Journal of Biomolecular NMR, 2 (1992) 557–572
ESCOM

J-Bio NMR 088

^{1}H resonance assignments, secondary structure and general topology of single-chain monellin in solution as determined by ^{1}H 2D-NMR

Milan T. Tomic[a], John R. Somoza[b], David E. Wemmer[a,*], Young Woo Park[c], Joon Myung Cho[c] and Sung-Hou Kim[a,*]

[a]*Department of Chemistry, University of California, and Lawrence Berkeley Laboratory, Berkeley, CA 94720, U.S.A.,*
[b]*Graduate Group in Biophysics, University of California, Berkeley, CA 94720, U.S.A.*
[c]*Lucky Biotech Corp., Emeryville, CA 94608, U.S.A.*

Received 23 June 1992
Accepted 16 September 1992

Keywords: Monellin; Sweet protein; ^{1}H assignments; Engineered loop

SUMMARY

We determined the resonance assignments, secondary structure and general topology of the 11-kDa sweet protein single-chain monellin (SCM), using two-dimensional proton nuclear magnetic resonance spectroscopy (2D-NMR). SCM is a genetically engineered protein whose design is based on the crystal structure of natural, two-chain monellin (Kim et al., 1989). Analysis of the NMR spectra shows that the secondary structure of SCM consists of a five-strand anti-parallel β-sheet and a 15-residue α-helix. Tertiary NOE constraints place the α-helix on the hydrophobic side of the β-sheet, and indicate that the sheet is partially wrapped around the helix. The general structural features determined for SCM are similar to those of native monellin (Ogata et al., 1987). Some differences between the SCM structure in solution and the crystal structure of monellin are discussed.

INTRODUCTION

In 1972 the sweet taste of the serendipity berry was traced to a protein. This protein, which was named monellin, is approximately 100 000 times more potent in sweetness than sucrose on a molar basis, and 2000 to 3000 times as sweet on a weight basis. Monellin consists of two peptide chains: an A chain of 45 residues, and a B chain of 50 residues. The native protein conformation is necessary for the sweet taste (Morris and Cagan, 1972).

* To whom correspondence should be addressed.

The design of a single-chain monellin protein (SCM) was based on the observation that in the crystal structure of monellin the carboxy-terminus of the B polypeptide chain was close to the amino-terminus of the A polypeptide chain (Ogata et al., 1987). In addition, three residues at each of these termini are partially disordered in the crystal structure, indicating flexibility in this part of the protein. By joining the A and B strands in this region, we hoped to make a single-chain protein that would fold into an overall conformation similar to that of natural monellin.

The redesigned single-chain protein is as sweet as natural monellin, and exhibits substantially increased thermal stability and increased renaturability after heat denaturation (Kim et al., 1989; Tomic and Somoza, unpublished results). We chose to study the solution conformation of single-chain monellin (SCM) as a starting point for understanding its folding and structural stability, and to identify the region responsible for its sweet taste by comparing it with structurally similar non-sweet mutant proteins.

Using the approach developed by Wüthrich and co-workers (Wüthrich et al., 1982; Wüthrich, 1986) and including the modifications proposed for larger proteins (Chazin and Wright, 1987; Chazin et al., 1988), we identified individual amino acid spin systems and assigned the majority to specific residues in the primary sequence of SCM. The secondary structure was determined using established connectivity patterns for antiparallel β-sheets and α-helices (Wüthrich, 1986; Englander et al., 1987). Tertiary structure was inferred from model building utilizing sequential and long-range, NMR-derived constraints. No structural information from the crystal structure of natural two-strand monellin was used in either the structure analysis or the model building.

The secondary structure of SCM consists of a five-strand antiparallel β-sheet that curves around a 15-residue α-helix. The large loop from the α-helix to the β-sheet shows some antiparallel β-strand features. Protons in the newly introduced genetically engineered loop do not show long-distance constraints to other parts of the protein, and several of the cross peaks are quite broad, suggesting that this segment of the protein is flexible. The structure determined for the fifth strand contains a bulge that may be necessary for protein stability. The general topology of the protein, as determined by model building using distance constraints obtained from NMR studies, is consistent with the natural monellin crystal structure.

MATERIALS AND METHODS

Sample preparation

The protein used in this study was prepared using expression and purification procedures that have been described previously (Kim et al., 1989). The samples used for 2D-NMR spectroscopy were approximately 1.5 mM in either 99.9% D_2O, or a 9:1 mixture of H_2O and D_2O. The solutions were buffered with 20 mM sodium phosphate, and the pH was adjusted to 3, 3.5, 4, or 5, using dilute hydrochloric acid. Amide exchange was qualitatively measured by using TOCSY (Bax and Davis, 1985) spectra at 33 °C taken immediately after dissolving a lyophilized protein sample at pH 3.5 in D_2O. Additionally DQF-COSY (Rance et al., 1983) spectra at 33 °C were taken after the sample had been allowed to equilibrate for 1 week at room temperature to identify the extremely slowly exchanging amide resonances that could indicate regions of stable secondary structure. Chemical shifts are relative to 3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid (TSP), which was used as an internal standard.

Purification of SCM

A small portion of the sample of the SCM protein used for 2D-NMR analysis was further purified by Pharmacia Mono-S 5/5 analytical and Mono-S 10/10 preparative column chromatography with a 0 to 1 M KCl gradient, buffered with 20 mM HEPES (pH 7). The analytical column was operated at a flow rate of 1 ml/min, and the preparative Mono-S 10/10 column was used at a flow rate of 5 ml/min. Elution was monitored by measuring the absorbance at 280 nm. Two peaks were eluted between 150 and 200 mM KCl. Dialyzed samples of the purified protein were sequenced (below) and were used for 1D-NMR spectroscopic studies.

Isoelectric focusing gel electrophoresis

Isoelectric focusing gel electrophoresis was carried out according to the instructions supplied with LKB Ampholine PAG plates (Pharmacia LKB Biotechnology AB, Bromma, Sweden) with the broad pl calibration kit obtained from Pharmacia (Pharmacia, Piscataway NJ). The gels were run at a constant 1500 V at 4 °C for 2 h.

Amino-terminal sequencing of SCM

Sequencing of the SCM samples was carried out at the Microchemical Facility, Cancer Research Laboratory, University of California, Berkeley, CA, with an Applied Biosystems automated sequencer (Applied Biosystems, Foster City, CA).

NMR data acquisition and processing

Data sets were collected on either a General Electric GN-500 spectrometer, operating at 500 MHz, or a Bruker AMX-600 spectrometer, operating at 600 MHz. The data were processed using the FTNMR software package (courtesy of Dr. Dennis Hare, Hare Research, Woodenville, WA). Before Fourier transformation, the time-domain data were apodized with phase-shifted skewed sine functions. For some spectra we used the cubic spline subroutine in FTNMR to remove baseline distortion arising from the residual water peak.

