Gros, P., Croop, J. M., & Housman, D. (1986a) *Cell 47*, 371.
Gros, P., Ben Neriah, Y., Croop, J., & Housman, D. (1986b) *Nature (London) 323*, 728.
Gros, P., Raymond, M., Bell, J., & Housman, D. (1988) *Mol. Cell. Biol. 8*, 2770.
Gros, P., Dhir, R., Croop, J., & Talbot, F. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 7289.
Hamada, H., & Tsuruo, T. (1988) *J. Biol. Chem. 263*, 1454.
Hammond, J., Johnstone, R. M., & Gros, P. (1989) *Cancer Res. 49*, 3867.
Haydon, D. A., & Hladky, S. B. (1972) *Q. Rev. Biophys. 5*, 187.
Hirata, H., Altendorf, K., & Harold, F. M. (1973) *Proc. Natl. Acad. Sci. U.S.A 70*, 1804.
Horio, M., Gottesman, M. M., & Pastan, I. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 3580.
Hsu, S. I. H., Cohen, D., Kirschner, L. S., Lothstein, L., Hartstein, M., & Horwitz, S. B. (1990) *Mol. Cell. Biol. 10*, 3596.
Kaback, H. R. (1990) *The Bacteria XII* (Krulwich, T. A., Ed.) p 151, Academic Press, Inc., New York.
Krogstad, D. J., Gluzman, I. Y., Kyles, D. E., Oduola, A. M. J., Martin, S. K., Milhous, W. K., & Schlesinger, P. H. (1987) *Science 238*, 1283.
Kushler, K., Sterne, R. E., & Thorner, J. (1989) *EMBO J. 13*, 3973.
Lichtenberg, H. C., Giebeler, H., & Höfer, M. (1988) *J. Membr. Biol. 103*, 255.
Lichtshtein, D., Kaback, H. R., & Blume, A. J. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 650.
Lombardi, F. J., Reeves, J. P., & Kaback, H. R. (1974) *Ann. N.Y. Acad. Sci. 227*, 312.
McGrath, J. P., & Varshavsky, A. (1989) *Nature (London) 340*, 400.
Monaco, J. J., Cho, S., & Attaya, M. (1990) *Science 250*, 1723.
Moscow, J. A., & Cowan, K. H. (1988) *J. Natl. Cancer Inst. 80*, 14.
Pearce, H. L., Safa, A. R., Bach, N. J., Winter, M. A., Cirtain, M. C., & Beck, W. T. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 5128.
Raviv, Y., Pollard, H. B., Bruggemann, E. P., Pastan, I., & Gottesman, M. M. (1990) *J. Biol. Chem. 265*, 3975.
Riordan, J. R., Rommens, J. M., Kerem, B. S., Alon, N., Rozmahel, R., Grzelczak, J., Zielenski, J., Lok, S., Plavsic, N., Chou, J. L., Drumm, M. L., Ianuzzi, F. S., Collins, F. S., & Tsui, L. C. (1989) *Science 245*, 1066.
Safa, A. R., Glover, C. J., Meyers, M. B., Biedler, J. L., & Felsted, R. L. (1986) *J. Biol. Chem. 261*, 6137.
Schuldiner, S., & Kaback, H. R. (1975) *Biochemistry 14*, 5451.
Schurr, E., Raymond, M., Bell, J., & Gros, P. (1989) *Cancer Res. 49*, 2729.
Skehan, P., Storeng, R., Scudiero, D., Monks, A., McMahon, J., Vistica, D., Warren, J., Bokesch, H., Kenney, S., & Boyd, M. R. (1989) *Proc. Am. Assoc. Cancer Res. 30*, 612.
Van Der Bliek, A. M., Bass, F., Ten Houte de Lange, T., Kooiman, P. M., Van Der Velde-Koerts, T., & Borst, P. (1987) *EMBO J. 6*, 3325.
Yoshimura, A., Kuwazuru, Y., Sumizawa, T., Ichikawa, M., Ikeda, S. I., Uda, T., & Akiyama, S. I. (1990) *J. Biol. Chem. 264*, 16282.

^{1}H NMR-Based Determination of the Secondary Structure of Porcine Pancreatic Spasmolytic Polypeptide: One of a New Family of "Trefoil" Motif Containing Cell Growth Factors†

Mark D. Carr
Laboratory of Molecular Structure, National Institute for Medical Research, The Ridgeway, Mill Hill, London, U.K.
Received July 17, 1991; Revised Manuscript Received November 5, 1991

ABSTRACT: Two-dimensional ^{1}H NMR spectroscopy has been used to obtain comprehensive sequence-specific resonance assignments for the putative cell growth factor porcine pancreatic spasmolytic polypeptide, which is a 106-residue protein containing two "trefoil" domains. The patterns of sequential ($i,i+1$), medium-range ($i,i<5$), and long-range NH to NH, αCH to NH, and αCH to αCH nuclear Overhauser effects clearly show that the protein's two trefoil domains adopt essentially the same secondary structure in solution. The main feature of each domain is a seven-residue helix followed by a short antiparallel β-sheet formed from two strands of four amino acids each. This is a novel supersecondary structure, which clearly identifies the trefoil motif as a new class of growth factor associated module, distinct from other types of highly disulfide cross-linked domains, such as those found in epidermal growth factor and insulin-like growth factor I.

