Articles

Type I Collagen α-1 Chain C-Telopeptide: Solution Structure Determined by 600-MHz Proton NMR Spectroscopy and Implications for Its Role in Collagen Fibrillogenesis†

Albin Otter,‡ Paul G. Scott,§ and George Kotovych*,‡

Department of Chemistry, University of Alberta, Edmonton, Alberta T6G 2G2, Canada, and Department of Oral Biology, University of Alberta, Edmonton, Alberta T6G 2N8, Canada

Received November 18, 1987; Revised Manuscript Received February 26, 1988

ABSTRACT: The solution conformation of the α-1 chain C-telopeptide has been studied by circular dichroism (CD) and 600-MHz ^{1}H NMR spectroscopy in 60% CD_3OH/40% H_2O solution. The C-telopeptide contains 27 amino acids which form the C-terminal end of the α-1 collagen polypeptide chain. By the combined application of various two-dimensional, phase-sensitive NMR techniques (COSY, RELAY, NOESY, ROESY), a nearly complete assignment of all proton resonances was achieved. Furthermore, the backbone conformation could be established, on the basis of coupling constant and NOE data. The spectroscopic evidence indicates that large sections of the peptide exist in a nonrandom, extended conformation and that there are two segments of higher mobility around the two Gly-Gly units in positions 2,3 and 20,21. Despite these hingelike, flexible sections no measurable fold-back of any of the extended parts was evident. On the basis of this structure, a model is proposed for the simultaneous interaction of the C-telopeptide with two adjacent collagen triple helices within the growing collagen fibril.

In its fibrillar form, type I collagen is the protein that gives mechanical strength to bones, teeth, skin, tendons, and ligaments. Fibrils closely resembling those seen in the tissues can be induced to precipitate from solutions of monomeric type I collagen by neutralization and warming (Gross et al., 1955; Jackson & Fessler, 1955). Consequently, it is believed that the information directing the assembly of the collagen fibril is contained within the amino acid sequence of the three polypeptide chains ("α-chains") of the collagen monomer. The two α-1 chains of the bovine type I collagen monomer carry 16- and 24-residue non-triple-helical extension peptides ("telopeptides") at the N- and C-terminal ends, respectively, while the α-2 chain carries 9- and 6-residue telopeptides at these locations (Glanville & Kühn, 1979). The capacity to form fibrils in vitro is impaired in collagen whose telopeptides have been selectively removed by proteolysis (Hodge et al., 1960; Rubin et al., 1963). It has been deduced from more recent studies on proteolytically modified type I collagen that the N-terminal telopeptides direct linear growth of the fibrils (Leibovich & Weiss, 1970), while the C-telopeptides promote lateral as well as linear growth (Helseth & Veis, 1981a; Capaldi & Chapman, 1982). These functions presumably depend on specific interactions between the telopeptides and the triple helices of collagen monomers adjacent to them in the growing fibrils. Because such interactions are likely to be strongly conformation dependent, we have begun to study, by spectroscopic techniques, the structures of these important regions of the collagen monomer.

A preliminary investigation (Scott, 1986) suggested that the α-1 chain C-telopeptide of bovine type I collagen in aqueous solution at neutral pH possessed little regular secondary structure. Nevertheless, the sensitivity of the circular dichroism spectrum to the presence of urea or guanidine hydrochloride precluded a completely random structure. The C-telopeptide could apparently be induced to fold up under certain conditions, although the resultant structure was not defined experimentally. Nuclear magnetic resonance techniques potentially can provide a detailed description of molecular conformation in solution. In the first study using this technique (Otter et al., 1987), we determined the structure of the synthetic peptide *N*-acetyl-L-Pro-L-Gln-L-Pro-L-Pro-L-Gln-NH_2. This sequence is found approximately in the middle of the α-1 chain C-telopeptide of type I collagen from all species for which data are available, namely, cow (Rauterberg et al., 1972), chick (Fuller & Boedtker, 1981), and rabbit (Becker et al., 1975), where we believe that it may play an important role in directing the overall conformation of the telopeptide. In the present paper we report the results of a high-field proton magnetic resonance study of the complete synthetic α-1 chain C-telopeptide and use these data in an attempt to explain the function of the C-telopeptide in fibrillogenesis.

Materials and Methods

Synthetic C-Telopeptide. The peptide (*N*-acetyl-Gly^{-3}-Pro^{-2}-Pro^{-1}-Ser^1-Gly^2-Gly^3-Tyr^4-Asp^5-Leu^6-Ser^7-Phe^8-Leu^9-Pro^{10}-Gln^{11}-Pro^{12}-Pro^{13}-Gln^{14}-Glu^{15}-Lys^{16}-Ala^{17}-His^{18}-Asp^{19}-Gly^{20}-Gly^{21}-Arg^{22}-Tyr^{23}-Tyr^{24}) was synthesized by the Alberta Peptide Institute (Department of Biochemistry, University of Alberta) and supplied as a crude product, which was purified by gel filtration on a 1 cm × 65 cm column of Sephadex G-50 Superfine, eluted with 0.1% (v/v) trifluoroacetic acid. Purity was assessed from the amino acid com-

† Supported by the Alberta Heritage Foundation for Medical Research, the Natural Sciences and Engineering Research Council of Canada, the Canadian Medical Research Council, and the University of Alberta.

‡ Department of Chemistry.

§ Department of Oral Biology.

0006-2960/88/0427-3560$01.50/0

Table I: Summary of Experimental Parameters Used in the Two-Dimensional NMR Experiments[a,b]

parameters	COSYPH	RELAY	NOESYPH	ROESYPH
sweep width in F_2 (Hz)	6410	6410	6410	6410
sweep width in F_1 (Hz)	3205	3205	3205	3205
matrix size ($F_1 \times F_2$) before zero filling	512 × 4K	256 × 4K	512 × 4K	512 × 4K
matrix size ($F_1 \times F_2$) after zero filling	1K × 8K	1K × 8K	1K × 8K	1K × 8K
evolution time				
initial value (μs)	1	1	1	1
increment (μs)	78	156	78	78
no. of scans (dummy scans)	16 (2)	32 (2)	48 (2)	32 (2)
acquisition time (s)	0.32	0.32	0.32	0.32
relaxation delay (s)	1.8	1.8	1.8	1.8
other delays (ms)		60[c]	150[d]	54[e]
window functions for 2D FT (F_1/F_2)	S/S	S/S	S/S	S/S
shifts of window functions in fractions of π (F_1/F_2)	4/8	0/0	3/4	3/4

[a] All spectra were obtained at 302 K at 5.2 mM concentration in 60% CD_3OH/40% H_2O. The pH was 7.05. For abbreviations of the various techniques, see Materials and Methods. S stands for a sine bell window function. [b] The phase-sensitive spectra were first phased using the second FID (first for NOESYPH). After the Fourier transformation, the phase was optimized in both dimensions by an additional phase correction applied to the entire matrix. [c] Coherence transfer time. [d] Mixing time. [e] Spin locking time, average radio-frequency (rf) field strength 3.3 kHz.

position and by chromatography on a C18 reverse-phase HPLC[1] column (Vydac 201TP104; Separations Group Inc.). These methods were described in detail previously (Scott, 1986) in a report of a spectroscopic study on the C-telopeptide isolated from type I collagen. Bacterial collagenase was used to prepare the material for that earlier study, and hence, it included at its N-terminal end the final Gly-Pro-Pro tripeptide from the triple-helical domain of the α-1 chain. This tripeptide sequence was included in the synthetic material. The one difference between the synthetic and natural products was at telopeptide position 16, which was lysine in the former and α-aminoadipic δ-semialdehyde (the product of the action of lysyl oxidase) in the latter.

