*Articles*

# Sequential $^1H$ NMR Assignment and Secondary Structure Determination of Salmon Calcitonin in Solution

Andrea Motta,*,‡ Maria Antonietta Castiglione Morelli,‡ Nagana Goud,§ and Piero Andrea Temussi∥

*Istituto per la Chimica di Molecole di Interesse Biologico del CNR, 80072 Arco Felice (Napoli), Italy, Bachem Inc., Torrance, California 90505, and Dipartimento di Chimica, Università di Napoli, 80134 Napoli, Italy*

*Received December 27, 1988; Revised Manuscript Received March 28, 1989*

ABSTRACT: Salmon calcitonin (sCT) has been investigated by NMR at 500 MHz in a 90% $DMSO_{d6}$–10% $^1H_2O$ (v/v) mixture at 278 K. All backbone and side-chain resonances of the hormone have been assigned by using high-resolution phase-sensitive two-dimensional techniques. Analysis of the type and magnitude of the observed sequential nuclear Overhauser effects, the NH–$\alpha$CH spin–spin coupling constants, and the $^1H/^2H$ exchange kinetics measured in 80% $DMSO_{d6}$–20% $^2H_2O$ (v/v) at 278 K enabled prediction of the secondary structure. Overall, an extended conformation is the dominant feature of the solution, but there are clear indications for a short double-stranded antiparallel $\beta$ sheet in the central region comprising residues 12–18, connected by a three-residue hairpin loop formed by residues 14–16. Two tight turns, made by residues 6–9 and 25–28, were also identified, but no evidence was found for the presence of a regular helical segment. The $\beta$ sheet favors an amphipathic distribution of the residues, orienting the predominantly hydrophilic Ser[13], Glu[15], and His[17] side chains above the plane of the sheet, and the predominantly hydrophobic Leu[12], Gln[14], and Leu[16] below it. This is interpreted as the "seed" of the amphipathic $\alpha$ helix postulated to be responsible for the interaction of sCT with lipids, a situation reminiscent of the folding mechanism of signal peptides in the interaction with membranes. The possible significance of the cis–trans Pro[23] isomerism is discussed.

Calcitonin (CT)[1] is a single polypeptide chain hormone secreted by the parafollicular cells of mammalian thyroid or by the ultimobranchial body in lower animals. In its common form, it consists of 32 amino acids with a disulfide bridge between the cysteine residues at sites 1 and 7, and ends with an amidated proline at the carboxyl terminus (Scheme I). Besides its regulatory function of the calcium–phosphorus metabolism (Austin & Heath, 1981), CT is postulated to be a neuromodulator and/or a neurotransmitter in the central nervous system. Indeed, calcitonin-like molecules have been found in the brain of many species as well as in human brain (Fisher et al., 1981; Rizzo & Goltzman, 1981).

Scheme I: Amino Acid Sequence of sCT

Cys[1]-Ser-Asn-Leu-Ser[5]-Thr-Cys-Val-Leu-Gly[10]-Lys-Leu-
Ser-Gln-Glu[15]-Leu-His-Lys-Leu-Gln[20]-Thr-Tyr-Pro-Arg-
Thr[25]-Asn-Thr-Gly-Ser-Gly[30]-Thr-Pro-$NH_2$

Although CTs from many species have been sequenced and characterized biologically, only few clear correlations have been made between structure and biological activity. CD studies in water find no significant presence of secondary structure (Epand, 1983), but the addition of structure-promoting solvents (such as trifluoroethanol or 2-chloroethanol) induces a conformational transition that could contain as much as 50% $\alpha$ helix both for sCT (Moe et al., 1983) and for pCT (Brewer & Edelhoch, 1970). Similar helix induction has been reported in the presence of SDS and various lipids (Epand, 1983).

A structural feature of sCT, common to several polypeptide hormones (Epand, 1983; Epand et al., 1983), is that the primary sequence has regularly spaced hydrophobic amino acids at every third or fourth position in the central region of the chain. The regular spacing of hydrophobic amino acid residues would allow the polypeptide to fold in an amphipathic helix in which one face is hydrophobic and the other hydrophilic. This feature was demonstrated to be important for the solubilization of lipids by several polypeptide hormones of diverse sequence (Epand, 1983) including sCT (Epand et al., 1983). Accordingly, CT analogues with amino acid replacement preserving the amphipathic helix have been synthesized and shown to retain biological activity (Moe et al., 1983; Moe & Kaiser, 1985).

Nevertheless, some doubts are cast on the role of helix-forming ability in determining the biological activity of sCT. Epand et al. (1985, 1986a,b, 1988) have indicated that the amphipathic helix is not required for the biological activity, which seems to be related to the conformational flexibility and "long-range" interactions between the amino- and carboxyl-terminal regions of the molecule affecting receptor binding

‡ Istituto per la Chimica di Molecole di Interesse Biologico del CNR.
§ Bachem Inc.
∥ Università di Napoli.

[1] Abbreviations: CT, calcitonin; CD, circular dichroism; sCT, salmon calcitonin; pCT, porcine calcitonin; SDS, sodium dodecyl sulfate; NMR, nuclear magnetic resonance; hCT, human calcitonin; $DMSO_{d6}$, perdeuterated dimethyl sulfoxide; NOE, nuclear Overhauser enhancement; TMS, tetramethylsilane; 1D, one dimensional; FID, free induction decay; 2D, two dimensional; DQ, two-dimensional double-quantum spectroscopy; DQF-COSY, two-dimensional double-quantum-filtered correlated spectroscopy; NOESY, two-dimensional nuclear Overhauser enhancement spectroscoopy; $^3J_{HN\alpha}$, vicinal coupling constant between NH and $\alpha$ protons; ppm, parts per million; $d_{\alpha N}$, distance between the $\alpha$CH of residue $i$ and the NH of residue $i + 1$; $d_{NN}$, distance between the NH protons of residues $i$ and $i + 1$.

0006-2960/89/0428-7996$01.50/0 

and/or conformational properties of the receptor-bound hormone.

Early NMR studies (Wuethrich, 1976) on hCT in aqueous and $DMSO_{d6}$ solutions have suggested that chemical shift data are consistent with a random-coil conformation.

