Unusual Helix-Containing Greek Keys in Development-Specific Ca^{2+}-Binding Protein S. ^{1}H, ^{15}N, and ^{13}C Assignments and Secondary Structure Determined with the Use of Multidimensional Double and Triple Resonance Heteronuclear NMR Spectroscopy†

Stefan Bagby,‡ Timothy S. Harvey,‡ Lewis E. Kay,§ Susan G. Eagle,∥ Sumiko Inouye,∥ and Mitsuhiko Ikura*,‡

Division of Molecular and Structural Biology, Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, 500 Sherbourne Street, Toronto, Ontario, Canada M4X 1K9, Protein Engineering Network Centres of Excellence and Departments of Medical Genetics, Biochemistry and Chemistry, University of Toronto, Toronto, Ontario, Canada M5S 1A8, and Department of Biochemistry, Robert Wood Johnson Medical School, Piscataway, New Jersey 08854

Received October 29, 1993; Revised Manuscript Received December 14, 1993®

ABSTRACT: Multidimensional heteronuclear NMR spectroscopy has been used to determine almost complete backbone and side-chain ^{1}H, ^{15}N, and ^{13}C resonance assignments of calcium loaded *Myxococcus xanthus* protein S (173 residues). Of the range of constant-time triple resonance experiments recorded, HNCACB and CBCA(CO)NH, which correlate C^{α} and C^{β} with backbone amide resonances of the same and the succeeding residue respectively, proved particularly useful in resolving assignment ambiguities created by the 4-fold internal homology of the protein S amino acid sequence. Extensive side-chain ^{1}H and ^{13}C assignments have been obtained by analysis of HCCH-TOCSY and ^{15}N-edited TOCSY-HMQC spectra. A combination of NOE, backbone amide proton exchange, $^{3}J_{NH\alpha}$ coupling constant, and chemical shift data has been used to show that each of the protein S repeat units consists of four β-strands in a Greek key arrangement. Two of the Greek keys contain a regular α-helix between the third and fourth strands, resulting in an unusual and possibly unique variation on this common folding motif. Despite similarity between two nine-residue stretches in the first and third domains of protein S and one of the Ca^{2+}-binding sequences in bovine brain calmodulin [Inouye, S., Franceschini, T., & Inouye, M. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80,* 6829–6833], the protein S topology in these regions is incompatible with an EF-hand calmodulin-type Ca^{2+}-binding site.

The Gram-negative soil bacterium *Myxococcus xanthus* has a complex life cycle (Kaiser et al., 1979; Zusman, 1984; Rosenberg, 1984; Dworkin & Kaiser, 1985; Shimkets, 1987, 1990) similar to that of eukaryotic slime molds (Kaiser, 1986; Loomis, 1988). When starved of nutrients on a solid surface, cells move toward an aggregation center to form a fruiting body. At this time, about 90% of cells undergo lysis (Wireman & Dworkin, 1977), while the remainder differentiate into dormant, environmentally resistant myxospores. From an early stage in the *M. xanthus* developmental cycle, protein S is synthesized and accumulated as a soluble protein in the cytoplasm (Inouye et al., 1979a,b). At the onset of lysis and sporulation, when it accounts for 15% of total protein synthesized, protein S is released from lysed cells, whence it diffuses freely to the surface of any nearby spores. Molecules of protein S then oligomerize to form a protective spore coat in a process which specifically requires Ca^{2+} (Inouye et al., 1979b, 1981; Teintze et al., 1985). The spore coat may also be involved in spore–spore interaction in the fruiting body (Inouye et al., 1979b). The role of protein S inside the developing *M. xanthus* cells is currently unknown.

Protein S is an acidic (p*I* 4.5), heat-stable Ca^{2+}-binding protein (Teintze et al., 1991). Its gene has been cloned and sequenced (Inouye et al., 1983a,b; Downard et al., 1984). The sequence indicates that protein S can be divided into four consecutive, homologous domains, in a manner reminiscent of other calcium-binding proteins such as calmodulin (Watterson et al., 1980) and troponin C (Collins et al., 1973). One molecule of protein S can bind two calcium ions, with dissociation constants in the range $2.7–7.6 \times 10^{-5}$ M (Teintze et al., 1988). Protein S has two regions with sequences similar to one of the calcium-binding sites of all-helical calmodulin (Inouye et al., 1983a). On the other hand, protein S has 19% identity throughout its sequence (Wistow et al., 1985) to the vertebrate eye lens $\beta\gamma$-crystallins (Wistow & Piatigorsky, 1988), which consist almost entirely of β-type secondary structure (Blundell et al., 1981; Wistow et al., 1983; Bax et al., 1990a; Lapatto et al., 1991). It has been proposed (Wistow, 1990) that protein S is evolutionarily related to the $\beta\gamma$-crystallins, which, like calmodulin and protein S, contain four internally homologous domains.

Although protein S crystals have been obtained (Inouye et al., 1980) both in the presence and absence of Ca^{2+}, no three-dimensional structure is available at present. In order to elucidate the structural relationship of protein S to calmodulin and the $\beta\gamma$-crystallins, the nature of its Ca^{2+}-binding sites and possible modes of multimerization, we have undertaken to characterize in detail the three-dimensional structure in solution of protein S. We report here backbone and side-chain resonance assignments, secondary structure, and topology of the molecule.

† This work was supported in part by grants to M.I. from the Medical Research Council of Canada and the Human Frontier Science Program Organization and by a grant to S.I. from the National Institutes of Health (GM26843). NMR spectrometers were bought with a Terry Fox team development grant from the National Cancer Institute of Canada and with funds from the Princess Margaret Hospital Research Foundation. S.B. and T.S.H. thank HFSPO and NATO for postdoctoral fellowships.

‡ Ontario Cancer Institute and Department of Medical Biophysics.

§ Protein Engineering Network Centres of Excellence and Departments of Medical Genetics, Biochemistry and Chemistry.

∥ Robert Wood Johnson Medical School.

® Abstract published in *Advance ACS Abstracts,* February 15, 1994.

0006-2960/94/0433-2409$04.50/0

EXPERIMENTAL PROCEDURES

Protein Expression, Purification, and NMR Sample Preparation. Protein S was overexpressed in *Escherichia coli* (*E. coli*)[1] by a derivative of the pET11a plasmid carrying a T7 promoter (Studier et al., 1990). Protein purification was carried out by $(NH_4)_2SO_4$ precipitation followed by ion exchange chromatography using PAE1000 (Amicon) and Mono S (Pharmacia). Sample homogeneity was greater than 95% as judged by SDS–PAGE and 2D NMR spectroscopy. Uniformly ^{15}N- and $^{15}N/^{13}C$-labeled protein was obtained by using $^{15}NH_4Cl$ and $^{15}NH_4Cl/[^{13}C_6]$-D-glucose as the sole nitrogen and nitrogen/carbon sources in M9 medium.

Uniformly ^{15}N-labeled protein was dissolved in 95% H_2O/5% D_2O and uniformly $^{15}N/^{13}C$-labeled protein was dissolved in either 95% H_2O/5% D_2O or 99.95% D_2O. NMR samples were adjusted to pH 6.7 without correction for isotope effects and contained 1.6 mM protein S, 10 mM $CaCl_2$, 100 mM KCl, and 0.5 mM NaN_3.

NMR Data Acquisition. All NMR spectra were recorded at 37 °C on a Varian UNITY-600 spectrometer equipped with a triple resonance probe head, unless otherwise stated. Carrier positions used in the various protein S spectra were as follows: ^{15}N, 119.5 ppm; ^{13}CO, 175 ppm; $^{13}C^{\alpha}$, 54 ppm; $^{13}C^{\alpha/\beta}$ 43 ppm; ^{1}H, 4.67 ppm. The 2D ^{15}N-^{1}H HSQC (Bodenhausen & Ruben, 1980) spectrum (Figure 1) was acquired on a UNITY+ 500-MHz spectrometer equipped with a pulsed field gradient (PFG) triple resonance probe, using the enhanced sensitivity method of Kay et al. (1992) which employs PFG selection of coherence transfer pathways. The numbers of complex points and acquisition times were as follows: ^{15}N (F_1) 400, 240 ms, ^{1}H (F_2) 512, 77 ms, 64 transients. The HMQC-J (Kay & Bax, 1990) spectrum was recorded with the following numbers of complex points and acquisition times: ^{15}N (F_1) 256, 128 ms, ^{1}H (F_2) 512, 64 ms, 128 transients.

In order to measure $^{15}N\{^{1}H\}$ NOEs (Kay et al., 1989), two types of spectrum were recorded, with and without the NOE effect. Each of these was recorded twice in a consecutive series, resulting in a total of four spectra. This permits averaging of peak intensities to help reduce errors in the measurements due to spectrometer instability. Each $^{15}N\{^{1}H\}$ NOE spectrum was acquired with the following numbers of complex points and acquisition times: ^{15}N (F_1) 180, 90 ms, ^{1}H (F_2) 512, 64 ms, 32 transients.

Triple resonance 3D spectra (Kay et al., 1990; Grzesiek & Bax, 1992a) correlating backbone amide protons were recorded on the uniformly $^{15}N/^{13}C$-labeled H_2O sample with the following numbers of complex points and acquisition times: HNCO (Kay et al., 1990), ^{15}N (F_1) 32, 22.9 ms, ^{13}CO (F_2) 64, 42.7 ms, ^{1}H (F_3) 512, 64 ms (16 transients); HNCA (Kay et al., 1990), ^{15}N (F_1) 32, 22.9 ms, $^{13}C^{\alpha}$ (F_2) 32, 8 ms, ^{1}H (F_3) 512, 64 ms (64 transients); HCA(CO)N (Kay et al., 1990), ^{15}N (F_1) 32, 21.3 ms, $^{13}C^{\alpha}$ (F_2) 32, 7.1 ms, ^{1}H (F_3) 512, 85 ms (16 transients); HN(CO)CA (Bax & Ikura, 1991) (recorded on a Varian UNITY-500 spectrometer) ^{15}N (F_1) 32, 21.3 ms, $^{13}C^{\alpha}$ (F_2) 32, 9.6 ms, ^{1}H (F_3) 512, 64 ms (32 transients); HACA(CO)NH, ^{15}N (F_1) 32, 22.9 ms, $^{1}H^{\alpha}$ (F_2) 48, 26.7 ms, ^{1}H (F_3) 512, 64 ms (32 transients); HNCACB (Wittekind & Mueller, 1993), $^{13}C^{\alpha/\beta}$ (F_1) 32, 3.5 ms, ^{15}N (F_2) 32, 22.9 ms, ^{1}H (F_3) 512, 64 ms (64 transients); CBCA(CO)NH (Grzesiek & Bax, 1992b), $^{13}C^{\alpha/\beta}$ (F_1) 42, 4.6 ms, ^{15}N (F_2) 32, 22.9 ms, ^{1}H (F_3) 512, 64 ms (64 transients). The ^{15}N chemical shift in all of the above triple resonance experiments was recorded in a constant-time manner (Bax et al., 1979; Rance et al., 1984; Powers et al., 1991; Grzesiek & Bax, 1992a; Palmer et al., 1992) and coherent proton decoupling [using WALTZ-16 (Shaka et al., 1983)] was employed to maintain in-phase heteronuclear magnetization where possible (Grzesiek & Bax, 1992a; Farmer et al., 1992). In the HNCACB experiment, CO decoupling during $^{13}C^{\alpha/\beta}$ chemical shift and $^{13}C^{\alpha/\beta}$-^{15}N scalar coupling evolution periods was achieved using a SEDUCE-1 ^{13}C homonuclear decoupling sequence (McCoy & Mueller, 1992a,b) generated using a separate frequency source centered at 175 ppm.