Circular Dichroism Spectroscopy. Samples were dissolved in water and diluted with water and/or methanol to a final concentration of 3 mg/mL. Spectra were recorded at ambient temperature in a cell of path length 0.0124 cm in a Jasco J-20A spectropolarimeter that had been previously calibrated with *d*-10-camphorsulfonic acid (DeTar, 1969). The mean residue weight was calculated from the amino acid composition. Other details are given in the legend to Figure 1.

Proton NMR Spectroscopy. A 7-mg sample of the purified solid material was dissolved first in water and the pH adjusted to 7.05. The sample was then freeze-dried and redissolved in 0.45 mL of a CD_3OH (60%)/H_2O (40%) mixture. A total of 50% of the water was D_2O to provide an internal deuterium lock signal during the NMR measurements. The sample was degassed prior to sealing by bubbling of argon through it for approximately 0.5 h. The concentration was 5.2 mM.

The NMR experiments were performed on a Bruker AM600 NMR spectrometer at 302 K. Data collection and data processing were controlled by an Aspect 3000 computer with an array processor using 1987 Bruker DISNMR software. The proton chemical shifts were determined relative to the undeuteriated fraction of the methyl group of CD_3OH (Merck Sharp & Dohme, MD-67, 99.2% atom % D) at 3.30 ppm with respect to trimethylsilane (TMS).

One-dimensional spectra were recorded with a sweep width of 6100 Hz and 32K (zero filled to 64K) data points. A total of 64 scans were accumulated to obtain a good signal-to-noise ratio even after treatment of the FID with an extensive resolution enhancement by means of Gaussian multiplication (line broadening factor −3 Hz). The two solvent signals (H_2O and OH group of CD_3OH at 4.72 ppm and the undeuteriated fraction of CD_3OH at 3.30 ppm) were both presaturated during a 3-s relaxation delay. This was achieved by changing the frequency of the decoupler every 0.25 s. To account for the big difference in intensity between the two solvent peaks, the water signal was presaturated twice as long as the methyl group. This relatively simple procedure made the use of more sophisticated composite pulse solvent suppression techniques (Hore, 1983) unnecessary. The effect of presaturation on the exchangeable protons of the molecule was carefully monitored, and no wipe out due to saturation transfer was observed.

The parameters used in the various two-dimensional NMR techniques (Benn & Günther, 1983; Bax, 1982) are summarized in Table I. The COSY (Aue et al., 1976), NOESY (Jeener et al., 1979), and ROESY (Bothner-By et al., 1984; Bax & Davis, 1985) experiments were all carried out in the phase-sensitive mode (Marion & Wüthrich, 1983; Bodenhausen et al., 1984), whereas the RELAY spectrum (Eich et al., 1982; Bolton & Bodenhausen, 1982) was recorded in the absolute value mode. Solvent suppression was achieved in the same way as described previously for the one-dimensional spectra. In addition, the irradiation with the decoupler on the water signal was contained at all times aside from data acquisition (Wider et al., 1983). The phase-sensitive experiments were first phased using the second FID (first FID for NOESY), and the phase constants thus determined were then applied to all spectra in the t_2 domain. The resulting 2D matrix was then rephased in both dimensions. It should be emphasized that the ROESY experiment requires a basic 90° phase correction in t_1 before the phase fine tuning can be done. As all the recorded two-dimensional spectra suffer from considerable t_1 noise and ridges (Mehlkopf et al., 1984), a great deal of caution must be exercised when trying to extract reliable information from the spectra. As long as only coupling connectivities, but not their actual values, were of interest, we found it useful to calculate the magnitude spectrum in the case of the COSY experiment. The ridges along t_1 can then be reduced dramatically in the following way. First of all, two spectra parallel to t_2 and outside the absorption range of the molecule (at approximately 0.3 and 9.5 ppm) were summed. The resulting one-dimensional spectrum contained the base line and ridges only but did not have any signals from the molecule under investigation. Two traces were used to average differences in the intensities of the ridges on the low-field and the high-field end of the spectrum. Second, the above-mentioned one-dimensional spectrum was subtracted from every

[1] Abbreviations: HPLC, high-performance liquid chromatography; 2D NMR, two-dimensional nuclear magnetic resonance; COSY, two-dimensional correlated spectroscopy; NOESY, two-dimensional nuclear Overhauser and exchange spectroscopy; ROESY, two-dimensional rotating-frame Overhauser enhancement spectroscopy; RELAY, two-dimensional relayed coherence transfer spectroscopy; CD, circular dichroism spectroscopy; FID, free induction decay.

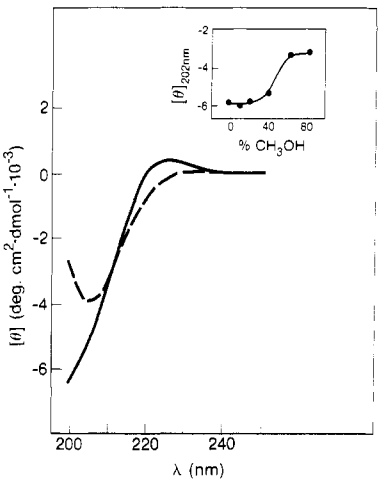

FIGURE 1: Circular dichroism spectra of synthetic C-telopeptide at 25 °C in water (—) and in 80% (v/v) methanol (– –). Each spectrum shown is the average of four, replotted after subtraction of the base line. The inset shows the effect of methanol on the mean residue ellipticity at 202 nm.

