w	w	s	w	m	X
Nleu $C_\delta H_{3,r}$−Pro $C_\gamma H_2$	w	m	−	m	s	m	−	m	w	X, ?[h]
Nleu $C_\delta H_{3,s}$−Pro $C_\gamma H_2$	m	−	m	w	w	−	s	m	s	
Pro $C_\gamma H_2$−Gly $C_\alpha H_2$	m	−	−	−	−	−	−	−	−	?[i]
Pro $C_\delta H_r$−Gly $C_\alpha H_2$	−	−	w	w	−	w	w	w	−	?[i]

[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 (−). Pseudoatom corrections are applied for the Gly $C_\alpha H_2$, Pro $C_\gamma H_2$, Nleu $C_\delta H_{3,r}$, and Nleu $C_\delta H_{3,s}$ protons. [b] The X indicates an observed NOE, while the ? symbol indicates an unresolved NOE cross-peak. [c] The Nleu $C_\gamma H$ and Pro $C_\beta H_h$ resonances overlap. [d] Overlap with the *trans*,*cis* and *cis*,*cis* set of resonances. [e] Too weak to be unambiguously detected. [f] NOE measured at 5 °C to avoid the overlap of the Pro $C_\alpha H$ resonance and the residual HDO signal. [g] At 5 °C, the Nleu $C_\beta H_{1,h}$ resonances partially overlap. [h] Overlap with the *trans*,*trans* set of resonances. [i] Overlap with the *cis*,*trans* set of resonances.