Our study reports on the characterization of the secondary structure of sCT in DMSO–water mixtures by high-resolution NMR. These mixtures, as well as other so-called cryoprotective solvents, have been used to investigate several enzyme-catalyzed reactions (Douzou & Petsko, 1984) and proteins systems at low temperatures (Fink, 1986; Tobias & Markley, 1986). One of the interesting properties of DMSO–water mixtures is the increase in viscosity compared with the viscosity of the individual solvent alone (Schichman & Amey, 1971), but the major advantage of this solvent system is that the viscosity, and thus the NOE percentage, can be modulated owing to the high sensitivity of DMSO/water to temperature changes (Motta et al., 1987; Fesik & Olejniczak, 1987). This has proved to be very useful in conformational studies (Motta et al., 1987, 1988; Fesik et al., 1987). Accordingly, in such a cryoprotective solvent, characteristic NOE patterns were observed, and they could be interpreted in terms of a nonrandom structure.

## Materials and Methods

sCT was prepared at Bachem Inc., Torrance, CA. For the acquisition of NMR spectra, sCT was dissolved in 90% $DMSO_{d6}$–10% $^1H_2O$ (v/v) to a concentration of $10^{-3}$ M, with isotopically labeled solvents originating from Aldrich (Milwaukee, WI).

$^1H$ spectra were recorded at 500 MHz on a Bruker WM-500 spectrometer interfaced to an Aspect 2000 computer, at 278 K, and referenced to internal TMS. 1D spectra were typically acquired in quadrature detection. The water signal was suppressed by low-power selective irradiation in the homogated mode. A 1D truncated-driven Overhauser experiment (Dubs et al., 1979) was run by preirradiating the transition for 550 ms before sampling. The frequency setting of the preirradiation pulse was alternated between off- and on-resonance spectral positions every eight scans, and the resulting FIDs were subtracted in computer memory to minimize long-term instability artifacts.

2D experiments were performed as follows. The DQ was carried out by using Mareci and Freeman's sequence (Mareci & Freeman, 1983), with a detection pulse of 135° and a delay $\tau = 1/8J$). A 32-step phase cycling was used to select specifically DQ coherences. A total of 96 transients were collected for each of the 512 $t_1$ increments.

For phase-sensitive DQF-COSY (Piantini et al., 1982; Shaka & Freeman, 1983; Rance et al., 1983) and NOESY (Jeener et al., 1979; Macura & Ernst, 1979; Bodenhausen et al., 1984) experiments, the time-proportional phase incrementation scheme was used (Bodenhausen et al., 1980; Marion & Wuethrich, 1983). A total of 512 experiments with 128 scans of 2048 points were performed over 6024 Hz of spectral width. Time domain data matrices were all zero-filled to 4K in both dimensions, yielding a digital resolution of 2.7 Hz/point; 10°-shifted sine-bell window functions were applied before transformation in both dimensions. Pure absorption NOESY spectra were obtained with different mixing times (50, 100, 200, and 400 ms) with a 10% random variation of mixing time to cancel scalar correlation effects (Macura et al., 1981).

Irradiation of the $^1H_2O$ resonance in 2D experiments was carried out during the relaxation time and, in the case of NOESY, during the mixing time.

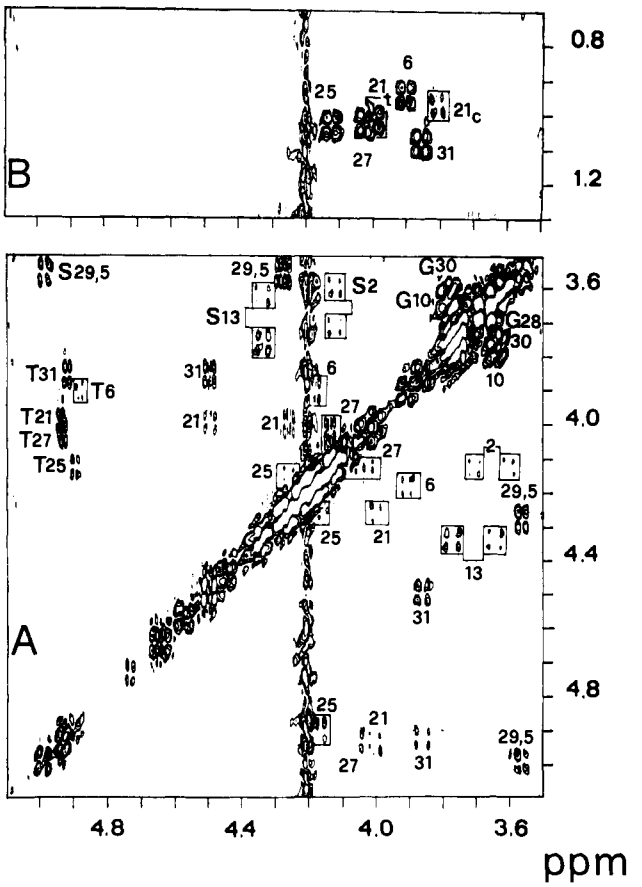


FIGURE 1: DQF-COSY spectrum of sCT in 90% $DMSO_{d6}$–10% $^1H_2O$ (v/v) at 278 K: identification of the spin systems of three Gly, four Ser, and five Thr residues. The standard one-letter code for amino acids has been used for labeling the cross-peaks. The sequence positions, obtained as a final result of the sequential assignment procedure, are also indicated. In panel A, weak cross-peaks, clearly seen at lower plot levels, are boxed. 21c and 21t (panel B) indicate $\beta H$–$\delta CH_3$ cross-peaks relative to cis and trans conformers of $Thr^{21}$.

The amide hydrogens which are relatively slow to exchange with the deuterium of the solvent were identified by dissolving protonated sCT in 80% $DMSO_{d6}$–20% $^2H_2O$ (v/v) at 278 K and recording 1D spectra over a period of 24 h.

## Results

The first step in the determination of protein structure by NMR is the complete sequence-specific assignments for all backbone and side-chain protons of the molecule (Wuethrich, 1986). Identification of complete spin systems of all 32 residues was based on DQF-COSY and DQ, and complemented with NOESY experiments when ambiguities arose. The identified amino acids were then sequentially ordered by resorting to NOEs between backbone protons and compared with the known primary sequence (Wuethrich, 1986). Finally, the pattern of NOEs was interpreted in structural terms. Identification of regular secondary structure elements relied on typical NOE patterns of sequential, medium-range, and long-range effects along different segments of the polypeptide (Billeter et al., 1982; Wuethrich et al., 1984), upon measurements of $^3J_{HN\alpha}$ coupling constants (Pardi et al., 1984), and upon studies of the backbone amide proton exchange rate (Wagner & Wuethrich, 1982).