HCACO (Kay et al., 1990), HCCH-TOCSY (Bax et al., 1990b), and ^{13}C-edited NOESY HMQC (Ikura et al., 1990a) spectra were recorded on the $^{15}N/^{13}C$-enriched D_2O sample with the following numbers of complex points and acquisition times: HCACO, ^{13}CO (F_1) 64, 42.7 ms, $^{13}C^{\alpha}$ (F_2) 32, 7.1 ms, ^{1}H (F_3) 512, 64 ms (16 transients); HCCH-TOCSY, ^{1}H (F_1) 128, 28.4 ms, $^{13}C^{\alpha}$ (F_2) 32, 8.9 ms, ^{1}H (F_3) 416, 52 ms (16 transients); two HCCH-TOCSYs were recorded, with mixing times of 8 ms and 16 ms respectively; ^{13}C-edited NOESY, ^{1}H (F_1) 128, 28.4 ms, $^{13}C^{\alpha}$ (F_2) 32, 8.9 ms, ^{1}H (F_3) 416, 52 ms (16 transients and a 100-ms NOESY mixing time).

Two ^{15}N-edited NOESY-HMQC (Zuiderweg & Fesik, 1989; Marion et al., 1989a) experiments were recorded with NOE mixing times of 50 and 100 ms, each with the following numbers of complex points and acquisition times: ^{1}H (F_1) 128, 19.7 ms, ^{15}N (F_2) 32, 22.9 ms, ^{1}H (F_3) 512, 64 ms (16 transients). A ^{15}N-edited TOCSY-HMQC (Marion et al., 1989b) experiment was recorded with a 44.4 ms mixing time using the DIPSI-2 mixing sequence (Shaka et al., 1988), and the following numbers of complex points and acquisition times: ^{1}H (F_1) 128, 19.7 ms, ^{15}N (F_2) 32, 22.9 ms, ^{1}H (F_3) 512, 64 ms (16 transients).

Solvent suppression was achieved either by selective presaturation of the H_2O resonance (Anil-Kumar et al., 1980; Wider et al., 1983) or by use of spin lock pulses which randomize magnetization of protons not attached to ^{15}N (Messerle et al., 1989), or both.

NMR Data Processing and Analysis. Three-dimensional data sets were processed on Sun Sparc2 and Sparc10 workstations using a combination of software written at National Institutes of Health (Kay et al., 1989) [for linear prediction and Fourier transform in the indirectly detected dimensions] and commercial software (NMR2, New Methods Research Inc., Syracuse, NY). The programs Capp and Pipp (Garrett et al., 1991), were used for peak picking and spectral analysis. Solvent suppression in experiments carried out in water was improved by convolution of time domain data (Marion et al., 1989c).

Two-dimensional data sets were processed using NMR2 (New Methods Research Inc., Syracuse, NY). A 60° phase-shifted sine bell and single zero-fill were applied in each of

[1] Abbreviations: *E. coli*, *Escherichia coli*; 2D, two-dimensional; 3D, three-dimensional HSQC, heteronuclear single quantum coherence; PFG, pulsed field gradient; HMQC, heteronuclear multiple quantum coherence; NOE, nuclear Overhauser effect; HNCO, amide proton to nitrogen to carbonyl carbon correlation; HNCA, amide proton to nitrogen to α-carbon correlation; HCA(CO)N, α-proton to α-carbon (via carbonyl carbon) to nitrogen correlation; HN(CO)CA, amide proton to nitrogen (via carbonyl carbon) to α-carbon correlation; HACA(CO)NH, α-proton to α-carbon (via carbonyl carbon) to nitrogen to amide proton correlation; HNCACB, amide proton to nitrogen to α/β-carbon correlation; CBCA(CO)NH, β-proton to α/β-carbon (via carbonyl carbon) to nitrogen to amide proton correlation; HCACO, α-proton to α-carbon to carbonyl carbon correlation; HCCH-TOCSY, proton–carbon–proton correlation using carbon total correlated spectroscopy; NOESY, nuclear Overhauser effect spectroscopy; WALTZ, wideband alternating-phase low-power technique for zero-residual-splitting; SEDUCE, selective decoupling using crafted excitation; DIPSI, decoupling in the presence of scalar interactions.

Table 1: Nuclei Correlated in Double and Triple Resonance 3D NMR Experiments[a]

expt	$C^{\beta(i-1)}$	$H^{\alpha(i-1)}$	$C^{\alpha(i-1)}$	$CO^{(i-1)}$	HN	N	C^{α}	H^{α}	C^{β}	CO
HNCO				■[144]	■	■				
HNCA			○[121]		■	■	■[151]			
HN(CO)CA			■[154]		■	■				
HACA(CO)NH		■[140]			■	■				
CBCA(CO)NH	■[134]		■[146]		■	■				
HNCACB	○[98]		○[110]		■	■	■[138]		■[122]	
^{15}N NOESY		■			■	■		■		
^{15}N TOCSY					■	■		■[124]		
HCACO							■	■		■[146]

[a] Symbols are as follows: ■, principal correlations; ○, additional, weaker correlations via two bond $J_{N(i)C^{\alpha}(i-1)}$ coupling. Numbers of correlations observed in protein S spectra are given as superscripts, where relevant.

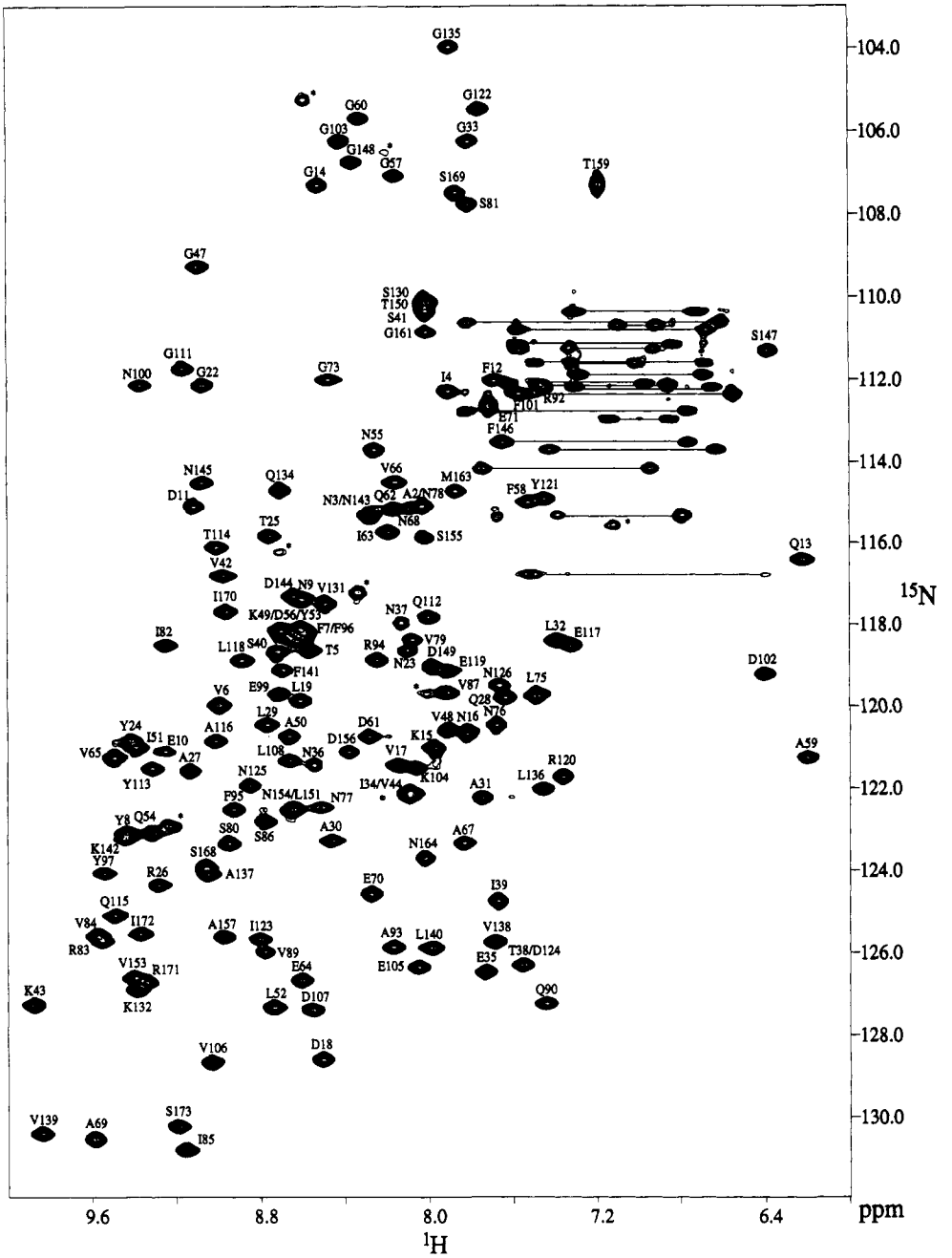

FIGURE 1: 2D ^{15}N–^{1}H HSQC spectrum of uniformly (>95%) ^{15}N-labeled protein S, recorded at 500-MHz ^{1}H frequency with pulsed field gradient selection of coherence transfer pathways and enhanced sensitivity. Asterisks indicate those backbone cross peaks which remain unassigned due to lack of signal in 3D NMR experiments. Pairs of cross peaks resulting from Asn and Gln side-chain NH_2 groups are connected by horizontal lines.

the dimensions, except in the case of the HMQC-J (Kay & Bax, 1990) experiment which requires stronger resolution enhancement in t_1 in order to allow measurement of splittings due to scalar coupling. For this purpose, Lorentzian-to-Gaussian filtering was employed with NMR2 parameters $G1$ = 14.0 Hz (Lorentzian line width to remove), $G2$ = 8.0 Hz

Table 2: ^{15}N, ^{13}C, ^{13}CO, and ^{1}H Assignments for Protein S at pH 6.7 and 37 °C[a]

