J. Mol. Biol. (1997) **267**, 418–432

JMB

High-resolution Solution Structure of Human pNR-2/pS2: A Single Trefoil Motif Protein

Vladimir I. Polshakov[1], Mark A. Williams[2], Angelo R. Gargaro[2] Thomas A. Frenkiel[3], Bruce R. Westley[4], Mark P. Chadwick[4] Felicity E. B. May[4] and James Feeney[2]*

[1]*Laboratory of Physical Methods, Chemical-Pharmaceutical Research Institute, Moscow 119815, Russia*

[2]*Molecular Structure Division National Institute for Medical Research, Mill Hill, London NW7 1AA, UK*

[3]*MRC Biomedical NMR Centre, National Institute for Medical Research, Mill Hill London NW7 1AA, UK*

[4]*Department of Pathology, Royal Victoria Infirmary University of Newcastle Newcastle NE1 4LP, UK*

**Corresponding author*

pNR-2/pS2 is a 60 residue extracellular protein, which was originally discovered in human breast cancer cells, and subsequently found in other tumours and normal gastric epithelial cells. We have determined the three-dimensional solution structure of a C58S mutant of human pNR-2/pS2 using 639 distance and 137 torsion angle constraints obtained from analysis of multidimensional NMR spectra. A series of simulated annealing calculations resulted in the unambiguous determination of the protein's disulphide bonding pattern and produced a family of 19 structures consistent with the constraints. The peptide contains a single ''trefoil'' sequence motif, a region of about 40 residues with a characteristic sequence pattern, which has been found, either singly or as a repeat, in about a dozen extracellular proteins. The trefoil domain contains three disulphide bonds, whose 1-5, 2-4 and 3-6 cysteine pairings form the structure into three closely packed loops with only a small amount of secondary structure, which consists of a short α-helix packed against a two-stranded antiparallel β-sheet. The structure of the domain is very similar to those of the two trefoil domains that occur in porcine spasmolytic polypeptide (PSP), the only member of the trefoil family whose three-dimensional structure has been previously determined. Outside the trefoil domain, which forms the compact ''head'' of the molecule, the N and C-terminal strands are closely associated, forming an extended ''tail'', which has some β-sheet character for part of its length and which becomes more disordered towards the termini as indicated by $^{15}N\{^1H\}$ NOEs. We have considered the structural implications of the possible formation of a native C58-C58 disulphide-bonded homodimer. Comparison of the surface features of pNR-2/pS2 and PSP, and consideration of the sequences of the other human trefoil domains in the light of these structures, illuminates the possible role of specific residues in ligand/receptor binding.

Keywords: protein structure; trefoil motif; pS2; pNR-2; nuclear magnetic resonance

Introduction

pNR-2/pS2 is a small secreted protein (6.5 kDa) that was discovered originally by virtue of its oestrogen regulation in cultured breast-cancer cells (Masiakowski *et al*., 1982; Prud'homme *et al*., 1985; May & Westley, 1986, 1988). Protein sequence studies show that it belongs to the ''trefoil'' family of peptides (Jørgensen *et al*., 1982a,b; Thim, 1989; Sands & Podolsky, 1996). These peptides consist of one or more trefoil or P domains, each of which comprises ~40 amino acid residues with well-con-

V.I.P. and M.A.W. contributed equally to this work.

Abbreviations used: COSY, correlated spectroscopy; DQF, double quantum filtered; HSQC, hetronuclear single quantum coherence; ^3J, three bond coupling constant; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; rms, root-mean-square; ROESY, rotating frame Overhauser effect spectroscopy; S_{rep}, representative structure; TOCSY, total correlation spectroscopy; UACL, ulcer-associated cell lineage; PSP, porcine pancreatic spasmolytic polypeptide; 3D, three-dimensional; 2D, two-dimensional; PFG, pulsed-field gradient.

0022–2836/97/120418–15 $25.00/0/mb970896

served sequence features. Members of the family are known that contain up to six trefoil domains. The most highly conserved features of the domains are six cysteine residues with essentially conserved spacings and a conserved arginine residue located between the first two cysteine residues. pNR-2/pS2 is one of the members of the family that has a single domain and a seventh cysteine residue close to the C terminus, capable of forming intermolecular disulphide bonds (Jakowlew *et al*., 1984; Piggott *et al*., 1991).

Three human trefoil peptides have been identified so far. Two with single domains, pNR-2/pS2 and hITF (human intestinal trefoil factor; Podolsky *et al*., 1993; Hauser *et al*., 1993), and one with a double domain, hSP (human spasmolytic polypeptide; Tomasetto *et al*., 1990). pNR-2/pS2 and hSP are expressed throughout the gastric mucosa in foveolar epithelial cells and in mucous neck cells (Rio *et al*., 1988; Hanby *et al*., 1993), whereas hITF is expressed predominantly in the goblet cells of the intestine (Podolsky *et al*., 1993; Hauster *et al*., 1993). Trefoil peptides are expressed at lower levels in other sites and in tissue related to various pathological conditions. All three are expressed in the ulcer-associated cell lineage (UACL), which is involved in the repair of mucosal damage in the gastrointestinal tract (Wright *et al*., 1990, 1993). pNR-2/pS2 is expressed at low levels in normal breast epithelium (Piggot *et al*., 1991; Hahnel *et al*., 1992) and is found at high levels in breast carcinoma, where it is a marker of hormonal responsiveness (Henry *et al*., 1989, 1991a; Schwartz *et al*., 1991). It is associated with a variety of other tumours, including those of the pancreas, large intestine, stomach, endometrium and ovary (Henry *et al*., 1991b; Lefebvre *et al*., 1996).

The functions of the trefoil peptides are not yet known. Early studies suggested that they were involved in controlling intestinal smooth muscle contraction and gastric acid secretion (Jørgensen *et al*., 1982b). Possible roles as growth factors have been suggested, based on their apparent growth-stimulating effects in certain cell lines (Hoosein *et al*., 1989). The observation that they are expressed at high levels in the UACL appeared to implicate them as possible agents used in the repair of damaged endodermal tissue (Wright *et al*., 1990, 1993). More recent studies have shown that trefoil peptides can modulate the effect of epidermal growth factor on epithelial ion transport (Chinery & Cox, 1995), may be involved in the control of cell migration (Dignass *et al*., 1994; Playford *et al*., 1995; Williams *et al*., 1996) and can protect against certain types of damage to the mucosal surface of the gastrointestinal tract (Chinery & Playford, 1995; Playford *et al*., 1996; Mashimo *et al*., 1996). The observation that they are copackaged with mucous-secreting cells (Podolsky *et al*., 1993) has prompted the suggestion that they are involved in the processing of mucin glycoproteins or mucous structures.

There is considerable interest in determining the three-dimensional structures of the trefoil peptides, since this information could help in elucidating their function by identifying conserved structural elements that may be involved in mediating their biological effects. The structure of porcine pancreatic spasmolytic polypeptide (PSP), which contains two trefoil domains in a single peptide, has been determined using X-ray crystallography (Gajhede *et al*., 1993; De *et al*., 1994) and NMR (Carr *et al*., 1994). These studies showed that the two domains have a similar tertiary structure, each containing a two-stranded antiparallel β-sheet and a short α-helix. Several conserved residues were found to be located on a single face of each domain. The N and C termini of PSP are linked by an interdomain disulphide bond, which helps to fix the relative orientation of the two domains.

In contrast, the structures of single-domain trefoil peptides are not known. In an earlier study we produced a recombinant pNR-2/pS2 and used NMR measurements to characterise some of its structural features (Polshakov *et al*., 1995b). An almost complete set of sequence-specific ^{1}H assignments was obtained and the protein was reported to contain secondary structure elements that are similar to those found for the trefoil domain of PSP, containing two short antiparallel β-strands (32 to 35 and 43 to 46), and an α-helix (25 to 30). The measurements were made on a variant of pNR-2/pS2 that had Cys58 replaced by serine (Chadwick *et al*., 1995). This residue is located outside the trefoil domain and the mutation was made to avoid complications arising from the presence of an unpaired cysteine residue. Here, we have analysed data from ^{1}H and ^{15}N-NMR measurements on this pNR-2/pS2 protein and determined its high-resolution three-dimensional structure in solution, including the pattern of the disulphide bonding. This is the first structure determination of a human trefoil peptide.

Results and Discussion

Resonance assignments

Previously, about 98% of the ^{1}H resonances of the pNR-2/pS2 protein had been assigned using homonuclear and ^{13}C-heteronuclear experiments carried out on samples with ^{13}C at natural abundance (Polshakov *et al*., 1995b). Here, the ^{1}H assignments have been extended to include many stereospecific assignments, and a complete set of ^{15}N amide resonance assignments has been obtained from analysis of 3D NOESY-HSQC and HNHB spectra.

Distance constraints

Three types of distance constraints were used in the final structure calculations; namely, NOE-based interproton distance constraints, hydrogen bonds and disulphide linkages. However, hydrogen

bonds and disulphide linkages were introduced into the calculations only after their presence had been established in the structures generated using sets of constraints that did not include them.