*Spin System Identification.* The analysis of the $\alpha H$–$\beta H$ cross-peaks was divided in three steps. Firstly, we analized the systems having unique chemical structure; second, the AMX systems; and, finally, those with long side chains.

In Figure 1, where the residues are labeled by sequence numbers, the three glycyl residues (sites 10, 28, and 30,

Table I: $^1H$ Chemical Shifts of sCT in 90% $DMSO_{d6}$–10% $^1H_2O$ (v/v) at 278 K[a]

| residue | shift of proton resonance (ppm) | | | |
|---|---|---|---|---|
| | NH | αH | βH | others |
| $C^1$ | | 4.11 | 3.01, 3.30 | |
| $S^2$ | 7.91 | 4.13 | 3.62, 3.69 | γOH 5.29 |
| $N^3$ | 7.61 | 4.43 | 2.62, 2.73 | $\delta NH_2$ 6.98, 7.18 |
| $L^4$ | 8.25 | 4.30 | 1.46, 1.46 | γCH 1.60 |
| | | | | $\delta CH_3$ 0.81, 0.85 |
| $S^5$ | 7.76 | 4.28 | 3.58, 3.58 | γOH 4.98 |
| $T^6$ | 7.60 | 4.18 | 3.90 | γOH 4.90 |
| | | | | $\gamma CH_3$ 0.96 |
| $C^7$ | 7.91 | 4.50 | 3.05, 3.21 | |
| $V^8$ | 7.72 | 4.06 | 1.99 | $\gamma CH_3$ 0.83, 0.83 |
| $L^9$ | 8.00 | 4.22 | 1.48, 1.48 | γCH 1.60 |
| | | | | $\delta CH_3$ 0.81, 0.85 |
| $G^{10}$ | 8.06 | 3.63, 3.79 | | |
| $K^{11}$ | 8.08 | 4.35 | 1.52, 1.52 | $\gamma CH_2$ 1.28, 1.28 |
| | | | | $\delta CH_2$ 1.48, 1.48 |
| | | | | $\epsilon CH_2$ 2.74, 2.74 |
| | | | | $\epsilon NH_2$ 7.64 |
| $L^{12}$ | 7.74 | 4.09 | 1.48, 1.48 | γCH 1.60 |
| | | | | $\delta CH_3$ 0.81, 0.85 |
| $S^{13}$ | 8.89 | 4.34 | 3.62, 3.76 | |
| $Q^{14}$ | 7.88 | 4.21 | 1.78, 1.88 | $\gamma CH_2$ 1.99, 2.07 |
| | | | | $\delta NH_2$ 6.82, 7.27 |
| $E^{15}$ | 8.16 | 4.13 | 1.79, 1.91 | $\gamma CH_2$ 2.14, 2.14 |
| $L^{16}$ | 7.49 | 4.11 | 1.56, 1.56 | γCH 1.55 |
| | | | | $\delta CH_3$ 0.82, 0.88 |
| $H^{17}$ | 7.82 | 4.48 | 3.03, 3.20 | C(2)H 8.35 |
| | | | | C(4)H 7.35 |
| | | | | N(3)H 14.11 |
| $K^{18}$ | 7.95 | 4.18 | 1.52, 1.52 | $\gamma CH_2$ 1.28, 1.28 |
| | | | | $\delta CH_2$ 1.31, 1.31 |
| | | | | $\epsilon CH_2$ 2.74, 2.74 |
| | | | | $\epsilon NH_2$ 7.64 |
| $L^{19}$ | 7.89 | 4.20 | 1.52, 1.52 | γCH 1.48 |
| | | | | $\delta CH_3$ 0.81, 0.85 |
| $Q^{20}$ | 8.00 | 4.22 | 1.78, 1.88 | $\gamma CH_2$ 1.99, 2.07 |
| | | | | $\delta NH_2$ 6.82, 7.27 |
| $T^{21}$ | 8.03t | 4.27t | 3.98t | γOH 4.95 |
| | 7.90c | 4.50c | 3.82c | $\gamma CH_3$c 0.99 |
| | | | | $\gamma CH_3$t 1.02 |
| $Y^{22}$ | 7.94t | 4.58t | 2.65t, 2.86t | C(2,6) 7.07t, 6.95c |
| | 8.18c | 4.56c | 2.56c, 2.79c | C(3,5) 6.63t, 6.66c |
| | | | | OH(4) 9.21t, 9.29c |
| $P^{23}$ | | 4.34 | 2.02, 2.07 | $\gamma CH_2$ 1.73, 1.86 |
| | | | | $\delta CH_2$ 3.57, 3.75 |
| $R^{24}$ | 7.96 | 4.28 | 1.53, 1.53 | $\gamma CH_2$ 1.52, 1.52 |
| | | | | $\delta CH_2$ 3.11, 3.11 |
| | | | | εNH 7.45 |
| $T^{25}$ | 8.03 | 4.27 | 4.13 | γOH 4.94 |
| | | | | $\gamma CH_3$ 1.02 |
| $N^{26}$ | 8.14 | 4.64 | 2.45, 2.59 | $\delta NH_2$ 7.02, 7.49 |
| $T^{27}$ | 7.74 | 4.13 | 4.04 | γOH 4.93 |
| | | | | $\gamma CH_3$ 1.04 |
| $G^{28}$ | 8.09 | 3.60, 3.76 | | |
| $S^{29}$ | 7.95 | 4.28 | 3.58, 3.58 | γOH 4.98 |
| $G^{30}$ | 8.17 | 3.62, 3.71 | | |
| $T^{31}$ | 7.90 | 4.49 | 3.87 | γOH 4.94 |
| | | | | $\gamma CH_3$ 1.10 |
| $P^{32}$ | | 4.74 | 2.08, 2.16 | $\gamma CH_2$ 1.72, 1.87 |
| | | | | $\delta CH_2$ 3.67, 3.77 |
| | | | | $NH_2$ 7.68, 7.70 |

[a] c and t refers to cis and trans isomers, respectively.