residue	^{15}N	CO	C^{α}	C^{β}	others
M1					
A2	115.3 (7.97)	171.4	51.6 (3.97)	19.8 (1.38)	
N3	115.2 (8.28)	177.3	51.6 (5.16)	40.1 (2.53, 2.39)	
I4	112.2 (7.90)	175.9	60.6 (4.95)	39.7 (2.16)	H^{γ} 1.4, 1.3; $C^{\gamma Me}$ 20.3 (1.03); C^{δ} 14.9 (0.75)
T5	118.5 (8.57)	176.4	62.0 (5.10)	71.6 (3.67)	C^{γ} 23.2 (0.66)
V6	119.9 (8.99)	173.2	59.3 (5.07)	32.8 (2.28)	C^{γ} 22.6 (0.95); 18.8 (0.95)
F7	118.1 (8.59)	177.9	56.5 (5.72)	40.8 (3.46, 3.33)	
Y8	122.9 (9.44)	175.6	59.4 (4.60)	37.8 (3.30)	H^{δ} 7.06; H^{ϵ} 6.77
N9	117.3 (8.60)	174.3	51.1 (5.39)	40.0 (3.03, 2.50)	
E10	121.0 (9.27)	175.3	57.8 (3.70)	(2.00, 1.90)	
D11	115.0 (9.11)	174.2	55.3 (3.09)	38.3 (2.67, 2.46)	
F12	111.9 (7.68)	174.1	55.3 (2.62)	35.2 (2.30, 2.16)	H^{δ} 6.75; H^{ϵ} 7.22
Q13	116.3 (6.21)	175.0	53.5 (4.82)	32.6 (2.28, 1.79)	C^{γ} 33.5 (2.39)
G14	107.2 (8.53)	172.7	44.1 (4.19, 3.84)		
K15	120.9 (7.98)	173.7	58.1 (3.83)	31.6 (1.23)	C^{γ} 24.9 (0.98); H^{δ} 0.61; H^{ϵ} 2.74
N16	120.6 (7.82)	174.8	51.7 (5.63)	33.0 (2.21, 1.96)	
V17	121.3 (8.13)	173.9	61.6 (4.11)	36.7 (1.56)	C^{γ} 21.1 (0.94); 21.5 (0.77)
D18	128.5 (8.51)	175.7	53.3 (5.45)	41.9 (2.62, 2.29)	
L19	119.8 (8.60)		51.0 (4.98)	42.0 (1.45, 1.01)	C^{γ} 25.9 (1.19); C^{δ} 26.7 (0.33), 23.6 (0.17)
P20			60.6 (5.20)	30.7 (2.45, 2.00)	C^{γ} 26.3 (1.98); C^{δ} 50.7 (3.87, 3.66)
P21		175.5	64.1 (4.21)	31.6 (2.29, 1.89)	$H\gamma$ 1.99; C^{δ} 49.9 (4.27, 3.53)
G22	112.1 (9.08)	170.1	44.6 (4.02, 3.76)		
N23	118.5 (8.10)	173.6	52.1 (5.16)	40.1 (2.62, 2.39)	
Y24	120.7 (9.41)	174.9	57.3 (5.11)	40.4 (3.17, 2.78)	H^{δ} 6.89; H^{ϵ} 6.76
T25	115.7 (8.77)	174.1	60.9 (4.08)	70.8 (5.12)	C^{γ} 22.5 (1.30)
R26	124.2 (9.28)	179.5	60.7 (4.11)	29.6 (2.04, 1.80)	H^{γ} 1.68; C^{δ} 43.8 (3.25)
A27	121.5 (9.14)	181.3	54.9 (4.20)	18.0 (1.46)	
Q28	119.7 (7.64)	179.3	58.7 (4.20)	29.1 (2.62, 2.30)	C^{γ} 35.2 (2.63, 2.53)
L29	120.4 (8.77)	179.2	58.1 (4.01)	39.8 (2.16, 1.40)	C^{γ} 26.6 (1.96); C^{δ} 25.2 (0.72), 23.9 (0.63)
A30	123.2 (8.46)	182.2	54.8 (4.11)	17.3 (1.51)	
A31	122.2 (7.74)	178.8	54.4 (4.23)	17.7 (1.62)	
L32	118.3 (7.38)	176.4	54.5 (4.42)	42.2 (1.90, 1.68)	C^{γ} 26.6 (1.92); C^{δ} 22.5 (0.97), 26.1 (0.87)
G33	106.2 (7.80)	173.6	45.0 (4.29, 3.70)		
I34	122.0 (8.08)	174.7	60.2 (4.29)	37.4 (1.66)	C^{γ} 27.5 (1.38, 1.31); $C^{\gamma Me}$ 17.6 (0.81); C^{δ} 15.0 (0.77)
E35	126.4 (7.73)	175.9	55.8 (4.20)	30.6 (2.14, 1.96)	C^{γ} 36.8 (2.36)
N36	121.3 (8.55)	174.6	54.0 (4.50)	38.3 (2.82, 2.68)	
N37	118.0 (8.08)		54.2 (4.00)	36.0 (3.04, 2.85)	
T38	126.3 (7.55)	176.7	– (4.57)		
I39	124.8 (7.67)	175.9	64.2 (3.89)	– (1.68)	H^{γ} 1.64, 1.05; $C^{\gamma Me}$ 17.5 (1.01); C^{δ} 14.3 (0.86)
S40	118.6 (8.72)	171.8	59.0 (5.08)	68.1 (3.92, 3.17)	
S41	110.2 (8.01)	173.6	59.9 (4.30)	64.5 (4.23, 4.14)	
V42	116.7 (8.97)	174.2	61.0 (5.68)	37.0 (2.24)	C^{γ} 22.4 (1.12); 22.9 (1.09)
K43	127.1 (9.87)	175.3	56.9 (4.54)	35.1 (1.76, 1.54)	H^{γ} 1.32; H^{δ} 1.44; C^{ϵ} 41.1 (2.30)
V44	122.1 (8.08)	172.0	59.0 (4.40)	33.2 (1.60)	C^{γ} 23.6 (0.79), 21.7 (0.79)
P45			61.9 (4.95)	30.4 (2.29, 2.04)	C^{γ} 26.5 (1.80); C^{δ} 51.3 (3.62, 3.55)
P46		177.3	63.6 (4.40)	32.6 (2.41, 1.80)	C^{γ} 28.3 (2.20); H^{δ} 3.71, 3.53
G47	109.1 (9.10)	174.4	44.2 (4.28, 3.82)		
V48	120.5 (7.90)	171.6	59.3 (5.27)	36.8 (1.87)	C^{γ} 22.4 (0.86); 22.4 (0.78)
K49	118.1 (8.70)	173.1	54.9 (4.77)	36.6 (1.99, 1.87)	C^{γ} 24.8 (1.43); H^{δ} 1.64; H^{ϵ} 3.11, 3.04
A50	120.6 (8.67)	174.7	50.2 (4.75)	23.5 (1.40)	
I51	120.9 (9.39)	173.9	61.1 (4.80)	39.8 (1.76)	H^{γ} 1.54, 1.01; $C^{\gamma Me}$ 18.5 (0.28); C^{δ} 13.9 (0.90)
L52	127.2 (8.74)	174.4	53.6 (4.80)	42.3 (1.93, 1.02)	C^{γ} 28.0 (1.80); C^{δ} 27.3 (0.90), 24.8 (0.84)
Y53	118.3 (8.63)	176.7	57.3 (5.12)	39.3 (3.43, 3.26)	H^{δ} 7.24; H^{ϵ} 6.68
Q54	122.9 (9.32)	176.4	57.1 (4.87)	30.6 (2.20, 2.12)	C^{γ} 33.3 (2.58, 2.44); H^{ϵ} 7.47, 7.21
N55	113.5 (8.24)	175.6	50.3 (5.28)	40.2 (2.95, 2.44)	
D56	118.2 (8.66)	176.7	53.2 (4.29)	40.6 (2.56, 2.50)	
G57	107.0 (8.15)	173.9	46.5 (3.44, 2.88)		
F58	114.9 (7.52)	173.4	55.7 (2.54)	34.8 (2.00, 1.88)	H^{δ} 6.78; H^{ϵ} 6.65
A59	121.2 (6.19)	175.5	50.2 (4.85)	22.3 (1.34)	
G60	105.6 (8.32)	173.7	44.1 (4.31, 3.82)		
D61	120.7 (8.29)	174.2	55.7 (4.53)	40.4 (2.55, 2.48)	
Q62	115.0 (8.16)	175.3	53.1 (5.76)	33.6 (1.80, 1.28)	C^{γ} 33.6 (2.34, 1.92)
I63	115.7 (8.20)	173.0	60.1 (4.50)	42.3 (1.47)	C^{γ} 26.5 (1.49, 1.02); $C^{\gamma Me}$ 17.6 (0.87); C^{δ} 13.7 (0.78)
E64	126.7 (8.61)	175.4	54.8 (5.11)	31.3 (2.14, 1.95)	C^{γ} 37.0 (2.18, 2.04)
V65	121.2 (9.50)	175.9	59.8 (5.32)	33.4 (2.44)	C^{γ} 24.0 (1.16); 20.5 (1.04)
V66	114.4 (8.15)	173.4	59.8 (4.73)	33.2 (2.58)	C^{γ} 21.3 (1.05); 17.5 (1.01)
A67	123.2 (7.84)	174.2	50.2 (4.65)	20.0 (1.03)	
N68	115.0 (8.09)	174.0	54.4 (4.94)	36.8 (3.09, 2.95)	
A69	130.4 (9.59)	175.0	50.8 (5.20)	19.3 (1.44)	
E70	124.4 (8.27)	174.2	57.9 (3.13)	30.1 (2.18, 1.92)	C^{γ} 36.2 (2.26)
E71	112.6 (7.70)	176.0	54.4 (4.99)	32.2 (2.05)	C^{γ} 35.3 (2.32)
L72		177.2	52.8 (4.43)	43.3 (1.43, 1.12)	C^{γ} 26.1 (1.36); C^{δ} 22.6 (0.64); 26.3 (0.60)
G73	111.9 (8.49)	175.1	45.4 (4.10, 3.71)		
F74		176.5	63.8 (4.45)	31.9 (2.37, 1.96)	C^{γ} 27.2 (2.07); C^{δ} 51.0 (3.96, 3.82)
L75	119.7 (7.48)	176.4	53.2 (4.24)	40.1 (1.78, 1.52)	C^{γ} 29.1 (1.70); C^{δ} 24.3 (0.75); 23.6 (0.70)
N76	120.3 (7.67)	176.4	54.7 (4.19)	36.2 (2.89, 2.70)	
N77	122.3 (8.53)	174.7	54.8 (3.78)	35.9 (2.95, 2.71)	

Table 2 (Continued)