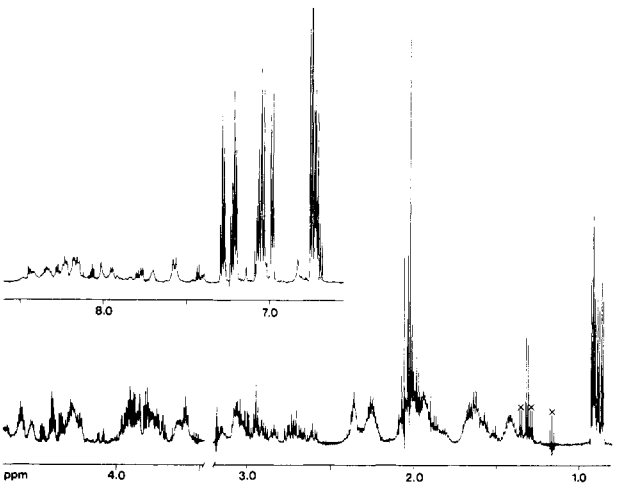

FIGURE 2: The 600-MHz ^1H spectrum of type I collagen α-1 chain C-telopeptide recorded at 5.2 mM in 60% CD_3OH/40% H_2O, pH 7.05, 302 K. The inset depicts the exchangeable and aromatic proton region, the lower trace shows the whole range from the H_α protons to the H_δ's of leucine. Impurities are denoted with an X.

spectrum of the 2D matrix parallel to t_2 by means of a 1987 Bruker DISNMR subtraction routine. The resulting two-dimensional spectrum is almost free of t_1 noise and ridges, but of course, caution must be used during this procedure to ensure that no artifacts are introduced or valuable cross-peaks are lost. We found that the overall quality of the COSY spectrum improved considerably and that connectivities could therefore be found much more easily. However, for the so-called "fingerprint" region (Wagner & Wüthrich, 1982) the phase sensitivity must be retained. Fortunately, as can be seen in Figure 3, this region is basically free of any obscuring ridges, and thus the coupling information can be obtained. t_1 noise was removed in the same way in the RELAY spectrum whereas the NOESY and ROESY spectra were kept unchanged. It is noteworthy that the NOESY spectrum had very few cross-peaks and hence did not allow a complete sequential assignment. In this regard the ROESY experiment yielded a much better result. In addition, we found that our experiment showed almost none of the commonly observed artifacts such as spurious resonances due to COSY and Hartmann–Hahn-type transfers (Marion, 1985; Neuhaus & Keeler, 1986). The selected carrier frequency at 4.72 ppm, a spin locking time of 54 ms, and a 30° flip angle for the hard pulse train (Kessler et al., 1987) apparently were good selections, and the only problem of considerable importance is the appearance of t_1 noise.

Results

The circular dichroism spectrum of the synthetic C-telopeptide in water (Figure 1) or in dilute phosphate buffer (not shown) is essentially the same as that previously reported for the material isolated from collagen (Scott, 1986). Comparison with published reference spectra for polypeptides (Greenfield & Fasman, 1969) or proteins (Chang et al., 1978) suggests that little if any regular secondary structure is present under these conditions. Addition of methanol to the aqueous telopeptide solution caused a change in conformation, as evidenced by changes in the CD spectrum which were first detectable at about 40% (v/v) methanol and complete by about 60%. Although the resulting spectrum was not clearly interpretable, we inferred that the presence of the organic solvent had stabilized a particular secondary structure, and therefore, the NMR investigation was carried out in the mixed solvent system: 60% methanol/40% water.

The spectral analysis of the telopeptide is not straightforward even at a high magnetic field of 14.1 T because many of its amino acids are of the same kind. Due to the presence of five Pro, five Gly, and three Tyr units, half of the peptide consists of only three different amino acids, thus causing a considerable overlap of resonances in all regions of interest (see Figure 2). Our sample was therefore subjected to a variety of two-dimensional NMR experiments suitable for structure elucidations of peptides (Wüthrich, 1986). First, a phase-sensitive COSY spectrum (Marion & Wüthrich, 1983) was recorded. The resulting contour plot of the fingerprint region (Wider et al., 1984), revealing NH–H_α connectivities, is shown in Figure 3. The digital resolution is ca. 1.5 Hz along the F_2 axis, and NH–H_α coupling constants can therefore be determined to an accuracy of about 0.8 Hz. Figure 3 dramatically illustrates the power of phase-sensitive COSY spectra in allowing the measurement of coupling constants from the cross-peaks rather than from the one-dimensional spectra, where severe overlapping makes such an attempt hopeless (see NH and α-region in one-dimensional spectra in Figure 3). The coupling constants derived from the COSY spectrum are summarized in Table II. The relayed COSY experiment (Eich et al., 1982; Bolton & Bodenhausen, 1982), which is an extension of the COSY, allows the observation of cross-peaks between spins that are not mutually coupled but belong to the same spin system, providing valuable information for the assignment of the NH groups in the aromatic amino acids Phe and Tyr (through NH–H_β cross-peaks). It also allowed us to distinguish the Gln and Glu moieties from other amino acids and from each other (through H_α–H_γ cross-peaks).

The combined results of the COSY and RELAY experiments allowed the unambiguous assignment of most of the resonances within each amino acid moiety. To answer the remaining question concerning the sequential assignment, we carried out two-dimensional NOE experiments (Jeener et al., 1979), namely, a phase-sensitive NOESY (Marion & Wüthrich, 1983) and a phase-sensitive rotating-frame NOE experiment (ROESY) (Bothner-By et al., 1984; Bax & Davis, 1985), both of which provide distance information based on the nuclear Overhauser effect (Noggle & Schirmer, 1971; Sanders & Mersh, 1982). The former experiment yielded a very limited number of reliable cross-peaks (that is, well above signal-to-noise) and could therefore not be used for a full sequential assignment. The ROESY experiment on the other