Scheme I) were easily identified by the large active coupling constants as the cross-peaks near the diagonal and centered at 3.7 ppm. The five threonyl residues (sites 6, 21, 25, 27, and 31) were recognized by the intense cross-peaks to methyl groups and the downfield shift of the βCH protons. For two Thr residues, successively assigned to $Thr^{25}$ and $Thr^{27}$, the αCH–βCH cross-peaks lay very close to the diagonal, and those for $Thr^{25}$ were clearly observed only at lower plot levels (boxes labeled 25 in Figure 1). The $Thr^{21}$ βH–$\gamma CH_3$ cross-peak, as well as those for the other side-chain protons, is split

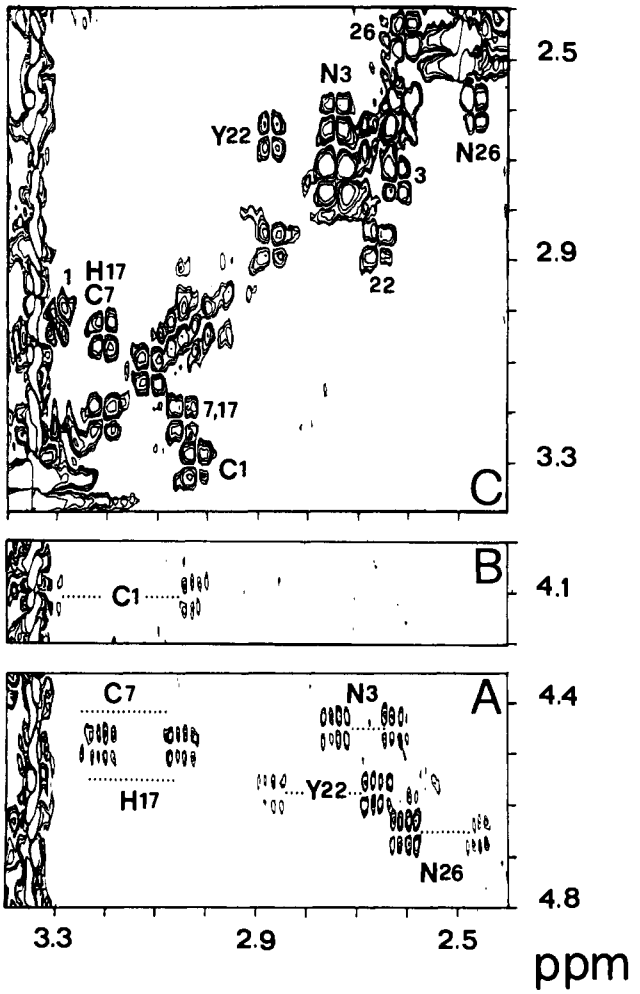


FIGURE 2: Region of the DQF-COSY spectrum of sCT in 90% $DMSO_{d6}$–10% $^1H_2O$ (v/v), 278 K, showing the αCH–$\beta CH_2$ identification of AMX spin systems of Asn, Cys, His, and Tyr (panels A and B) and the βH–βH′ cross-peaks (panel C). Cross-peaks are labeled with their sequence positions. The strong vertical band at 3.35 ppm stems from the residual $^1H_2O$ signal.

into two components appearing at 0.99, 3.82 ppm and 1.02, 3.98 ppm. Similar splitting is observed for all protons of the unique $Tyr^{22}$ and the remaining protons of $Thr^{21}$ (Table I). This conformational heterogeneity is ascribed to slow exchange between the cis and trans forms of the peptide bond preceding $Pro^{23}$. Similar observations, spanning two residues back in the sequence, on a large peptide have been reported for corticotropin, a 32 amino acid peptide (Toma et al., 1978). These observations indicate that sCT in DMSO–water exists in two forms which interchange slowly on the NMR time scale (Dwek, 1973). The separate signals for the cis and trans conformers are indicated by $21_c$ and $21_t$, respectively, in Figure 1B. All five Thr hydroxylic protons were observed (Figure 1A), thus completing the identification of the side chains of those residues. Seryl residues (sites 2, 5, 13, and 29) were recognized as being the remaining residues whose αH–βH cross-peaks are in the region depicted in Figure 1. Except for $Ser^{13}$ and $Ser^2$, the complete shifts of the side chains of the other two residues, including the hydroxylic protons, coincide (Table I). The single $Val^8$ was assigned by the high intensity of the resonance corresponding to the two degenerate methyls coupled to a single βCH proton at about 2 ppm. A remote peak in the DQ spectrum (not shown) between αH and $\gamma CH_3$ confirmed the assignment.

The second group relative to AMX systems contains Asn, Cys, His, and Tyr (sCT does not contain Asp, Phe, and Trp).

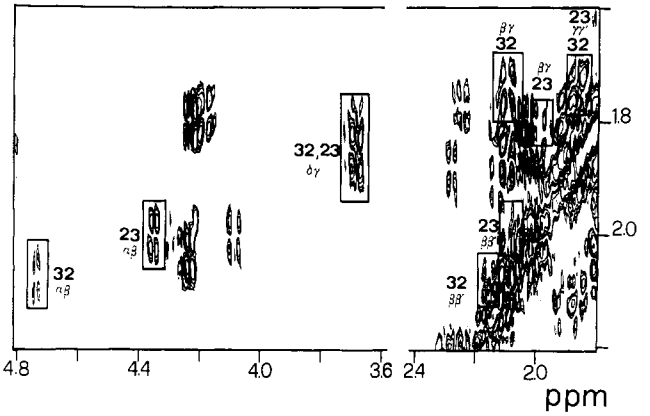


FIGURE 3: Identification of the proline spin systems in the DQF-COSY spectrum of sCT in 90% $DMSO_{d6}$–10% $^1H_2O$ (v/v), 278 K. Cross-peaks corresponding to the rings of $Pro^{23}$ and $Pro^{32}$ are boxed and labeled.

The cross-peaks relative to $\alpha CH$–$\beta CH_2$ are typical AMX coupling networks and are usually observed at ppm values below 2.5 ppm (Figure 2). Further distinction within these groups was achieved by a combination of NOESY and DQF-COSY spectra. The single $His^{17}$ and $Tyr^{22}$ were assigned by NOEs between the ring protons and the aliphatic $\beta CH_2$ protons. The two asparagines (sites 3 and 26) were distinguished from the two cysteines (sites 1 and 7) by NOEs from the side-chain protons to $\beta CH_2$. The $\alpha$ and $\beta$ protons of $Cys^7$ and $His^{17}$ have almost identical shifts (Figure 2A), but they could be distinguished by DQ spectroscopy since the chemical shift value of the double-quantum coherence in the $\omega_1$ dimension corresponds to the sum of the frequencies of the single-quantum coherences of the coupled nuclei (Mareci & Freeman, 1983).