residue	^{15}N	CO	C^{α}	C^{β}	others
N78	115.1 (8.02)	173.0	53.7 (5.03)	41.2 (2.76, 2.17)	
V79		176.8	62.8 (3.98)	33.1 (2.10)	C^{γ} 21.3 (0.98); 21.0 (0.96)
S80	123.1 (8.94)	171.3	59.4 (4.92)	66.4 (4.10, 3.20)	
S81	107.7 (7.81)	173.3	59.1 (4.80)	65.4 (4.20, 3.87)	
I82	118.4 (9.24)	175.1	61.4 (5.59)	45.8 (1.75)	C^{γ} 28.5 (1.61, 1.20); $C^{\gamma Me}$ 17.7 (0.97); C^{δ} 13.7 (0.67)
R83	125.6 (9.55)	174.4	56.1 (5.32)	33.0 (1.87, 1.80)	H^{γ} (1.71, 1.61); H^{δ} (3.12, 2.80)
V84	125.4 (9.57)	175.8	62.7 (4.65)	32.8 (2.68)	C^{γ} 22.2 (1.02, 0.98)
I85	130.7 (9.16)	174.7	60.3 (4.40)	42.0 (1.53)	C^{γ} 27.2 (1.39, 1.05); $C^{\gamma Me}$ 17.3 (0.93); C^{δ} 14.3 (0.81)
S86	122.7 (8.78)	173.7	59.2 (4.83)	63.1 (3.91, 3.83)	
V87	119.6 (7.91)	172.8	58.1 (4.81)	32.8 (2.27)	C^{γ} 22.4 (0.86); 17.6 (0.54)
P88		177.9	62.7 (4.38)	31.8 (2.36, 1.95)	C^{γ} 27.7 (2.08, 1.96); C^{δ} 50.0 (3.78, 3.61)
V89	125.9 (8.79)	175.4	65.2 (3.41)	31.4 (2.00)	C^{γ} 22.1 (1.01); 21.3 (0.84)
Q90	127.2 (7.44)	173.9	52.5 (5.03)	30.9 (2.18, 1.80)	C^{γ} 33.5 (2.38)
P91		172.4	62.1 (4.56)	32.2 (2.43, 1.84)	C^{γ} 27.2 (2.05); C^{δ} 50.5 (3.89, 3.65)
R92	112.1 (7.45)	175.4	54.4 (4.65)	33.0 (1.74, 1.61)	H^{γ} 1.50
A93	125.8 (8.16)	175.0	51.0 (4.06)	20.4 (1.16)	
R94	118.8 (8.24)	173.4	54.7 (4.63)	33.0 (1.76, 1.61)	H^{γ} 1.70
F95	122.4 (8.94)	174.0	57.7 (4.71)	40.8 (3.21, 3.10)	H^{δ} 7.30; H^{ϵ} 7.10
F96	118.0 (8.61)	177.9	56.5 (5.88)	40.1 (3.65, 3.46)	H^{δ} 7.04; H^{ϵ} 7.20
Y97	124.0 (9.52)		60.6 (5.90)	39.9 (3.68, 3.44)	H^{δ} 7.30; H^{ϵ} 7.06
K98		175.9	54.3 (5.02)	35.5 (2.30, 2.09)	H^{γ} 1.40
E99	119.6 (8.71)	175.9	57.3 (3.75)	30.4 (2.01, 1.95)	C^{γ} 37.4 (2.39)
N100	112.1 (9.40)	174.0	55.7 (2.60)	(2.08, 1.97)	
F101	112.3 (7.57)	173.4	55.3 (2.90)	34.8 (1.81, 1.65)	H^{δ} 6.79; H^{ϵ} 7.22; H^{ζ} 7.04
D102	119.1 (6.39)	175.0	51.5 (5.20)	44.1 (2.58, 2.50)	
G103	106.2 (8.41)	172.8	44.1 (4.08, 3.92)		
K104	121.5 (8.06)	173.4	58.2 (3.90)	32.2 (1.31, 1.20)	H^{γ} 0.83; H^{δ} 1.00; H^{ϵ} 2.94, 2.84
E105	126.2 (8.05)	175.7	52.8 (5.27)	31.4 (1.79, 1.26)	C^{γ} 34.9 (1.77)
V106	128.5 (9.03)	173.5	61.8 (4.27)	36.5 (1.65)	C^{γ} 21.3 (1.02); 20.94 (0.99)
D107	127.3 (8.56)	174.5	50.9 (6.16)	40.8 (2.59, 2.34)	
L108	121.2 (8.67)		50.8 (4.96)	45.6 (1.41, 1.00)	C^{γ} 26.2 (1.16); C^{δ} 22.3 (0.13); 24.0 (–0.30)
P109		173.6	61.1 (5.34)	30.9 (2.28, 1.79)	H^{γ} 2.08; C^{δ} 49.7 (3.71, 3.55)
P110		174.9	64.0 (4.13)	32.0 (2.23, 1.86)	H^{γ} 2.00; C^{δ} 50.1 (4.24, 3.44)
G111	111.6 (9.17)	170.6	44.5 (4.18, 3.77)		
Q112	117.7 (8.00)	174.0	54.0 (5.14)	29.9 (1.77, 1.68)	C^{γ} 32.6 (2.19, 2.09); H^{ϵ} 6.83, 6.65
Y113	121.4 (9.31)	174.9	57.3 (5.24)	42.2 (2.93, 2.89)	H^{δ} 7.34; H^{ϵ} 7.00
T114	116.0 (9.01)	174.2	61.0 (4.11)	70.8 (5.09)	C^{γ} 22.5 (1.38)
Q115	124.9 (9.48)	179.5	61.0 (3.97)	29.0 (1.84, 1.57)	C^{γ} 33.0 (2.40, 2.20)
A116	120.7 (9.02)	181.1	54.7 (4.17)	18.3 (1.41)	
E117	118.4 (7.32)	177.8	59.4 (3.79)	30.0 (2.30, 2.09)	C^{γ} 37.8 (2.62, 2.46)
L118	118.8 (8.89)	178.9	58.6 (3.91)	39.7 (1.56, 1.50)	C^{γ} 26.4 (1.88); C^{δ} 25.1 (0.62); 24.0 (0.50)
E119	119.0 (7.91)	181.0	59.8 (3.70)	29.3 (2.17, 2.09)	C^{γ} 36.7 (2.48, 2.29)
R120	121.7 (7.35)	176.9	60.1 (3.70)	30.1 (1.87, 1.80)	C^{δ} 44.4 (3.12, 2.80)
Y121	114.8 (7.44)	175.2	56.7 (5.49)	37.8 (3.60, 2.70)	H^{δ} 7.14; H^{ϵ} 7.00
G122	105.4 (7.75)	175.5	45.2 (4.31, 4.00)		
I123	125.6 (8.80)	172.6	61.1 (3.97)	37.7 (1.60)	C^{γ} 27.7 (1.50, 0.70); $C^{\gamma Me}$ 16.9 (0.64); C^{δ} 14.0 (0.43)
D124	126.1 (7.51)	176.0	53.8 (4.57)	41.0 (2.65, 2.60)	
N125	121.6 (8.86)	175.3	54.5 (4.39)	38.5 (2.80, 2.48)	
N126	119.3 (7.70)		54.2 (4.67)	41.1 (2.71, 2.58)	
T127			(4.60)	–(3.13)	C^{γ} 22.1 (0.82)
I128			62.8 (3.93)		$C^{\gamma Me}$ 18.9 (1.28); C^{δ} 12.8 (0.78)
S129			60.5 (4.45)	64.3 (4.08)	
S130	109.8 (7.99)	173.9	60.5 (4.43)	64.6 (3.46, 3.26)	
V131	117.4 (8.50)	174.4	61.4 (5.30)	38.6 (2.05)	C^{γ} 23.7 (1.21); 20.8 (0.54)
K132	126.8 (9.38)	176.3	53.6 (5.15)	34.3 (1.87, 1.76)	H^{γ} 1.47; H^{δ} 1.69; C^{ϵ} 41.8 (2.99, 2.90)
P133		174.9	64.2 (4.01)	31.7 (2.59, 1.80)	C^{γ} 26.2 (2.00); H^{δ} 4.30, 3.28
Q134	114.6 (8.71)	175.5	55.6 (3.70)	26.8 (2.15)	C^{γ} 34.4 (2.58, 2.12)
G135	103.9 (7.90)	174.4	45.3 (4.11, 3.68)		
L136	121.9 (7.45)	176.9	53.3 (4.54)	42.3 (1.83, 1.40)	C^{γ} 26.5 (1.55); C^{δ} 22.6 (0.84); 25.4 (0.72)
A137	124.0 (9.05)	175.2	51.2 (4.82)	18.2 (1.39)	
V138	125.6 (7.68)	175.5	61.3 (4.79)	33.4 (2.52)	C^{γ} 22.2 (1.16); 22.8 (0.97)
V139	130.3 (9.84)	173.8	63.0 (4.38)	33.6 (1.95)	C^{γ} 21.3 (0.85); 22.8 (0.20)
L140	125.8 (7.99)	175.3	53.6 (4.68)	40.8 (1.86, 1.26)	C^{γ} 26.9 (1.96); C^{δ} 23.3 (1.07); 27.7 (1.02)
F141	119.0 (8.70)	176.9	56.8 (5.14)	39.5 (3.48, 3.43)	H^{δ} 7.25; H^{ϵ} 7.20; H^{ζ} 6.87
K142	123.1 (9.46)	176.6	58.2 (4.39)	34.6 (1.90, 1.86)	H^{γ} 1.47; H^{δ} 1.68; H^{ϵ} 2.69
N143	115.6 (8.29)	174.6	50.3 (5.34)	39.3 (2.95, 2.80)	
D144	117.2 (8.66)	175.9	52.9 (4.26)	40.5 (2.76, 2.72)	
N145	114.4 (9.08)	173.7	54.4 (3.17)	36.7 (2.73, 2.51)	
F146	113.5 (7.64)	174.1	56.4 (2.50)	34.5 (1.85, 1.40)	H^{δ} 6.76; H^{ϵ} 6.70
S147	111.2 (6.37)	172.9	56.4 (4.98)	66.7 (3.92, 3.74)	
G148	106.7 (8.36)	173.2	44.3 (4.27)		
D149	118.9 (7.98)	174.3	55.6 (4.48)	41.1 (2.74, 2.42)	
T150	110.1 (8.00)	174.8	58.9 (5.34)	72.2 (3.63)	C^{γ} 22.9 (0.93)
L151	122.5 (8.65)		52.1 (4.97)	46.6 (1.88, 1.28)	C^{γ} 27.0 (1.46); C^{δ} 23.2 (1.03); 27.5 (0.92)
P152		176.2	61.6 (5.08)	31.8 (2.36, 1.98)	C^{γ} 27.0 (2.25); C^{δ} 51.2 (3.85, 3.68)
V153	126.5 (9.39)	173.9	62.0 (4.24)	34.0 (1.92)	C^{γ} 22.6 (1.02); 20.5 (0.89)
N154	122.5 (8.67)	174.5	52.8 (5.26)	40.4 (3.00, 2.68)	

Table 2 (Continued)

residue	^{15}N	CO	C^{α}	C^{β}	others
S155	115.8 (8.01)	171.9	56.5 (4.62)	64.9 (3.97, 3.93)	
D156	121.0 (8.39)	176.1	57.3 (4.62)	40.4 (2.75, 2.60)	
A157	125.6 (8.98)	174.0	48.5 (5.41)	20.2 (1.49)	
P158			(4.12)	(2.16, 2.05)	H^{γ} 1.84; C^{δ} 50.8 (4.33, 3.77)
T159	107.2 (7.18)		(4.83)	(3.90)	H^{γ} 1.28
L160		177.6	52.7 (4.60)	41.3 (1.40)	C^{γ} 25.9 (1.44); C^{δ} 22.4 (0.73)
G161	110.9 (8.03)	176.7	46.8 (3.92, 3.80)		
A162		178.4	(4.23)	(1.48)	
M163	114.7 (7.88)	176.7	52.8 (4.64)	31.1 (2.09)	H^{γ} 2.75, 2.46; H^{ϵ} 2.13
N164	123.8 (8.02)		54.4 (4.31)	37.8 (3.18, 2.72)	
N165					
N166					
T167		175.5	67.1 (4.11)	69.3 (4.03)	C^{γ} 22.6 (1.27)
S168	123.5 (9.02)	169.9	59.0 (4.97)	66.4 (4.15, 3.20)	
S169	107.4 (7.86)	173.4	59.7 (4.70)	65.0 (4.30, 3.89)	
I170	117.6 (8.97)	174.7	61.6 (5.59)	45.5 (1.60)	C^{γ} 28.5 (1.55, 1.10); $C^{\gamma Me}$ 18.2 (1.00); C^{δ} 13.8 (0.13)
R171	126.7 (9.35)	173.8	56.2 (5.26)	34.2 (2.06, 1.89)	H^{γ} 1.68, 1.59; C^{δ} 43.2 (3.18)
I172	125.4 (9.37)	175.0	61.0 (5.10)	41.3 (2.19)	C^{γ} 28.6 (1.95, 1.18); $C^{\gamma Me}$ 19.9 (0.95); C^{δ} 16.1 (1.12)
S173	130.1 (9.19)	176.2	59.8 (4.64)	65.9 (3.94)	

[a] ^{15}N and ^{13}C chemical shifts are given first and the attached ^{1}H chemical shift is in parentheses. The chemical shift reference used for ^{1}H and ^{13}C is 3-(trimethylsilyl)propionate sodium salt. ^{15}N chemical shifts are reported relative to external liquid NH_3.

(Gaussian line width to create), and $G3 = 0.0$ (corresponding to a maximum at the first point in the FID). $^{3}J_{NH\alpha}$ coupling constants were then measured using spectral simulation (Redfield & Dobson, 1990; Smith et al., 1991) to correct for the discrepancy between measured and actual $^{3}J_{NH\alpha}$ values arising from finite line widths, and for dispersive contributions to the line shape in F_1. $^{3}J_{NH\alpha} < 5$ Hz was taken to indicate the presence of α secondary structure, and β secondary structure was inferred for $^{3}J_{NH\alpha} > 9$ Hz (Pardi et al., 1984).