Two-dimensional nuclear Overhauser effect spectroscopy (NOESY, Wider et al., 1984) and double-quantum filtered COSY (DQF-COSY, Piantini et al., 1982; Rance et al., 1983; Shaka and Freeman, 1983) were collected at 11, 22, 25, 33, 35, 40, and 45 °C. The mixing times used for the NOESY experiments ranged from 100 ms to 300 ms. In addition, TOCSY (Braunschweiler and Ernst, 1983; Bax and Davis, 1985) experiments were performed at 25, 33 and 35 °C, with a mixing time of 35 ms to 75 ms, and relay COSY experiments (RCOSY, Wagner, 1983; Chazin and Wüthrich, 1987) were collected with delay times of 20 and 35 ms.

By systematically collecting data at different temperatures, we were able to resolve degeneracies in the sequential cross peaks and in the cross peaks obscured by the residual water peak. Suppression of the water peak in most of the experiments was accomplished by saturating the water resonance with a low power RF pulse (presaturation). We also obtained good results by selective excitation using a 'jump-return' sequence for the detection pulse of the NOESY (JR-NOESY, Plateau and Guéron, 1982) and the SCUBA peak recovery method for the DQF-COSY experiment (Brown et al., 1988).

All of the spectra in this study were collected in phase-sensitive mode, using time-proportional phase incrementation (TPPI, Rance et al., 1983). In the t_2 dimension, 1024 points were collected, and 48 to 80 scans were averaged. In the t_1 dimension, 512 increments (rows) for NOESY, and 800

increments for COSY were usually collected and zero-filled to create 1-k by 1-k data files. The spectral width was 6192 Hz for the data collected at 500 MHz and 7143 Hz for the data collected at 600 MHz.

Modeling

Physical models of the SCM protein were built with the HGS Protein Molecular Model Set (Maruzen Co., Ltd., Tokyo, Japan) incorporating constraints obtained from nuclear Overhauser effect measurements (NOEs). We determined over 500 distance constraints for SCM and obtained

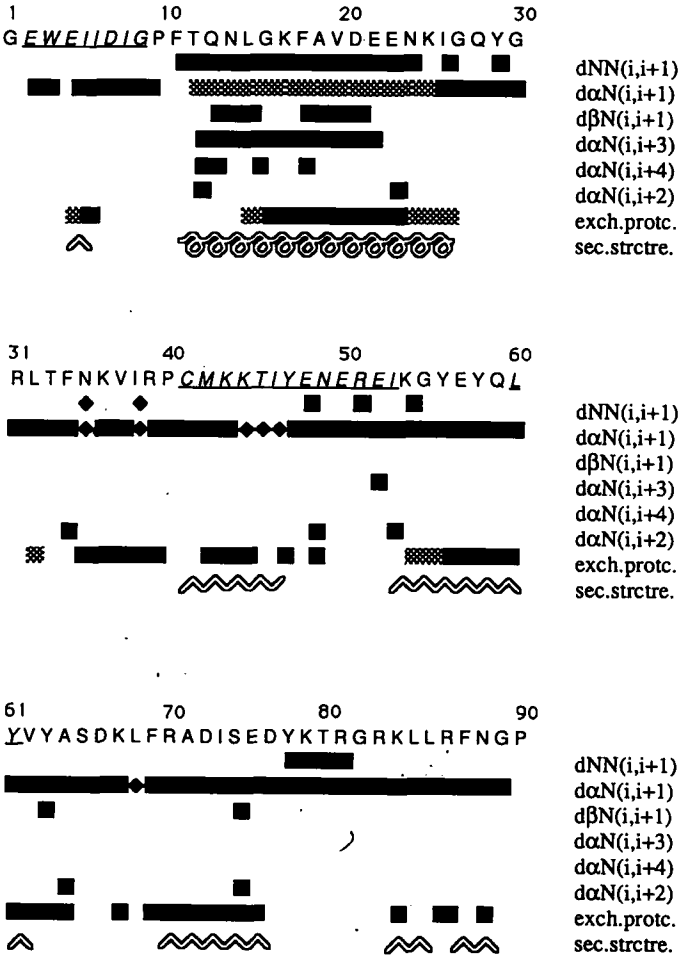

Fig. 1. Sequence of the single-chain sweet protein (SCM) used in this study: amino acids are represented with standard one-letter nomenclature. The short-range sequential NOE connectivity pattern determined for HN, Hα and Hβ in SCM is shown. Weak connections are indicated with 'diamonds'. Regions of regular secondary structure are also indicated; a coil indicates the position of the α-helix, and a 'zig-zag' line indicates a β-strand. The 22 residues where cross peaks show chemical-shift doubling are underlined and italicized (see text for explanation). Residues protected from exchange are indicated; gray bars are used to represent residues weakly protected from exchange. Gray bars are also used in the helix to show weak $H\alpha_i$-HN_{i+1} connectivity. The carboxy-terminal VPPP sequence (residues 91 through 94) is not shown in the figure.

initial structures with the X-PLOR program (Brünger, 1990) which are similar in topology to the monellin crystal structure. Determination of the solution structure of SCM with X-PLOR is in progress.

RESULTS

Spin-system identification

The first step in the assignment of the SCM proton spectra was the identification of spin systems (Wüthrich, 1986) determined from COSY, RCOSY, DQF-COSY and TOCSY spectra. Some of the relayed connectivities were not observed because they originate from broad resonances or involve small scalar coupling constants. These connections were inferred from a combination of DQF-COSY and NOESY spectra. Because of overlap of cross peaks in the α-proton (Hα) and in the $\omega_1 = 2.5$ to 1.0 ppm regions, we used relayed peaks to the backbone amide proton for determining some of our spin systems (Chazin and Wright, 1987). This strategy was shown to be very effective for spin-system assignments of French Bean Plastocyanin (Chazin et al., 1988), which is