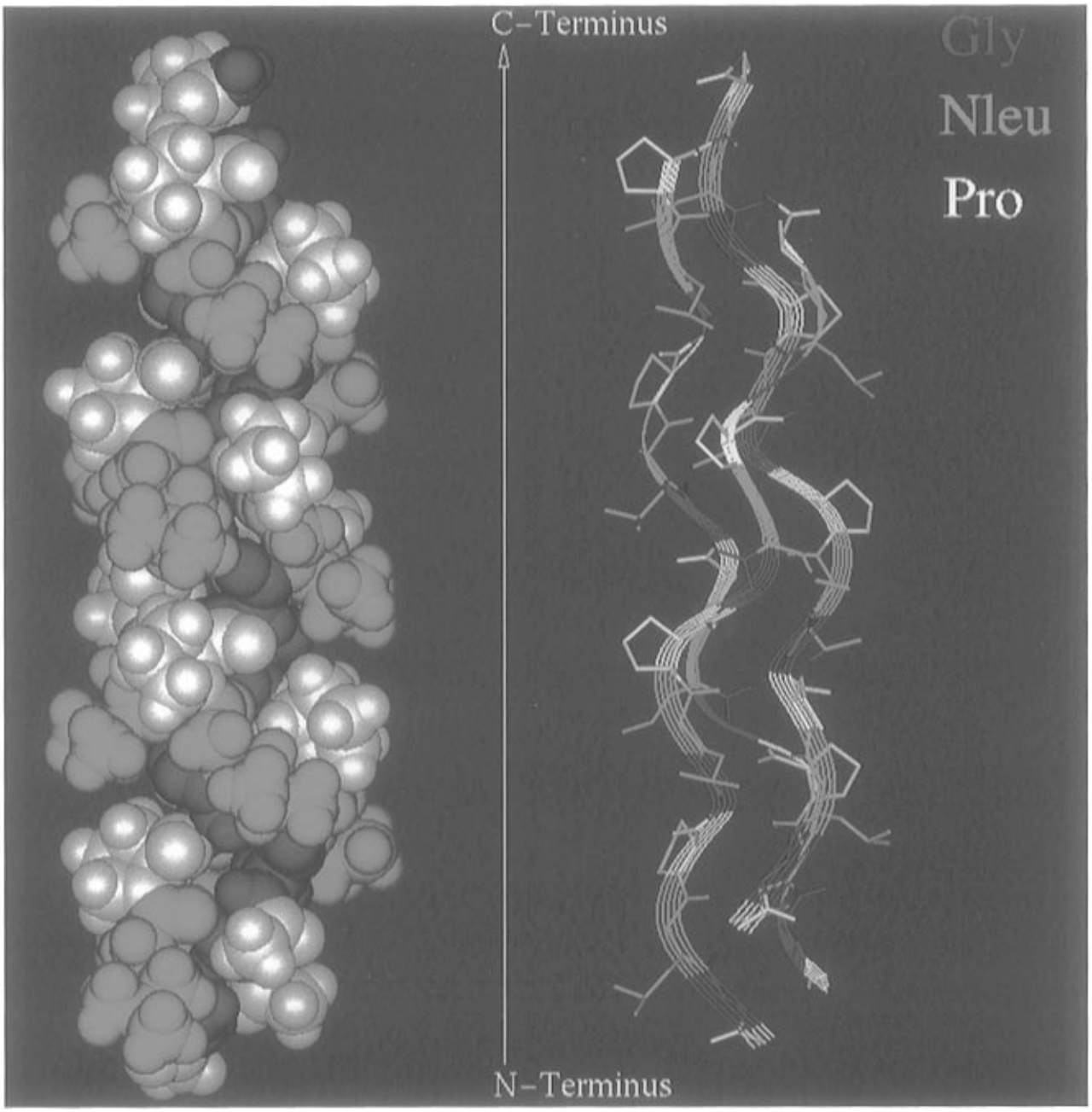

FIGURE 4: Lowest-energy triple-helical conformation of Ac-(Gly-Nleu-Pro)$_4$-NH_2 (cluster **u2** of Table 1 of the Supporting Information) represented as a CPK model (left) and as a heavy atom stick model (right). For the stick model, the backbone ribbon diagram is shown for each of the three chains, which are staggered by one residue. The Gly, Nleu, and Pro residues are color coded with red, green, and white, respectively.

does not. This difference in sequence can affect the extent of triple helicity.

Comparison between Chirooptical and NMR Results. Both the minimum chain length required for triple helicity and the melting temperatures determined by NMR are in full agreement with the results independently obtained by circular dichroism (CD) and optical rotation measurements (Feng et al., 1997). In addition, the magnitude of the melting transition monitored by optical rotation can be used as an indicator of the relative triple-helix percentage for compounds having the same sequence and chain length (Feng et al., 1997). For instance, the transition magnitudes observed in water solution for Ac-(Gly-Nleu-Pro)$_6$-NH_2 [2.0 mg/mL, Figure 1D of Feng et al. (1997)] and KTA-[Gly-(Gly-Nleu-Pro)$_6$-NH_2]$_3$ [Figure 4B, curve a, of Feng et al. (1997)] indicate that the triple-helix percentage of Ac-(Gly-Nleu-Pro)$_6$-NH_2 is approximately 80% of that of KTA-[Gly-(Gly-Nleu-Pro)$_6$-NH_2]$_3$. This result is in full agreement with the relative triple-helix estimate obtained by NMR for these compounds of 86% (Table 1). Furthermore, the triple-helix percentages obtained by NMR are absolute values and can be used for comparisons between compounds with different sequences as discussed above and/or different chain lengths [i.e. Ac-(Gly-Nleu-Pro)$_{10}$-NH_2 and Ac-(Gly-Nleu-Pro)$_6$-NH_2 in Table 1]. Finally, the poly(proline-II)-like backbone which accounts for the CD spectra is also supported by NMR and molecular modeling studies, as explained below.