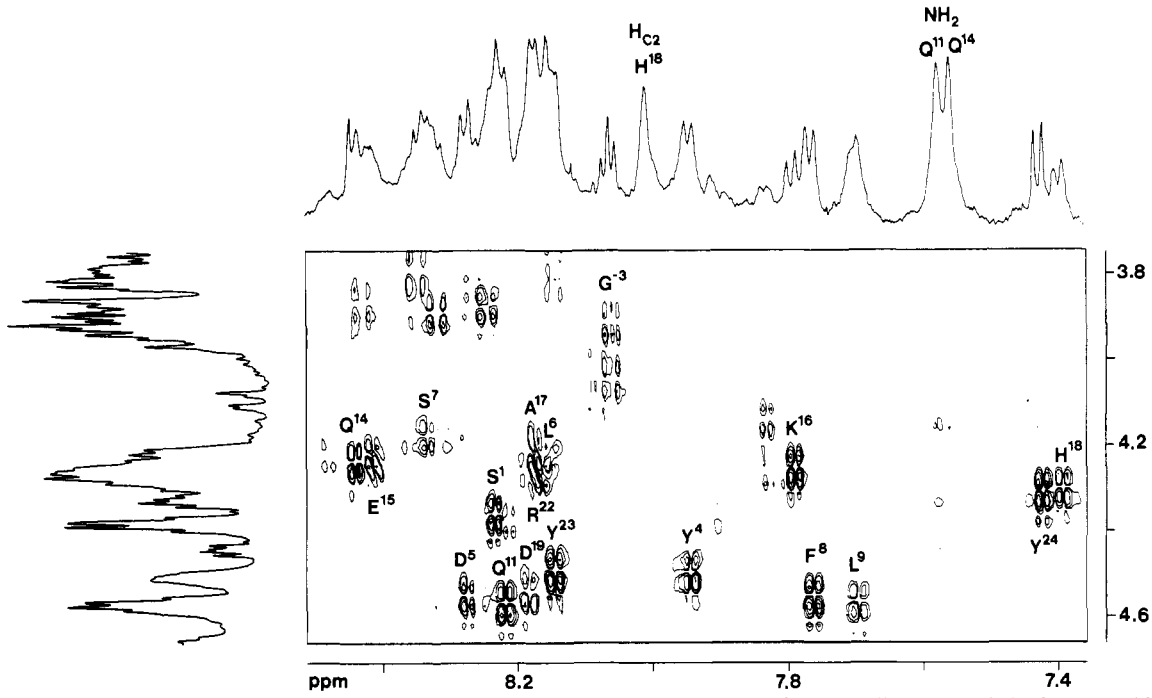

FIGURE 3: Two-dimensional contour plot of a 600-MHz phase-sensitive COSY experiment of type I collagen α-1 chain C-telopeptide recorded under the same experimental conditions as described in Figure 2. Only the fingerprint region and the corresponding one-dimensional spectra are shown. The digital resolution along the F_2 axis is 1.5 Hz. Positive and negative contour levels are plotted without distinction, and the spectrum is not symmetrized.

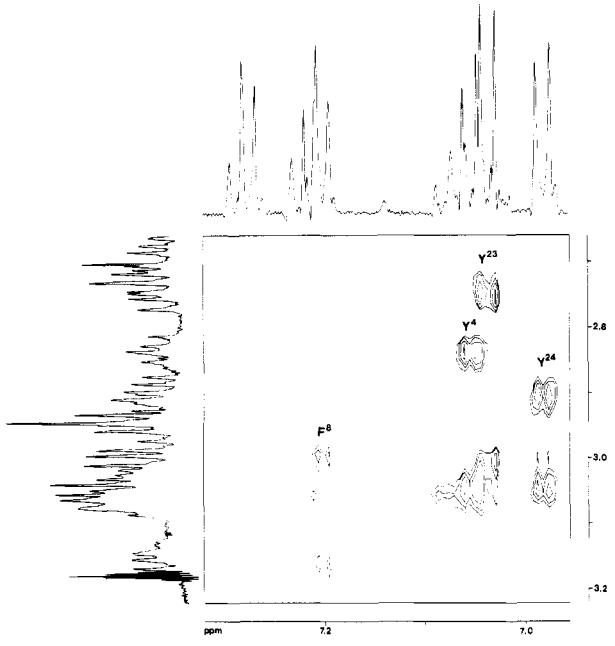

FIGURE 4: $H_{aromatic}$–H_β region of a 600-MHz phase-sensitive ROESY experiment of type I collagen α-1 chain C-telopeptide together with the corresponding sections of the one-dimensional spectrum. Same experimental conditions as in Figure 2; spin locking time was 54 ms. The contour levels are equidistant and of opposite sign with respect to the diagonal peaks. The spectrum is unsymmetrized.

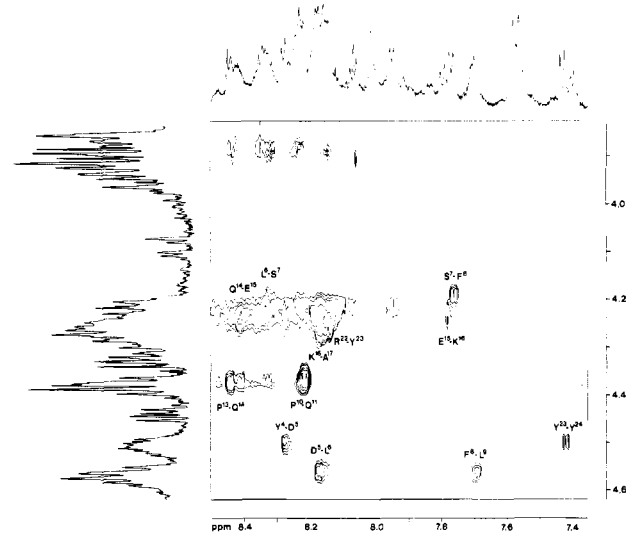

FIGURE 5: NH–H_α region of the ROESY experiment of type I collagen α-1 chain C-telopeptide as described in Figure 4.

hand provided much more valuable information and was therefore the basis for all the sequencing. It must be pointed out, however, that the results of the NOESY and ROESY experiments are by no means contradictory. The only differences are the larger intensities of the information-bearing cross-peaks in the ROESY experiment with respect to its NOESY counterpart, thus allowing the definition of peaks otherwise obscured due to thermal noise. By recording the ROESY spectrum in the phase-sensitive mode, spurious resonances due to COSY and Hartmann–Hahn-type transfers (Marion, 1985; Neuhaus & Keeler, 1986) could be identified easily by means of their phase, which is the same as for the diagonal signals. Under the selected experimental conditions (see Materials and Methods), the resulting two-dimensional spectrum showed very few such peaks. Some relevant sections of the ROESY experiment are shown in Figures 4 and 5. The identification of the four aromatic amino acids in the telopeptide through H_β/H_{ortho} connectivities was particularly useful (Figure 4). Together with the information obtained from the unique amino acids and the knowledge of the previously studied Pro^{10} to Gln^{14} segment (Otter et al., 1987), these aromatic amino acids were "landmarks" from where sequential assignment could be started. Cross-peaks containing sequential information were found between H_α and NH of adjacent amino acids (Figure 5) and between H_α protons and $H_\delta/H_{\delta'}$ of proline units (not shown). The corresponding distances are referred to as $d_{\alpha N}(i,j)$ and $d^{XP}_{\alpha\delta}(i,j)$, respectively (Wüthrich et

Table II: ^{1}H Chemical Shifts[a] and Coupling Constants[b] of Type I α-1 Chain C-Telopeptide[c]