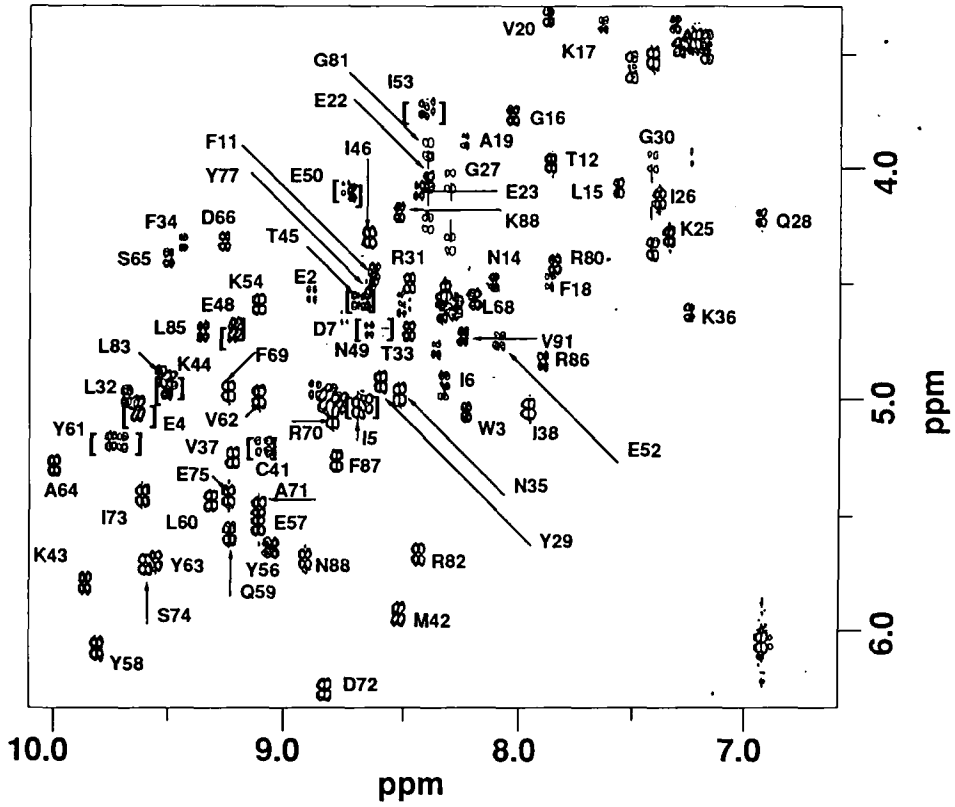

Fig. 2. SCUBA DQF-COSY data obtained at 600 MHz and 45 °C. The HN-Hα cross-peak region for the SCM protein is labeled to show the sequentially determined resonance assignments. Bracketed residues are the most significantly doubled. Hα and Hα′ cross peaks are indicated for some glycines (*e.g.*, G27, G30, G81) with vertical dashed lines. Not all the assigned cross peaks are labeled.

TABLE I CHEMICAL SHIFT POSITION OF ASSIGNED PROTON RESONANCES FOR SCM[a]

Residue	Amide	Alpha	Beta	Gamma	Delta	Epsilon	Eta
E 2	8.24	4.44	1.83/1.53	3.19/"	-	-	-
E' 2	8.82	4.34	?	?	-	-	-
W 3	8.29	4.79	3.14/"	-	7.35	10.10	-
W' 3	8.18	4.90	3.16/"	-	7.42	10.13	-
E 4	9.52	4.88	2.20/2.12	2.46/"	-	-	-
E' 4	9.54	4.91	2.20/2.12	2.46/"	-	-	-
I 5	8.60	4.89	2.11	1.03/?	?	-	-
I' 5	8.58	4.87	2.11	1.03/?	?	-	-
I 6	8.24	4.75	2.82	2.61 m/2.69/"	2.58m	-	-
D 7	8.68	4.45	1.79/"	-	-	-	-
I 8	8.45	4.40	?	?/?	?	-	-
G 9	8.66	4.16/4.08	-	-	-	-	-
F 11	8.58	4.27	3.29/3.12	-	7.15/"	?	?
T 12	7.75	3.76	4.38	1.50m	-	-	-
Q 13	8.54	4.25	2.21/"	2.63/"	-	-	-
N 14	8.03	4.30	2.85/2.69	-	-	-	-
L 15	7.43	3.86	1.87/"	1.08	0.19m/-0.05m	-	-
G 16	7.90	3.53/3.53	-	-	-	-	-
K 17	7.53	3.10	1.22/1.18	0.9/"	1.63/1.61	3.07/"	7.22n
F 18	7.74	4.30	3.28/"	-	7.22/"	6.93/"	?
A 19	8.15	3.66	0.97m	-	-	-	-
V 20	7.80	3.06	2.13	1.19m/0.81m	-	-	-
D 21	8.55	4.31	2.83/2.59	-	-	-	-
E 22	8.31	3.81	0.94/"	1.66/1.54	-	-	-
E 23	8.38	3.88	2.16/"	2.36/"	-	-	-
N 24	8.81	4.63	2.98/2.71	-	-	-	-
K 25	7.21	3.95	?	?	?	?	?
I 26	7.25	3.90	1.93	1.03m/0.87/"	1.14m	-	-
G 27	8.20	4.13/3.83	-	-	-	-	-
Q 28	6.82	3.98	1.76/"	2.03/1.88	-	-	-
Y 29	8.51	4.76	3.50/2.37	-	6.88/"	6.60/"	-
G 30	7.31	4.14/3.76	-	-	-	-	-
R 31	8.44	4.30	1.80/1.66	1.48/"	3.17/"	7.15	-
L 32	9.63	4.84	1.74/"	1.57	1.00m/0.28m	-	-
T 33	8.40	4.52	4.01	1.25m	-	-	-
F 34	9.40	4.11	3.09/3.00	-	7.06/"	7.23/"	7.70
N 35	8.43	4.81	2.80/2.59	-	-	-	-
K 36	7.14	4.44	?	?	-	-	-
V 37	9.20	5.12	2.08	1.35m/1.26m	-	-	-
I 38	7.84	4.89	2.11	2.20m/1.11/"	1.01m	-	-
C 41	9.12	5.04	3.03/2.88	2.08	-	-	-
C' 41	9.10	5.04	3.03/2.86	2.10	-	-	-
M 42	8.37	5.86	1.92/"	2.48/2.26	-	?	-
M' 42	8.35	5.84	1.92/"	2.48/2.26	-	?	-
K 43	9.78	5.69	1.77/1.70	1.10/0.88	1.45/"	2.57/2.41	5.69n
K 44	9.41	4.79	?	?	?	?	?
K' 44	9.38	4.74	?	?	?	?	?
T 45	8.60	4.39	3.86	0.58m	-	-	-
T' 45	8.64	4.39	3.86	0.58m	-	-	-
I 46	8.56	4.08	1.43	0.78m/1.18/"	0.62m	-	-
Y 47	8.80	4.85	2.98/2.64	-	6.89/"	6.64/"	-
Y' 47	8.84	4.79	2.98/2.64	-	6.89/"	6.64/"	-
E 48	9.13	4.50	1.91/"	2.3/"	-	-	-
E' 48	9.15	4.53	1.92/"	2.3/"	-	-	-
N 49	8.57	4.51	3.07/2.98	-	-	-	-

TABLE 1 (continued)