Conformational Characterizations. The assignment of the assembled set of resonances to triple-helical structures is also supported by the NOESY spectra which were analyzed according to a procedure previously used for the NMR study of other collagen-like sequences (Li et al., 1993; Melacini et al., 1996a,b). This approach relies on the distiction between intra- and interchain NOEs on the basis of triple-

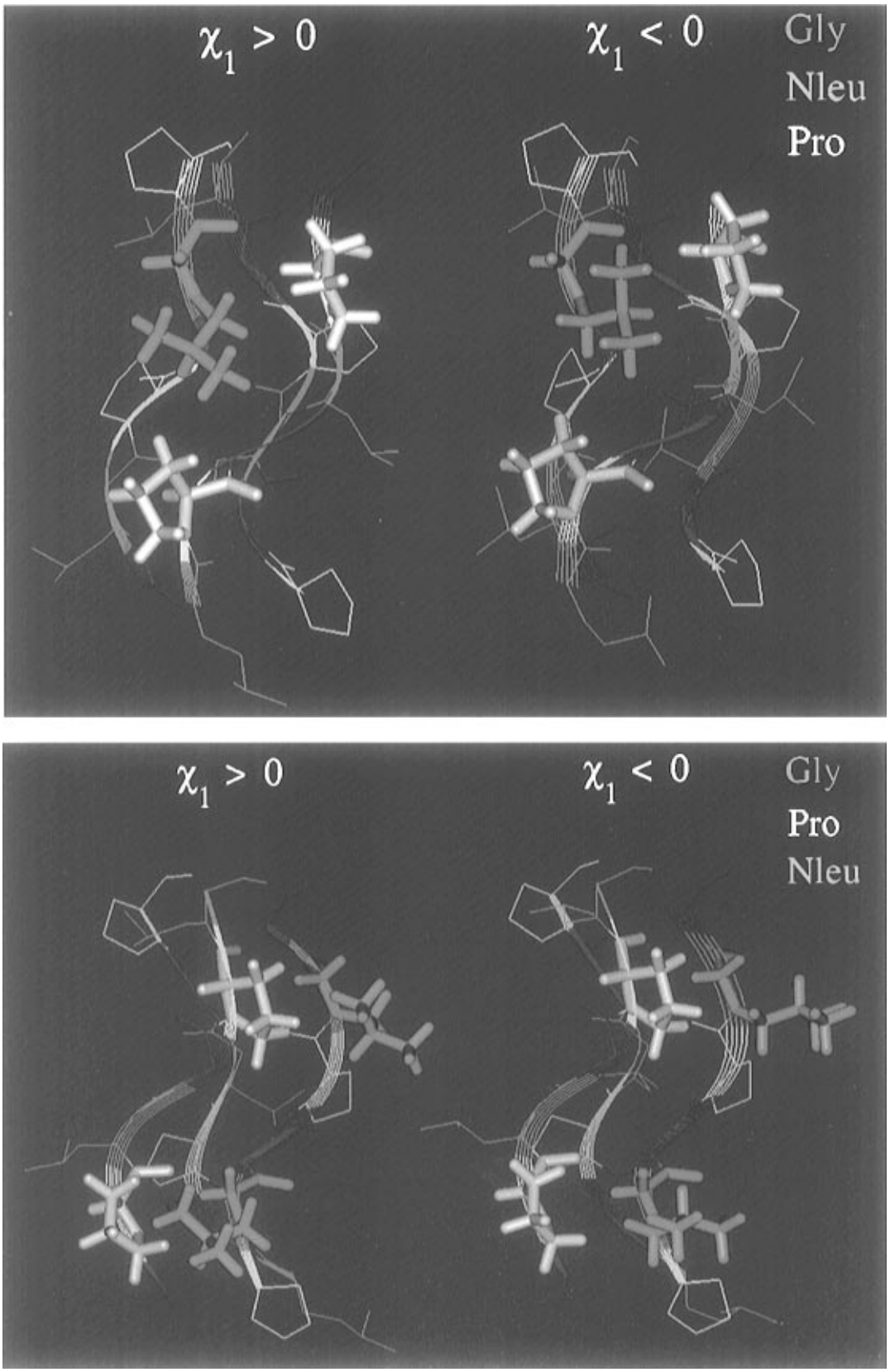

FIGURE 5: View of the enlarged central region of representative triple-helical conformations with different Nleu side chain orientations. The backbone ribbon is shown, and the Gly, Nleu, and Pro residues are color coded as in Figure 4. (A, top) Ac-(Gly-Nleu-Pro)$_4$-NH$_2$ (clusters **u1** and **u2** of Table 1 of the Supporting Information; note that cluster **u1** with a χ_1 of $><0$ has a higher energy than cluster **u2** with a χ_1 of $><0$). The Pro and Nleu residues of a central cross section are highlighted, showing that the side chain of each Nleu residue of a given chain is close to two Pro residues of another chain. (B, bottom) Ac-(Gly-Pro-Nleu)$_4$-NH$_2$ [see Melacini et al. (1996a)]. The highlighted Pro and Nleu residues show that each Nleu side chain of a given chain is close to one Pro residues of another chain.

helical modeled structures. In order to distinguish reliably between intra- and interchain NOEs and in order to avoid conformational averaging artifacts, it is important to search systematically for all the accessible triple-helical conformations. In particular, our conformational search for triple-helical structures composed of Gly-Nleu-Pro repeats resulted in nine minimum energy conformational families (see the Supporting Information). All nine conformations adopt a

poly(proline-II)-like backbone, while the major contribution to conformational heterogeneity comes from the Pro and Nleu side chains; five conformational families are characterized by the up Pro puckering (**u1**−**u5**) and the remaining four conformational families adopt the down Pro puckering (**d1**−**d4**). In clusters **u1** and **d1**, the Nleu χ_1 torsion, which describes the relative orientation of the Nleu side chain and the chain backbone, is negative, indicating that the $C_\beta-C_\gamma$ bond of the Nleu side chain points toward the Pro ring belonging to the same cross section of the triple helix. In the remaining clusters, Nleu χ_1 is positive, indicating that the $C_\beta-C_\gamma$ bond of the Nleu side chain points in the opposite direction (Figures 4 and 5A).