	NH	H_α	H_β	H_γ	H_δ	others	J_{NH-H_α}
G^{-3}[d]	8.06	4.09/3.96[e]					5.8/5.8[f]
P^{-2}		4.44	2.25/1.94	(2.02)[g]	3.65/3.57		
P^{-1}		4.38	2.22/1.87	(2.00)	3.89/3.74		
S^1	8.23	4.37	3.90/3.82				7.0
G^2[h]	8.35	3.83					5.8/5.8
G^3[h]	8.24	3.91					5.6/5.6
Y^4	7.95	4.51	3.03/2.83			H_2/H_6 7.05; H_3/H_5 6.73	7.8
D^5	8.28	4.56	2.72/2.60				7.0
L^6	8.16	4.24	1.63/1.61	1.68	0.92/0.86		7.0
S^7	8.33	4.20	3.76/3.71				7.0
F^8	7.76	4.56	3.16/2.99			H_2/H_6 7.20; H_3/H_5 7.28; H_4 7.22	7.8
L^9	7.70	4.57	1.57/1.52	1.57	0.91/0.88		9.7
P^{10}		4.37	2.22/1.87	(1.98)	3.57		
Q^{11}	8.22	4.58	2.06/1.86	2.37		NH_2 7.58/6.83	7.8
P^{12}		4.66	2.28/1.92	(1.95)	3.75/3.58		
P^{13}		4.38	2.26/1.92	(2.03)	3.79/3.63		
Q^{14}	8.44	4.26	2.07/1.94	2.36		NH_2 7.56/6.75	7.0
E^{15}	8.41	4.24	1.92/1.65	2.25			7.0
K^{16}	7.79	4.27	1.80/1.69	1.41	1.65	H_ϵ 2.94	7.8
A^{17}	8.17	4.22	1.32				7.0
H^{18}	7.40	4.31	3.13/2.92			H_2 8.02; H_4 7.07	7.8
D^{19}	8.18	4.54	2.69/2.64				7.8
G^{20}[h]	8.43	3.90					6.2/6.2
G^{21}[h]	8.32	3.92					6.0/6.0
R^{22}	8.16	4.28	(1.84/1.72)	1.47/1.42	3.04		7.0
Y^{23}	8.14	4.50	3.00/2.75			H_2/H_6 7.03; H_3/H_5 6.73	7.8
Y^{24}	7.43	4.33	3.04/2.89			H_2/H_6 6.98; H_3/H_5 6.69	7.8

[a] Chemical shifts in ppm relative to methyl group of CD_3OH (3.30 ppm downfield from TMS). [b] Coupling constants in Hz. [c] All spectra were measured at 302 K at a concentration of 5.2 mM in 60% CD_3OH/40% H_2O. The pH was 7.05. [d] Data for *N*-acetyl: 2.02 ppm. [e] Splitting due to cis/trans isomerism of Gly^{-3}–Pro^{-2} peptide bond (see text). [f] Same coupling constants in both isomers. [g] Data in parentheses could not be determined very accurately due to severe spectral overlap. [h] The Gly units G^2, G^3, G^{20}, and G^{21} might be interchanged.

al., 1984). The phase of the observed peaks in the NOESY experiment indicates that the effect is negative. By comparison with the effects observed in one-dimensional quantitative NOE experiments (Neuhaus, 1983; Kinns & Sanders, 1984) of the aforementioned pentapeptide fragment, the magnitude of relevant NOEs bearing conformational information is in the 5–10% range. The unambiguous sequential assignment could be achieved for the large majority of amino acids, and the resulting chemical shifts are summarized in Table II.

Discussion

The desirable conformational information for the C-telopeptide is the determination of torsion angles (IUPAC–IUB, 1970) in the peptide backbone, thereby testing recently published theoretical predictions of the conformation (Jones & Miller, 1987; Dion & Myers, 1987) against NMR data. The combination of nuclear Overhauser enhancements and H_α–NH coupling constants allows an almost complete definition of all the important rotational angles and thus the entire backbone conformation.

In the first place, the torsion angle ϕ, describing the rotation about the N–C_α bond, is considered. It is related to the dihedral angle θ, which itself is related to the observed coupling constants (Pardi et al., 1984) by the well-known Karplus equation (Karplus, 1963). Although various sets of constants exist for this equation (Becker et al., 1979), a relatively narrow limit for θ is found for all three observed J values with four different sets of Karplus constants (Jardetzky & Roberts, 1981). By the very nature of the ϕ/θ relationship, always two values for ϕ result from one θ (J, respectively) value:

7.0 Hz: $-78° < \phi < -85°$ or $-155° < \phi < -162°$
7.8 Hz: $-81° < \phi < -94°$ or $-146° < \phi < -159°$
9.7 Hz: $-95° < \phi < -120°$ or $-120° < \phi < -145°$

The first set of ϕ values suggests that the NH protons and the β-protons of the same amino acid should be very close to each other in space. Consequently, they should exhibit NOE cross-peaks which were clearly not observed. Therefore, we conclude that the second set of torsion angles represents reality much better. This is in perfect agreement with earlier observations made on the pentapeptide P^{10}-Q^{11}-P^{12}-P^{13}-Q^{14} (Otter et al., 1987). For further reference, we use approximately average values of −158°, −152°, and −133°, respectively. The Ψ angles cannot be derived from ^{1}H coupling constants (Bystrov, 1976), but they can be estimated to a satisfactory degree from the $H_\alpha(i)$–$NH(i+1)$ and $H_\alpha(i)$–$H_{\delta,\delta'}(i+1)$ NOEs (Figure 6). Assuming trans peptide bonds ($\omega = 180°$), it can be easily shown from a CPK model that the existence of the above-mentioned NOEs requires the C_α–H_α bond (i) to be approximately trans to the carbonyl group ($i+1$). This corresponds to a Ψ angle of approximately +120° for the adjacent amino acids exhibiting such nuclear Overhauser effects. For some sections of the molecule such effects are not present, however. As no other NOEs are observed, this indicates increased flexibility in these particular segments, around the Gly-Gly subunits. Finally, the peptide bond angle ω has to be considered. Due to its partial double bond

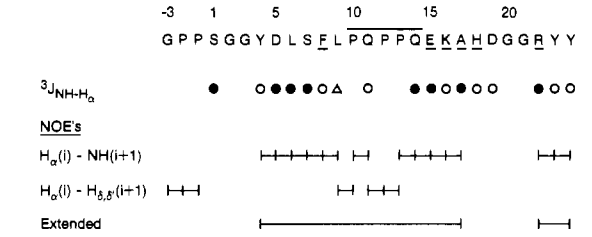

FIGURE 6: Amino acid sequence of type I α-1 chain C-telopeptide with unique residues underlined. The previously investigated fragment P^{10}–Q^{14} (Otter et al., 1987) is denoted with a bar. For ^3J coupling constants the symbols are used in the following way: filled circles for values of 7.0 ± 0.4 Hz, open circles for 7.8 ± 0.4 Hz, and a triangle for 9.7 ± 0.4 Hz. The NOEs allowing the sequential assignment are indicated by lines.