Residue	Amide	Alpha	Beta	Gamma	Delta	Epsilon	Eta
N' 49	8.49	4.53	3.12/3.03	-	-	-	-
E 50	8.67	3.85	2.17/2.11	2.37/2.34	-	-	-
E' 50	8.64	3.89	2.17/2.11	2.37/2.34	-	-	-
R 51	8.23	4.32	1.96/1.81	1.57/"	3.17/"	7.15	-
E 52	7.98	4.57	2.00/"	2.40/2.27	-	-	-
E' 52	8.00	4.55	2.00/"	2.40/2.27	-	-	-
I 53	8.36	3.47	0.67	0.11m/0.33/"	0.54m	-	-
I' 53	8.32	3.44	0.63	0.10m/0.31/"	0.54m	-	-
K 54	9.04	4.39	1.38/"	1.48/"	2.96/"	?	-
G 55	7.57	4.65/3.55	-	-	-	-	-
Y 56	8.99	5.54	2.61/2.45	-	7.06/"	6.85/"	-
E 57	9.01	5.42	1.90/1.80	2.43/"	-	-	-
Y 58	9.70	6.01	2.92/2.76	-	6.95/"	?	-
Q 59	9.14	5.74	1.61/1.58	2.03/"	-	-	-
L 60	9.27	5.32	1.74/1.65	1.78	0.95m/0.85m	-	-
L' 60	9.26	5.31	1.74/1.65	1.78	0.95m/0.85m	-	-
Y 61	9.72	5.02	2.85/2.63	-	7.02/"	6.69/"	-
Y' 61	9.67	5.02	2.85/2.63	-	7.02/"	6.69/"	-
V 62	9.16	4.80	1.89	1.02m/?"m	-	-	-
Y 63	9.46	5.59	3.18/2.78	-	6.74/"	6.76/"	-
A 64	9.90	5.14	1.41m	-	-	-	-
S 65	9.44	4.20	4.34/"	-	-	-	-
D 66	9.22	4.11	3.06/"	-	-	-	-
K 67	8.27	4.61	1.95/"	1.44/"	1.20/"	?	?
L 68	8.12	4.35	1.99/"	1.15	0.61m/0.13m	-	-
F 69	9.01	4.86	2.04/1.76	-	6.72/"	6.90/"	?
R 70	8.70	4.94	1.67/1.31	1.07/"	2.71/2.59	6.61	-
A 71	8.97	5.36	1.26m	-	-	-	-
D 72	8.74	6.20	2.94/2.67	-	-	-	-
I 73	9.50	5.30	1.89	1.07m/1.32/"	?	-	-
S 74	9.52	5.60	4.14/3.86	-	-	-	-
E 75	9.09	5.28	2.33/"	2.12/"	-	-	-
D 76	8.70	4.79	2.77/"	-	-	-	-
Y 77	8.58	4.19	2.73/2.43	-	6.97/"	6.90/"	-
K 78	8.44	3.97	2.06/1.96	1.5/"	1.81/1.72	3.12	7.74n
T 79	8.26	4.39	3.91	?	-	-	-
R 80	7.72	4.22	2.19/2.11	1.63/1.57	3.17/"	7.09	-
G 81	8.32	4.01/3.62	-	-	-	-	-
R 82	8.40	5.57	1.04/0.91	1.22/"	2.50/2.29	7.32	-
K 83	8.69	4.84	1.86/"	1.46/"	1.71/"	3.02/"	7.54n
L 84	9.49	4.74	2.16/"	?	?/?	-	-
L 85	9.27	4.53	1.78/"	1.53	1.18m/0.90m	-	-
R 86	7.77	4.66	2.00/1.90	1.69/"	3.26/"	7.33	-
F 87	8.72	5.13	3.38/2.76	-	6.82/"	5.95/"	6.13
N 88	8.81	5.59	2.81/2.68	-	-	-	-
G 89	8.20	4.36/3.61	-	-	-	-	-
V 91	8.15	4.59	2.01	0.98m/?"m	-	-	-

[a] Residues that show resonance doubling are grouped and underlined; m indicates a methyl resonance, geminal protons are separated by a /, if degenerate by ', and a ? indicates a resonance that is not yet assigned. Residues in the first strand are underlined because of small chemical-shift doubling for either the alpha or amide protons. Prolines are not assigned yet and are excluded from the table. Chemical shifts are from a sample in H_2O at 35 °C and pH 3.5, and are relative to an internal TSP standard (20 mM phosphate buffer, see methods).

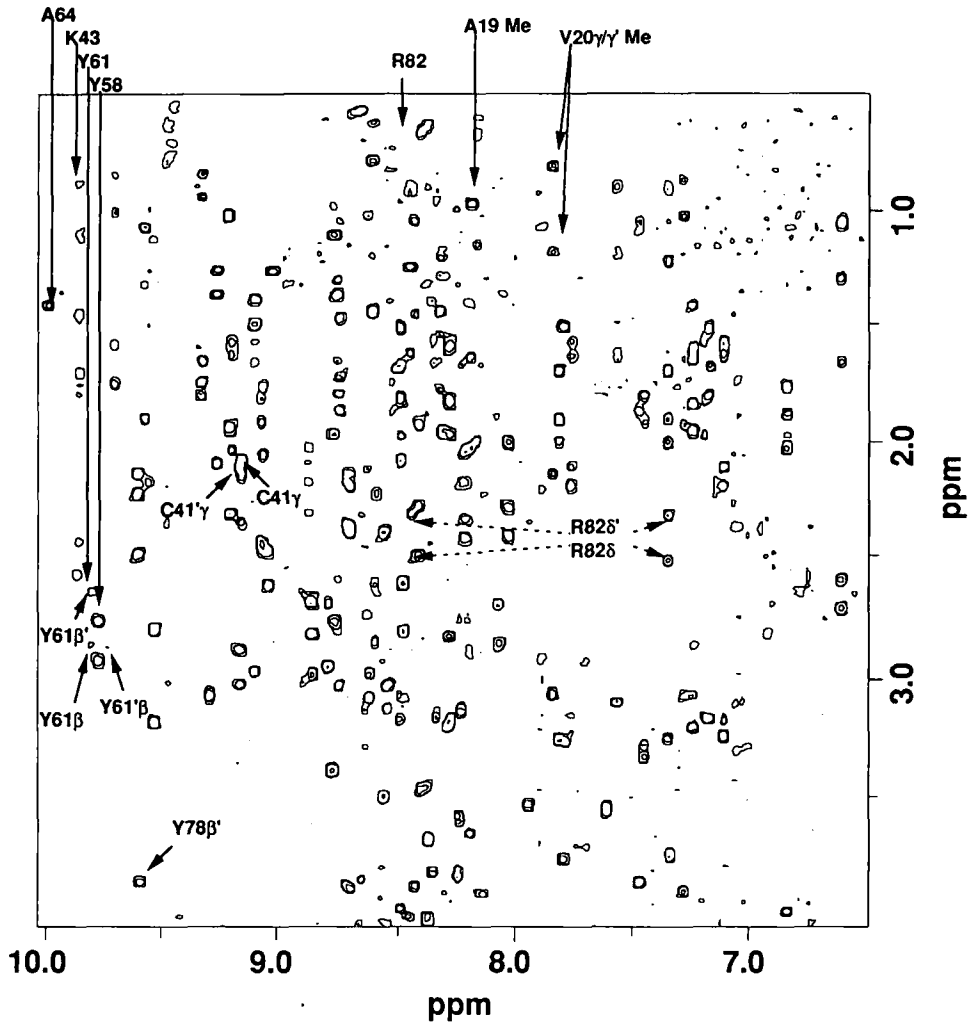

Fig. 3. Symmetrized TOCSY amide to side-chain protons relay region for SCM taken at 35 °C and 70 ms mixing time. Positions of several HN chemical shifts for the second β-strand are indicated, and the chemical shifts of the Y62(HN)–Y62(Hβ/Hβ') and Y62'(HN)–Y62'(Hβ/Hβ') cross peaks are labeled. Also indicated are the C41 and C41' HN relay peaks to the Hγ. Relay peaks to the R82(Hδ) from both the R82(HN) and the R82(Hε) chemical shifts are indicated with dashed lines; relay peaks to the methyl protons of both A19 and V20 from their respective amide protons are also indicated.

similar in size to SCM. Assigned chemical-shift positions are shown in Table 1 for most amino acids (except prolines which are in regions of high spectral overlap).