The long side chains of Leu, Lys, Arg, Pro, Gln, and Glu (sCT does not contain Met) were finally analyzed. Although some of them represent unique chemical networks (Wuethrich, 1986), the fine structures relative to $\alpha H$–$\beta H$ cross-peaks are often incompletely resolved, being intrinsically more complicated than those for the AMX systems. The combination of DQF-COSY, DQ, and NOESY was useful in the identification. The five Leu (sites 4, 9, 12, 16, and 19) connectivity patterns were interrupted at the $\gamma H$–$\beta H_2$ cross-peaks because they fall on the diagonal in a region crowded by the peaks from the five Leu and the two Lys (sites 11 and 18). Intra-side-chain NOEs from the $\delta$-methyl resonance to the $\alpha H$ suggested patterns which were accepted only after all the assignments were established to be consistent. More difficult was the identification of the network of the $Glu^{15}$, due to the similarity of its $\beta$ and $\gamma$ proton shifts with those of Gln (sites 14 and 20). The absence of a NOESY cross-peak between $NH_2$ and $\gamma$ protons (as detected for Gln) unambiguously identified the Glu (and the Gln) side chains. For the two Lys residues and $Arg^{24}$, we made the assumption that the $\epsilon CH_2$ in Lys and the $\delta CH_2$ in Arg resonated not far from the expected random-coil chemical shift values in neat $DMSO_{d6}$ (Wuethrich, 1976). The presence of NH–$\delta CH_2$ coupling was useful for the identification of the Arg side chain. One such cross-peak was found, providing evidence of degeneracy of $Arg^{24}$ $\delta$ protons. Accordingly, the COSY connectivities of the two Lys could be traced from $\epsilon CH_2$ to $\alpha CH$. Some ambiguities arose in distinguishing between seven possible $\alpha CH$–$\beta CH_2$ cross-peaks around 1.5 ppm, but five could be ascribed to Leu and two to Lys. This group of amino acids accounted for seven residues with overlapping COSY cross-peaks and essentially similar connectivity networks.

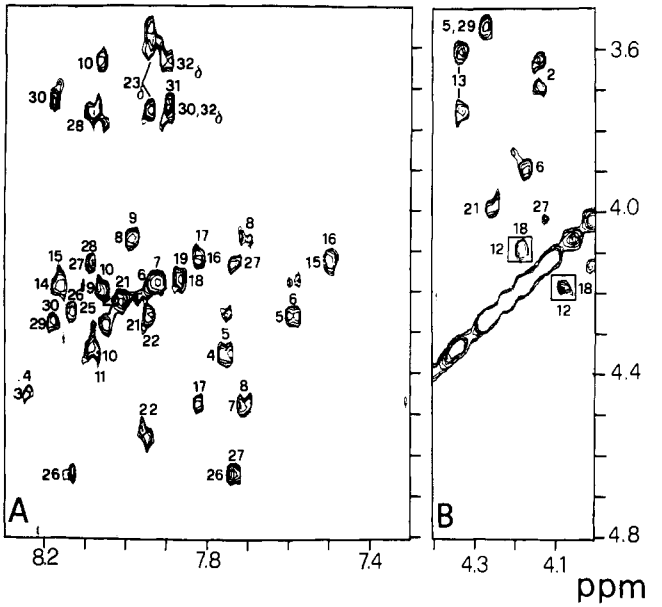


FIGURE 4: NOESY spectrum at 400-ms mixing time of sCT in 90% $DMSO_{d6}$–10% $^1H_2O$ (v/v), 278 K. (A) Fingerprint region showing sequential $d_{\alpha N}$ NOEs labeled by two sequence numbers. Amide resonances are marked either directly above or directly below each cross-peak while a single number labels the intraresidue effect from the amide proton. (B) Region ($\omega_1$ = 3.5–4.8 ppm; $\omega_2$ = 4.0–4.4 ppm) containing both intraresidue connectivities $\alpha H$–$\beta H$ for Ser and Thr and interresidue $\alpha H$–$\alpha H$. The $\alpha$–$\alpha$ effect between residues 12 and 18 is boxed.

Due to the absence of an NH amide proton, complete Pro spin system identification is necessary for the assignment of the sequentially preceding residue based on NOEs with $\delta CH_2$ (Billeter et al., 1982). The COSY cross-peaks for the two prolines residues (sites 23 and 32) in sCT are shown in Figure 3. Only one $\alpha H$–$\beta H_2$ cross-peak was observed for both Pro, but the identification of the $\beta H$–$\beta H'$ could be achieved from cross-peaks very close to the diagonal (boxes labeled 23 $\beta\beta'$ and 32 $\beta\beta'$ in Figure 3), and from the presence of remote peaks in the DQ spectrum, confirming the chemical shifts of $\beta H_2$ reported in Table I. From one $\beta$ proton, we obtained one $\gamma$ proton for both Pro residues, while the second $\gamma$ proton was identified by the presence of a $\gamma H$–$\gamma H'$ cross-peak. From one of them, only one $\delta$ proton was detected for each Pro, while the $\gamma$ protons are coincident for both residues (box 23, 32 $\gamma\gamma'$). The position of the $\delta H'$ of each Pro was found in a NOESY spectrum from the preceding NH amides of $Tyr^{22}$ and $Thr^{31}$ for $Pro^{23}$ and $Pro^{32}$, respectively.

*Sequential Assignments.* Figure 4A shows the NH–$\alpha CH$ fingerprint region of a NOESY spectrum at 400-ms mixing time, with all peaks labeled. The sequence-specific assignments for sCT relied on observation of interresidue NOE relations, mainly $d_{\alpha N}$ and $d_{NN}$ (Wuethrich et al., 1982; Billeter et al., 1982). By comparison with DQF-COSY, we were able to rule out NH–$\alpha CH$ intraresidue effects in NOESY spectra (cross-peaks labeled by a single sequence number in Figure 4). NOESY cross-peaks, located at the intersection of the horizontal lines through the $\alpha CH$ protons (labeled by sequence numbers at the right or left side) with the vertical lines through the NH resonances (labeled by sequence numbers above or below), identified the amino acid which precedes the one bearing the NH under observation. The starting point was $Ser^{13}$. An NH signal at 8.89 ppm was easily recognized as stemming from a Ser (Table I): from it, according to the procedure described above, a NOESY peak was observed with the $\alpha CH$ of a Leu, whose NH shows an effect with the $\alpha CH$ of a Lys. Since $Lys^{11}$-$Leu^{12}$-Ser-$^{13}$ is the only such tripeptide