Porcine pancreatic spasmolytic polypeptide (PSP)[1] is a monomeric 106-residue protein (M_r = 11 700) produced in large quantities in the pancreas and is believed to act as a cell growth factor (Thim et al., 1985; Hoosein et al., 1989; Thim, 1989). The protein has been shown to inhibit both gastrointestinal motility and secretion of gastric acids (Jorgensen et al., 1982). However, recent studies of the human analogue hPSP (Tomasetto et al., 1990) suggest that the major role of the protein in vivo is to promote the healing of damaged en-

† This work was funded by the U.K. Medical Research Council.

[1] Abbreviations: 2D, two dimensional; 3D, three dimensional; DQF-COSY, double-quantum filtered correlation spectroscopy; EGF, epidermal growth factor; HOHAHA, homonuclear Hartmann-Hahn spectroscopy; IGF-I, insulin-like growth factor I; NMR, nuclear magnetic resonance, NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; PSP, porcine pancreatic spasmolytic polypeptide.

0006-2960/92/0431-1998$03.00/0

doderm tissue, such as ulcers in the gut (Wright et al., 1990).

Several families of small proteins with cell growth stimulating effects have now been identified and structurally characterized in some detail, such as those typified by epidermal growth factor (EGF) (Gregory & Preston, 1977; Cooke et al., 1987) and insulin-like growth factor I (IGF-I) (Rinderknecht & Humbel, 1978; Cooke et al., 1991). PSP is composed of two highly homologous domains of about 40 amino acids (nearly 50% identity), which each contain three disulfide bonds, resulting in the production of a three-looped structure termed the "trefoil" motif (Thim et al., 1985; Thim, 1989). The disulfide-bond configuration determined for PSP is distinct from that found for the other previously characterized classes of growth factors. In addition, amino acid sequence comparisons detect no significant sequence homology with PSP (Thim et al., 1985; Thim, 1989). Hence, it has been proposed that the trefoil domain proteins represent a new family of growth factors.

To date there are two other trefoil domain proteins known, in addition to PSP and hPSP, namely, the human protein pS2 (Masiakowski et al., 1982; Brown et al., 1984; Rio et al., 1988a; Mori et al., 1988) and the *Xenopus* polypeptide spasmolysin (Hoffmann, 1988). The 60-residue protein pS2 is particularly interesting since it appears to be expressed at fairly high levels in all estrogen-dependent breast tumor cells (Brown et al., 1984; Foekens et al., 1990), whereas there is no significant expression in normal breast tissue (Rio et al., 1987, 1988b). Clearly, understanding the role of pS2 here may be of great clinical importance. However, pS2 is not available in sufficient quantities for structural characterization; thus, a detailed study of the closely related protein PSP has been undertaken.

In recent years it has been conclusively shown that NMR spectroscopy can provide detailed three-dimensional structural information for small proteins ($M_r < 20\,000$) in solution, at a resolution comparable to that obtained by X-ray crystallography (Hard et al., 1990; Schwabe et al., 1990; Breg et al., 1990; Billeter et al., 1990; Clore et al., 1990, 1991). In the study reported here, 2D ^{1}H NMR spectroscopy has been used to obtain comprehensive sequence-specific ^{1}H resonance assignments for PSP. This has allowed us to determine the secondary structure of the protein and therefore of the two trefoil domains, on the basis of the characteristic patterns of short-range (<0.5 nm) through-space NOE effects observed in 2D NOESY spectra (Wüthrich, 1986).

Materials and Methods

Purified PSP was obtained as a gift from Dr. L. Thim of Nova Nordisk. The NMR experiments were carried out on 0.6-mL samples of 5 mM PSP adjusted to pH* 6.2, dissolved in either 100% D_2O or 90% H_2O/10% D_2O as appropriate (pH* values refer to the actual pH meter readings uncorrected for deuterium isotope affects).

The ^{1}H NMR experiments were carried out on a Varian Unity 600 spectrometer operating at a ^{1}H frequency of 600 MHz. All the 2D ^{1}H NMR spectra were acquired in the phase-sensitive mode using the method of States et al. (1982). The NMR measurements were made at 40 °C. The following experiments were carried out on both D_2O and H_2O samples of PSP: DQF-COSY (Rance et al., 1983), HOHAHA (Braunschweiler & Ernst, 1983; Davis & Bax, 1985), using isotropic mixing times of between 55 and 100 ms, and NOESY (Jeener et al., 1979; Macura et al., 1981) with NOE buildup periods of 75 and 150 ms. The HOHAHA spectra were acquired using a spin-lock field of around 8 kHz produced by an MLEV17 pulse sequence (Bax & Davis, 1985).

Water suppression in 2D ^{1}H NMR spectra recorded from samples of PSP dissolved in 90% H_2O was achieved by the use of selective on-resonance presaturation. In addition, to prevent the recovery of the solvent signal during the mixing time of NOESY experiments, a 180° pulse was applied at the center of this period. In order to observe cross peaks involving protons resonating at, or very close to, the chemical shift of the water, the 2D experiments also employed a 60-ms SCUBA sequence immediately after the presaturation period (Brown et al., 1988). The D_2O samples of PSP were prepared from protein repeatedly freeze-dried from D_2O; thus, in most cases it was not necessary to suppress the very small residual HOD signal.

The 2D ^{1}H NMR experiments were typically carried out over a period of 20–40 h, collecting 300 pairs of t_1 increments, 48–112 scans per increment, and 8192 points per scan. The spectra from D_2O samples of PSP were acquired with a spectral width of 6500 Hz, while for H_2O samples a value of 8000 Hz was used.

The 2D data were processed on a SUN SPARC-330 workstation using the Varian VNMR software package. The original data were usually zero filled to 2048 complex points in F_1 prior to Fourier transformation, and mild resolution enhancement was achieved by applying a $\pi/3$-shifted sine-squared apodization function in both dimensions.