Interproton distance constraints were obtained from NOEs measured from 2D NOESY and ROESY and 3D ^{15}N-^{1}H-NOESY-HSQC spectra. Initially, unambiguous assignments were made using chemical shift data alone. A set of about 500 distance constraints derived from these assignments was then used in the generation of a set of initial structures. The initial structures were used to assist in assigning more NOEs by identifying nuclei separated by less than 6.0 Å. Further assignments were made in an iterative manner, calculating structures, checking for violations of constraints, and adding new constraints when the structures suggested new possible assignments. Ambiguous NOE distance constraints were often used in this iterative process and were especially important for the refinement of two regions of the protein structure around residues 3 to 6 and 49 to 52, and 15 to 18 and 44 to 45, where many of the H^{α} and H^{β} signals have degenerate chemical shifts and where the residues are close to each other.

Torsion angle constraints and stereospecific assignments

The ϕ, ψ and χ torsion constraints and the stereospecific assignments for most of the prochiral atoms and groups of the pNR-2/pS2 protein were obtained from the analysis of ^{3}J and NOE-based distance data using the program AngleSearch (Polshakov *et al*., 1995a). A combination of information from preliminary structures and the results of the AngleSearch calculations (see Materials and Methods for details) allowed us to obtain 137 torsion angle constraints (46 ϕ, 49 ψ, 38 χ_1, 2 χ_2, 1 χ_3 and 1 χ_4 values) and stereospecific assignments for three of the four glycine residues, 25 of the 40 pairs of β-methylene protons and ten pairs of H^{γ} and H^{δ} protons. Most of the residues for which stereospecific assignments were not obtained either have methylene protons with degenerate chemical shifts or are situated in the random-coil part of the protein (residues 53 to 60).

For some residues, such as Arg14 and Ile50, all the side-chain torsion angles could be determined. For the important Arg14 residue, all the protons gave well-resolved signals, and many intraresidue NOEs could be collected for use in the AngleSearch calculations. In this way, the side-chain conformations and stereospecific assignments for all side-chain methylene protons of the Arg14 residue were determined. Additionally, an HNHA experiment provided estimates of $^{3}J(H^{\varepsilon}, H^{\delta})$ and an HNHB experiment gave estimates of the $^{3}J(N^{\varepsilon}, H^{\gamma})$ and $^{3}J(N^{\eta}, H^{\varepsilon})$ coupling constants. The χ_2, χ_3 and χ_4 torsion angles obtained for Arg14 using all the available data indicated that its side-chain is in an extended *trans* conformation.

Hydrogen bonds and disulphide linkage constraints

Hydrogen bonds were used as constraints only after the acceptor partners for the slowly exchanging NH protons (Cys17, Cys27, Ala28, Asn29, Cys33, Asp35, Phe45 and Asn48; Polshakov *et al*., 1995b) had been identified from preliminary families of structures calculated without these hydrogen bond constraints. Eight hydrogen bonds were included in the final structure calculations as 16 distance constraints.

The cysteine partners in the disulphide bonds were determined to be Cys7-Cys33, Cys17-Cys32 and Cys27-Cys44 by examining the intersulphur distances for all six Cys residues in 20 structures obtained using the 714 constraints determined at that time, but without using disulphide linkage constraints. The Cys7-Cys33 and Cys27-Cys44 S-S distances are <3.5 Å in, respectively, 19 and 15 of the structures in this set with no other close partner for either residue 7 or 27. Consequently, although Cys17 is close to both Cys32 and Cys33 in several structures, the Cys17-Cys32 linkage is the only remaining possibility. The disulphide bonds determined in this way were then introduced into the final stages of the structure refinement. The pattern of disulphide linkages determined in these NMR experiments is the same as that previously determined for PSP using chemical methods (Thim, 1989) and X-ray crystallography (Gajhede *et al*., 1993; De *et al*., 1994). However, this pattern of disulphide bonds differs from that found for a sample of pNR-2/pS2 prepared by peptide synthesis and folded *in vitro* in the presence of glutathione (Rye *et al*., 1994). This indicates that the synthetic peptide has not folded in the same way as the recombinant Ser58 or the native pNR-2/pS2 peptide.

Family of calculated structures

A total of 776 constraints was used in the final structure calculations, consisting of 582 unambiguous and 41 ambiguous NOE interproton distance constraints, together with 137 torsion angle constraints and 16 distance constraints associated with the eight hydrogen bonds. This corresponds to an average of 17.3 NOE-derived distance constraints per residue. However, as can be seen in Figure 1a, the NOE-derived distance constraints are concentrated in the region of the trefoil motif, resulting in an average of 21.6 constraints per residue in the 7 to 47 range.

Nineteen of the 20 structures generated using our simulated annealing protocol and these constraints, satisfied our acceptance criteria (see Materials and Methods for details). A stereoview of the backbones of these 19 structures of pNR-2/pS2 is given in Figure 2. The structural statistics and atomic rms differences for the ensemble of the final family of simulated annealing structures are given in Table 1. The backbone rms differences were cal-

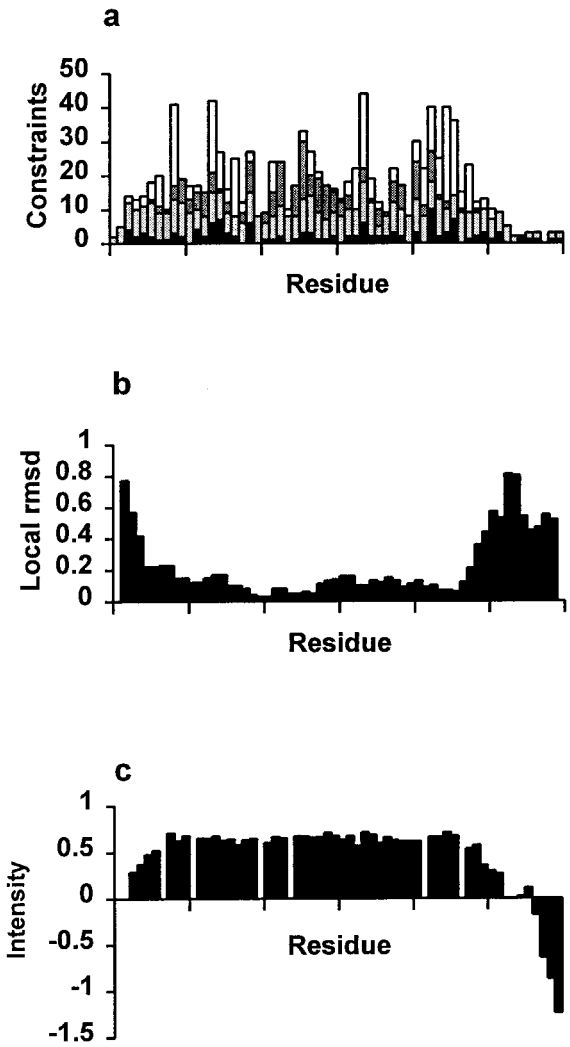

Figure 1. Summary of structural data *versus* residue number (ticks on the horizontal axes mark every 10th residue). a, The distribution of unambiguous NOE constraints for the residues of pNR-2/pS2. Intraresidue, sequential, medium- $(1 < |i - j| < 5)$ and long-range $(|i - j| > 4)$ constraints are represented by black, light grey, dark grey and white blocks, respectively. b, the local 3 residue average rms difference for the heavy backbone atoms (N, C^{α} and C) across the final family of structures (Å). c, The measured $^{15}N\{^{1}H\}$NOE values at 298 K and 14.1 Tesla (600 MHz for ^{1}H) for ^{15}N in the amide groups of pNR-2/pS2. Lower values indicate higher mobility.

culated for the trefoil region (residues 7 to 47) of all pairs of structures in the family and a representative structure, S_{rep}, was chosen that had the minimum mean rms difference from all other members of the family. When the backbone atoms in the peptide sequence 7 to 47 for each of the members of the family are superimposed on this representative structure, the mean atomic rms difference is small (0.48 Å), indicating that the structure is well defined in the region of the trefoil motif. The rms difference between structures is increased substantially with the inclusion of residues 3 to 6 and 48 to 52 (1.12 Å) and no sensible superposition can be made of the full sequence because of the highly disordered termini. The close correspondence between the degree of positional disorder of each residue of the terminal strands across the family of structures (Figure 1b) and their actual highly mobile behaviour observed *via* the $^{15}N\{^{1}H\}$ NOE data (Figure 1c) suggests that the structural disorder observed in the tail is not merely an artifact due to lack of assigned NOEs for these residues, but reflects real diversity in the solution structure of the protein.

The Ramachandran plot of the ϕ and ψ dihedral angles for the residues in the representative structure S_{rep} shows that 87% of its residues have angles in the most favourable regions of the plot (Figure 3). Analysis of the family of 19 structures identified no residue in the structured region of the protein with dihedral angles (ϕ, ψ, χ_1 and χ_2) in disallowed regions for any member of the family. PROCHECK (Laskowski *et al.*, 1993) analysis of S_{rep} indicates that all its overall structural characteristics, except the pooled standard deviation of the χ_2 angles, are as good or better than a typical 2.0 Å resolution crystal structure.