For all 3D spectra, a 60° phase-shifted sine bell and single zero-fill were applied in each of the t_2 and t_3 dimensions prior to Fourier transformation, and linear prediction (Barkhuijsen et al., 1985) was used to double the number of planes in the t_1 (t_2 of HNCACB) dimension of all triple resonance experiments, except where otherwise stated, from 64 to 128. In the case of the CBCA(CO)NH experiment, linear prediction was used in both t_1 and t_2 dimensions. This required a different processing sequence to that used for other triple resonance experiments. The data was first Fourier transformed in t_3 and then subjected to a smoothing window function in t_2 which has a nonzero value at the end of the interferogram. This was followed by Fourier transformation in t_2 and then linear prediction and Fourier transformation in t_1. An inverse Fourier transform was then applied to F_2, followed by the reverse of the original t_2 window function and then linear prediction in t_2. Fourier transformation in t_2 yielded the final data set. For the HNCACB data, an inverse cosine function tuned to 35 Hz ($^{1}J_{C\alpha\beta}$) was applied in the carbon dimension prior to linear prediction in order to improve the spectral resolution in F_1 (Wittekind & Mueller, 1993).

Slowly exchanging amide protons were identified by recording a series of ^{15}N-^{1}H HSQC (Bodenhausen & Ruben, 1980) experiments at a range of time intervals beginning immediately after dissolving lyophilized protein in D_2O. These data, together with observation of sequential NOEs characteristic of helical or extended secondary structure, allowed identification of 80 hydrogen bonds.

RESULTS AND DISCUSSION

Resonance Assignment. The methodology used here is similar to those described in some previous NMR studies of higher molecular weight (15 kDa and above) proteins, such as calmodulin (Ikura et al., 1990b), calmodulin complexed with a 26-residue peptide (Ikura et al., 1991), interleukin-4 (Powers et al., 1992), and interferon-γ (Grzesiek et al., 1992). This method relies on recording a range of three-dimensional NMR spectra which utilize large heteronuclear scalar couplings to correlate different combinations of nuclei. Table 1 illustrates how such a series of experiments, pairs of which have a common correlated nucleus other than amide N and NH, can be used to connect residues through their backbone and C^{β} resonances. The sensitivity and line widths obtained in the NMR spectra indicate that protein S is monomeric under the conditions used for the NMR experiments, despite the fact that its function involves multimerization on the myxospore surface. Figure 1 summarizes the backbone amide ^{1}H and ^{15}N assignments in the ^{15}N-^{1}H HSQC (Bodenhausen & Ruben, 1980; Kay et al., 1992) spectrum of protein S. Backbone and side-chain ^{1}H, ^{15}N, and ^{13}C chemical shifts are reported in Table 2.

In order to increase both sensitivity and resolution in the ^{15}N dimension, most of the standard triple resonance experiments [HNCO, HNCA, HCACO, HN(CO)CA, and HCA(CO)N (Kay et al., 1990; Bax & Ikura, 1991)] were modified by introduction of a constant-time ^{15}N evolution period (Bax et al., 1979; Rance et al., 1984; Powers et al., 1991; Grzesiek & Bax, 1992a; Palmer et al., 1992) and in-phase heteronuclear magnetization was maintained through the use of coherent ^{1}H decoupling (Grzesiek & Bax, 1992a; Farmer et al., 1992). Despite these improvements, the signal to noise ratio obtained in the HCA(CO)N experiment was too low to yield useful data. The resulting lack of the ^{15}N junction point between adjacent residues caused some difficulty in making sequential connectivities. This was exacerbated by similarities in chemical shift and NOE patterns between the four internally homologous motifs. For example, two stretches (FQGKNVDLPPGNYT, 12–25, and FDGKEVDLPPGQYT, 101–114) are found in motifs 1 and 3 with very similar sequences and close matching of backbone chemical shifts for most pairs of corresponding residues. Experiments correlating amide proton and nitrogen resonances with C^{α} and C^{β} resonances (Grzesiek & Bax, 1992b; Wittekind & Mueller, 1993) were particularly useful in resolving these ambiguities. A pair of such experiments was recorded after approximately 50% of sequential backbone assignments had already been obtained. The CBCA(CO)NH technique (Grzesiek & Bax, 1992b)

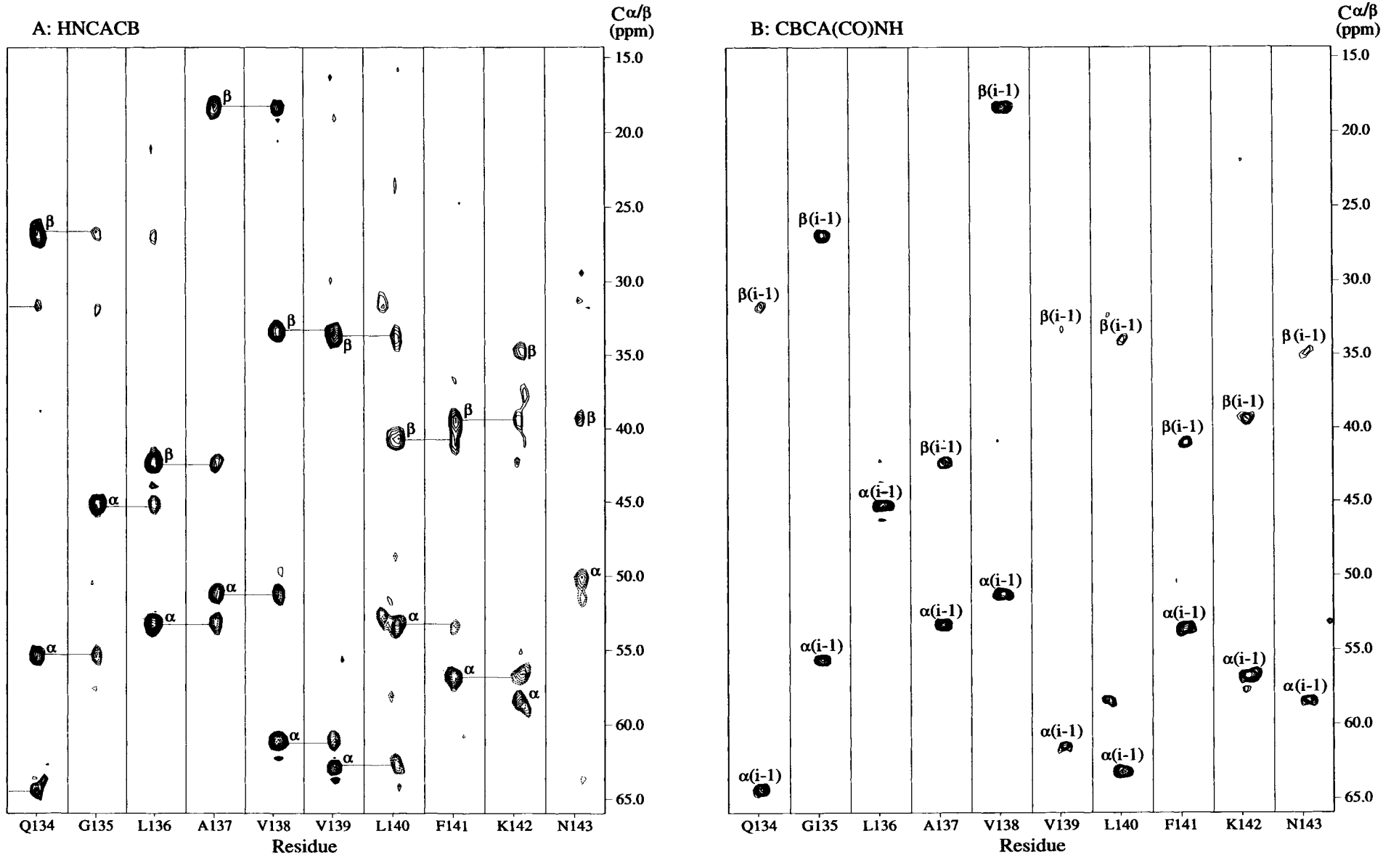

FIGURE 2: F_1 ($^{13}C^{\alpha/\beta}$) strips for residues Gln 134 to Asn 143 from (A) the 3D HNCACB spectrum and from (B) the 3D CBCA(CO)NH spectrum of protein S. Strips are extracted at the F_3/F_2 ($^1H/^{15}N$) frequencies of the backbone amide of each residue. In each strip of A, the intraresidue $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ correlations are labeled. Each strip (except that for Asn 143) also contains correlations of the amide proton of residue i with the C^{α} and C_{β} atoms of the previous ($i - 1$) residue. These are emphasized by horizontal lines connecting the peaks in adjacent strips. The interresidue correlations are less intense than the intraresidue correlations. Distinction between C^{α} and C^{β} resonance is aided in this experiment by their opposite phase. In this case, C^{α} resonances are negative (dotted lines), C^{β} resonances positive. The CBCA(CO)NH spectrum (B) contains only interresidue correlations between the amide proton of residue i and the C^{α} and C^{β} atoms of the previous ($i - 1$) residue. Matching of ^{13}C chemical shifts between these two spectra facilitates sequential assignment for $^{13}C/^{15}N$-labeled proteins.

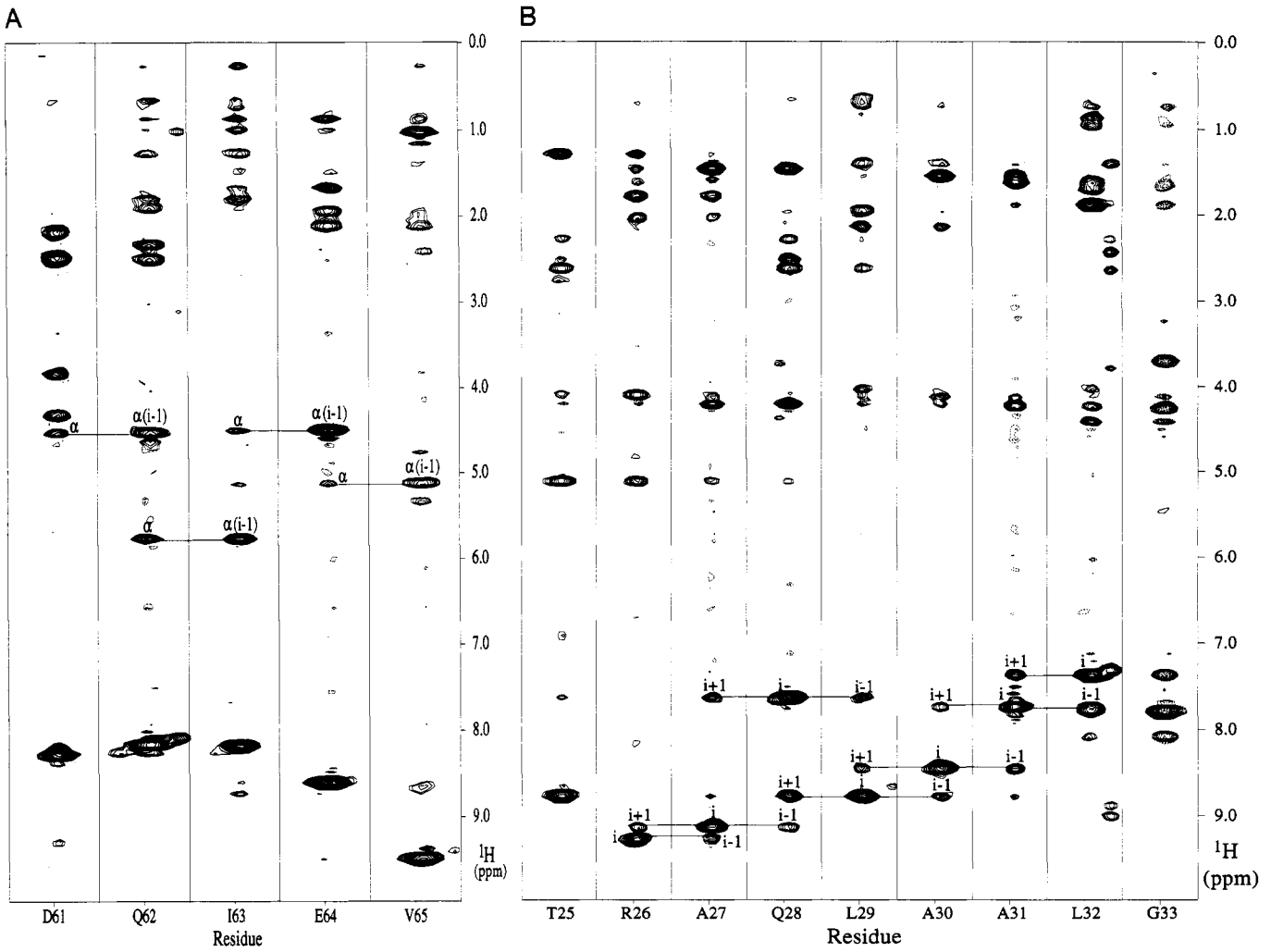

FIGURE 3: Strip plots extracted from the 100-ms mixing time ^{15}N-edited NOESY-HMQC spectrum of protein S. Examples from (A) a β-strand region, residues Asp 61 to Val 65, showing characteristic intense $d_{\alpha N}$ NOE connectivities and (B) from one of the two α-helical sections, showing a string of intense sequential D_{NN} NOE connectivities. Cross peaks depicted using dotted lines have negative intensity.