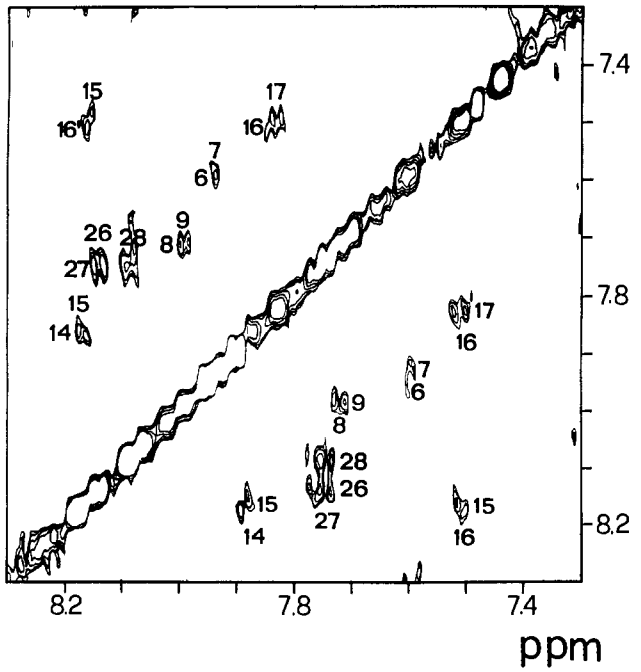


FIGURE 5: Spectral region of the NOESY of sCT in 90% $DMSO_{d6}$–10% $^1H_2O$ (v/v), 278 K ($\tau_m$ = 400 ms), showing the $d_{NN}$ connectivities. Cross-peaks are labeled with their sequence numbers.

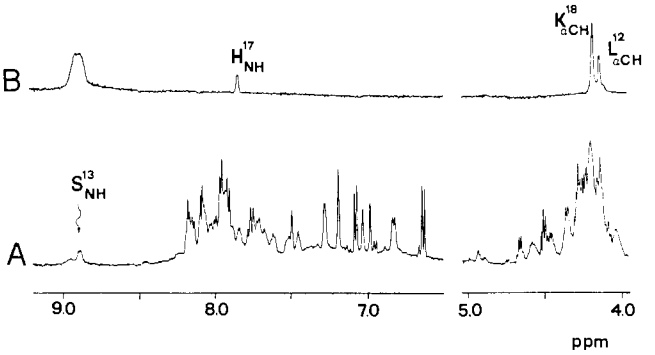


FIGURE 6: Truncated-driven NOE for sCT in 90% $DMSO_{d6}$–10% $^1H_2O$ (v/v), 278 K. In the reference spectrum (A), the preirradiated line of $Ser^{13}$ NH is identified with a wavy arrow (irradiation time, 550 ms). The difference spectrum (B) averages 4000 scans.

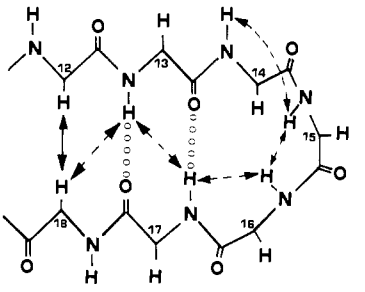

FIGURE 7: Schematic representation of the $\beta$-sheet secondary structure in sCT as determined by NMR. For each residue, the backbone atoms are drawn, and the sequence position if given. Interstrand $^1H$–$^1H$ NOEs are indicated by arrows. Solid arrow, $d_{\alpha\alpha}$; broken arrows, $d_{\alpha N}$ and $d_{NN}$. Circles represent hydrogen bonds.

present in sCT (Scheme I), the assignment was achieved. Using an identical procedure, we traced connectivities from $Ser^{13}$ down to $Asn^3$. The connectivity was lost between $Asn^3$ NH and $Ser^2$ $\alpha$CH. The NH protons of $Cys^1$ were not detectable, presumably due to fast exchange with the water solvent. Similarly, $d_{\alpha N}$ connectivities were identified between adjacent residues from the unique $Tyr^{22}$, identified from its effect with $\delta$CH of $Pro^{23}$ (Figure 4A), down to $Lys^{18}$. The sequential relation between $Lys^{18}$ and $His^{17}$ was lost, but the $His^{17}$ NH was obtained by intraresidue NOE to its $\beta$ protons. The assignment of $His^{17}$ allowed the identification of the amino acids from $His^{17}$ to $Ser^{13}$. Independent support for the segment $His^{17}$ to $Asn^{14}$ was obtained by $d_{NN}$ effects (Figure 5). Evidence for the presence of a certain amount of an isomer with a cis peptide bond was found for $Pro^{23}$. Its presence (21%) was revealed by a small effect (seen at a lower plot level in Figure 4B) between $\alpha$CH protons of $Tyr^{22}$ and $Pro^{23}$. The trans form (79%) was identified by the effect between $Tyr^{22}$ NH and $\delta$CH protons of $Pro^{23}$ (Figure 4A). Resonances sensitive to the geometry of the $Pro^{23}$ peptide bond are those of $Tyr^{22}$ and $Thr^{21}$. Both groups of protons give separate signals for the cis and trans form, allowing the evaluation of the percentage for the two isomers. $Pro^{32}$ was found to be predominantly trans since only the effect between $Thr^{31}$ NH and $\delta$CH protons of $Pro^{32}$ was found (Figure 4A). For the remaining residues, the sequential $d_{\alpha N}$ effect was lost between $Thr^{25}$ and $Arg^{24}$, but $Arg^{24}$ NH was found from its $\alpha$CH. The complete sequence-specific assignment of resonances is reported in Table I.