Results and Discussion

(*i*) *Sequence-Specific Assignments.* In the case of PSP, sequence-specific ^{1}H resonance assignments were obtained using the well-proven sequential assignment procedure developed by Wüthrich and co-workers, which relies upon correlating amino acid sequence and NMR data (Wagner and Wüthrich, 1982; Wüthrich, 1986). Thus, signals were initially grouped into amino acid spin systems, assigned to particular types or classes of residues, and then sequential NOEs were used to identify stretches of neighboring amino acid spin systems, which could be compared to the known protein sequence in order to obtain sequence-specific assignments.

The spin-system assignments obtained for PSP were based mainly on the through-bond connectivities observed in DQF-COSY (Rance et al., 1983) or HOHAHA (Braunschweiler & Ernst, 1983; Davis & Bax, 1985) spectra and on the fine structures of DQF-COSY cross peaks. For example, it was possible to identify the βCH to γ/γ'CH$_3$, αCH to γ/γ'CH$_3$, and NH to αCH connectivities for all seven valine residues, while six of the seven also gave rise to NH to βCH and NH to γ/γ'CH$_3$ relayed cross peaks. Complete spin-system assignments for the aromatic residues, though, also relied upon the identification in NOESY (Jeener et al., 1979; Macura et al., 1981) spectra of NOEs between the aromatic ring signals and their corresponding NH, αCH, or βCH resonances. To give some indication of the quality of the PSP data and types of cross peaks observed, Figures 1 and 2 show the aromatic portion of a HOHAHA spectrum and the high-field region from a DQF-COSY experiment. From the selection of spectra recorded, it proved possible to characterize fully the following 61 amino acid spin systems from PSP, which were subsequently used as anchor points in the sequential assignment procedure: Alanine (6/6), threonine (3/3), valine (7/7), leucine (1/1), isoleucine (3/3), tyrosine (2/2), phenylalanine (7/7), tryptophan (2/2), lysine (3/4), proline (2/12), glycine (1/6), serine (6/9), cysteine, aspartate or asparagine (15/24), and glutamate, glutamine, or methionine (3/14). In addition, partial spin-system assignments were obtained for one histidine, four glycines, two prolines, and one cysteine, aspartate, or asparagine residue.

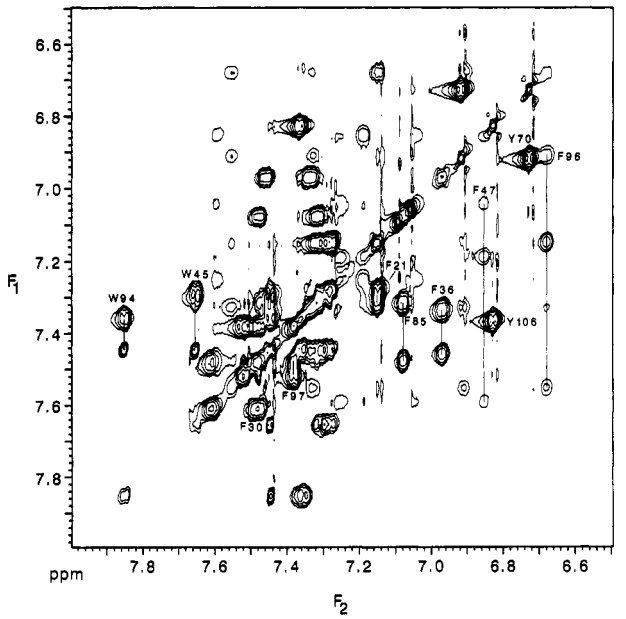

FIGURE 1: Aromatic region from a HOHAHA spectrum of PSP acquired with a 55-ms mixing period. Sequence-specific assignments for the tyrosine, phenylalanine, and tryptophan resonances are indicated at the 2,6 to 3,5 (Y70, Y106, F21, F30, F36, F85, and F97), 2 to 3 (F47 and F96), or 4 to 5 (W45 and W94) cross peaks, with the vertical lines shown leading to peaks arising from relayed connectivities.

In common with the other signals from the protein, the dispersion of the chemical shifts of the NH and αCH resonances of PSP is particularly good, as illustrated by the fingerprint region of the DQF-COSY spectrum shown in Figure 3. Thus, there are fewer degenerate signals from backbone protons than might be expected for a typical protein of 106 amino acids. Consequently, the identification of the sequential NH to NH [NN (i,i+1)], αCH to NH [αN (i,i+1)], and βCH to NH [βN (i,i+1)] NOEs required to make sequence-specific assignments was relatively straightforward, allowing long stretches of neighboring amino acids to be determined (up to 19 residues), as illustrated by the NN/αN (i,i+1) sequential walk from glutamate 76 to aspartate 87 shown on the NOESY spectrum in Figure 4. Even in cases where chemical shift degeneracies produced branch points in the sequential assignment pathways, it was usually possible to discriminate between the possibilities on the basis of either additional supporting sequential NOEs or compatibility with the amino acid sequence of the protein. Resonance assignments for a few residues of PSP, such as those in the region from glycine 20 to threonine 25, are based upon the identification of stretches of only two or three neighboring amino acids, principally due to signal overlap precluding the complete tracing of proline spin systems. However, these regions contain many unique dipeptide and tripeptide sequences involving amino acids with readily identifiable spin systems. Thus, at