Structure of pNR-2/pS2

Of the 60 residues of pNR-2/pS2, only the 50 between positions 3 and 52 have any ordered structure in solution (Figure 4). The three disulphide bonds in the trefoil domain (7 to 47) form it into three sequential loops, which are packed together with the third loop sandwiched between the first and second. The protein has only a small amount of secondary structure, which consists of an α-helix (23 to 31) packed against a two-stranded antiparallel β-sheet (33 to 35, 43 to 45). To this extent the secondary structure of this trefoil domain confirms that determined earlier from limited data (Polshakov *et al.*, 1955b). The present data have additionally revealed the detailed structure of the turns and terminal regions. Residues 10 to 14 are found to form either a composite turn or a 3_{10}-helix in roughly equal proportion among members of the family, a type-I β-turn is formed by residues 19 to 22, and the turn between the two β-strands is predominantly a type-VI β-turn (residues 38 to 42). Outside the trefoil domain, which forms the compact ''head'' of the molecule, the N and C-terminal strands are closely associated, forming an extended ''tail'', which has some β-sheet type interactions between residues in strands 3 to 7 and 47 to 51, and which becomes more disordered towards the termini.

Interstrand NOEs between the α-protons of residues 7 and 47, 5 and 49, and 3 and 51 have been identified in both NOESY and ROESY spectra, indicating the presence of some β-sheet type interactions between these two strands. The H^{α}-H^{α} NOE intensity for the 7-47 interactions is large and furthermore the amide NH group of Asn48 has

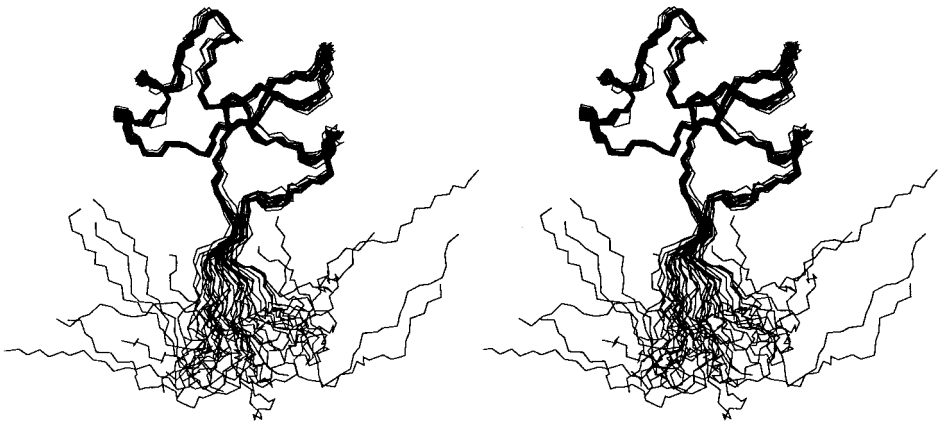

Figure 2. Stereoview of the superimposed backbone atoms (N, C^{α} and C) for the final family of 19 structures. The superimpositions were made onto the backbone atoms of the core residues (7 to 47) of a representative structure (S_{rep}) chosen to have the minimum mean atomic rms difference from all other members of the family. The conformation of the C terminus is not well defined with respect to the rest of the structure.

been shown to be protected against exchange with water protons (Polshakov *et al.*, 1995b), consistent with it forming a hydrogen bond with the carbonyl oxygen atom of Thr6. The intensities of the inter-strand H^{α}-H^{α} NOEs for the pairs 5-49 and 3-51 are about 50% and 25%, respectively, of the values expected if the structure adopted β-sheet interactions of the usual kind and it seems likely that there is an equlibrium involving exchange between a population of β-sheet structures and other conformations. None of the amide NH protons of the residues 3, 4, 5, 49, 50 or 51 are protected against exchange and this would be consistent with the equilibrium mixture proposed, since exchange would occur in the non-β-sheet population. Additionally the $^{3}J_{HN\alpha}$ coupling constants of these residues lie in the range 6.3 to 8.3, consistent with an extended, but not perfect, β conformation. These β-sheet interactions were not noted in the earlier

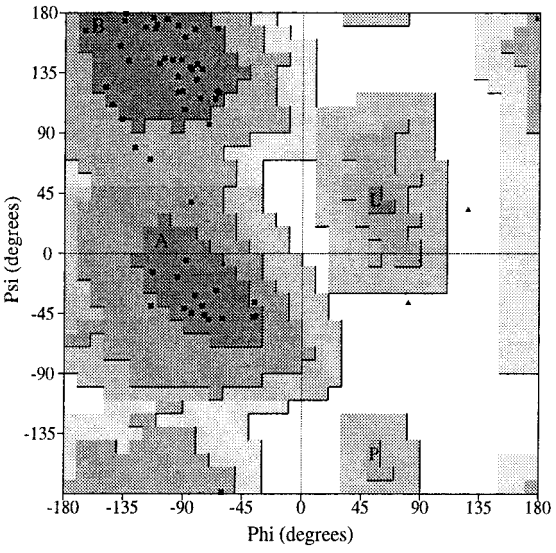

Figure 3. A Ramachandran plot of the ϕ and ψ torsion angles in the S_{rep} structure of pNR-2/pS2 generated using PROCHECK (Laskowski *et al.*, 1993). The background of the plot is shaded according to the probability of observing a non-Gly and non-Pro residue in that locality in a high-resolution crystal structure, regions with the most heavily shaded background are most favoured, and the lightest regions are regarded as disallowed. Small triangles represent glycine residues.

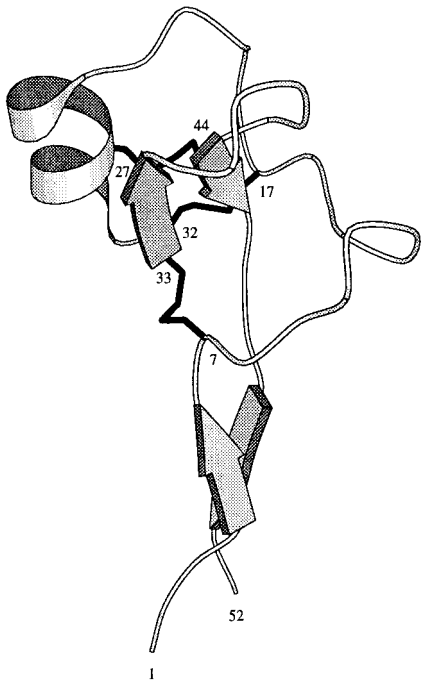

Figure 4. A representation of the structure of pNR-2/pS2 (residues 1 to 52) produced using MOLSCRIPT (Kraulis, 1991). The short anti-parallel β-sheet shown here between residues 4 to 6 and 48 to 50 is probably present in only a fraction of the solution structures at any one time (see the text for discussion). The cysteine side-chains involved in the three disulphide bonds are shown in black.

Table 1. NMR constraints and structural statistics for pNR-2/pS2

A. *Constraints used in the final structure calculation*		
Unambiguous NOES	Long-range ($\|i-j\|>4$)	152
	Medium-range ($1<\|i-j\|\leqslant 4$)	87
	Sequential ($\|i-j\|=1$)	219
	Intraresidue	124
	Total	582
Ambiguous NOES		41
Dihedral angles	ϕ	46
	ψ	49
	χ	42
Hydrogen bonds		8
Disulphide bonds		3
B. *Constraint violations in the final ensemble of 19 structures*[a]		
Number of NOE constraint violations per structure		17 ± 3 (23)
Number of NOE constraint violations >0.2 Å per structure		0
Cumulative NOE constraint violation per structure (Å)		0.40 ± 0.12 (0.61)
Number of dihedral angle violations >3° per structure		0
XPLOR energies (kcal mol^{-1})[b]	E_{NOE}	0.74 ± 0.22 (1.14)
	E_{CDIH}	0.14 ± 0.09 (0.42)
	E_{TOTAL}	84 ± 3 (93)
C. *Structural statistics for final ensemble*		
PROCHECK analysis		
% of residues in most favourable region of Ramachandran plot		83
% of residues in most favourable region of Ramachandran plot in S_{rep}		87
Number of ''significant'' bond length and angle violations in S_{rep}		0
Number of ''bad contacts'' in S_{rep}		1
Pairwise superposition (Å)[a]		
Backbone rmsd of the trefoil domain (residues 7 to 47)		0.59 ± 0.19 (1.38)
Heavy-atom rmsd of the trefoil domain		1.24 ± 0.19 (1.92)
Backbone rmsd of structured region (residues 3 to 52)		1.35 ± 0.47 (2.30)
Heavy-atom rmsd of structured region		1.91 ± 0.41 (2.86)
Superposition onto the chosen representative structure S_{rep} (Å)[a]		
Backbone rmsd of the trefoil domain		0.48 ± 0.16 (1.00)
Heavy-atom rmsd of the trefoil domain		1.17 ± 0.15 (1.67)
Backbone rmsd of structured region		1.12 ± 0.22 (1.53)
Heavy-atom rmsd of structured region		1.75 ± 0.21 (2.12)

[a] Mean ± standard deviation (maximum value)

[b] The force constants used to calculate E_{NOE} were 50 kcal mol^{-1} Å^{-2} for ''normal'' NOE constraints and 200 kcal mol^{-1} Å^{-2} for those representing hydrogen bonds. The force constant used to calculate E_{CDIH} was 200 kcal mol^{-1} rad^{-2}. E_{TOTAL} was calculated with the parameters given in Materials and Methods.

work because the NOEs defining these regions of the sequence became unambiguous only during the iterative structure-refinement procedure mentioned earlier.