*Secondary Structure Determination.* Information on the secondary structure can be obtained from sequential NOE connectivities. As previously established (Wuethrich et al., 1984), the prevalence of generally strong $d_{\alpha N}$ connectivities along nearly the entire polypeptide chain (Figure 4A) and, with few exceptions, the weak or absent $d_{NN}$ (Figure 5) contacts are typical for a polypeptide backbone in an extended conformation. Support for this conclusion is lent by the coupling constants $^3J_{HN\alpha} > 7$ Hz observed for the majority of the amino acid residues which certainly exclude the presence of a regular $\alpha$ helix (Pardi et al., 1984). A strong effect is observed between the $\alpha$ protons of $Leu^{12}$ and $Lys^{18}$ (boxed in Figure 4B). Such an effect is unique for antiparallel $\beta$ sheets (Wuethrich et al., 1984) and calls for a hairpin connection between the adjacent antiparallel $\beta$ strands (Richardson, 1981). Such a loop could be identified by $d_{NN}$ connectivities between residues 14–17 in a NOESY with a mixing time of 400 ms (Figure 5). Additional information for a structure determination can be drawn from amide exchange studies, since NH groups involved in intramolecular hydrogen bonds are expected to exchange with the slowest rates (Woodward et al., 1982; Englander & Kallenbach, 1985). In an 80% $DMSO_{d6}$ –20% $^2H_2O$ (v/v) solution of sCT at 278 K, slow exchange was observed for $Ser^{13}$ and $His^{17}$ amide protons, with half-lives of 5.5 and 7.8 h, respectively. For a polypeptide as small as sCT, it is unlikely that the slow exchange is only justified by the inaccessibility to the solvent without the formation of a hydrogen bond. Since the width of the $Ser^{13}$ NH resonance at 8.89 ppm hampers unequivocal NOESY identification, its environment was tested by 1D NOE. Figure 6 depicts a truncated NOE from $Ser^{13}$ NH with 550-ms irradiation time. The difference spectrum of Figure 6B shows that the $\alpha$ protons of $Leu^{12}$ and $Lys^{18}$ are perturbed together with the $His^{17}$ amide proton at 7.82 ppm. The one on $Leu^{12}$ $\alpha$CH is a sequential effect, while the other two can be justified by the presence of a hydrogen bond between the NH of $Ser^{13}$ and the carbonyl of $His^{17}$. Accordingly, the NH of $His^{17}$ is hydrogen bonded to the carbonyl of $Ser^{13}$. The fact that residues 13 and 17 are hydrogen bonded to each other indicates that it is a three-residue hairpin turn (Chothia & Lesk, 1987) which forms the same-end connection as found in the short double-stranded $\beta$ sheet. According to these findings, we propose that the extended polypeptide segments form a short double-stranded antiparallel $\beta$ sheet in the central region, linked by

a three-residue hairpin turn, as schematically shown in Figure 7, in which the observed NOEs are indicated by arrows. Two $\beta$ turns were identified by $d_{\alpha N}$ and $d_{NN}$ connectivities in the N-terminal region of residues 6–9, and in the carboxyl-terminal part of the molecule, involving residues 25–28.

## Discussion

In this paper, we described the complete assignment and the secondary structure determination of sCT in a 90% DMSO–10% water (v/v) mixture. All resonances have been identified and assigned to specific residues by the sequential method. The size of the Overhauser effects could be adjusted owing to the sensitivity of the viscosity, and thus of the NOE percentage, of the DMSO–water mixture to temperature changes. The experimental conditions were chosen by monitoring the half-height line width of the resonances vs temperature. A good compromise was obtained by setting the temperature at 278 K. In these conditions, large negative NOEs were observed (Figure 4). Secondary structure elements were identified from sequential NOEs and from additional long-range NOE contacts. According to the found NOE pattern, we propose that in the mixture sCT adopts a short double-stranded antiparallel $\beta$ sheet with a three-residue hairpin loop connection in the central part (Figure 7) and forms two $\beta$ turns made by residues 6–9 and 25–28. The choice of such a solvent is not unusual since cryoprotective mixtures have been widely used in biochemical work (Douzou & Petsko, 1984). Our conditions, characterized by a fairly high viscosity with respect to bulk water [ca. 6 cP at 278 K compared with ca. 1 cP for water at room temperature (Schichman & Amey, 1971)], were motivated, in part, by the possibility of simulating the medium at the interface between bulk water and the membrane (Motta et al., 1988). In fact, under these experimental conditions, the reduced hydrogen-bonding capability of the DMSO may recall the interface (Sargent & Schwyzer, 1986) between the membrane and the transport fluid. In water, no signficiant secondary structure was found at pH 7.4 by CD studies (Epand et al., 1983) and by NMR (our unpublished results). In contrast, the interaction with various lipids gave rise to a consistent increase in the $\alpha$-helix content, as monitored by CD (Epand et al., 1983). This helix, amphipathic in nature, forms stable complexes with the phospholipids. The lipid affinity thus promotes the partitioning of this peptide hormone into biological membranes and increases the binding affinity to cell-surface receptors. Our study offers insights into the conformation of sCT in an intermediate state in which the polypeptide starts to assume some preferential conformation(s). On going from bulk water to DMSO–water mixtures, a $\beta$ sheet has been observed (Figure 7). This kind of $\beta$ structure can be relevant for the interaction of the peptide with a lipid phase. It has been shown, in fact, that signal sequences involved in protein export across membranes assume this conformation upon approaching the membrane and in the initial step of the insertion (Brigg et al., 1986). The actual finding of a $\beta$ structure is consistent with this hypothesis. Furthermore, as a consequence of the $\beta$-sheet formation, the side chains are oriented above and below the plane of the sheet in an alternating fashion (Richardson, 1981). By examining the nature of the residues, the $\beta$ sheet locates the predominantly hydrophilic $Ser^{13}$, $Glu^{15}$, and $His^{17}$ above the sheet, while $Leu^{12}$, $Gln^{14}$, and $Leu^{16}$ side chains (hydrophobic except for $Gln^{14}$, which is a polar but neutral residue) are below it. The seventh side chain, i.e., that of $Lys^{18}$, being at the border of the amphipathic $\beta$ sheet, may be oriented toward bulk solvent. This recalls the amphipathic distribution of the $\alpha$ helix invoked to be fundamental in increasing lipid affinity and biological potency of membrane-active peptides (Epand et al., 1988). It is even more relevant that the $\beta$ sheet has been described as one of the crucial structures for the interaction of signal peptides with the lipid phase (Briggs et al., 1986). It is then conceivable that the $\beta$ sheet evolves into the $\alpha$ helix when sCT is inside the membrane, in agreement with the catalytic role of the membrane proposed by Schwyzer and co-workers (Gremlich et al., 1984; Gysin & Schwyzer, 1984; Erne et al., 1985).

Evidence for a cis isomer has been found for $Pro^{23}$. A survey of the sequences of membrane-spanning polypeptides and proteins shows that Pro residues frequently occur within the membrane-spanning region of transport proteins (Brandl & Deber, 1986). It has been suggested that local structural changes stemming from the cis–trans isomerism of the Xaa–Pro bond may provide a mechanism for gating of channels formed by such proteins. Recently, Lolkema et al. (1988) have reported that in the *lac* permease of *Escherichia coli*, a Pro (at site 327) is not essential for transport function. By replacing $Pro^{327}$ with Ala, Gly, or Leu, they have demonstrated that there is no relationship between permease activity and the helix-breaking (Pro and Gly) or helix-forming (Ala and Leu) properties of the residue at site 327. They have suggested that primarily a chemical property of the side chain at position 327 is important for activity and that neither cis–trans isomerization nor the presence of a kink at that position is important (Lolkema et al., 1988). In sCT, it is surprising that the cis–trans isomerism is detected for $Pro^{23}$ but not for the carboxyl-terminal $Pro^{32}$, which is predominantly trans. Such a difference can be ascribed to the fact that the half-life for Xaa-Pro isomerization is highly dependent on the peptide chain, and thus on the chemical properties of the site in which the bond is found (Brandl & Deber, 1986).