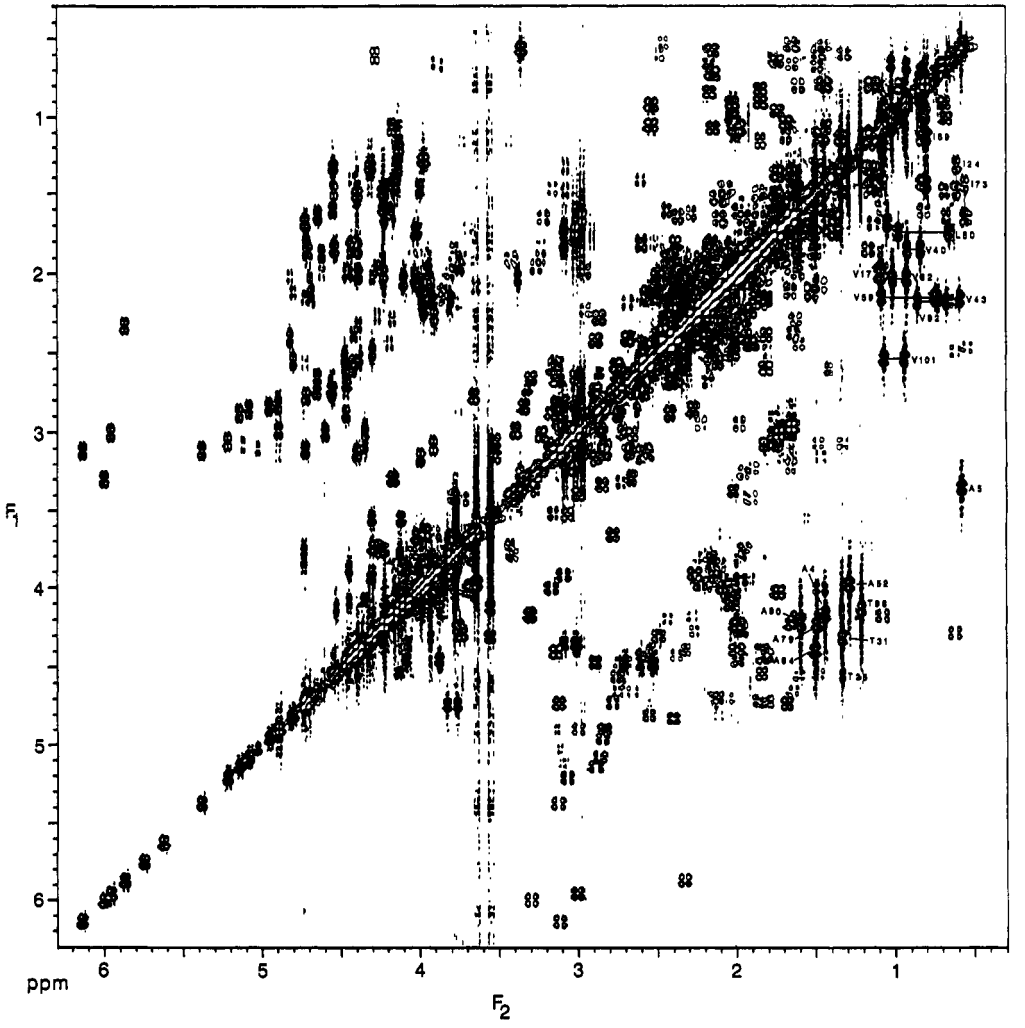

FIGURE 2: Aliphatic portion from a DQF-COSY spectrum of PSP. The relatively well-resolved connectivities involving the terminal methyl groups of alanine, threonine, valine, leucine, and isoleucine residues have been labeled, in order to give some indication of the ease with which particular types of cross peaks can be identified to obtain amino acid spin-system assignments.

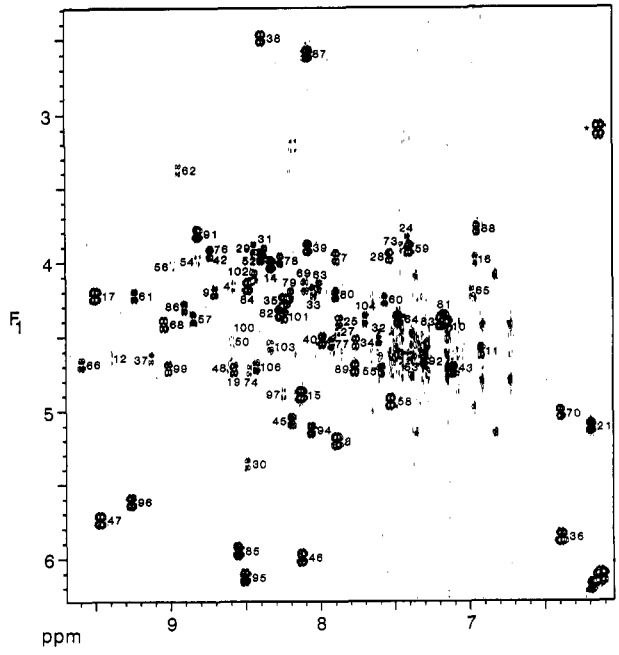

FIGURE 3: Fingerprint (NH to αCH) region from a DQF-COSY spectrum of PSP, illustrating the good resolution of signals from the backbone protons. The NH to αCH cross peaks are labeled according to the position of the amino acid residue in the protein sequence. The peak marked with an asterisk arises from an αCH to βCH correlation.

the end of the sequential assignment process, when signals from only these few residues remained to be identified, sequence-specific assignments could be made with confidence. In addition, the assignments obtained for these residues have been recently confirmed by cross peaks observed in 3D NOESY-HOHAHA spectra (M. D. Carr and C. J. Bauer, unpublished results). Hence, it was possible to obtain complete sequence-specific assignments for 84 residues and partial assignments for a further 21 (Table I) using the sequential NOEs identified in 2D NOESY spectra, which are summarized in Figure 5. The only residue for which no sequence-specific assignments could be made is serine 26.