The presence of the β-sheet interactions is consistent with the results of the $^{15}N\{^1H\}$ NOE measurements made for the backbone amide NH groups of each residue (Figure 1c). The residues of the protein within the structurally well defined core (residues 7 to 47) have steady-state NOEs in the range of 0.6 to 0.7, which is in good agreement with the values expected for the rigid part of a protein having a rotational correlation time $\tau_c = 4.5$ (±0.5) ns (at 298 K; Polshakov *et al.*, 1995b). In the terminal regions of pNR-2/pS2, the steady-state NOEs are smaller, indicating greater mobility of the backbone. The $^{15}N\{^1H\}$ NOEs progressively decrease along the sequences 7 to 3 and 47 to 52, indicating that the mobility increases as one moves away from the core. The pairs of residues involved in interstrand NOEs appear to have similar $^{15}N\{^1H\}$ NOEs, indicating similar mobilities, as expected for residues in an extended structure separated from the core by the same distance. The NOEs continue to decrease beyond the β-sheet-like region, and are indeed negative for residues 56 to 60, indicating that the C terminus of the protein is highly mobile.

The NOE data indicated that all the proline peptide bonds have *trans* configurations. In earlier work (Polshakov *et al.*, 1955b), Pro47 was erroneously reported to have a *cis* configuration based on misassignments of some Cys17 and Tyr46 proton signals. The AMX spin-systems of these residues were incorrectly assigned and the assignments given earlier have been interchanged.

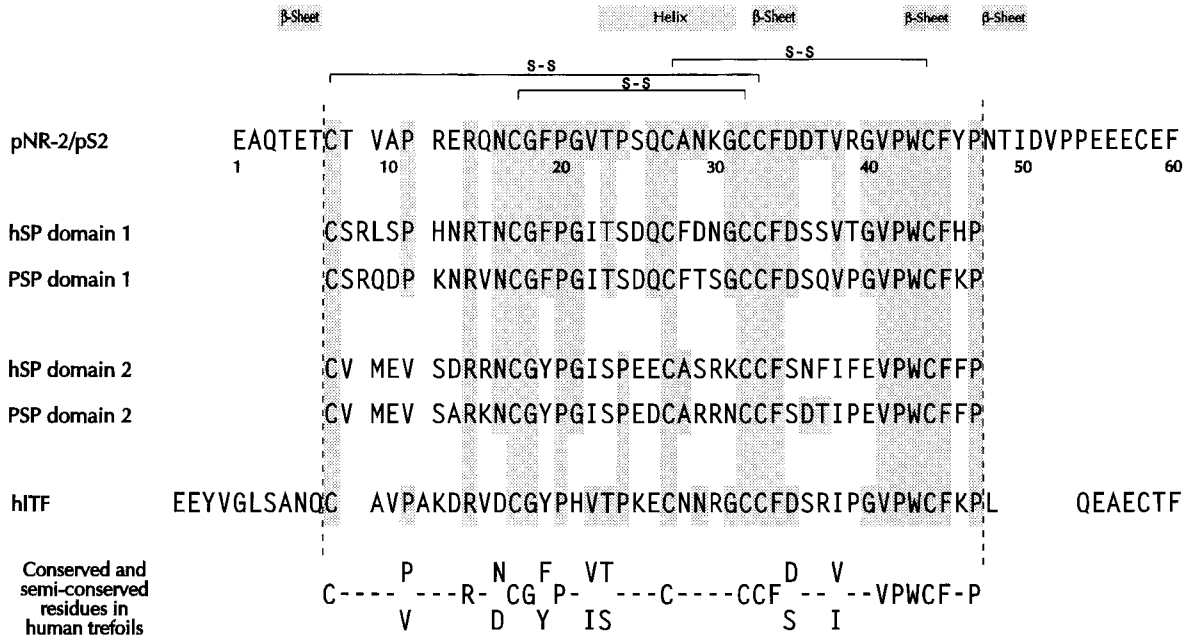

Figure 5. The amino acid sequence of pNR-2/pS2 aligned with corresponding residues in the trefoil motifs of the mammalian trefoil peptides PSP, hSP and hITF. The disulphide bonds and secondary structure elements determined in this present work are shown at the top of the diagram. Residues identical with those in pNR-2/pS2 have been shaded in the diagram. A consensus sequence motif derived from the mammalian trefoil peptides is given at the bottom of the diagram. The sequences of porcine, human, rat and mouse spasmolytic polypeptide, human and rat intestinal trefoil factor, and human and mouse pS2 were used to produce the consensus.

Comparison with the PSP structure

It is useful to compare the pNR-2/pS2 structure with that of the two trefoil domains of PSP, which is the only other trefoil protein for which detailed structural information is available. There is a high degree of sequence conservation across the mammalian trefoil peptides generally, and in particular there are 17 residues conserved in all three of these domains, with 25 conserved between pNR-2/pS2 and domain 1 of PSP, 21 conserved between pNR-2/pS2 and domain 2 of PSP, and several other residues that are conservatively mutated (Figure 5). The superposition of the trefoil domain of the pNR-2/pS2 solution structure onto both domains of the crystal structure of PSP (Gajhede *et al.*, 1993) reveals that this high degree of sequence conservation is reflected in the very similar folds of the three domains. The antiparallel β-sheet and the α and 3_{10}-helices observed in pNR-2/pS2 are found in both domains of PSP. However, no turns have identical sequences to those of pNR-2/pS2 and only the turn in domain 1 of PSP corresponding to residues 38 to 42 of pNR-2/pS2 is of the same type. The backbone rms differences for residues 9 to 47 of the trefoil domain in pNR-2/pS2 superimposed on domain 1 (1.5 Å) and on domain 2 (0.9 Å) of PSP shows the overall conformation to be more similar to that of domain 2. The principal reason for this difference is that the α-helix and its preceding turn are shifted by approximately 4 Å in domain 1 of PSP with respect to the position of the corresponding residues (17 to 31) of pNR-2/pS2 (Figure 6). The separation of the second and third loops is consequently greater in domain 1 of PSP than in pNR-2/pS2, producing a more ''open'' structure with greater solvent exposure of the conserved side-chains at the centre of the putative binding site (see later). The probable cause of the open conformation of domain 1 of PSP is the mutation in this domain of the residue equivalent to Ala28 in pNR-2/pS2 to phenylalanine. If domain 1 of PSP were to adopt the less open conformation of pNR-2/pS2, then this phenylalanine residue would be exposed to solvent, whereas in the open conformation the aromatic ring is able to interact with the residues of the C and N termini (Figure 6) to bury a substantial amount of hydrophobic surface. The energetic benefit of burying the ring locks domain 1 in the open conformation. In contrast, the second domain of PSP, in which the residue is alanine as in pNR-2/pS2, adopts a conformation closer to that of pNR-2/pS2.

Conserved features of the trefoil peptides

The amino acid sequence alignment of pNR-2/pS2, hITF (the other human single-domain trefoil peptide), PSP and its human analogue hSP shows that there is a great deal of sequence similarity between the proteins in the region of the trefoil motif (Figure 5). If we consider the characteristic physical properties (hydrophobic, hydrogen bond donor and/or acceptor, positively or negatively charged) of each of the residues of the trefoil motif in all

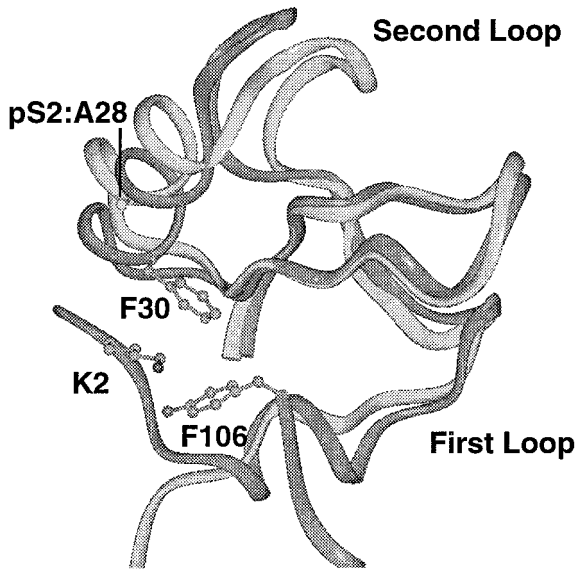

Figure 6. A superimposition of the structure for pNR-2/pS2 (light ribbon) on the corresponding part of the crystal structure of domain-1 of PSP (dark ribbon) reported by Gajhede *et al*. (1993). The superimposition was carried out using the residues of the first and third loops of the structures, and consequently clearly shows the different positions of the second loop.

mammalian trefoil peptide sequences, then at 24 positions at least one of the characteristic physical properties is absolutely conserved. Of these positions, 15 are residues with an absolutely conserved identity, seven have one of two alternative identities (these conserved and semi-conserved residues are given as a consensus sequence at the foot of Figure 5), one residue is always a hydrogen bond donor (R12 in pNR-2/pS2), and one is always able to act as an acceptor (Q26 in pNR-2/pS2). The conservation of a residue or a physical feature at a particular position suggests an essential structural and/or functional role for that residue or property.