correlates amide signals with the C^α and C^β resonances of the preceding residue and the HNCACB experiment (Wittekind & Mueller, 1993) correlates amide signals with intraresidue C^α and C^β resonances and also, in most cases, with C^α and C^β of the preceding residue. The data obtained from these two experiments helped to resolve a number of uncertainties in assignment due to degeneracy or close similarity of backbone chemical shifts. For example, the clear distinction between C^β chemical shifts of N23 (40.1 ppm) and Q112 (29.9 ppm) aided in placing the stretches consisting of residues 12–25 and 101–114 in their correct sequence locations. Further facilitation of the assignment process comes from the distinctive C^α chemical shifts usually given by glycine residues, and characteristic C^α and C^β chemical shifts of alanine, serine, and threonine. This often allows immediate identification of these residues in CBCA(CO)NH and HNCACB spectra. In some cases, once assignments were made for one of a pair of internally homologous residues, the assignments could be used to help in identification of the other member of the pair. Similar situations were reported for sequential resonance assignment of other multidomain proteins such as calmodulin (Ikura et al., 1990b) and a fragment of troponin C (Findlay & Sykes, 1993). Figure 2 shows several C^β(i-1)–C^α(i-1)–N(i)–HN(i) connectivities identified in the CBCA(CO)NH spectrum together with the corresponding C^β(i)–C^α(i)–N(i)–HN(i) connectivities found in the HNCACB spectrum. The assignments obtained are in good agreement with the sequential NOE analysis using ^{15}N-edited NOESY-HMQC (Zuiderweg & Fesik, 1989; Marion et al., 1989a). Examples of H^α(i-1)–HN(i) and HN(i-1)–HN(i) NOE connectivities from extended and helical portions of protein S are shown in Figure 3.

Side-chain assignments were obtained by analysis of two HCCH-TOCSY spectra, one recorded with an 8-ms mixing time, the other with a 16-ms mixing time. The former provides direct correlations, the latter both direct and relayed correlations. For certain spin systems, the ^{15}N-edited TOCSY-HMQC spectrum could also be used for side-chain proton assignment, although this experiment was used mainly to provide direct correlations between backbone amide NH and C^αH protons.

The Secondary Structure and Topology of Protein S. Sequential and short-range ($|i-j| < 5$) NOEs, $^3J_{NH\alpha}$ coupling constants and slowly exchanging backbone amide protons are summarized in Figure 4. As deduced from patterns of hydrogen bonds and long range ($|i-j| > 5$) backbone–backbone interstrand NOEs (Figure 5), protein S consists of two α/β barrel domains (residues 1–86 and 92–173) connected by a short linker. Each domain consists of two Greek key motifs, one of which contains a regular α-helix in addition to the classical four antiparallel β-strands (termed a, b, c, and d in each motif), such that the two motifs in each domain are topologically inequivalent. In each domain of protein S, the Greek keys form two β-sheets, each sheet consisting of strands

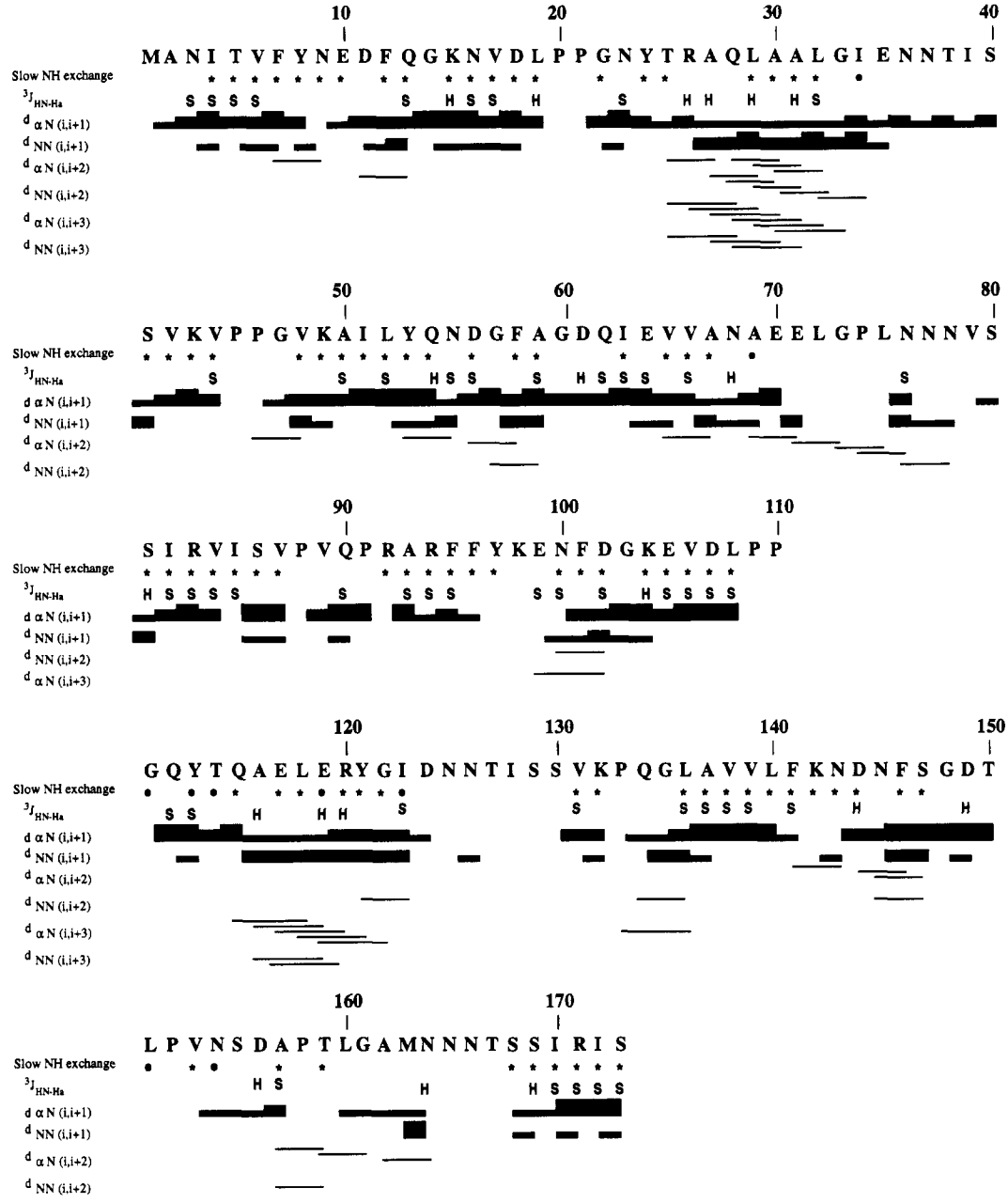

FIGURE 4: Summary of the sequential and short ($1 < |i-j| \leq 5$) range NOEs involving backbone NH and H^{α} atoms, backbone amide exchange, $^3J_{HN\alpha}$ coupling constant and deduced secondary structure of protein S. Line thickness for $d_{\alpha N}$ and d_{NN} sequential NOE connectivities reflects the intensities of the cross peaks. Residues marked with an asterisk are those which produced cross peaks in a ^{15}N–^1H HSQC spectrum of 45 min duration recorded immediately after dissolving lyophilized protein S in D_2O. Residues marked with S or H are those for which the $^3J_{HN\alpha}$ coupling constant value indicates extended (strand) or helical secondary structure.

a, b, and d from one motif and strand c from the other motif. The strand residue ranges are 3–8, 15–20, 22–24, and 41–44 in Greek key one, 48–53, 61–65, 68–70, and 81–86 in Greek key two, 92–97, 104–109, 111–114, and 130–133 in Greek key three, and 137–141, 149–154, 156–158, and 169–173 in the fourth and final Greek key. The topologies of the N-terminal domains of protein S and γB-crystallin (Blundell et al., 1981; Wistow et al., 1983) are compared in Figure 6. There are differences in strand lengths and locations, but the major distinguishing feature is the α-helix of protein S. Alignment of sequences indicates that γB-crystallin has a five residue gap at the location of the protein S helix (also see later). The two regions of protein S (residues 35–43 and 124–132) which have been previously noted to be similar to Ca^{2+}-binding sequences in calmodulin, appear to consist of six residues of the irregular loops following strand *c* in motifs 1 and 3, and also three residues of the subsequent strand d. This suggests that there is no higher order structural similarity between these regions of protein S and the Ca^{2+}-binding sites of calmodulin.