(*ii*) *Secondary Structure Determination.* The elements of regular secondary structure found in proteins give rise to characteristic patterns of sequential ($i,i+1$), medium-range ($i,<i+5$), and long-range NOEs involving backbone protons (Wüthrich, 1986), the majority of which are identified as part of the sequential assignment procedure. For example, helices are characterized by relatively large NH to NH ($i,i+1$) and weak αCH to NH ($i,i+3$) NOEs, whereas the extended backbone conformation found in β-sheets gives rise to large sequential αCH to NH, as well as long-range NH to NH, αCH to NH, and αCH to αCH, NOEs.

The sequential and medium-range backbone to backbone and backbone to side chain NOEs identified for PSP are summarized in Figure 5. Residues 27–33 from domain I and 76–82 from domain II, which occupy equivalent positions in the two domains, are characterized by NN ($i,i+1$), NN ($i,i+2$), αN ($i,i+3$), and αβ ($i,i+3$) NOEs and therefore clearly form two seven-residue helices. In contrast, observation of a significant number of long-range backbone to backbone NOEs, linking together two analogous four-residue stretches within each domain, shows that residues 34–37 and 45–48 from domain I as well as 83–86 and 94–97 from domain II form two short antiparallel β-sheets, as illustrated in Figure 6. The arrangement of the two β-strands in each sheet is further supported by the fact that the amide protons of residues 35, 47, 84, and 96 persist for many months in D_2O solutions,

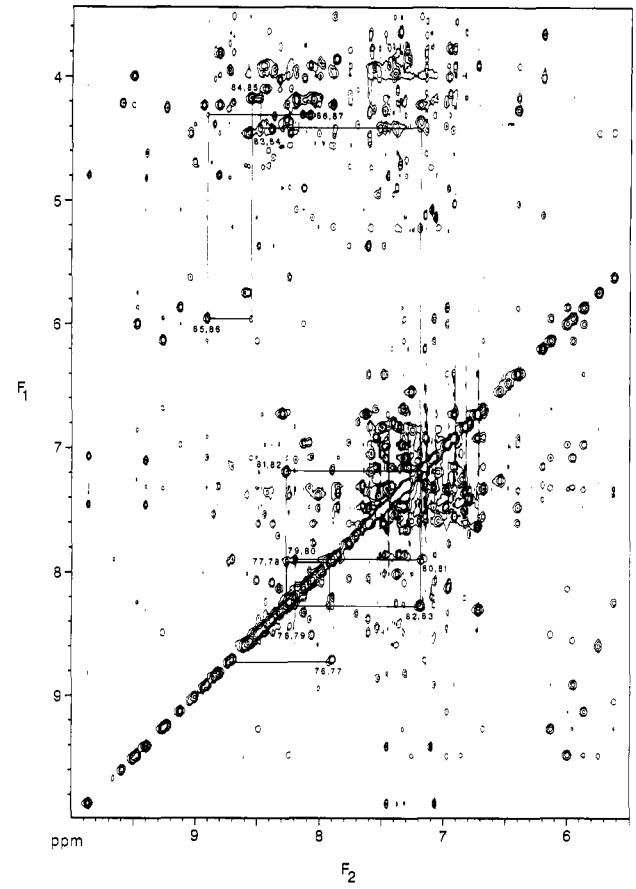

FIGURE 4: Selected region from a 150-ms mixing time NOESY spectrum of PSP dissolved in 90% H_2O. The labeled cross peaks arise from the sequential NN and αN NOEs, which lead to the sequence-specific assignment of resonances from residues 76–87. The continuous, relatively intense, NH to NH sequential NOEs observed from E76 to N82 clearly show that this stretch of the protein forms a short helix.

consistent with their involvement in the interstrand hydrogen bonds shown in Figure 6. Similarly, the low-field αCH shifts of phenylalanine 36 (5.86 ppm), cysteine 46 (5.99 ppm), phenylalanine 47 (5.74 ppm), phenylalanine 85 (5.95 ppm), cysteine 95 (6.12 ppm), and phenylalanine 96 (5.61 ppm) support their location in an antiparallel β-sheet (Clayden & Williams, 1982; Pastore & Saudek, 1990).

The other notable element of secondary structure that can identified in both trefoil domains of PSP, on the basis of the sequential and medium-range NOE patterns, is a type I β-turn located near the tip of loop 1. The precise position of the turn in the two domains differs by one residue. However, this is probably simply because of some local structural rearrangements, necessitated by the fact that loop 1 from domain I contains one more residue than the trefoil domain concensus sequence.