Conserved structural features of pNR-2/pS2

As previously discussed, the conserved cysteine residues form the peptide chain into three loops, which we can consider as running from 7 to 17, 18 to 32 and 33 to 44, and which stack together with the third loop sandwiched between the first and second (Figure 4). Several of the conserved and semi-conserved residues are buried in the interfaces of these loops and are responsible for holding them in the observed compact three-dimensional arrangement. The three residues equivalent to Arg14, Phe45 and Val22 can be regarded as buried, each having less than 10% side-chain solvent accessibility in all three of the known trefoil domain structures. The low accessibility and high degree of conservation of these residues indicates that they are particularly important to the integrity of the trefoil fold.

The Arg14 NH^{η} protons are observed to be non-equivalent in the spectra, indicating the presence of hindered rotation in the guanidino group and suggesting strong interactions of some of the NH^{η} protons with neighbouring groups. The downfield shift of the protons of one of the Arg14 NH_2 groups and the upfield shift of the Arg14 NH^{ε} proton indicates strong polarisation of the guanidino group, most likely arising from an interaction of the NH_2 group with a negatively charged side-chain (Gargaro *et al*., 1996). In a few of the final family of structures the $NH^{\eta 2}$ protons are within hydrogen bonding distance of the Asp35 γ-carboxylate oxygen atoms, an interaction seen in both domains of the crystal structure of PSP (Gajhede *et al*., 1993). If present in pS2/pNR-2, this charge-charge interaction would largely explain the observed chemical shifts. Additionally, the NH^{ε} proton of Arg14 is close to the carbonyl oxygen atom of Val9 in the majority of structures, suggesting a hydrogen bond not seen in the PSP structure. None of these side-chain hydrogen bonds was used as a constraint in the structure calculations.

The aliphatic part of the side-chain of Arg14 is packed against the aromatic ring of Phe45. It was noted earlier (Polshakov *et al*., 1995b) that the aromatic ring of Phe45 is not undergoing the rapid ring-flipping at 298 K that is often observed for buried aromatic rings in proteins. The corresponding phenylalanine residues in PSP were also found not to be undergoing rapid ring-flipping at this temperature (Carr *et al*., 1994). A possible explanation for this is that hydrogen bonding interactions between Arg14 and its neighbours contribute to creating part of a very rigid pocket for the binding of the aromatic ring of Phe45, which prevents the easy transient fluctuations in the protein structure required for aromatic ring-flipping.

The buried Phe45 residue with its aromatic ring in a well-packed environment together with the interacting Arg14/Asp35 residues are obviously important contributors to the stabilisation of the common core structure found for the trefoil proteins between the first and third loops. In the interface between the second and third loops there are also many conserved residues; however, only Val22 and its corresponding residues in PSP are buried in this region, and can be confidently regarded as having a structural role. Additionally, it seems reasonable to suppose that Phe34, also in the interface between the second and third loops, is structurally important as it is the only other residue to have a side-chain solvent accessibility consistently less than 25%.

Conserved surface features of pNR-2/pS2

It is likely that the degree of conservation of the surface features of the proteins reflects the import-

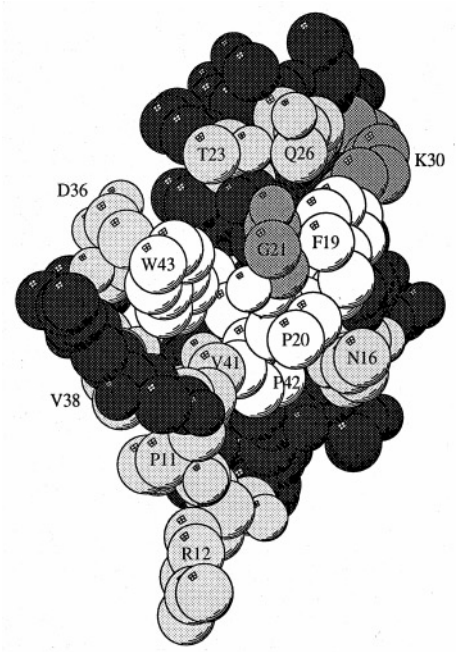

Figure 7. A CPK-type representation of pNR-2/pS2 showing the surface residues on one face of the protein. The conserved residues Phe19, Pro20, Pro42 and Trp43 postulated to form a ligand/receptor binding site are indicated in white. The physical characteristics of the flanking residues indicated in light grey are preserved in all mammalian trefoil peptides.

ance of these features in ligand/receptor recognition events. There are three residues (Pro20, Pro42 and Trp43) that are fully conserved in the known mammalian trefoil peptides, and one residue (Phe19), which is conservatively replaced by Tyr in some domains, that form a contiguous hydrophobic patch on the surface of the protein (Figure 7). Several of the surface positions surrounding this patch, while not conserving residue identity, have features that are fully conserved (i.e. hydrophobic residues are always present at the positions corresponding to Pro11, Val38 and Val41 of pNR-2/pS2; a hydrogen bond donor is present at the position of Arg12; hydrogen bond acceptors always replace Asn16, Thr23, Gln26 and Asp36). These residues are indicated in Figure 7. The high degree of conservation of the physical properties of this region of the protein surface strongly suggests that the different mammalian trefoil domains bind similar target molecules in this region.

It has been suggested that the cleft between loops 2 and 3 of PSP, which contains the conserved hydrophobic patch, could provide a binding site for an oligosaccharide chain of the type found in mucin proteins (Gajhede *et al.*, 1993). However, the actual targets of the trefoil domains are not known, and the cleft could just as easily accommodate a hydrophobic interaction with protein side-chains such as Phe or Trp and thus form part of a protein binding site. The different position of the second loop in pNR-2/pS2 results in the cleft being narrower in pNR-2/pS2 than in PSP (the distance between the C^{α} atoms of Trp43 and Pro20 is typically 6 Å for pNR-2/pS2 *versus* 10 Å for domain 1 and 8 Å for domain 2 of PSP). A contribution to the specificity of binding could come from the differences in conformation of these binding sites.

Two other well-conserved residues, Gly21 and Lys30, are on the edge of this conserved region and could have a role in determining the specificity of binding. In pNR-2/pS2, hSP and PSP there is a glycine residue at the position corresponding to Gly21 and this would allow easy access to the hydrophobic cleft. The side-chain of any other residue, such as the histidine residue found at that position in hITF, would impede such access. The residues equivalent to Lys30 are positively charged in all mammalian trefoil domains, except for the second domain of the spasmolytic polypeptides (PSP, hSP, etc.). The mutation of a charged residue to a neutral residue at this position could be important for producing differences in binding specificity of the two domains of the spasmolytic polypeptides. Overall, the surface of pNR-2/pS2 around the region containing the conserved residues is more similar to the corresponding regions in hSP and PSP than to that of hITF, and if we assume that the second loop in all domains, apart from domain 1 of the spasmolytic polypeptides, is able to change its position, then the binding sites of domain 1 of the spasmolytic polypeptides (PSP/hSP) are most similar to that of pNR-2/pS2.

It would appear that the residues Phe19, Pro20, Pro42 and Trp43 form a constant part of a largely hydrophobic ligand/receptor-binding site in each of the human trefoil proteins. Several residues flanking these four, and usually having conserved physical features, provide additional interactions conferring specificity to the binding of the different ligands/receptors. However, the detailed nature of the binding process will remain an open question until specific binding ligands or receptors are identified.

Implications of the pNR-2/pS2 structure for dimer formation

Several workers have pointed out that the presence of the seventh Cys residue at position 58 provides an opportunity for dimerisation of the native pNR-2/pS2 through formation of a disulphide bond. In our earlier studies we showed that the variant of pNR-2/pS2 with Cys58 replaced by serine exists as a monomer in solution, which suggests that there is no strongly interacting interface that would aid the dimerisation. The fact that PSP exists as a pair of covalently linked trefoil domains that have very similar structures to the trefoil domains of pNR-2/pS2 has encouraged speculation that, *in vivo*, a pNR-2/pS2 homodimer could adopt an overall structure similar to that of PSP. Chinery and co-workers have used structural information from the trefoil domain structures in

PSP to construct a model of such a dimer (Chinery *et al.*, 1995). The pNR-2/pS2 structures reported here now allow us to make a more detailed evaluation of this possibility. When two copies of the pNR-2/pS2 structure, S_{rep}, are superimposed on the domains of PSP in its crystal structure, a sensible structure can be achieved only by allowing a structural rearrangement of the pNR-2/pS2 structure to take place in the region of residues 3 to 5 and 51 to 49, where the β-sheet type interactions were detected in the structural studies. This observation does not rule out the possibility that the active form of pNR-2/pS2 is a dimer resembling PSP, as such a rearrangement is possible (e.g. with the residues in the disrupted β-sheet making additional interactions between the peptides). However, there is a distinct possibility that the pNR-2/pS2 dimer has a spatial arrangement of its domains that is different from PSP. A detailed model for the dimerised form of pNR-2/pS2 can be obtained only from direct structural studies on the covalently linked domains.