character, it is generally found to be either 180° (trans) or 0° (cis), with the former value being preferred. Considering the presence of five proline moieties, which can undergo cis/trans isomerization quite easily (Deslauriers & Smith, 1978; Otter & Kotovych, 1987), and considering the fact that two X-Pro-Pro subunits are present, for which cis isomers were found (London et al., 1978), the "all-trans" assumption has to be examined very carefully. Our NMR data indicate that the peptide bond between Gly^{-3} and Pro^{-2} (the first X-Pro-Pro unit) exists in a cis/trans equilibrium. Figure 3 shows clearly that the NH(Gly) resonance at 8.06 ppm, rather than all the other NH's of Gly, exhibits two widely separated NH/H_α cross-peaks at 4.09 and 3.96 ppm, respectively. The signal at higher field (3.96 ppm) exhibits a considerable NOE into the H_β(Pro) region, indicating close proximity between H_α-(Gly^{-3}) and $H_{\delta,\delta'}$(Pro^{-2}). As this proximity is only possible in the trans conformation [pp 123–125 of Wüthrich (1986)], we conclude that the signal at 3.96 is related to the trans isomer whereas the resonance at 4.09 ppm is assigned to the cis form. The determination of the cis/trans ratio is more difficult than that in shorter peptides where both isomers could be observed simultaneously (Otter et al., 1986). Because of the severe overlapping of resonance lines, this is not possible for the telopeptide, and the cis/trans ratio can only be estimated roughly from an examination of the cross section through the NH(Gly^{-3}) resonance at 8.06 ppm. This reveals that the cis form is predominant (ca. 60%). Bearing in mind the variety of factors influencing peak intensities in two-dimensional NMR cross-peaks, we assess the cis/trans ratio at 60:40 (±20%). Other than the cis/trans isomerization discussed above, we find no indication of any other such effects in all our data. Consequently, the $\omega = 180°$ assumption is fully acceptable.

Concerning side-chain rotamer populations (Jardetzky & Roberts, 1981), clearly resolved H_α–H_β coupling constants are a prerequisite. For the vast majority of amino acids in the telopeptide these were not observed because of overlapping resonance lines. For the five glycine moieties the task is unnecessary, and finally, for the five proline units a complete seven-spin spectral simulation would be needed, thus requiring that a large majority of proline lines be accessible. Only for the two Ser, two Asp, the Phe^8, and one of the three Tyr units a complete, unambiguous set of coupling constants can be obtained. However, when analyzed with literature values for the three classical rotamers (Feeney, 1976), none of these amino acids revealed a highly dominant population of one single rotamer, and thus fixed side-chain conformations for the above-mentioned amino acids can be excluded. With regard to the aromatic amino acids, Phe and Tyr, it must be emphasized that the observation of an NOE cross-peak between the $H_{\beta,\beta'}$ and the H_{ortho} protons of the ring system does not allow the conclusion that the ring's rotation is hindered because for all torsion angles χ_i^2, at least one ^1H–^1H distance between $H_{\beta,\beta'}$ and the ring is considerably shorter than 3.0 Å (Billeter et al., 1982). For His^{18} the situation would be different for an H(C4)–$H_{\beta,\beta'}$ interaction, but the absence of any such NOE again indicates fairly free side-chain rotation.

The final consideration with regard to the conformation of the telopeptide is the evaluation of possible hydrogen bonds. From a study of the temperature dependence of the NH resonances (data not shown), we conclude that no such interaction is present as the values for the changes in chemical shifts as a function of temperature were all greater than 3×10^{-3} ppm/K. These are typical effects for protons which are not hydrogen bonded or otherwise shielded from interacting with the solvent (Hruby, 1974).

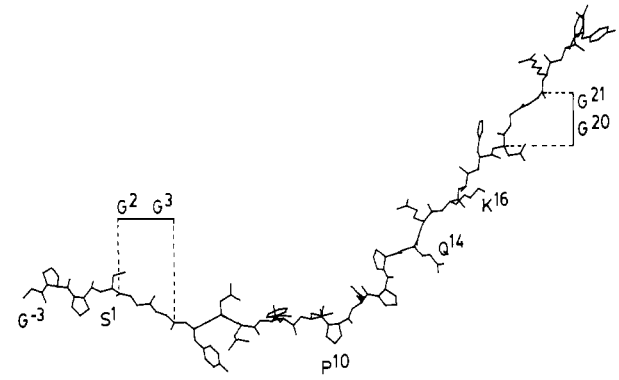

FIGURE 7: Model of the C-telopeptide constructed with the NMR data and the program PCMODEL (Henkel & Clarke, 1986). Hydrogen atoms are omitted for clarity. All Ψ angles have been set to 120°. φ angles were set according to the relationship between coupling constant (Figure 6) and θ (see text). Side chains are shown in the fully extended conformation (all χ angles 180°). Letters identifying specific amino acid residues are placed directly below or to the side of the C_α of the residue to which they refer. The peptide is oriented with C_α of Gly^{-3} and Ser^1 along the x axis and in the x–y plane. Bars delineate the flexible segments around the Gly-Gly units (shown with $\phi = -160°$ and $\Psi = 120°$).

The conformation consistent with the above data (Figure 7) can best be described as extended for large sections of the molecule. This does not imply the existence of one fixed structure, which thermodynamic considerations render extremely unlikely (Creighton, 1984), but rather a strong preference for extended over folded conformations for much of the sequence. The five residue segment P^{10}–Q^{14}, which has the same backbone conformation in the intact C-telopeptide (present data) and in isolation (Otter et al., 1987), may play a special role in directing this conformational preference. The circular dichroism of the synthetic pentapeptide *N*-acetyl-Pro-Gln-Pro-Pro-Gln-NH_2 is quite insensitive to temperature and to the presence of organic cosolvents (methanol or trifluoroethanol) or denaturants such as urea (Otter et al., 1987). Together with the NMR evidence, this suggests an especially stable conformation that is extended, within the limitations imposed by the proline rings (see below). By contrast, the two Gly-Gly sequences may provide hingelike flexibility at important positions in the telopeptide. None of the existing theoretical predictions (Helseth & Veis, 1981b; Dion & Myers, 1987; Jones & Miller, 1987) could be confirmed by our NMR data, indicating that results from probability-based methods must be viewed with caution.

The use of a methanol/water mixture as solvent for the NMR studies was based on the observation that the addition of methanol to a solution of the peptide in water caused the CD spectrum to change discontinuously toward one suggesting regular secondary structure. Although the structuring effects on peptides of organic cosolvents such as methanol and trifluoroethanol appear not to be well-understood, the increased viscosity (Noda et al., 1982; Dizechi & Marschall, 1982) and/or reduced dielectric constant may be involved. Water is extruded from growing collagen fibrils; hence, it could be argued that organic solvent/water mixtures may provide an environment mimicking to some extent that present during this process.