Conclusions. The ^1H NMR studies of PSP clearly show that the protein's two trefoil domains adopt essentially the same supersecondary structure in solution, the main feature being a seven-residue helix at the C-terminus of loop 2 immediately followed by the first four amino acid strand of an antiparallel β-sheet. The two strands of the β-sheet are separated by a seven-residue loop, which appears to have no regular secondary structure. Previously, on the basis of the disulfide configuration determined for PSP and amino acid sequence comparisons, the trefoil domain was proposed to be a new independent structural motif, distinct from several other small, disulfide cross-linked, growth factor associated domains, such as those found in EGF and IGF-I (Thim, 1989). The NMR-deter-

Table I: ^{1}H Resonance Assignments for PSP[a]

residue	NH	α	α′	β	β′	γ	γ′	δ	δ′	ε	other
p-Glu 1		4.29		2.49	2.01	2.38					
Lys 2	8.14	4.53									
Pro 3		4.44									
Ala 4	8.57	4.15		1.43							
Ala 5	8.20	3.34		0.57							
Cys 6	9.66										
Arg 7	7.89	3.96		2.05							
Cys 8	7.89	5.20		3.51	3.06						
Ser 9	8.70	4.19		3.81	3.68						
Arg 10	7.14	4.42									
Gln 11	6.92	4.60		1.98	1.86	2.19	2.13				8.01, 7.37 (NH_2)
Asp 12	9.40	4.64		2.75	2.65						
Pro 13								4.04	3.76		
Lys 14	8.33	4.01		1.73		1.37		1.65		2.96	
Asn 15	8.13	4.89		3.14	2.83						8.30, 6.72 (NH_2)
Arg 16	6.96	3.99		1.46							
Val 17	9.50	4.22		1.99		1.08					
Asn 18	8.81	4.45		2.70							
Cys 19	8.57	4.72		3.12	2.56						
Gly 20		4.00	3.65								
Phe 21	6.19	5.11		3.18	3.08						7.15 (2,6), 7.28 (3,5), 7.31 (4)
Pro 22								3.94	3.80		
Gly 23		4.27	3.72								
Ile 24	7.41	3.85		0.68		1.47	1.30	0.60			0.91 (γCH_3)
Thr 25	7.87	4.40		4.54		1.33					
Ser 26											
Asp 27	7.89	4.46		2.61	2.53						
Gln 28	7.53	3.96		2.22	1.90	2.43					7.50, 6.83 (NH_2)
Cys 29	8.44	3.90		3.08	2.81						
Phe 30	8.49	5.36		3.12	2.94						7.61 (2,6), 7.48 (3,5)
Thr 31	8.37	3.92		4.30		1.33					
Ser 32	7.61	4.52		4.13	4.05						
Gly 33	8.04	4.19									
Cys 34	7.76	4.54		3.31	2.71						
Cys 35	8.24	4.26		2.15	0.62						
Phe 36	6.39	5.86		2.33	1.93						6.97 (2,6), 7.34 (3,5), 7.46 (4)
Asp 37	9.13	4.64		3.01	2.66						
Ser 38	8.38	2.49									
Gln 39	8.08	3.90		2.26	2.16	2.46	2.31				7.26, 6.54 (NH_2)
Val 40	7.98	4.52		1.83		0.91	0.83				
Pro 41		4.81		2.40	2.00	2.10		3.83			
Gly 42	8.73	3.95	3.62								
Val 43	7.11	4.72		2.14		0.72	0.58				
Pro 44		4.18		2.43	1.77	1.86		3.73	3.42		
Trp 45	8.19	5.07		2.86	2.26						9.40 (1), 7.09 (2), 7.65 (4), 7.29 (5), 7.31 (6), 7.45 (7)
Cys 46	8.13	5.99		3.29	2.66						
Phe 47	9.47	5.74		3.16	2.59						6.85 (2), 7.04 (3), 7.19 (4), 7.59 (5), 7.25 (6)
Lys 48	8.60	4.70		1.97	1.84	1.60	1.49	1.78		3.08	
Pro 49		4.45		1.99	1.90	2.10	1.71	3.67			
Leu 50	8.58	4.53		1.53	1.47	1.73		0.97	0.65		
Pro 51		4.41		2.38	1.80	2.08		3.89	3.64		
Ala 52	8.39	3.96		1.28							
Gln 53	7.40	4.65		1.64		2.35					
Glu 54	8.81	3.98		2.08		2.31					
Ser 55	7.59	4.73		3.83	3.77						
Glu 56	8.98	4.02		2.02		2.29	2.22				
Glu 57	8.84	4.37									
Cys 58	7.53	4.94		3.33	2.84						
Val 59	7.40	3.91		2.13		1.08	0.73				
Met 60	7.57	4.25		2.25	1.95						
Gln 61	9.24	4.22		2.16	1.97	2.68	2.40				
Val 62	8.94	3.37		2.01		1.01	0.92				
Ser 63	8.01	4.16		4.00	3.87						
Ala 64	7.48	4.38		1.50							
Arg 65	6.99	4.20		1.65		1.94		3.25			
Lys 66	9.59	4.68		2.10		1.52		1.72		3.01	
Asn 67	8.88	4.46		2.89	2.74						
Cys 68	9.04	4.41		2.59	1.41						
Gly 69	8.10	4.16	3.29								
Tyr 70	6.39	5.02		3.11							6.92 (2,6), 6.73 (3,5)
Pro 71								3.92	3.78		
Gly 72		4.24	3.74								
Ile 73	7.45	3.90		0.67		1.63	1.43	0.55			1.01 (γCH_3)
Ser 74	8.47	4.72									
Pro 75		4.37		2.31							
Glu 76	8.73	3.93		1.97		2.39					
Asp 77	7.92	4.54		2.75	2.57						