Conclusions

The availability of a high-quality three-dimensional structure for pNR-2/pS2 confirms the consistency of the trefoil fold suggested by the sequence conservation amongst members of the family, and gives confidence in homology modelling of other trefoil domains. The structure allows consideration of the role of surface residues as potential binding sites for receptors or ligands. The structure also provides a firm basis for the design of site-directed mutants with altered properties, which will be fully realised only when the molecular targets of pNR-2/pS2 have been defined. The structural features defined outside the trefoil motif indicate the problems of using the structure of the spasmolytic polypeptides as a basis for modelling the biologically active dimer of a protein containing a single trefoil motif. The structure of the monomer will provide a basis for interpretation of the structural and dynamic properties of the dimer when they are determined.

Materials and Methods

Expression of ^{15}N-pNR-2/pS2

The pNR-2/pS2 expression vector was electroporated into the *Escherichia coli* strain JM109, which grows well in minimal media. A yield of 3.5 mg/l was obtained, which is ~6-fold lower than the yield obtained from HB101 cells grown in rich medium. The bacteria were grown in 10 l of modified M9 medium (42.3 mM Na_2HPO_4, 22.1 mM KH_2PO_4, 8.6 mM NaCl, 1 mM $MgCl_2$, 100 mM $CaCl_2$, 2×10^{-4}% (w/v) thiamine, 0.2% (w/v) glucose, 50 mg/ml ampicillin) containing 18.7 mM $(^{15}NH_4)_2SO_4$ as the sole nitrogen source. The bacterial cultures were never diluted more than 50-fold in going from inoculation to large-scale culture. The purification procedure used was identical with that described previously for unlabelled pNR-2/pS2 (Polshakov *et al.*, 1995b; Chadwick *et al.*, 1995) except that a resource Q (Pharmacia) column was used for the ion-exchange chromatography. The final freeze-dried sample was dissolved in water (90% H_2O/10% ^2H_2O) at a concentration of approximately 1.3 mM.

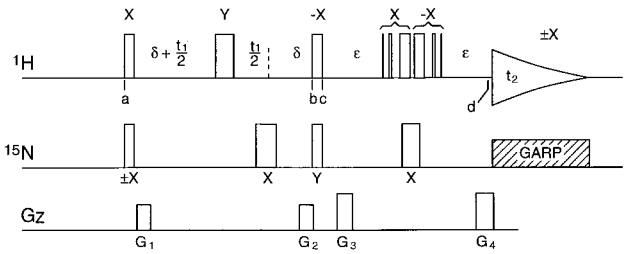

Figure 8. The pulse sequence used for the ^{15}N{^{1}H}NOE experiment.

NMR experiments

The ^{15}N-labelled sample used for the NMR experiments contained 1.3 mM protein in 90% H_2O/10% ^2H_2O, pH* = 6 (pH* refers to pH meter readings uncorrected for deuterium isotope effects). The NMR experiments were performed at 283 K and 298 K using Varian Unity spectrometers operating at 500 and 600 MHz. A series of NMR experiments (DQF-COSY, TOCSY, NOESY, ROESY, ^{13}C-^{1}H HSQC and HSQC-TOCSY) had previously been carried out on a non-labelled sample of pNR-2/pS2 (Polshakov *et al.*, 1995b). In the present study, several additional experiments were conducted on the sample of the uniformly ^{15}N-labelled protein. These included an experiment to measure ^{15}N{^{1}H} steady-state NOEs and experiments to record 3D ^{15}N-^{1}H HNHA, HNHB and NOESY-HSQC spectra. The HNHB and NOESY-HSQC experiments used standard sequences (Kuboniwa *et al.*, 1994; Archer *et al.*, 1991; Marion *et al.*, 1989; Bodenhausen & Ruben, 1980) modified in both cases by the conversion of the final ^{15}N to ^{1}H polarisation transfer step to include a WATERGATE sequence (Sklenar *et al.*, 1993) for water-suppression and no presaturation used. In the case of the HNHA experiment, a WATERGATE element was appended to the standard sequence (Vuister & Bax, 1993), and no presaturation was used. The NOESY-HSQC experiment was recorded with a mixing time of 150 ms and additional pulsed-field gradient pulses were used for coherence pathway rejection (Wider & Wüthrich, 1993; Bax & Pochapsky, 1992).

The heteronuclear ^{15}N{^{1}H} NOE measurements were made with the new pulse sequence shown in Figure 8. This is based on the method proposed by Kay *et al.* (1989) in which magnetisation is transferred from ^{15}N to ^{1}H, and the NOE is determined by comparing the intensity of signals in the absence and presence of saturation of the ^{1}H nuclei. The accuracy of this approach depends critically upon there being no unwanted perturbation of the ^{1}H magnetisation in the non-saturated spectrum. Several groups have identified chemical exchange between water protons and amide protons as a cause of such unwanted saturation effects (for example, see Kay *et al.*, 1989; Clore *et al.*, 1990; Stone *et al.*, 1992), either because the water signal has been deliberately presaturated or, in sequences that use pulsed-field gradients (PFGs), because the water is transiently saturated by a combination of the PFGs and the radiofrequency pulses. Sev-

eral different solutions to this problem have been proposed. These solutions are based either on the use of PFGs for water suppression together with long recovery delays (Li & Montelione, 1994) or, more efficiently, on the use of selective pulses to actively restore the water magnetisation to its equilibrium value (Grzesiek & Bax, 1993; Dayie & Wagner, 1994). The new sequence achieves active restoration of the water magnetisation without the use of selective pulses, making it extremely convenient to implement. The sequence is conventional apart from the first ^{1}H 90° pulse and the WATERGATE sequence, which has been incorporated into the ^{15}N to ^{1}H polarisation transfer step. The heteronuclear NOEs are calculated from the intensity of signals that start as ^{15}N Z magnetisation at time *a*. The evolution of this magnetisation is exactly as in the previously reported sequences and will not be discussed further. However, this sequence differs from earlier ones in the evolution of the proton magnetisation. In scans acquired without proton saturation the system starts with proton Z magnetisation at time *a*. The first ^{1}H pulse transfers this magnetisation into the transverse plane, after which it is dephased by gradient pulse G_1, thus preventing radiation damping. The subsequent ^{1}H 180° pulse and a gradient pulse G_2 (equal to G_1) refocus the proton magnetization at time *b*. The phase of the second ^{1}H 90° pulse is chosen to return the proton magnetisation to the $+Z$ axis at time *c*. Since the 3-9-19-19-9-3 refocussing pulse used in the WATERGATE sequence has no net effect on the water magnetisation, any water magnetisation that was present along the Z axis at time *a* will be returned to $+Z$ at time *d*. This method of restoring the water magnetisation without using shaped pulses has been used in heteronuclear correlation sequences (Stonehouse *et al*., 1994; Mori *et al*., 1995). The restoration is not perfect, since a small fraction of the water magnetisation fails to refocus properly due to imperfections in the proton 180° pulse and due to diffusion between the first and second gradient pulses. However, these gradient pulses need not be particularly intense and for the values used in this work the loss due to diffusion is calculated to be 8% (Van Zijl & Moonen, 1990) at the maximum value of t_1.

The NMR spectra were processed and analysed on Silicon Graphics or Sun workstations using the programs VNMR 5.1 (Varian), FELIX (Molecular Simulations), XEASY (Bartels *et al*., 1995) and software written in-house. The final size of all 3D data matrices was $1024 \times 512 \times 128$ real points. Mild resolution enhancement was obtained by applying a $\pi/2.5$ shifted sine-squared apodisation function in all dimensions. Zero-filling was employed in all indirectly detected dimensions. Linear prediction was used in some cases to extend the data by a quarter in the heteronuclear dimensions. The ^{1}H spectra were referenced to sodium 2,2-dimethyl-2-silapentane-5-sulphonate (DSS) and the ^{15}N specta to liquid NH_3 using the method described by Live *et al*. (1984) with the gyromagnetic ratios given by Wishart *et al*. (1995).

Distance constraints

Interproton distance constraints were obtained from NOEs measured from 2D NOESY and ROESY and 3D ^{15}N-^{1}H-NOESY-HSQC spectra. The assigned NOEs were placed in one of five classes, corresponding to the upper limit distance constraints of 2.8, 3.2, 4.0, 5.0 and 5.5 Å, on the basis of their intensity (number of contour levels). The successive classes correspond to intensity changes of a factor of 2 or more and are easily distinguished in the spectra. In view of the possibility of spin-diffusion significantly affecting the intensities of some signals in the NOESY spectra, where possible the relative intensities of signals were checked for consistency with those in the ROESY spectrum. Otherwise such signals were treated with caution and were intially given a larger upper limit distance constraint than strictly warranted by their contour level in the NOESY spectra. The upper limit was then reduced if consistent with the intermediate structures. The conversion of contour levels to upper limit constraints was calibrated by reference to non-overlapping cross-peaks in regions of secondary structure, which corresponded to well-known interproton distances. In all cases, the lower limits were set at 0 Å to achieve good sampling during the high-temperature stage of the calculations (Hommel *et al*., 1992). All non-stereospecifically assigned protons were treated with a $1/R^6$ sum averaging procedure (Nilges, 1995). Two distance constraints were used for each NH…O hydrogen bond; one between the hydrogen and the oxygen atoms of 1.7 to 2.3 Å, and the other between the donor nitrogen atom and the acceptor oxygen atom of 2.7 to 3.3 Å.