In spite of the presence of the hingelike Gly-Gly segments, we found no evidence for any stable folded tertiary structure such as that suggested by earlier data on the C-telopeptide isolated from acid-soluble type I collagen (Scott, 1986). This may be due to the substitution of α-aminoadipic δ-semialdehyde by lysine in the synthetic C-telopeptide, since the UV

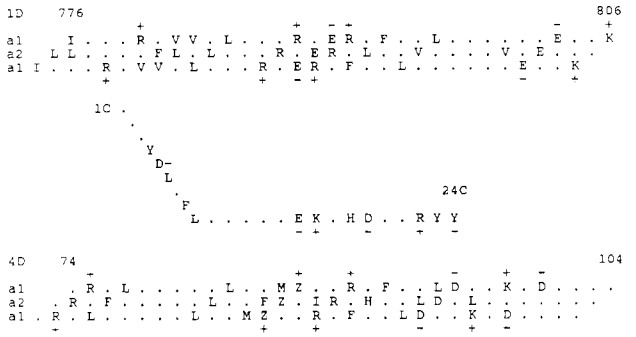

FIGURE 8: Alignment of C-telopeptide sequence with sequences (Chapman & Hulmes, 1984) of triple helices displaced axially by 1D (234 residues) and 4D (936 residues). Hydroxylysine is represented by Z. The α-2 chain is assumed to be in the middle (reference) position in the triple helix, as suggested by Traub and Fietzek (1976).

difference and circular dichroism spectra of these materials differ somewhat in the response to the presence of perturbants (P. G. Scott, unpublished work).

The results of this study have implications for the role of the C-telopeptide in fibrillogenesis in vitro. Its removal by proteolysis results in a collagen preparation that assembles primarily through the formation of long filaments (Capaldi & Chapman, 1982). This implies a capacity of the intact C-telopeptide to interact with collagen monomers that are axially related to it by some periodicity less than 4D (which would give rise to the long filaments). The 4D interaction is, however, the only one for which there is direct evidence from covalent cross-linking [reviewed by Miller (1984)].

Inspection of the amino acid sequences in collagen triple helices which would be close to the C-telopeptide in the different D-staggered arrangements reveals a strongly hydrophobic cluster at 1D, a slightly less hydrophobic cluster at 2D, and no such clustering at 3D or 4D (P. G. Scott, unpublished data). Considered in these same terms, the C-telopeptide has two clearly demarked segments, residues 6–9 being strongly hydrophobic and residues 15–22 strongly hydrophilic, separated by the sequence Pro-Gln-Pro-Pro-Gln for which we have presented evidence of rigidity. As might be anticipated, this proline-rich sequence imposes on the peptide backbone a change in direction that can be clearly understood from Figure 7.

We suggest that the hydrophobic segment of the telopeptide binds preferentially to 1D-staggered monomers, through hydrophobic interactions and perhaps also through charge–charge interaction between telopeptide residue 5 (Asp) and helix residue(s) 780 (Arg). The requirement that lysine at position 16 in the telopeptide be sufficiently close to hydroxylysine at position 87 in the helix of an adjacent 4D-staggered collagen monomer to permit the formation of a covalent intermolecular cross-link imposes some severe constraints on the conformation that the telopeptide can adopt in the growing collagen fibril, as noted by other authors (Capaldi & Chapman, 1982; Bender et al., 1983). Assuming a fixed origin, 16 telopeptide residues must be accommodated in an axial length of about 27 Å, with a rise per residue of 1.7 Å. This is close to half that for a fully extended structure. Electron-optical (Capaldi & Chapman, 1982; Bender et al., 1983), X-ray diffraction (Hulmes et al., 1977), and neutron diffraction (Hulmes et al., 1980) data obtained on fibrillar forms of collagen have been interpreted as suggesting a "condensation" of the N-terminal part of the C-telopeptide. This requirement could be accommodated within the present model by adjustments to the ϕ and/or Ψ angles of Gly^2 and Gly^3 such that the peptide backbone of residues 4–10 remains extended but at an angle to the axis of the triple helix from which it originates. The potential for hydrophobic interactions with a 1D-staggered triple helix would seem to be maximized by this arrangement, which is illustrated schematically in Figure 8. Residues 15–24 would then be extended more or less axially along a 4D-staggered triple helix, as suggested by Capaldi and Chapman (1982), with their side chains interacting with those on the 4D triple helix (and possibly also the 1D triple helix), primarily through charge–charge interactions. The transition from "oblique" to "axial" orientations would be given by Pro^{10}–Gln^{14}. As noted above, the 4D axial stagger is then stabilized by covalent intermolecular cross-linking.

Although the physical data presented here pertain only to the conformation of the α-1 chain C-telopeptide in solution, we suggest that the demonstrated structure provides, at this level, a satisfactory explanation for the observed role of this region of the type I collagen molecule in fibrillogenesis in vitro. A more rigorous, and realistic, three-dimensional analysis of the theoretical problem than that presented here would have to consider two α-1 chain C-telopeptides, one α-2 chain C-telopeptide, and associated triple helices. However, given the large number of degrees of freedom, only somewhat reduced by the available physical data, a unique solution to this problem may prove elusive.

ACKNOWLEDGMENTS

We gratefully acknowledge Bruker (Germany) and especially Dr. Tony Keller for providing access to the AM 600 system. We are indebted to Drs. Manfred Spraul and Wolfgang Bermel (Bruker, Rheinstetten), Dr. Hans Vogel (University of Calgary), and Elena Edwards for Technical assistance.

REFERENCES

Aue, W. P., Bartholdi, E., & Ernst, R. R. (1976) *J. Chem. Phys. 64*, 2229–2246.

Bax, A. (1982) *Two-Dimensional Nuclear Magnetic Resonance in Liquids*, Reidel, Boston.

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 63*, 207–213.

Becker, E. L., Bleich, H. E., Day, A. R., Freer, R. J., Glasel, J. A., & Visintainer, J. A. (1979) *Biochemistry 18*, 4656–4668.

Becker, U., Fietzek, P. P., Furthmayr, H., & Timpl, R. (1975) *Eur. J. Biochem. 54*, 359–366.

Bender, E., Silver, F. H., & Hayashi, K. (1983) *Collagen Relat. Res. 3*, 407–418.

Benn, R., & Günther, H. (1983) *Angew. Chem., Int. Ed. Engl. 22*, 350–380.

Billeter, M., Braun, W., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 321–346.

Bodenhausen, G., Kogler, H., & Ernst, R. R. (1984) *J. Magn. Reson. 58*, 370–388.