Conformational studies of sCT, as well as hCT, in a membranelike environment to assess the folding of the amphipathic region, and thus the Xaa-Pro conformational state(s), are under way in our laboratory.

## Acknowledgments

We especially thank Consorzio Richerche Sardegna (Progetto I. Agr. In.) and Prof. M. L. Ganadu (University of Sassari) for their warm hospitality.

**Registry No.** sCT, 47931-85-1; DMSO, 67-68-5; L-Pro, 147-85-3.

## References

Austin, L. A., & Heath, H. (1981) *N. Engl. J. Med. 304*, 269–278.

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

Bodenhausen, G., Vold, R. L., & Vold, R. R. (1980) *J. Magn. Reson. 37*, 93–106.

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

Brandl, C. J., & Deber, C. M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 917–921.

Brewer, H. B., & Edelhoch, H. (1970) *J. Biol. Chem. 245*, 2402–2408.

Briggs, M. S., Cornell, D. G., Dluhy, R. A., & Gierash, L. M. (1986) *Science 233*, 206–208.

Chothia, C., & Lesk, A. M. (1987) *J. Mol. Biol. 196*, 901–917.

Douzou, P., & Petsko, G. A. (1984) *Adv. Protein Chem. 36*, 245–361.

Dubs, A., Wagner, G., & Wuethrich, K. (1979) *Biochim. Biophys. Acta 577*, 177–194.

Dwek, R. (1973) *Nuclear Magnetic Resonance in Biochemistry*, Oxford University Press, London.

Englander, S. W., & Kallenbach, N. R. (1984) *Q. Rev. Biophys. 16*, 521–655.
Epand, R. M. (1983) *Mol. Cell. Biochem. 57*, 41–47.
Epand, R. M., Epand, R. F., Orlowski, R. C., Schlueter, R. J., Boni, L. T., & Hui, S. W. (1983) *Biochemistry 22*, 5074–5084.
Epand, R. M., Epand, R. F., & Orlowski, R. C. (1985) *Int. J. Pept. Protein Res. 25*, 105–111.
Epand, R. M., Epand, R. F., Orlowski, R. C., Seyler, J. K., & Colescott, R. L. (1986a) *Biochemistry 25*, 1964–1968.
Epand, R. M., Stahl, G. L., & Orlowski, R. C. (1986b) *Int. J. Pept. Protein Res. 27*, 501–506.
Epand, R. M., Epand, R. F., & Orlowski, R. C. (1988) *Biochem. Biophys. Res. Commun. 152*, 203–207.
Erne, D., Sargent, D. F., & Schwyzer, R. (1985) *Biochemistry 24*, 4261–4263.
Fesik, S. W., & Olejniczak, E. T. (1987) *Magn. Reson. Chem. 25*, 1046–1048.
Fesik, S. W., Bolis, G., Sham, H. L., & Olejniczak, E. T. (1987) *Biochemistry 26*, 1851–1857.
Fink, A. L. (1986) *Methods Enzymol. 131*, 173–185.
Fisher, J. A., Tobler, P. H., Kaufmann, M., Born, W., Henke, H., Cooper, P. E., Sagar, S. M., & Martin, J. B. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 7801–7805.
Gremlich, H. U., Fringeli, H. P., & Schwyzer, R. (1984) *Biochemistry 23*, 1808–1810.
Gysin, B., & Schwyzer, R. (1984) *Biochemistry 23*, 1811–1818.
Jardetzky, O., & Roberts, G. C. K. (1981) *NMR in Molecular Biology*, Academic Press, New York.
Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.
Lolkema, J. S., Puettner, I. B., & Kaback, H. R. (1988) *Biochemistry 27*, 8307–8310.
Macura, S., & Ernst, R. R. (1979) *Mol. Phys. 41*, 95–101.
Macura, S., Huang, Y., Suter, D., & Ernst, R. R. (1981) *J. Magn. Reson. 43*, 259–281.
Mareci, T. H., & Freeman, R. (1983) *J. Magn. Reson. 51*, 531–538.
Marion, D., & Wuethrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Moe, G. R., & Kaiser, E. T. (1985) *Biochemistry 24*, 1971–1976.
Moe, G. R., Miller, R. J., & Kaiser, E. T. (1983) *J. Am. Chem. Soc. 105*, 4100–4102.
Motta, A., Picone, D., Tancredi, T., & Temussi, P. A. (1987) *J. Magn. Reson. 75*, 364–370.
Motta, A., Picone, D., Tancredi, T., & Temussi, P. A. (1988) *Tetrahedron 44*, 975–990.
Pardi, A., Billeter, M., & Wuthrich, K. (1984) *J. Mol. Biol. 180*, 741–751.
Piantini, U., Sørensen, O. W., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800–6801.
Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wuethrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.
Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167–339.
Rizzo, A. J., & Goltzman, D. (1981) *Endocrinology 108*, 1672–1677.
Sargent, D. F., & Schwyzer, R. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5774–5778.
Schichman, S. A., & Amey, R. L.(1971) *J. Phys. Chem. 75*, 98–102.
Shaka, A., & Freeman, R. (1983) *J. Magn. Reson. 51*, 169–173.
Tobias, B., & Markely, J. L. (1986) *J. Magn. Reson. 69*, 381–385.
Toma, F., Fermandjan, S., Low, M., & Kisfaludy, L. (1978) *Biochim. Biophys. Acta 534*, 112–122.
Wagner, G., & Wuethrich, K. (1982) *J. Mol. Biol. 160*, 343–361.
Woodward, C. K., Simon, I., & Tuechsen, E. (1982) *Mol. Cell. Biochem. 48*, 135–160.
Wuethrich, K. (1976) in *NMR in Biological Research: Peptides and Proteins*, pp 65–72, North-Holland, Amsterdam.
Wuethrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Wuethrich, K., Wider, G., Wagner, G., & Braun, W. (1982) *J. Mol. Biol. 155*, 311–319.
Wuethrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715–740.