As was observed for Gly-Pro-Nleu triple helices, neither the differences in the average relative energies (see the Supporting Information) nor the absent NOEs (Bruschweiler et al., 1991; Blackledge et al., 1993) allowed us to reliably rule out the possibility of conformational equilibria involving two or more of the nine conformational clusters. The conservative approach of assuming that in solution an equilibrium conformational ensemble is present in which each of the nine minimum energy triple-helical clusters may have a non-zero population was therefore taken. On the basis of this equilibrium conformational ensemble of triple helices composed of the Gly-Nleu-Pro repeat, 27 NOEs are anticipated to arise uniquely from interchain interactions [ensemble interchain NOEs; see the definition of Melacini et al. (1996a)]. These ensemble interchain NOEs are listed in Table 3 (the stereospecific assignments of the nondegenerate methylene resonances of the assembled residues are being published elsewhere) (Melacini et al., 1997). Table 3 shows that the following general features clearly emerge from the conformational complexity discussed above.

(a) As shown in Table 3 and Figure 3, the expected ensemble interchain NOEs are consistent with the NOESY spectra, providing a further critical test for the triple-helical array composed of three poly(proline-II)-like backbones.

(b) Conformations with a positive Nleu χ_1 side chain torsion (Figures 4 and 5A) are unambiguously demonstrated by the four NOEs Nleu $C_\beta H_{s,r}$−Pro $C_\alpha H$ and Nleu $C_\beta H_{s,r}$−Pro $C_\delta H_r$ which cannot arise from any other type of conformations (Table 3). This result is fully consistent with our molecular modeling which shows that the triple-helical structures with positive Nleu χ_1 values always have lower energies than the corresponding conformations with negative Nleu χ_1 values (see the Supporting Information and Figure 5A).