Sequential assignment

Identified spin systems were connected into groups of four to five spin systems using sequential NOE cross peaks (Wüthrich, 1986) and unambiguously placed in the SCM primary sequence. Sequential assignments were confirmed with the main-chain directed methods (Englander and Wand, 1987). The primary sequence of SCM is shown in Fig. 1. Early in the assignment process

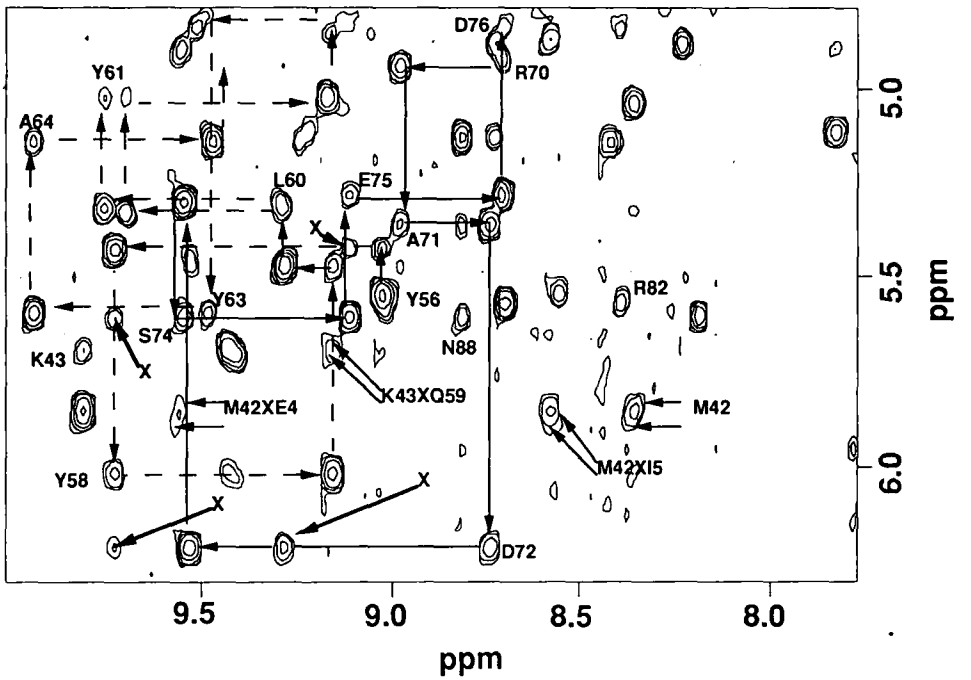

Fig. 4. A section of the HN to downfield Hα NOESY cross-peak region is shown, collected in water at 600 MHz, 35 °C, with a 150-ms mixing time. The third and fourth β-strand sequential connectivities that were observed in this region are indicated with dashed and solid lines, respectively. HN- Hα cross peaks between these two antiparallel strands are indicated with an **X**. Intraresidue NOE cross peaks observed in this region are labeled with a standard one-letter code and numbered according to their position in the primary sequence.

we noted that there was more than one NH and/or αH chemical-shift position for some residues, which considerably complicated the sequential assignment and structure determination. In some cases, chemical-shift doubling was also observed for protons in the side chain (Table 1).

By systematically collecting COSY and NOESY data at different temperatures and pH, we were able to resolve most of the overlap in the spectra and to determine the sequential connectivities of SCM. Assigned Hα-HN cross peaks for SCM are shown in Fig. 2. Chemical-shift doubling was localized to the first and second β-strands, and the newly introduced connecting loop. An example of the chemical-shift doubling observed for a HN to Hβ cross peak in a 2D-TOCSY experiment is shown in Fig. 3. Doubling in the NOESY cross peaks was also observed for the characteristic β-sheet cross peaks shown for residue M42 in Fig. 4. Since resonance doubling can arise from either a mixture of different amino acid sequences, or from conformational equilibria, additional chemical methods were used to analyze the sample.

Identification of N-terminal sequence heterogeneity

Isoelectric focusing gel electrophoresis of the NMR sample separated two proteins with slightly different isoelectric points in the pI range of 9.3 to 9.7 (data not shown). Cation exchange FPLC of the NMR sample, under non-denaturing conditions, separated two closely eluting proteins with a peak area ratio of approximately 2 to 1 (data not shown).

One-dimensional NMR (1D-NMR) spectra collected for the FPLC purified protein samples showed only one set of resonances in the resolvable downfield amide proton region (9 to 10 ppm). Moreover, we did not observe chemical exchange cross peaks, or re-equilibration of cross-peak intensities at any of the temperatures or concentrations used in our studies. We therefore have no indication of dimerization or alternative structures for SCM in solution.

Amino-terminal sequencing of the FPLC-purified samples revealed two distinct amino-terminal sequences for the NMR sample of SCM used in these studies. The larger peak, which eluted first, was found to have the amino-terminal sequence expected for SCM: G1-E2-W3-E4-I5-. The smaller and broader peak eluting later in the gradient (SCM′) was found to have three additional amino acids at the amino-terminus: M1′-K2′-D3′-G1-E2-W3-E4-I5-. These sequencing results support our conclusion that the chemical-shift doubling observed for the residues in the first two β-strands is attributable to a sequence difference at the amino-terminus of the protein. The synthetic DNA used to obtain the clone that produced this protein does not contain an additional upstream promoter, so it is not clear at this time how the extra three amino-terminal residues observed in SCM′ arise. However, these results show that sequence heterogeneity can occur even for recombinant expressed proteins.

Secondary structure assignment

Regions of antiparallel β-sheet were identified by their characteristic sequential and interstrand connectivities (Wüthrich, 1986; Englander and Wand, 1987). Strong sequential $H\alpha_i$-HN_{i+1} NOEs, expected for an extended β-strand structure, were observed throughout the sheet (Fig. 4). Additional cross-strand NOEs characteristic of an antiparallel β-sheet structure were also observed: strong Hα-Hα and medium Hβ-Hβ and HN-H_α NOEs. The expected pattern of slowly exchanging amide protons was observed for the core of the antiparallel sheet structure. Residues protected from exchange are indicated in Fig. 1. The secondary structure of the β-sheet determined from NMR-derived constraints is shown schematically in Fig. 5. The core of the β-sheet structure was composed of residues C41–Y47 in the second strand, K54–V62 in the third strand, and F69–E75 in the fourth strand (Fig. 5). The fifth strand started at residue R82 and appeared to be regular through residue L85, then R86 formed a β-bulge (indicated by a $H\alpha_i$-HN_{i+2} NOE), and the strand continued for another three residues (F87 through G89).