Table I (Continued)

residue	NH	α	α′	β	β′	γ	γ′	δ	δ′	ε	other
Cys 78	8.26	3.98		3.16	2.88						
Ala 79	8.20	4.22		1.49							
Ala 80	7.90	4.22		1.59							
Arg 81	7.17	4.37		1.99	1.83	1.89	1.77	3.23			
Asn 82	8.27	4.34		3.09	3.01						7.41, 6.79 (NH_2)
Cys 83	7.19	4.40		3.13	2.54						
Cys 84	8.48	4.16		1.96	1.07						
Phe 85	8.55	5.95		3.00	2.71						7.08 (2,6), 7.32 (3,5), 7.48 (4)
Ser 86	8.91	4.30		4.12	3.56						
Asp 87	8.08	2.59		1.33	1.14						
Thr 88	6.95	3.78		4.12		1.21					
Ile 89	7.76	4.71		1.67		1.43	1.12	0.79			1.04 (γCH_3)
Pro 90		4.79		2.55	2.11	2.16		3.97	3.87		
Glu 91	8.82	3.80		2.15		2.24					
Val 92	7.30	4.66		2.16		0.85	0.67				
Pro 93		4.17		2.28	1.61	1.89		3.78	3.40		
Trp 94	8.06	5.13		2.88	2.41						9.86 (1), 7.06 (2), 7.85 (4), 7.36 (5), 7.36 (6), 7.44 (7)
Cys 95	8.51	6.12		3.12	2.66						
Phe 96	9.26	5.61		3.13	2.60						6.68 (2), 6.91 (3), 7.15 (4), 7.55 (5), 7.33 (6)
Phe 97	8.24	4.88		3.39	3.00						7.38 (2,6), 7.52 (3,5), 7.48 (4)
Pro 98		4.68		2.04				4.10	3.95		
Met 99	9.01	4.71									
Ser 100	8.41	4.44		4.08	3.87						
Val 101	8.24	4.36		2.52		1.06	0.93				
Glu 102	8.43	4.09		2.04		2.30					
Asp 103	8.33	4.56		2.67							
Cys 104	7.70	4.39		3.14	2.68						
His 105		4.59		3.36	3.26						8.65 (2), 7.39 (4)
Tyr 106	8.42	4.70		3.63	2.78						7.37 (2,6), 6.83 (3,5)

[a] The ^1H chemical shifts (ppm) are referenced to DSS (sodium 4,4-dimethyl-4-silapentane-1-sulfonate).

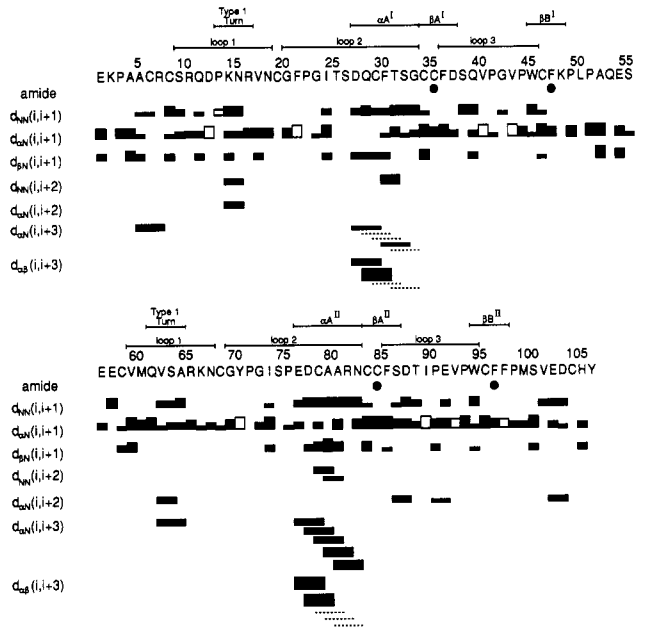

FIGURE 5: Summary of the sequential ($i,i+1$) and medium-range ($i,i<5$) NOEs identified for PSP. The height of the bar used to represent an NOE is an indication of its intensity, while open bars are shown for NOEs involving proline δ-protons. The dashed lines represent possible medium-range NOEs which could not be unambiguously identified due to resonance overlap. Closed circles below residues denote backbone amide protons that persist for many months in D_2O solutions. The positions of the main elements of regular secondary secondary in PSP, identified from the characteristic patterns of sequential, medium-range, and long-range NOEs, are shown above the protein sequence, as are the locations of the three disulfide-closed loops present in the trefoil domains.

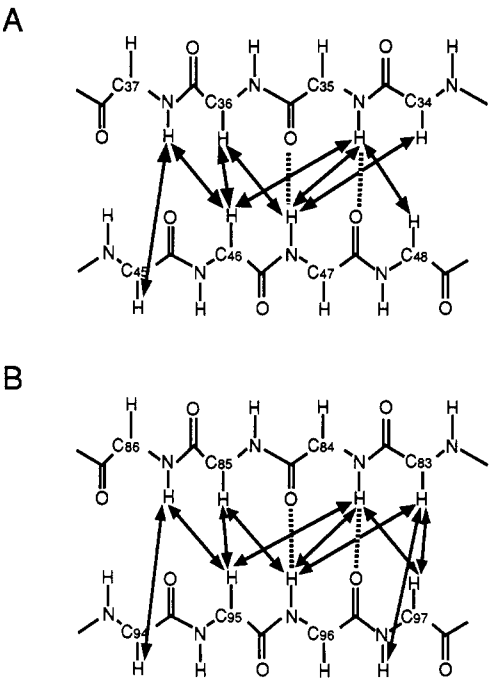

FIGURE 6: Schematic representation of the arrangement of the short antiparallel β-sheets found in domains I (A) and II (B) of PSP. The observed interstrand NOEs are indicated by arrows, while hydrogen bonds implied by very slowly exchanging amide protons and NOEs are denoted by broken lines.

mined secondary structure of PSP clearly shows this to be the case. The sequence-specific ^1H resonance assignments reported here will now form the basis of a complete three-dimensional structural determination of PSP, based on analysis of 2D NOESY and 3D NOESY-HOHAHA (Oschkinat et al., 1988; Vuister et al., 1988) spectra.