The residues connecting strands c and d of the Greek keys exhibit slightly different characteristics in the C-terminal domain (residues 123–129 and 159–168) compared to those (residues 33–40 and 72–80) in the N-terminal domain. None of these sets of residues appear to form any regular secondary structure element. Residues 127–129 and 165–167 gave no useful signal in any 3D experiment, preventing their sequential assignment (Table 2). The corresponding residues of the N-terminal domain produced sufficient signal to allow almost complete sequential assignments to be obtained (Table 2). There is evidence from $^{15}N\{^1H\}$ NOE experiments that protein S is, overall, a rigid molecule, and that its most flexible portions are the residues connecting strands c and d in each motif. Of these, the C-terminal regions would seem to be more flexible

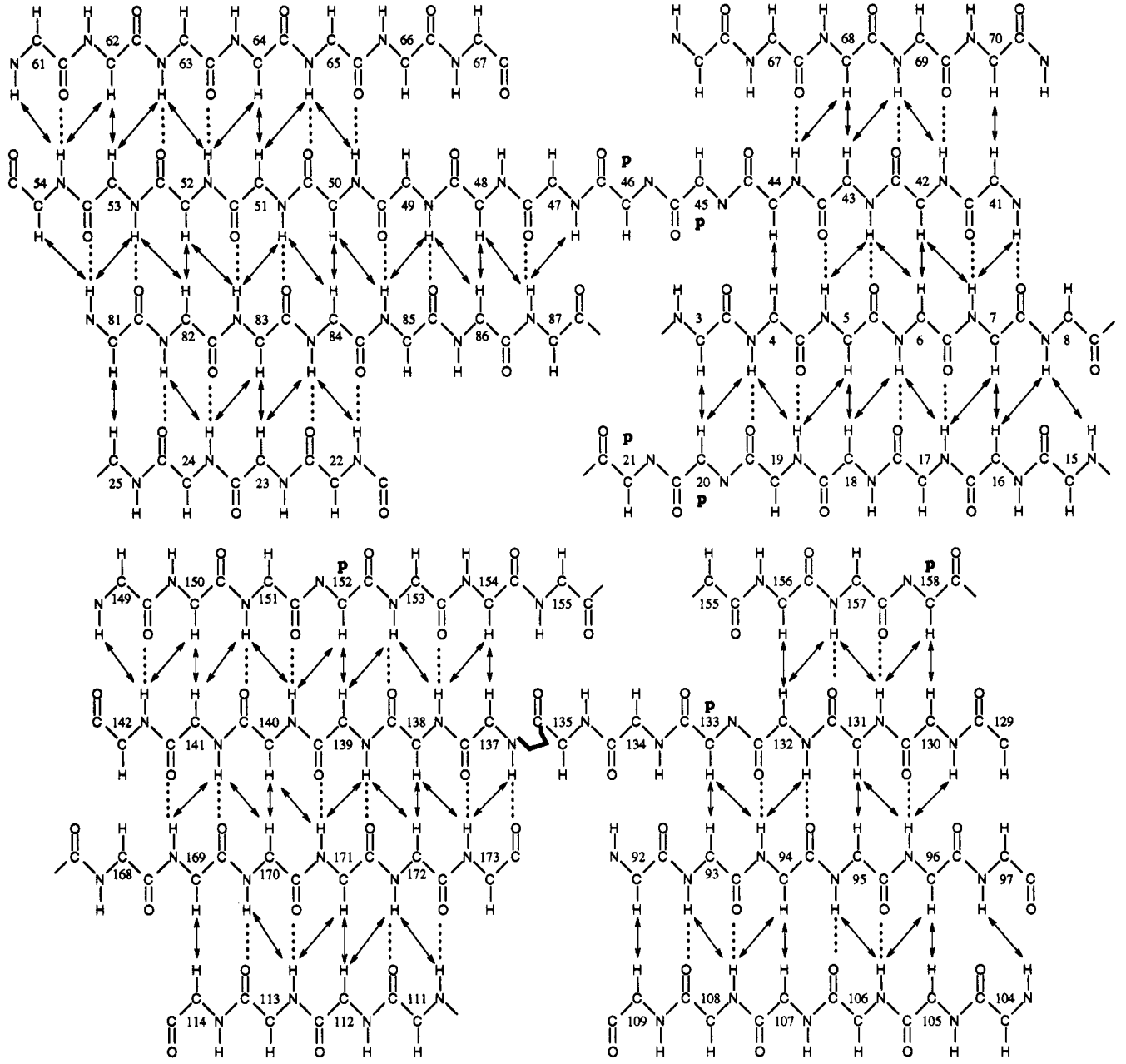

FIGURE 5: Secondary structure of the two domains of protein S as deduced from long-range backbone–backbone interstrand NOEs and amide protein exchange data. Hydrogen bonds (dashed lines) and long-range interstrand NOEs (arrows) are shown. The two α-helices (residues 26-32 and 115–122) are now shown. Proline residues are indicated by the letter P.

than their N-terminal counterparts. There are at least six weak cross peaks in both the gradient-enhanced HSQC spectrum (Figure 1) and in the $^{15}N\{^1H\}$ NOE spectra which must derive from unassigned residues (including 127–129 and 165–167). These signals all have negative or relatively low $^{15}N\{^1H\}$ NOE values. The two which could be measured (0.62 $\pm$ 0.15 and 0.55 $\pm$ 0.09) are significantly lower than any other observed for protein S. These results support the notion of high flexibility in these regions (127–129 and 165–167) of protein S. It may be relevant to note here that tryptic digestion of protein S yields a stable N-terminal fragment only (Inouye et al., 1983a). Also, preliminary results (S. Eagle & S. Inouye, unpublished observation) suggest that a fragment consisting of just the C-terminal domain is less stable than one comprising the N-terminal domain during overexpression in *E. coli*. γ-Crystallin shows the same relative domain stabilities (Rudolph et al., 1990).

Internal Structural Homology. Protein S can be divided into four homologous regions (motifs), consisting of residues 1–47, 48–86, 92–135, and 136–173, each forming a Greek key as described above. Analysis of these repeated sequences shows that motifs 1 and 3 and motifs 2 and 4 have the highest identities to one another (55% and 46%) (Wistow et al., 1985), which is consistent with the finding that motifs 1 and 3 each contain a regular α-helix, whereas motifs 2 and 4 do not. This pattern of homologies presumably resulted from two successive gene duplication events, as is found in calmodulin (Watterson et al., 1980), troponin C (Collins et al., 1973), and $\beta\gamma$-crystallins (Wistow, 1990). A previous comparison of protein S with γB-crystallin (Wistow et al., 1985) indicated that motifs 2 and 4 of protein S most closely match motif 1 of γB-crystallin, and motifs 1 and 3 correspond most closely to motifs 2 and 4 of γB-crystallin.

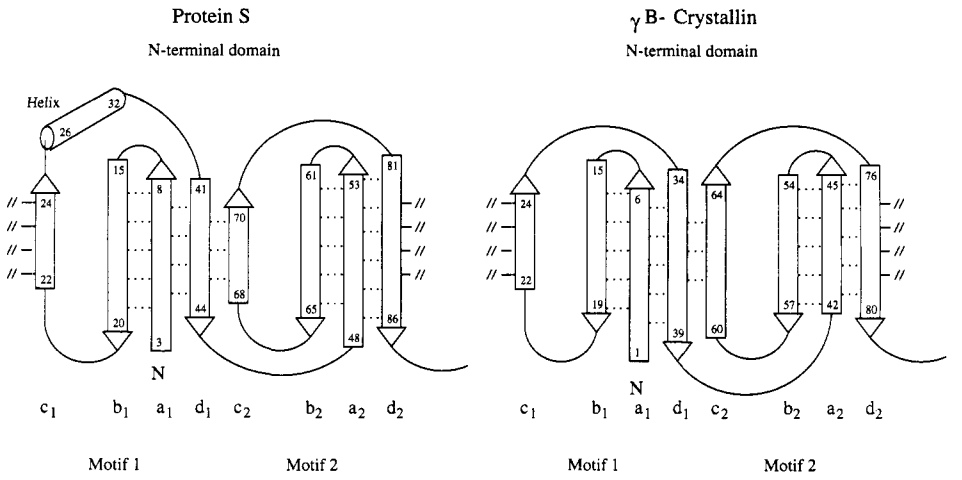

FIGURE 6: Topology of the N-terminal domains of protein S and γB-crystallin. Sequence locations of the β-strands and α-helix are indicated. The C-terminal domains have essentially the same topology.

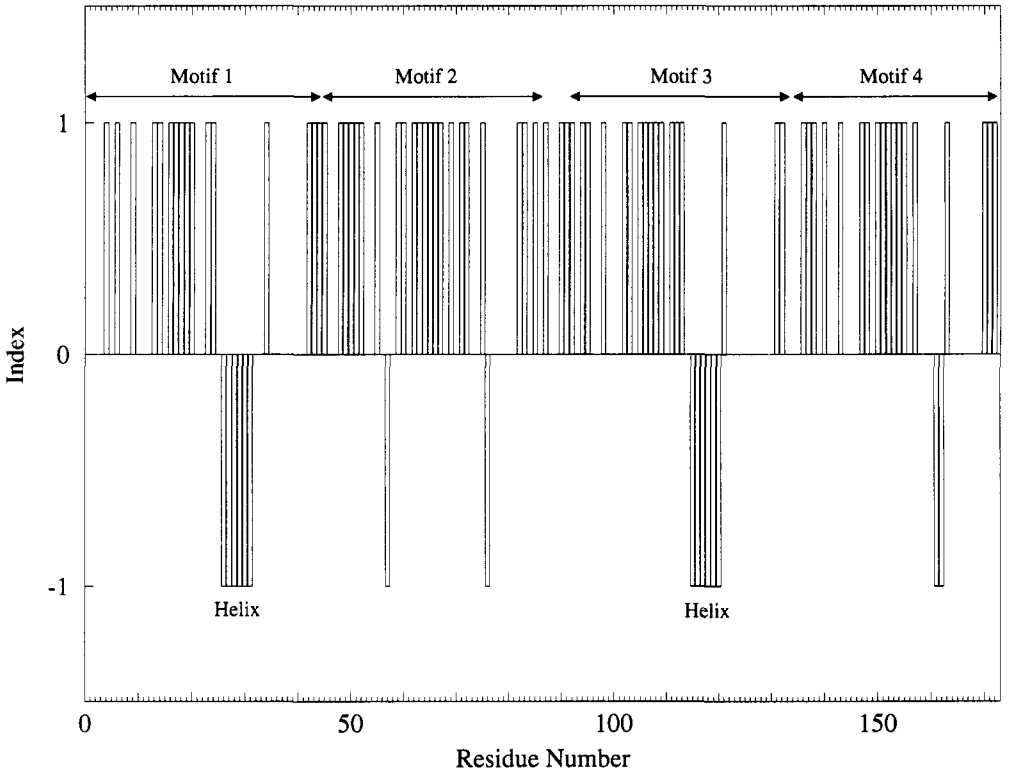

FIGURE 7: A consensus chemical shift index (Wishart & Sykes, 1993) plot of protein S. The consensus value in this case is given by the sum of the chemical shift indices for C^{α} and CO minus the index for H^{α}. Consensus values of ±1 were zeroed for the purposes of this plot, while values of ±2 and ±3 were normalized. Residues for which insufficient chemical shift information is available to calculate a consensus chemical shift index were assigned a value of zero.

Structural similarities between the motifs are evidenced by repetition of chemical shift, hydrogen bond, and NOE patterns (Figure 4 and 5). The consensus (sum of the chemical shift indices for C^{α} and CO minus that for H^{α}) chemical shift index (Wishart et al., 1992; Wishart & Sykes, 1993a,b) for protein S is shown in Figure 7. There is good correspondence between the type of secondary structure and the consensus value of the index. Most residues have a positive consensus index, as expected for a protein which is dominated by β-type secondary structure. Residues 33–40, 72–80, 123–129, and 159–168, which appear to form long loops, constitute the longest sequences of zero consensus index values. This represents a combination of incomplete assignments and a tendency toward random coil in these regions, with potentially relatively high flexibility [see above (Wishart & Sykes, 1993b]. The four nonhelical residues which have a negative value of the consensus chemical shift index (Gly 57, Asn 76, Gly 161, and Ala 162) all occur outside of the β-strands either in folded hairpin or long loops.

The H^{α} and C^{β} resonances of Phe 12, Phe 58, Phe 101, and Phe 146 all have upfield chemical shifts (Table 2). These residues occur at positions in the sequence which correspond to the folded hairpins of γB-crystallin (Blundell et al., 1981; Wistow et al., 1983), which each have two aromatic rings in close proximity. An identical sequence separation of pairs of aromatic residues occurs at these positions in protein S, leading to the inference that the upfield chemical shifts of Phe 12, Phe 58, Phe 101, and Phe 146 are due to the proximity of a second aromatic ring, Phe 7, Phe 53, Phe 96, and Phe 141, in each case. Confirmation awaits determination of a detailed three-dimensional structure of protein S.