Sequential NOEs were weak for the first 10 amino-terminal residues in the first β-strand, suggesting that this region is irregular in solution. We were able to identify the AMPTX-AMX-AMPTX (Wüthrich, 1986) spin systems for residues E2, W3 and E4. W3 was identified by weak Hα and strong Hβ NOEs to the ring Hδ; W3(Hδ) was clearly identified by a strong NOE cross peak to the downfield shifted tryptophan imino proton (Fig. 6). The spin system for I6 was clearly identified by the TOCSY experiment (Fig. 3), and through sequential connectivities we were able to trace the strand to G9 (Fig. 2). Register of this strand was determined from a strong Hα-Hα NOE between position 5 in the first strand and M42 in the second strand. The antiparallel orientation of the first strand was inferred from the NOEs observed from W3(Hδ) to the Hα of residues K43 and K44. We also observed a strong NOE between E4 (HN) and M42 (Hα). This NOE was not characteristic of a canonical antiparallel β-strand structure and indicated that the first strand may not form part of the regular β-sheet in solution. The amide of the residue in the fifth position further indicated the structural irregularity of the first strand, as the amide was weakly protected from exchange. We did not observe the extensive exchange protection of the amide protons of the

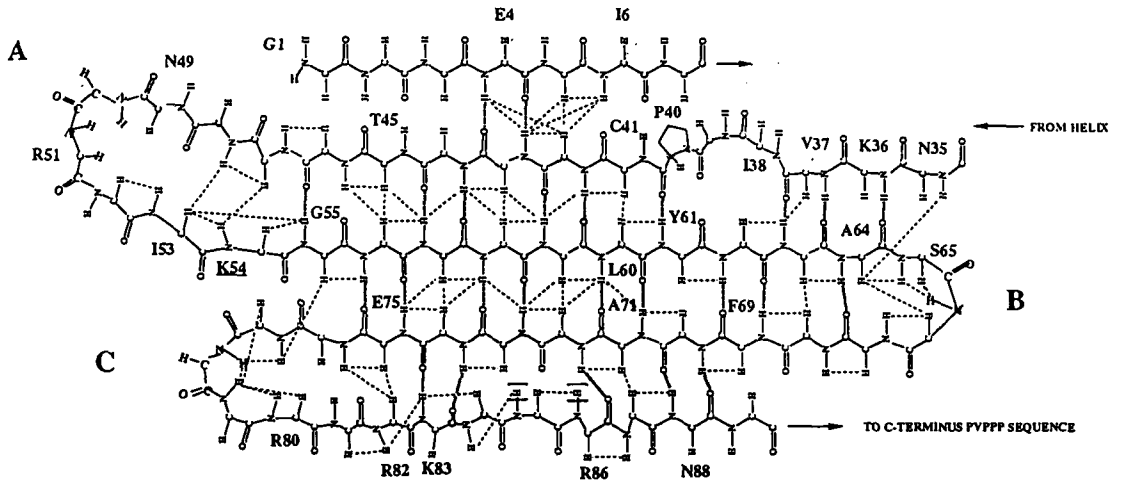

Fig. 5. Schematic drawing of the β-sheet with some of the observed NOE connectivities determined for SCM. A is the engineered loop between residue 45 of the B chain and residue 2 of the A chain of monellin. Loop B is located between the third and fourth antiparallel β-strands, and loop C connects the fourth and fifth β-strands. The helix and the loop-back to the β-sheet are not shown. Hydrogen bonds determined by D_2O exchange data are indicated by hatched lines, and proton-proton NOEs are indicated by dashed lines. Amide protons that are protected from exchange that could not be unambiguously assigned to a hydrogen-bond acceptor are bracketed. (Not all determined cross-peak connections are shown in this figure).

first strand that would be predicted from the regularized hydrogen bonding determined for the crystal structure of native monellin (Jiang, 1990; Jiang et al., unpublished results; Somoza et al., unpublished results).

The presence of a helix encompassing residues 11 to 26 was clearly indicated by the characteristic HN_i-HN_{i+1}, $H\beta_i$-HN_{i+1} and $H\alpha_i$-HN_{i+3} NOE connectivities. In Fig. 6 the expanded downfield region of a 150-ms NOESY is annotated to show some of the sequential HN-HN cross peaks characteristic of the helix. Most of the expected $H\alpha_i$-HN_{i+4} cross peaks were also observed for residues 11 through 22. Sequential NOEs ($H\alpha_i$-HN_{i+1}) were observed for several residues: L15 and G16, K17 and F18, N24, K25 and I26. In the 150-ms NOESY, the termini of the helix also showed weak $H\alpha_i$-HN_{i+2} cross peaks: between T12(Hα) and N14(HN), and between E23(Hα) and K25(HN) (data not shown). We did not see any obvious spin-diffusion pathways for these cross peaks at this mixing time, and since we did not observe similar peaks for the rest of the α-helix, we believe that these NOEs are not due to spin diffusion. Observation of these connectivities is unusual for regular α-helix structures (Wüthrich, 1986). Twelve helical amides were protected from exchange with the solvent (Fig. 1): the amide protons of N24, K25 and I26 were weakly protected and completely exchanged with solvent in 12 to 24 h. Our data suggest some fraying at the carboxy-terminus of the SCM helix and could explain the discrepancy in helix length between the solution and crystal structures.

During the sequential and secondary structure assignment, we were able to identify the thiol, C41(Hγ), in both COSY (data not shown) and TOCSY (Fig. 3) experiments. To our knowledge, this proton has not been observed previously in proteins, possibly because cysteines are often exposed to the solvent and the Hγ proton can exchange. From our secondary structure analysis, the