Structural Comparison between (Gly-Nleu-Pro)$_n$ with (Gly-Pro-Nleu)$_n$ Triple Helices. It has been shown that both (Gly-Nleu-Pro)$_n$ and (Gly-Pro-Nleu)$_n$ sequences can be assembled into collagen-like triple helices with poly(proline-II)-like backbone structures (Goodman et al., 1996a; Feng et al., 1996; Melacini et al., 1996a). In both of these sequences, the triple helices are stabilized by interchain interactions between hydrophobic residues as suggested by the observed denaturation upon the addition of sodium dodecyl sulfate (SDS) to the water solution (Feng et al., 1996). The presence of interchain interactions between the hydrophobic Nleu and Pro residues is also consistent with the dramatically increased stability of (Gly-Nleu-Pro)$_n$ and (Gly-Pro-Nleu)$_n$ triple helices as compared to the stabilities of those composed of (Gly-Sar-Pro)$_n$ and (Gly-Pro-Sar)$_n$ sequences (Ananthanaratanan et al., 1976; Feng et al., 1996). These results are supported by the molecular models proposed for (Gly-Nleu-Pro)$_n$ and (Gly-Pro-Nleu)$_n$ triple helices (Figure 5A,B) which show the vicinity between Nleu and Pro residues belonging to different chains.

In spite of the similarities between (Gly-Nleu-Pro)$_n$ and (Gly-Pro-Nleu)$_n$ sequences mentioned above, the modeled triple-helical structures (Figure 5A,B) show that the side chains of these two sequences adopt different arrays. Figure 5A shows two representative Nleu side chain orientations accessible in (Gly-Nleu-Pro)$_n$ triple helices, indicating that each Nleu side chain is arrayed between two Pro residues. One Pro residue belongs to the same triple-helix cross section as Nleu, while the other Pro residue belongs to the same polypeptide chain as the first Pro residue and precedes it by two residues. On the contrary, in (Gly-Pro-Nleu)$_n$ triple helices (Figure 5B), the accessible conformations allow the Nleu side chain to be in the proximity of only the Pro residue belonging to the same triple-helix cross section. These differences between the side chain arrays of Gly-Nleu-Pro and Gly-Pro-Nleu triple helices offer a basis for computing the relative free energy of triple-helix formation in Gly-Pro-Nleu and Gly-Nleu-Pro sequences. However, a detailed evaluation of triple-helix stability is a complex task because it requires detailed knowledge of the conformational states not only of the triple-helical structures but also of the flexible single chains. In addition, the problem is further complicated by the difficulties in evaluating the conformational entropies of each system.

CONCLUSIONS

On the basis of computer simulations and ^{1}H-NMR, a molecular model is proposed for Gly-Nleu-Pro triple helices which shows significant differences in the array of the Nleu and Pro side chains as compared to that of the previously characterized Gly-Pro-Nleu triple helices (Melacini et al., 1996a). The structural differences outlined represent the first step toward the rational explanation of the sequence control of triple helicity in these peptoid-containing collagen-like compounds. In addition, the distinct conformational features of (Gly-Nleu-Pro)$_n$ and (Gly-Pro-Nleu)$_n$ triple helices will be useful in the study of higher-order structures such as collagen fibrils incorporating peptoid residues.

SUPPORTING INFORMATION AVAILABLE

Assignments of the unassembled resonance sets, molecular modeling-grid search strategy, and torsions of triple-helical structures composed of Gly-Nleu-Pro repeats (4 pages). Ordering information is given on any current masthead page.

REFERENCES

Ananthanarayanan, V. S., Brahmachari, S. K., Rapaka, R. S., & Bhatnagar, R. S. (1976) *Biopolymers 15*, 707.
Aue, W. P., Bartholdi, E., & Ernst, R. R. (1976) *J. Chem. Phys. 64*, 2229.
Bax, A., & Freeman, R. (1981) *J. Magn. Reson. 44*, 542.
Bax, A., & Davis, D. (1985) *J. Magn. Reson. 65*, 355.
Blackledge, M. J., Bruschweiler, R., Griesinger, C., Schmidt, J. M., Xu, P., & Ernst, R. R. (1993) *Biochemistry 41*, 10960.
Bothner-By, A. A., Stephens, R. L., Lee, J., Warren, C. D., & Jeanloz, R. W. (1984) *J. Am. Chem. Soc. 106*, 811.
Braunsweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521.
Brodsky, B., Li, M. H., Long, G. G., Apigo, J., & Baum, J. (1992) *Biopolymers 32*, 447.
Bruschweiler, R., Blackledge, M. J., & Ernst, R. R. (1991) *J. Biomol. NMR 1*, 3.
Crippen, G. M. (1981) in *Distance Geometry and and Conformational Calculations* (Bawden, D., Ed.) Research Studies Press (Wiley), New York.