Bolton, P. H., & Bodenhausen, G. (1982) *Chem. Phys. Lett. 89*, 139–144.

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

Bystrov, V. F. (1976) *Prog. Nucl. Magn. Reson. Spectrosc. 10*, 41–81.

Capaldi, M. J., & Chapman, J. A. (1982) *Biopolymers 21*, 2291–2313.

Chang, C. T., Wu, C.-S. C., & Yang, J. T. (1978) *Anal. Biochem. 91*, 13–31.

Chapman, J. A., & Hulmes, D. J. S. (1984) in *Ultrastructure of the Connective Tissue Matrix* (Ruggeri, A., & Motta, P. M., Eds.) pp 1–33, Martinus-Nijhoff, Boston.

Creighton, T. E. (1984) *Proteins*, Freeman, New York.

Deslauriers, R., & Smith, I. C. P. (1976) *Topics in Carbon-13 NMR Spectroscopy* (Levy, G. C., Ed.) pp 2–66, Wiley-Interscience, New York.
DeTar, Del. F. (1969) *Anal. Chem. 41*, 1406–1408.
Dion, A. S., & Myers, J. C. (1987) *J. Mol. Biol. 193*, 127–143.
Dizechi, M., & Marschall, E. (1982) *J. Chem. Eng. Data 27*, 358–359.
Eich, G., Bodenhausen, G., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 3731–3732.
Feeney, J. (1976) *J. Magn. Reson. 21*, 473–478.
Fuller, F., & Boedtker, H. (1981) *Biochemistry 20*, 996–1006.
Glanville, R. W., & Kühn, K. (1979) *Fibrous Proteins: Scientific, Industrial and Medical Aspects* (Parry, D. A. D., & Cramer, L. K., Eds.) Vol. 1, pp 133–150, Academic, London.
Greenfield, N. J., & Fasman, G. D. (1969) *Biochemistry 8*, 4108–4116.
Gross, J., Highberger, J. H., & Schmitt, F. O. (1955) *Proc. Natl. Acad. Sci. U.S.A. 41*, 1–7.
Helseth, D. L., Jr., & Veis, A. (1981a) *J. Biol. Chem. 256*, 7118–7128.
Helseth, D. L., Jr., & Veis, A. (1981b) *Chemistry and Biology of Mineralized Connective Tissues* (Veis, A., Ed.) pp 85–92, Elsevier/North-Holland, New York.
Henkel, J. G., & Clarke, F. H. (1986) *Molecular Graphics on the IBM PC Microcomputer*, Academic, Orlando, FL.
Hodge, A. J., Highberger, J. H., Deffner, G. G. J., & Schmitt, F. O. (1960) *Proc. Natl. Acad. Sci. U.S.A. 46*, 197–206.
Hore, P. J. (1983) *J. Magn. Reson. 55*, 283–300.
Hruby, V. J. (1974) *Chem. Biochem. Amino Acids, Pept., Proteins 3*, 1–188.
Hulmes, D. J. S., Miller, A., White, S. W., & Doyle, B. B. (1977) *J. Mol. Biol. 110*, 643–666.
Hulmes, D. J. S., Miller, A., White, S. W., Timmins, P. A., & Berthet-Colominas, C. (1980) *Int. J. Biol. Macromol. 2*, 338–346.
IUPAC–IUB Commission on Biochemical Nomenclature (1970) *Biochemistry 9*, 3471–3479.
Jackson, D. S., & Fessler, J. H. (1955) *Nature (London) 176*, 69–70.
Jardetzky, O., & Roberts, G. C. K. (1981) *NMR in Molecular Biology*, pp 143–186, Academic, New York.
Jeener, J., Meier, R., Bachman, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.
Jones, E. Y., & Miller, A. (1987) *Biopolymers 26*, 463–480.
Karplus, M. J. (1963) *J. Am. Chem. Soc. 85*, 2870–2871.
Kessler, H., Griesinger, Ch., Kerssebaum, R., Wagner, R., & Ernst, R. R. (1987) *J. Am. Chem. Soc. 109*, 607–609.
Kinns, M., & Sanders, J. K. M. (1984) *J. Magn. Reson. 56*, 518–520.
Leibovich, S. J., & Weiss, J. B. (1970) *Biochim. Biophys. Acta 214*, 445–454.
London, R. E., Matwiyoff, N. A., Stewart, J. M., & Cann, J. R. (1978) *Biochemistry 17*, 2277–2283.
Marion, D. (1985) *FEBS Lett. 192*, 99–103.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Mehlkopf, A. F., Korbee, D., & Tiggelmann, T. A. (1984) *J. Magn. Reson. 58*, 315–323.
Miller, E. J. (1984) in *Extracellular Matrix Biochemistry* (Piez, K. A., & Reddi, A. H., Eds.) pp 41–81, Elsevier, New York.
Neuhaus, D. (1983) *J. Magn. Reson. 53*, 109–114.
Neuhaus, D., & Keeler, J. (1986) *J. Magn. Reson. 68*, 568–574.
Noda, K., Ohashi, M., & Ishida, K. (1982) *J. Chem. Eng. Data 27*, 326–327.
Noggle, J. H., & Schirmer, R. E. (1971) *The Nuclear Overhauser Effect*, Academic, New York.
Otter, A., & Kotovych, G. (1987) *J. Magn. Reson. 74*, 293–307.
Otter, A., Kotovych, G., & Stewart, J. M. (1986) *J. Biomol. Struct. Dyn. 3*, 1121–1131.
Otter, A., Scott, P. G., & Kotovych, G. (1987) *J. Am. Chem. Soc. 109*, 6995–7001.
Pardi, A., Billeter, M., & Wüthrich, K. (1984) *J. Mol. Biol. 180*, 741–751.
Rauterberg, J., Fietzek, P. P., Rexrodt, F., Becker, U., Stark, M., & Kühn, K. (1972) *FEBS Lett. 21*, 75–79.
Rubin, A. L., Pfahl, D., Speakman, P. T., Davison, P. F., & Schmitt, F. O. (1963) *Science (Washington, D.C.) 139*, 37–39.
Sanders, J. K. M., & Mersh, J. D. (1982) *Prog. Nucl. Magn. Reson. Spectrosc. 15*, 353–371.
Scott, P. G. (1986) *Biochemistry 25*, 974–980.
Traub, W., & Fietzek, P. P. (1976) *FEBS Lett. 68*, 245–249.
Wagner, G., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 347–366.
Wider, G., Hosur, R. V., & Wüthrich, K. (1983) *J. Magn. Reson. 52*, 130–135.
Wider, G., Macura, S., Kumar, A., & Ernst, R. R. (1984) *J. Magn. Reson. 56*, 207–234.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley-Interscience, New York.
Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715–740.