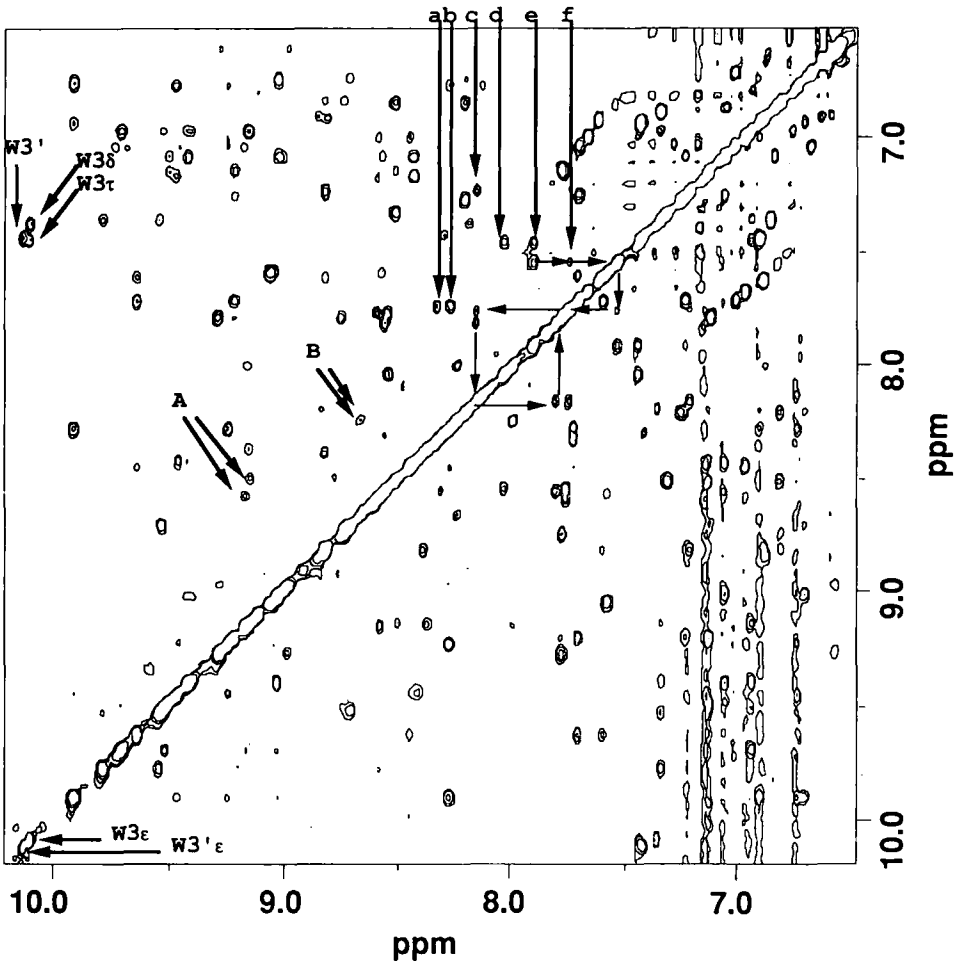

Fig. 6. Expanded plot of the HN-HN cross-peak region of 150-ms NOESY at 35 °C showing helical HN_i-HN_{i+1} NOESY cross peaks, and HN-HN cross peaks observed for the loop between the second and third β-strands. Chemical shifts of W3(Hε) and W'3(Hε) and cross peaks indicating the position of W3(Hδ) for each of the sequence variants are indicated (see text for a complete explanation); W3(Hε)-W3(Hζ) cross peak is also indicated. The cross peak from E48(HN) to N49(HN) is indicated for both sequences with an uppercase A. B indicates the position of E50(HN) to E52(HN) cross peak. These cross peaks illustrate that the chemical-shift separation of the doubled residues is variable (see also Table 1). NOEs of bulged residues in the fifth β-strand are indicated with lower case letters, a and b, and represent G81(HN) to R80(HN) and T79(HN) to R80(HN) cross peaks, respectively. Lower case letters, c, d, e and f, indicate the chemical shifts of helical residues; c = A19(HN), d = N14(HN), e = G16(HN), f = F18(HN). Arrows on either side of the diagonal indicate the HN_i-HN_{i+1} cross-peak connectivity that is observed for G16, K17, F18, A19, and V20.

side chain of C41 was on the same side of the β-sheet as the helix, whereas according to the protein topology of SCM (discussed below) and the crystal structure of natural monellin (Ogata et al., 1987), it is under the helix and probably protected from exchange with the solvent.

Protein topology

A schematic drawing of the SCM structure is shown in Fig. 7, with some residues involved in

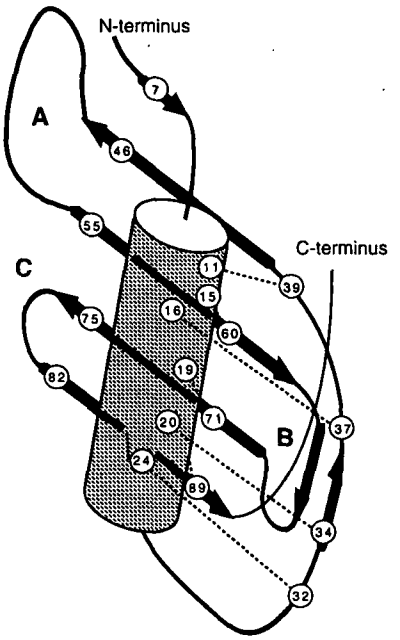

Fig. 7. Schematic representation of the topology determined for SCM (see text for explanation). The approximate location of the α-helix is indicated by a shaded cylinder; loops are labeled as in Fig. 5. Residues are labeled as in Fig. 1 and the approximate location of some of the helix-sheet NOEs is indicated with dashed lines (see text for explanation). The location of the carboxyl-terminal PVPPP sequence (residues 90 through 94) has not been determined and may interact with other parts of the protein.

tertiary contacts indicated. These are discussed in subsequent paragraphs. The SCM helix has some amphiphilic character. Residues F11, L15 and A19 are on the same side of the α-helix and interacted with the area of the SCM β-sheet that was also predominantly hydrophobic. Residue L15 in the helix interacted with a hydrophobic area of the sheet which included residues Y58 and L60. NOE connectivities indicated that residues F69, A71 and F87 in the β-sheet formed a hydrophobic 'pocket' that interacted with the helical residues A19 and V20 (data not shown). The β-bulge at R86 in the fifth β-strand appears to be necessary to allow the ring of F87 to orient on the side of the β-sheet that is interacting with the predominantly hydrophobic side of the helix. We also observed cross peaks from the side chain of V20 to the side chain of F34 (see discussion below). According to our structural determination, at least one turn of the helix would be accessible to the solvent; this turn would include residues E23, N24, K25, and I26.

The connecting loop between the helix and the second β-strand (here referred to as the 'loop-back') encompassed residues 26 to 40, and included several amide protons that were protected from exchange: N35, K36, V37, and I38. We also observed NOEs from the Hα and HN protons in the loop-back region to backbone (Hα and HN) protons of residues Y61 through S65 (data not shown). Several of these NOEs were consistent with the formation of a short anti-parallel β-strand between residues N35 to I38 of this sequence and Y63 to S65 in the third β-strand. The cross peaks identified suggest that the β-sheet in this region is not regular, and indicated that the

Loop B (Fig. 5) may be twisted toward the fifth β-strand in order to increase contacts with the loop-back sequence.

A strong ($H\alpha_i$-HN_{i+2}) NOE between F34 and K36 indicated that the loop-back region has a kink at residue N35. We also identified several NOESY cross peaks between F34 and resides G16, K17, and V20 in the helix (data not shown). The extensive protection observed for residues N35 to R39 supports the idea that part of this sequence forms a antiparallel β-sheet with residues 61 to 65. Residues 33 through 39 must be at least partially buried by the α-helix, which protects some of the amide protons (those not forming antiparallel β-strand contacts) from exchange.

Analysis of the cross peaks for the backbone protons in the core of the sheet showed some weak NOEs in the 150-ms NOESY which were unusual for a regular antiparallel β-sheet. These were between the third and fourth β-strands: R82(Hα) to S74(HN), and E57(Hα) to I73(HN). A similar backbone NOE was also observed between the second and third β-strands: E57(HN) to K43(Hα). These long-range cross peaks are observed when the β-sheet is slightly distorted or curved. In spite of the size of this protein, these cross peaks were probably not due to spin diffusion at this mixing time. Since we observed very regular hydrogen protection and cross strand NOEs for the core of the sheet, the most likely explanation for these NOEs is that the sheet is curved. An indication of which side the sheet may curve toward came from the analysis of the side chain of R82. The Hε and Hη protons of this side chain showed well-resolved NOE cross peaks to the Hβ of F18, and to the methyl groups of L15 in the helix, suggesting that the β-sheet curves around the helix.