Crippen, G. M., & Havel, T. F. (1988) in *Distance Geometry and Molecular Conformation* (Bawden, D., Ed.) Research Studies Press (Wiley), New York.
Feng, Y. (1996) Ph.D. Thesis, University of California, San Diego, CA.
Feng, Y., Melacini, G., Taulane, J. P., & Goodman, M. (1996) *Biopolymers 39*, 859.
Feng, Y., Melacini, G., & Goodman, M. (1997) *Biochemistry 36*, 8716−8724.
Goodman, M., Melacini, G., & Feng, Y. (1996a) *J. Am. Chem. Soc. 118*, 10928.
Goodman, M., Feng, Y., Melacini, G., & Taulane, J. P. (1996b) *J. Am. Chem. Soc. 118*, 5156.
Hagler, A. T. (1985) in *The Peptides* (Udenfriends, S., Meienhofer, J., & Hruby, V. J., Eds.) pp 214−296, Academic Press, Orlando, FL.
Havel, T. F., & Kuntz, I. D. (1983) *Bull. Math. Biol. 45*, 665.
Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546.
Kemp, D. S., & Petrakis, K. S. (1981) *J. Org. Chem. 46*, 5140.
Kumar, A., Ernst, R. R., & Wuthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1.
Li, M., Fan, P., Brodsky, B., & Baum, J. (1993) *Biochemistry 32*, 7377.
Long, C. G., Braswell, E., Zhu, D., Apigo, J., Baum, J., & Brodsky, B. (1993) *Biochemistry 32*, 11688.
Marion, D., & Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967.
McCammon, J. A., Wolynes, P. G., & Karplus, M. (1979) *Biochemistry 18*, 927.
Melacini, G., & Goodman, M. (1997) *Chirality* (in press).
Melacini, G., Feng, Y., & Goodman, M. (1996a) *J. Am. Chem. Soc. 118*, 10725.
Melacini, G., Feng, Y., & Goodman, M. (1996b) *J. Am. Chem. Soc. 118*, 10359.
Miller, M. H., & Scheraga, H. A. (1976) *J. Polym. Sci., Polym. Symp. 54*, 171.
Miller, M. H., Nemethy, G., & Scheraga, H. A. (1980a) *Macromolecules 13*, 910.
Miller, M. H., Nemethy, G., & Scheraga, H. A. (1980b) *Macromolecules 13*, 470.
Nemethy, G. (1981) *Biochemie 63*, 125.
Nemethy, G., & Scheraga, H. A. (1982) *Biopolymers 21*, 1535.
Nemethy, G., Miller, M. H., & Scheraga, H. A. (1980) *Macromolecules 13*, 914.
Nemethy, G., Gibson, K. D., Palmer, K. A., Yoon, C. N., Paterlini, G., Zagari, A., Rumsey, S., & Scheraga, H. A. (1992) *J. Phys. Chem. 96*, 6472.
Otting, G., Widmer, H., Wagner, G., & Wuthrich, K. (1986) *J. Magn. Reson. 66*, 187.
Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wuthrich, K. (1984) *Biochem. Biophys. Res. Commun. 117*, 479.
Redfield, A. G., & Kuntz, S. D. (1975) *J. Magn. Reson. 19*, 250.
Shaka, B., & Freeman, R. (1983) *J. Magn. Reson. 51*, 169.
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. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 9367.
Wider, G., Macura, S., Kumar, A., Ernst, R. R., & Wuthrich, K. (1984) *J. Magn. Reson. 56*, 207.
Zagari, A., Nemethy, G., & Scheraga, H. A. (1990) *Biopolymers 30*, 967.

BI962981R