Each of the β-sheets has a similar pattern of long-range (involving pairs of residues more than five sequence positions apart) interstrand hydrogen bonds and backbone–backbone NOEs (Figure 5). In the C-terminal domain, however, the β-sheet formed by strands b_3, a_3, d_3, and c_4 appears to contain two hydrogen bonds fewer than the other three protein S β-sheets. The missing hydrogen bonds are Ser130 HN–Phe96 CO and Thr159 HN–Ser129 CO. In the former case, Ser130 HN is not slowly exchanging. The reason for the absence of the latter hydrogen bond is not clear at this point.

The pair of α-helices in protein S is a second structural feature which gives rise to symmetry of chemical shift and NOE patterns between the two domains. Thus, the two helical sections result in two regions of downfield secondary shift of C^{α} and upfield secondary shift of C^{β} resonances (Table 2) and stretches of negative consensus chemical shift index (Figure 7), together with two sequences of characteristic $d_{NN}(i, i+1)$ (Figure 3B), $d_{\alpha N}(i, i+3)$, and $d_{\alpha\beta}(i, i+3)$ NOEs (Figure 4). The corresponding sections of both γ- and β-crystallins do show some propensity toward helical secondary structure (Blundell et al., 1981; Wistow et al., 1983; Lapatto et al., 1991). βB2-crystallin contains short 3_{10} helices in these two sections (Lapatto et al., 1991), although γB-crystallin has significant deletions here relative to protein S (five and three residues in the N- and C-terminal domains of γB-crystallin respectively). Thus each domain of protein S comprises a pair of motifs made topologically inequivalent by the presence of regular α-helical portions. This contrasts from the more symmetric motifs found in each domain of the βγ-crystallins (Figure 6).

CONCLUSIONS

Two- and three-dimensional double and triple resonance NMR techniques have been successfully applied to obtain almost complete ^{1}H, ^{15}N, ^{13}C, and ^{13}CO assignments of a 19-kDa protein. This is despite the 4-fold internal homology of protein S which causes extra difficulty in resonance assignment due to chemical shift similarities of homologous residues. The data obtained in this study will permit a detailed three-dimensional structure to be determined. From its secondary structure and topology, it is clear that protein S belongs to the βγ-crystallin superfamily and has little or no structural similarity to calmodulin and related proteins. The nonlinear motif correspondence (Wistow et al., 1985) indicates that protein S and the βγ-crystallins have evolved by distinct pathways involving different orders of gene fusion (Wistow, 1990). This divergence is further evidenced by the presence of the α-helices, one in each domain of protein S, which give rise to a topology distinct from the classical Greek key motif.

NOTE ADDED IN PROOF

The three-dimensional solution structure of Ca^{2+}-loaded protein S has been determined recently in this laboratory and appears in *Structure 2* (1994), pp 107–122.

ACKNOWLEDGMENT

We thank Dan Garrett of NIH for supplying the programs Pipp and Capp and Masayori Inouye for stimulating discussions and continuous encouragement.

REFERENCES

Anil-Kumar, Wagner, G., Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 96*, 1156–1163.

Barkhuijsen, H., De Beer, R., Bovee, W. M. M. J., & Van Ormondt, D. (1985) *J. Magn. Reson. 61*, 465–481.

Bax, A., & Ikura, M. (1991) *J. Biomol. NMR 1*, 99–104.

Bax, A., Mehlkopf, A. F., & Smidt, J. (1979) *J. Magn. Reson. 35*, 167–169.

Bax, B., Lapatto, R., Nalini, V., Driessen, H., Lindley, P. F., Mahadevan, D., Blundell, T. L., & Slingsby, C. (1990a) *Nature 347*, 776–780.

Bax, A., Clore, G. M., & Gronenborn, A. M. (1990b) *J. Magn. Reson. 88*, 425–431.

Blundell, T., Lindley, P., Miller, L., Moss, D., Slingsby, C., Tickle, I., Turnell, B., & Wistow, G. (1981) *Nature 289*, 771–777.

Bodenhausen, G., & Ruben, D. J. (1980) *Chem. Phys. Lett. 69*, 185–189.

Collins, J. H., Potter, J. D., Horn, M. J., Wilshire, G., & Jackman, N. (1973) *FEBS Lett. 36*, 268–272.

Downard, J. S., Kupfer, D., & Zusman, D. R. (1984) *J. Mol. Biol. 175*, 469–492.

Dworkin, M., & Kaiser, D. (1985) *Science 230*, 18–24.

Farmer, B. T., Venters, R. A., Spicer, L. D., Wittekind, M. G., & Müller, L. (1992) *J. Biomol. NMR 2*, 195–202.

Findlay, W. A., & Sykes, B. D. (1993) *Biochemistry 32*, 3461–3467.

Garrett, D. S., Powers, R., Gronenborn, A. M., & Clore, G. M. (1991) *J. Magn. Reson. 95*, 214–220.

Grzesiek, S., & Bax, A. (1992a) *J. Magn. Reson. 96*, 432–440.

Grzesiek, S., & Bax, A. (1992b) *J. Am. Chem. Soc. 114*, 6291–6293.

Grzesiek, S., Döbeli, H., Gentz, R., Garotta, G., Labhardt, A. M., & Bax, A. (1992) *Biochemistry 31*, 8180–8190.

Howarth, O. W., & Lilley, D. M. J. (1978) *Progr. NMR Spectrosc. 12*, 1–40.

Ikura, M., Kay, L. E., Tschudin, R., & Bax, A. (1990a), *J. Magn. Reson. 86*, 204–209.

Ikura, M., Kay, L. E., & Bax, A. (1990b) *Biochemistry 29*, 4659–4667.

Ikura, M., Kay, L. E.; Krinks, M., & Bax, A. (1991) *Biochemistry 30*, 5498–5504.

Inouye, M., Inouye, S., & Zusman, D. R. (1979a) *Dev. Biol. 68*, 579–591.

Inouye, M., Inouye, S., & Zusman, D. R. (1979b) *Proc. Natl. Acad. Sci. U.S.A. 76*, 209–213.

Inouye, S., Inouye, M., McKeever, B., & Sarma, R. (1980) *J. Biol. Chem. 255*, 3713–3714.

Inouye, S., Harada, W., Zusman, D., & Inouye, M. (1981) *J. Bacteriol. 148*, 678–683.

Inouye, S., Franceschini, T., & Inouye, M. (1983a) *Proc. Natl. Acad. Sci. U.S.A. 80*, 6829–6833.

Inouye, S., Ike, Y., & Inouye, M. (1983b) *J. Biol. Chem. 258*, 38–40.

Kaiser, D. (1986) *Ann. Rev. Genet. 20*, 539–566.

Kaiser, D., Manoil, C., & Dworkin, M. (1979) *Annu. Rev. Microbiol. 33*, 595–639.

Kay, L. E., & Bax, A. (1990) *J. Magn. Reson. 86*, 110–126.

Kay, L. E., Torchia, D. A., & Bax, A. (1989) *Biochemistry 28*, 8972–8979.

Kay, L. E., Ikura, M., Tschudin, R., & Bax, A. (1990) *J. Magn. Reson. 89*, 496–514.

Kay, L. E., Keifer, P., & Saarinen T. (1992) *J. Am. Chem. Soc. 114*, 10663–10665.

Lapatto, R., Nalini, V., Bax, B., Driessen, H., Lindley, P. F., Blundell, T. L., & Slingsby, C. (1991) *J. Mol. Biol. 222*, 1067–1083.

Loomis, W. F. (1988) *Dev. Genet. 9*, 549–559.

McCoy, M. A., & Mueller, L. (1992a) *J. Am. Chem. Soc. 114*, 2108–2112.

McCoy, M. A., & Mueller, L. (1992b) *J. Magn. Reson. 98*, 674–679.

Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., & Bax, A. (1989a) *J. Am. Chem. Soc. 111*, 1515–1517.

Marion, D., Driscoll, P. C., Kay, L. E., Wingfeld, P. T., Bax, A., Gronenborn, A. M., & Clore, G. M. (1989b) *Biochemistry 28*, 6150–6156.
Marion, D., Ikura, M., & Bax, A. (1989c) *J. Magn. Reson. 84*, 425–430.
Messerle, B. A., Wider, G., Otting, G., Weber, C., & Wüthrich, K. (1989) *J. Magn. Reson. 85*, 608–613.
Palmer, A. G., Fairbrother, W. J., Cavanagh, J., Wright, P. E., & Rance, M. (1992) *J. Biomol. NMR 2*, 103–108.
Pardi, A., Billeter, M., & Wüthrich, K. (1984) *J. Mol. Biol. 180*, 741–751.
Powers, R., Gronenborn, A. M., Clore, G. M., & Bax, A. (1991) *J. Magn. Reson. 94*, 209–213.
Powers, R., Gronenborn, A. M., Clore, G. M., & Bax, A. (1992) *Biochemistry 31*, 4336–4346.
Rance, M., Wagner, G., Sørensen, O. W., Wüthrich, K., & Ernst, R. R. (1984) *J. Magn. Reson. 59*, 250–261.
Redfield, C., & Dobson, C. M. (1990) *Biochemistry 29*, 7201–7214.
Rosenberg, E. (1984) *Myxobacteria: Development and Cell Interaction.* Springer-Verlag, New York.
Rudolph, R., Siebendritt, R., Nesslaüer, G., Sharma, A. K., & Jaenicke, R. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 4625–4629.
Shaka, A. J., Keeler, J., Frenkiel, T., & Freeman, R. (1983) *J. Magn. Reson. 52*, 335–338.
Shaka, A. J., Lee, C. J., & Pines, A. (1988). *J. Magn. Reson. 77*, 274–293.
Shimkets, L. J. (1987) *CRC Crit. Rev. Microbiol. 14*, 195–225.
Shimkets, L. J. (1990) *Microbiol. Rev. 54*, 473–501.
Smith, L. J., Sutcliffe, M. J., Redfield, C., & Dobson, C. M. (1991) *Biochemistry 30*, 986–996.
Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60–89.
Teintze, M., Thomas, R., Inouye, M., & Inoyue, S. (1985) in *Spores IX: the Molecular Biology of Microbial Differentiation* (Hoch, J., & Setlow, P., Eds.) pp 253–260, American Society for Microbiology, Washington, DC.
Teintze, M., Inouye, M., & Inouye, S. (1988) *J. Biol. Chem. 263*, 1199–1203.
Teintze, M., Inouye, M., & Inouye, S. (1991) in *Novel Calcium-Binding Proteins* (Heizmann, C. W., Ed) pp 437–446, Springer-Verlag, New York.
Watterson, D. M., Sharief, F., & Vanaman, T. C. (1980) *J. Biol. Chem. 255*, 962–975.
Wider, G., Hosur, R. V., & Wüthrich, K. (1983) *J. Magn. Reson. 52*, 130–135.
Wireman, J. W., & Dworkin, M. (1977) *J. Bacteriol. 129*, 796–802.
Wishart, D. S., & Sykes, B. D. (1993a) *J. Biomol. NMR* (in press).
Wishart, D. S., & Sykes, B. D. (1993b) *Methods Enzymol.* (in press).
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1992) *Biochemistry 31*, 1647–1651.
Wistow, G. (1990) *J. Mol. Evol. 30*, 140–145.
Wistow, G. J., & Piatigorsky, J. (1988) *Annu. Rev. Biochem. 57*, 479–504.
Wistow, G., Turnell, B., Summers, L., Slingsby, C., Moss, D., Miller, L., Lindley, P., & Blundell, T. (1983) *J. Mol. Biol. 170*, 175–202.
Wistow, G., Summers, L., & Blundell, T. (1985) *Nature 315*, 771–773.
Wittekind, M., & Mueller, L. (1993) *J. Magn. Reson. 101*, 201–205.
Zuiderweg, F. R. P., & Fesik, S. W. (1989) *Biochemistry 28*, 2387–2391.
Zusman, D. R. (1984) *Quart. Rev. Biol. 59*, 119–138.