We were not able unambiguously to assign NOEs that could clearly position either the engineered loop (Loop A) or Loop C (Fig. 5) relative to the rest of the protein. Lack of constraints and protected residues, as well as the weak intensities of cross peaks observed for some residues, in these loops indicated that they may be flexible and are accessible to solvent. The pattern of observed NOEs was similar for both chemical-shift positions for the newly introduced Loop A. We interpreted this as an indication that the engineered loop was near the heterogeneous amino-terminus, and that this proximity resulted in the multiple chemical shifts observed in this region for the sequentially equivalent residues.

DISCUSSION

In designing SCM, the primary goal was to obtain a protein with structural and biochemical properties similar to those of natural two-peptide monellin. Our NMR data indicate that the major elements of the secondary structure of SCM, and the general tertiary topology are indeed very similar to those observed in the refined natural monellin crystal structure (Jiang et al., unpublished results), a five-stranded antiparallel β-sheet curving around an α-helix.

In solution, however, we found no evidence for the extensive hydrogen bonding between the first and second strands that was observed in the crystal structure of monellin. It is possible that the introduction of the engineered loop (Loop A) disrupted the interaction between these two strands, or that the contacts between monomers in the crystal were stabilizing this interaction. Additionally, we did not observe any evidence of exchange cross peaks that would be indicative of monomer-dimer equilibria or of slow-exchange of conformational isomers. We are in the process of further refining the conformation of the first strand of SCM and other mutant proteins.

Tertiary NOE constraints showed that the hydrophobic side of the amphiphilic helix (encompassing residues 11 through 26) interacted with the hydrophobic region of the β-sheet. The car-

boxyl-terminus of the helix was shown to extend to residue I26, and the position of the helix relative to the sheet indicated that the last turn of the helix extends beyond the sheet. This analysis is in agreement with the crystal structure of natural monellin (Ogata et al., 1987; Jiang et al., unpublished results). Our results indicate that the helix in SCM is 15 residues in length, as compared to the 16 residues of the crystal structure.

Several long-range and backbone NOEs indicated that the β-sheet was curved around the helix and that some residues near Loop C were on the same side of the β-sheet as the helix. These results are in agreement with the refined crystal structures (Jiang et al., unpublished results; Somoza et al., unpublished results). The hydrogen bonding and strand register determined for the fifth strand of SCM in solution indicated that a β-bulge helped to form a hydrophobic pocket and allowed the F87 side chain to interact with the helix. We propose that elimination of R86 would not affect the overall folding of SCM significantly, and that this mutation may be useful for evaluating the effect of this region on sweet taste. Furthermore, NMR structural assignments provide a starting point for studying the structure–taste relationship of many point-mutants of SCM which have already been constructed and are being characterized.

CONCLUSION

From the results presented in this paper it is clear that the heterogeneity observed for the sample used arose from the presence of two sequences. In theory, the assignments and the structure of SCM could have been obtained more easily with isotopically labeled samples; however, in practice we found it difficult to obtain isotopically labeled samples in significant quantities.

The present NMR data clearly showed that the overall fold of SCM is the same as that of monellin, and that the introduction of the engineered loop did not affect the stability of the secondary structure of the second and third antiparallel β-strands. There are indications of minor differences between the solution and crystal structures, which will be resolved when the fully refined solution structure of SCM is available. Work is also in progress to determine the solvent-exchange rates and folding of SCM and mutant proteins.

ACKNOWLEDGEMENTS

The authors wish to thank Dr. Rosie Kim for her assistance with some of the gels and Dr. Vince Powers for his help in setting up the FPLC. This work was supported by the Director, Office of Energy Research, Office of Biological and Environmental Research, General Life Sciences Division of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098; through instrumentation grants from the U.S. Department of Energy, DE-FG05-86ER75281, and the National Science Foundation, DMB 86-09305 and BBS 87-20134, to David E. Wemmer; and through National Institute of Health grant DC 00145 to Sung-Hou Kim.

REFERENCES

Bax, A. and Davis, D.G. (1985) *J. Magn. Reson.*, **65**, 355–360.
Braunschweiler, L. and Ernst, R.R. (1983) *J. Magn. Reson.*, **53**, 521–528.

Brown, S.C., Webber, P. and Mueller, L. (1988) *J. Magn. Reson.*, **77**, 166–169.
Brünger, A.T. (1990) In *X-PLOR, A System for Crystallography* and *NMR*, Yale University, New Haven, CT, p. 255.
Chazin, W.J. and Wright, P.E. (1987) *Biopolymers*, **26**, 973–977.
Chazin, W.J. and Wüthrich, K. (1987) *J. Magn. Reson.*, **72**, 358–363.
Chazin, W.J., Rance, M. and Wright, P.E. (1988) *J. Mol. Biol.*, **202**, 603–622.
Englander, S.W. and Wand, A.J. (1987) *Biochemistry*, **26**, 5958–5962.
Jiang, F. (1990) Ph.D. thesis, University of California, Berkeley.
Kang, C.-H. (1988) Ph.D. thesis, University of California, Berkeley.
Kim, S.-H., Kang, C.-H., Kim, R., Cho, J.M. and Lee, T.-K. (1989) *Prot. Eng.*, **2**, 571–575.
Morris, J.A. and Cagan, R.H. (1972) *Biochim. Biophys. Acta*, **261**, 114–122.
Ogata, C., Hatada, M., Tomlinson, G., Shin, W.-C. and Kim, S.-H. (1987) *Nature*, **328**, 739–742.
Piantini, U., Sørensen, O.W. and Ernst, R.R. (1982) *J. Am. Chem. Soc.*, **104**, 6800–6801.
Plateau, P. and Guéron, M. (1982) *J. Am. Chem. Soc.*, **104**, 7310–7311.
Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R. and Wüthrich, K. (1983) *Biochem. Biophys. Res Comm.*, **117**, 479–485.
Reisfeld, R.A., Lewis, U.J. and Williams, D.E. (1962) *Nature*, **195**, 281–283.
Shaka', A.J. and Freeman, R. (1983) *J. Magn. Reson.*, **51**, 169–173.
Wagner, G. (1983) *J. Magn. Reson.*, **55**, 151–156.
Wider, G., Macura, S., Anil-Kumar, Ernst, R.R. and Wüthrich, K. (1984) *J. Magn. Reson.*, **56**, 207–234.
Wüthrich, K., Wider, G., Wagner, G. and Braun, W. (1982) *J. Mol. Biol.*, **155**, 311–319.
Wüthrich, K., Billeter, M. and Braun, W. (1984) *J. Mol. Biol.*, **180**, 715–740.
Wüthrich, K. (1986) in *NMR* of *Proteins and Nucleic Acids*, Wiley-Interscience, New York, NY, pp. 130.