Torsion angle constraints

The program AngleSearch (Polshakov *et al*., 1995a) was used in conjunction with preliminary structures to determine the ranges of torsion angles that are consistent with the input ^3J coupling constant and NOE distance data. Coupling constant related data were extracted from spectra obtained from the following experiments: HNHA (^3J(HN, Hα)), HNHB (^3J(N, Hβ) and $^3J(H\alpha^{i-1}$, N)), and DQF-COSY (^3J(Hα, Hβ)): Archer *et al*. (1991); Vuister & Bax (1993). Distance data for intraresidue and sequential proton-proton contacts were obtained from spectra from 2D ROESY and 3D ^{15}N-^{1}H NOESY-HSQC experiments.

The data for each residue were analysed using two models available in the AngleSearch program, one based on a fixed χ_1 rotamer and the other on a mixture of χ_1 rotamers. Experimental data for most of the residues in the core of the protein were found to be consistent with the fixed rotamer model with some existing as non-staggered rotamers.

In the cases where very narrow ranges of ϕ, ψ and χ torsion angles were determined by the AngleSearch analysis, the values were increased to give a broader range (±30°) for the values used as corresponding constraints in the simulated annealing calculations. The results of the AngleSearch calculations were compared to the torsion angles found in the families of structures obtained from the simulated annealing calculations and in some cases this allowed some of the alternative AngleSearch solutions to be eliminated. In this way torsion angles that were under-determined in the initial AngleSearch calculation could be better defined. A similar approach was used previously for making stereospecific assignments of glycine H^α protons (Polshakov *et al*., 1995a). Here, this has been extended to the determination of stereospecific assignments of methylene protons in other residues and to calculations of improved torsion angle ranges.

Structure calculations

Three-dimensional structures were calculated with the program X-PLOR 3.1 (Brünger, 1992), using the dynamical simulated annealing protocol of Nilges *et al*. (1988)

with minor modifications. Calculations were carried out using the *topallhdg.pro* topology file and the *parallhdg.pro* parameter file modified to correct the planarity of the Arg guanidino group, to free the proline χ_1 angle, and to bring the backbone bond lengths and angles in line with the measurements made by Engh & Huber (1991). The structure calculations were carried out in two stages. In the first stage, a set of 20 structures was generated using simulated annealing starting from an extended chain structure, using a set of constraints derived from NOE and ^3J coupling data and in the absence of disulphide bonds. The protocol of the calculation included a high-temperature stage (40 ps at 1000 K), a linear cooling stage (over 20 ps from 1000 K to 100 K), and 250 steps of Powell minimisation. The time-step used was 4 fs and the bond lengths were kept fixed by the SHAKE algorithm (Ryckaert *et al*., 1977) during the dynamics. This stage was repeated several times and the structures generated were used to assign further NOEs and finally used to determine the disulphide bonding pattern.

In the second stage of calculation these 20 structures were submitted to further refinement, in which the disulphide bonds and hydrogen bonds were explicitly included. This refinement stage involved a high-temperature molecular dynamics simulation with linear cooling from 1000 K to 100 K in 20 ps with a 4 fs time-step, followed by 500 steps of Powell minimisation. This second stage was repeated several times and the structures generated each time allowed improvements to be made to the NOE derived constraints and hydrogen bond list. From the final set of 20 structures, 19 that had cumulative NOE constraint violations of less than 1 Å were selected for further analysis. The structures (1ps2) and NMR derived constraints (r1ps2mr) have been deposited with the Brookhaven Protein Data Bank.

Structure analysis

The analysis of the final structures in terms of their Ramachandran plots, deviations from ideal structural parameters and satisfaction of the NMR constraints, was performed with the programs PROCHECK (Laskowski *et al*., 1993) and AQUA/PROCHECK-NMR (Laskowski *et al*., 1997). Hydrogen bonding was analysed using HBPLUS (McDonald & Thornton, 1994) and structural features characterised with PROMOTIF (Hutchinson & Thornton, 1996) using the IUPAC-IUB (1970) convention 6.3 to define elements of secondary structure. Visual inspections and superposition of structures were performed with theprograms InsightII (Molecular Simulations) and MOLMOL (Koradi *et al*., 1996), and illustrations created with GRASP (Nicholls *et al*., 1991) and MOLSCRIPT (Kraulis, 1991). The solvent accessibilities were calculated using the program NACCESS (S. Hubbard, University College, London, UK).

Acknowledgements

The NMR experiments were made using the facilities at the Medical Research Council Biomedical NMR Centre, V.I.P. acknowledges the award of a Howard Hughes International Scholarship. M.A.W. and M.P.C. thank the Medical Research Council for the award of a Research Fellowship and a Training Scholarship, respectively. We are grateful for the contribution made by the pupils of Newcastle upon Tyne Church High School towards the purchase of a FPLC system.

References

Archer, S. J., Ikura, M., Torchia, D. A. & Bax, A. (1991). An alternative 3D NMR technique for correlating backbone ^{15}N with side chain Hβ resonances in larger proteins. *J. Magn. Reson,* **95**, 636–641.

Bartels, C., Xia, T.-H., Billeter, M., Güntert, P. & Wüthrich, K. (1995). The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR,* **5**, 1–10.

Bax, A. & Pochapsky, S. S. (1992). Optimised recording of heteronuclear multidimensional NMR spectra using pulse field gradients. *J. Magn. Reson.* **99**, 638–643.

Bodenhausen, G. & Ruben, D. J. (1980). Natural abundance ^{15}N NMR by enhanced heteronuclear spectoscopy. *Chem. Phys. Letters,* **69**, 185–189.

Brünger, A. T. (1992). *X-PLOR 3.1. A System for X-ray Crystallography and NMR,* Yale University Press, New Haven, CT.

Carr, M. D., Bauer, C. J., Gradwell, M. J. & Feeney, J. (1994). Solution structure of a trefoil-motif-containing cell growth factor, porcine spasmolytic protein. *Proc. Natl Acad. Sci. USA,* **91**, 2206–2210.

Chadwick, M. P., May, F. E. B. & Westley, B. R. (1995). Production and comparison of mature single-domain 'trefoil' peptides pNR-2/pS2 Cys^{58} and pNR-2/pS2 Ser^{58}. *Biochem. J.* **308**, 1001–1007.

Chinery, R. & Cox, H. M. (1995). Modulation of epidermal growth factor effects on epithelial ion transport by intestinal trefoil factor. *Br. J. Pharmacol.* **115**, 77–80.

Chinery, R. & Playford, R. J. (1995). Combined intestinal trefoil factor and epidermal growth factor is prophylactic against indomethecin-induced gastric damage in the rat. *Clin. Sci.* **88**, 401–403.

Chinery, R., Bates, P. A., De, A. & Freemont, P. S. (1995). Characterisation of the single copy trefoil peptides intestinal trefoil factor and pS2 and their ability to form covalent dimers. *FEBS Letters,* **357**, 50–54.

Clore, G. M., Driscoll, P. C., Wingfield, P. T. & Gronenborn, A. M. (1990). Analysis of the backbone dynamics of interleukin-1β using 2-dimensional inverse detected heteronuclear ^{15}N-^{1}H NMR spectroscopy. *Biochemistry,* **29**, 7387–7401.

Dayie, K. T. & Wagner, G. (1994). Relaxation-rate measurements for ^{15}N-^{1}H groups with pulse-field gradients and preservation of coherence pathways. *J. Magn. Reson. ser. A,* **111**, 121–126.

De, A., Brown, D. G., Gorman, M. A., Carr, M. D., Sanderson, M. R. & Freemont, P. S. (1994). Crystal structure of a disulfide-linked ''trefoil'' motif found in a large family of putative growth factors. *Proc. Natl Acad. Sci. USA,* **91**, 1084–1088.

Dignass, A., Lynch-Devaney, K., Kindon, H., Thim, L. & Podolsky, D. K. (1994). Trefoil peptides promote epithelial migration through a transforming growth factor β-independent pathway. *J. Clin. Invest.* **94**, 376–383.

Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallog. sect. A,* **47**, 392–400.

Gajhede, M., Petersen, T. N., Henriksen, A., Petersen, J. F. W., Dauter, Z., Wilson, K. S. & Thim, L. (1993).

Pancreatic spasmolytic polypeptide: first three-dimensional structure of a member of the mammalian trefoil family of peptides. *Structure,* **1**, 253–262.