ACKNOWLEDGMENTS

The NMR spectra were recorded using facilities at the Medical Research Council Biomedical NMR Centre, National Institute for Medical Research, Mill Hill. I thank Drs. A. Lane, N. Wright, and L. Thim for many useful discussions about PSP and Drs. C. Bauer and T. Frenkiel for help in

optimizing the NMR experiments. Finally, I also thank Dr. J. Feeney for his constant encouragement and support.

Supplementary Material Available

Two figures showing the aliphatic and NH to aliphatic regions from HOHAHA spectra of PSP (2 pages). Ordering information is given on any current masthead page.

References

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.
Billeter, M., Qian, Y., Otting, G., Muller, M., Gehring, W. J., & Wüthrich, K. (1990) *J. Mol. Biol. 214*, 183–197.
Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521–528.
Breg, J. N., Van Opheusden, J. H. J., Burgering, M. J. M., Boelens, R., & Kaptein, R. (1990) *Nature 346*, 586–589.
Brown, A. M. C., Jeltsch, J. M., Roberts, M., & Chambon, P. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 6344–6398.
Brown, S. C., Weber, P. L., & Mueller, L. (1988) *J. Magn. Reson. 77*, 166–169.
Clayden, N. G., & Williams, R. J. P. (1982) *J. Magn. Reson. 49*, 383–396.
Clore, G. M., Appella, E., Yamada, M., Matsushima, K., & Gronenborn, A. M. (1990) *Biochemistry 29*, 1689–1696.
Clore, G. M., Wingfield, P. T., & Gronenborn, A. M. (1991) *Biochemistry 30*, 2315–2323.
Cooke, R. M., Wilkinson, A. J., Baron, M., Pastore, A., Tappin, M. J., Campbell, I. D., Gregory, H., & Sheard, B. (1987) *Nature 327*, 339–341.
Cooke, R. M., Harvey, T. S., & Campbell, I. D. (1991) *Biochemistry 30*, 5484–5491.
Davis, D. G., & Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2820–2821.
Foekens, J. A., Rio, M.-C., Seguin, P., Van Putten, W. L. J., Fauque, J., Nap, M., Klijn, J. G. M., & Chambon, P. (1990) *Cancer Res. 50*, 3832–3837.
Gregory, H., & Preston, B. M. (1977) *Int. J. Pept. Protein Res. 9*, 107–118.
Hard, T., Kellenbach, E., Boelens, R., Maler, B. A., Dahlman, K., Freedman, L. P., Carlstedt-Duke, J., Yamamoto, K. R., Gustafsson, J.-A., & Kaptein, R. (1990) *Science 249*, 157–160.
Hoffmann, W. (1988) *J. Biol. Chem. 263*, 7686–7690.
Hoosein, N. M., Thim, L., Jorgensen, K. H., & Brattain, M. G. (1989) *FEBS Lett. 247*, 303–306.
Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.
Jorgensen, K. D., Diamant, B., Jorgensen, K. H., & Thim, L. (1982) *Regul. Pept. 3*, 231–243.
Macura, S., Huong, Y., Suter, D., & Ernst, R. R. (1981) *J. Magn. Reson. 43*, 259–281.
Masiakowski, P., Breathnach, R., Bloch, J., Gannon, F., Krust, A., & Chambon, P. (1982) *Nucleic Acids Res. 10*, 7895–7903.
Mori, K., Fujii, R., Kida, N., Ohta, M., & Hayaski, K. (1988) *Biochem. Biophys. Res. Commun. 155*, 366–372.
Oschkinat, H., Griesinger, C., Kraulis, P. J., Sorensen, O. W., Ernst, R. R., Gronenborn, A. M., & Clore, G. M. (1988) *Nature 332*, 374–376.
Pastore, A., & Saudek, V. (1990) *J. Magn. Reson. 90*, 165–176.
Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.
Rinderknecht, E., & Humbel, R. E. (1978) *J. Biol. Chem. 253*, 2769–2776.
Rio, M.-C., Bellocq, J. P., Gairard, B., Rasmussen, U. B., Krust, A., Koehl, C., Calderoli, H., Schiff, V., Renaud, R., & Chambon, P. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 9243–9247.
Rio, M.-C., Lepage, P., Diemunsch, P., Roitsch, C., & Chambon, P. (1988a) *C. R. Acad. Sci., Ser. 2 307*, 825–832.
Rio, M.-C., Bellocq, J. P., Daniel, J. Y., Tomasetto, C., Lathe, R., Chenard, M. P., Batzenschlager, A., & Chambon, P. (1988b) *Science 241*, 705–708.
Schwabe, J. W. R., Neuhaus, D., & Rhodes, D. (1990) *Nature 348*, 458–461.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.
Thim, L. (1989) *FEBS Lett. 250*, 85–90.
Thim, L., Thomsen, J., Christensen, M., & Jorgensen, K. H. (1985) *Biochim. Biophys. Acta 827*, 410–418.
Tomasetto, C., Rio, M.-C., Gautier, C., Wolf, C., Hareuveni, M., Chambon, P., & Lathe, R. (1990) *EMBO J. 9*, 407–414.
Vuister, G. W., Boelens, R., & Kaptein, R. (1988) *J. Magn. Reson. 80*, 176–185.
Wagner, G., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 347–366.
Wright, N. A., Poulsom, R., Stamp, G. W. H., Hall, P. A., Jeffery, R. E., Longcroft, J. M., Rio, M.-C., Tomasetto, C., & Chambon, P. (1990) *J. Pathol. 162*, 279–284.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.