Gargaro, A. R., Frenkiel, T. A., Nieto, P. M., Birdsall, B., Polshakov, V. I., Morgan, W. D. & Feeney, J. (1996). NMR detection of arginine-ligand interactions in complexes of *Lactobacillus casei* dihydrofolate reductase. *Eur. J. Biochem.* **238**, 435–439.

Grzesiek, S. & Bax, A. (1993). The importance of not saturating H_2O in protein NMR. Application to sensitivity enhancement NOE measurements. *J. Am. Chem. Soc.* **115**, 12593–12594.

Hahnel, E., Joyce, R., Sterrett, G., Harvey, J. & Hahnel, R. (1992). Detection of estradiol-induced messenger RNA (pS2) in uninvolved breast tissue from mastectomies for breast cancer. *Breast Cancer Res. Treat.* **20**, 167–176.

Hanby, A. M., Poulsom, R., Singh, S., Elia, G., Jeffery, R. E. & Wright, N. A. (1993). Human spasmolytic polypeptide (hSP) is a major antral peptide. The distribution of the trefoil peptides hSP and pS2 in the stomach. *Gastroenterology,* **105**, 1110–1116.

Hauser, F., Poulsom, R., Chinery, R., Rogers, L. A., Hanby, A. M., Wright, N. A. & Hoffmann, W. (1993). hP1.B, a human P-domain peptide homologous with rat intestinal trefoil factor, is expressed also in the ulcer-associated cell lineage and the uterus. *Proc. Natl Acad. Sci. USA,* **90**, 6961–6965.

Henry, J. A., Nicholson, S., Hennessy, C., Lennard, T. W. J., May, F. E. B. & Westley, B. R. (1989). Expression of the estrogen-regulated pNR-2 mRNA in human breast cancer: relation to estrogen receptor mRNA levels and response to tamoxifen therapy. *Br. J. Cancer,* **61**, 32–38.

Henry, J. A., Piggott, N. H., Mallick, U. K., NIcholson, S., Farndon, J. R., Westley, B. R. & May, F. E. B. (1991a). pNR-2/pS2 immunohistochemical staining in breast cancer: correlation with prognostic factors and endocrine response. *Br. J. Cancer,* **63**, 615–622.

Henry, J. A., Bennett, M. K., Piggott, N. H., Levett, D. L., May, F. E. B. & Westley, B. R. (1991b). Expression of the pNR-2/pS2 protein in diverse human epithelial tumours. *Br. J. Cancer,* **64**, 677–682.

Hommel, V., Harvey, T. S., Driscoll, P. C. & Campbell, I. D. (1992). Human epidermal growth factor: high resolution solution structure and comparison with human transforming growth factor a. *J. Mol. Biol.* **227**, 271–282.

Hoosein, N. M., Thim, L., Jørgensen, K. H. & Brattain, M. G. (1989). Growth stimulatory effect of pancreatic spasmolytic polypeptide on cultured colon and breast tumor cells. *FEBS Letters,* **247**, 303–306.

Hutchinson, E. G. & Thornton, J. M. (1996). PROMOTIF: a program to identify and analyze structural motifs in proteins. *Protein Sci.* **5**, 212–220.

IUPAC-IUB Commission on Biochemical Nomenclature, (1970). Abbreviations and symbols for the description of the conformation of polypeptide chains. *J. Mol. Biol.* **52**, 1–17.

Jakowlew, S. B., Breathnach, R., Jeltsch, J. M., Masiakowski, P. & Chambon, P. (1984). Sequence of the pS2 mRNA induced by estrogen in the human breast cancer cell line MCF-7. *Nucl. Acids Res.* **12**, 2861–2878.

Jørgensen, K. H., Thim, L. & Jacobsen, H. E. (1982a). Pancreatic spasmolytic polypeptide (PSP): I. preparation and initial chemical characterization of a new polypeptide from porcine pancreas. *Regul. Pept.* **3**, 207–219.

Jørgensen, K. D., Diamant, B., Jørgensen, K. H. & Thim, L. (1982b). Pancreatic spasmolytic polypeptide (PSP). III. Pharmacology of a new porcine pancreatic polypeptide with spasmolytic and gastric acid secretion inhibitory effects. *Regul. Pept.* **3**, 231–243.

Kay, L. E., Torchia, D. A. & Bax, A. (1989). Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy. Application to staphylococcal nuclease. *Biochemistry,* **28**, 8972–8979.

Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.

Kuboniwa, H., Grzesiek, S., Delagio, F. & Bax, A. (1994). Measurement of H^N-H^α J couplings in calcium-free calmodulin using new 2D and 3D water-flip-back methods. *J. Biomol. NMR,* **4**, 871–878.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.

Lasowski, R. A., Rullman, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1997). AQUA and PROCHECK-NMR: programs for checking the qaulity of protein structures solved by NMR. *J. Biomol. NMR,* **8**, 477–486.

Lefebvre, O., Chenard, M.-P., Masson, R., Linares, J., Dierich, A., LeMeur, M., Wendling, C., Tomasetto, C., Chambon, P. & Rio, M.-C. (1996). Gastric mucosa abnormalities and tumorigenesis in mic lacking the pS2 trefoil protein. *Science,* **274**, 259–262.

Li, Y. C. & Montelione, G. T. (1994). Overcoming solvent saturation-transfer artifacts in protein NMR at neutral pH. Application of pulse field gradients in measurements of ^1H-^{15}N Overhauser effects. *J. Magn. Reson. ser. B,* **105**, 45–51.

Live, D. H., Davis, D. G., Agosta, W. C. & Cowburn, D. (1984). Obsevation of 1000-fold enhancement of ^{15}N NMR via proton detected multiquantum coherences: studies of large peptides. *J. Am. Chem. Soc.* **106**, 1939–1941.

Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A. & Bax, A. (1989). Three-dimensional heteronuclear NMR of ^{15}N-labelled proteins. *J. Am. Chem. Soc.* **111**, 1515–1517.

Mashimo, H., Wu, D.-C., Podolsky, D. K. & Fishman, M. C. (1996). Impaired defense of intestinal mucosa in mice lacking intestinal trefoil factor. *Science,* **274**, 262–265.

Masiakowski, P., Breathnach, R., Bloch, J., Gannon, F., Krust, A. & Chambon, P. (1982). Cloning of cDNA sequences of hormone-regulated genes from the MCF-7 human breast cancer cell line. *Nucl. Acids Res.* **24**, 7895–7903.

May, F. E. B. & Westley, B. R. (1986). Cloning of estrogen-regulated messenger RNA sequences from human breast cancer cells. *Cancer Res.* **46**, 6034–6040.

May, F. E. B. & Westley, B. R. (1988). Identification and characterization of estrogen-regulated RNAs in human breast cancer cells. *J. Biol. Chem.* **263**, 12901–12908.

McDonald, I. K. & Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Biol. Chem.* **238**, 777–793.

Mori, S., Abeygunawardana, C., O'Neil Johnson, M. & van Zijl, P. C. M. (1995). Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *J. Magn. Reson. ser. B*, **108**, 94–98.

Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.

Nilges, M. (1995). Calculation of protein structures with ambiguous distance constraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J. Mol. Biol.* **245**, 645–660.

Nilges, M., Gronenborn, A. M., Brünger, A. T. & Clore, G. M. (1988). Determination of three-dimensional structures of proteins by simulated annealing with interproton distance constraints: application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Protein Eng.* **2**, 27–38.

Piggott, N. H., Henry, J. A., May, F. E. B. & Westley, B. R. (1991). Antipeptide antibodies against the pNR-2 oestrogen-regulated protein of human breast cancer cells and detection of pNR-2 expression in normal tissues by immunohistochemistry. *J. Pathol.* **163**, 95–104.

Playford, R. J., Marchbank, T., Chinery, R., Evison, R., Pignatelli, M., Boulton, R. A., Thim, L. & Hanby, A. M. (1995). Human spasmolytic polypeptide is a cytoprotective agent that stimulates cell migration. *Gastroenterology,* **108**, 108–116.

Playford, R. J., Marchbank, T., Goodlad, R. A., Chinery, R. A., Poulsom, R., Hanby, A. M. & Wright, N. A. (1996). Transgenic mice that overexpress the human trefoil peptide pS2 have an increased resistance to intestinal damage. *Proc. Natl Acad. Sci. USA,* **93**, 2137–2142.

Podolsky, D. K., Lynch-Devaney, K., Stow, J. L., Oates, P., Murgue, B., DeBeaumont, M., Sands, B. E. & Mahida, Y. R. (1993). Identification of human intestinal trefoil factor. Goblet cell-specific expression of a peptide targeted for apical secretion. *J. Biol. Chem.* **268**, 6694–6702.

Polshakov, V. I., Frenkiel, T. A., Birdsall, B., Soteriou, A. & Feeney, J. (1995a). Determination of stereospecific assignments, torsion-angle constraints, and rotamer populations in proteins using the program AngleSearch. *J. Magn. Reson. ser. B,* **108**, 31–43.

Polshakov, V. I., Frenkiel, T. A., Westley, B., Chadwick, M., May, F., Carr, M. D. & Feeney, J. (1995b). NMR-based structural studies of the pNR-2/pS2 single domain trefoil peptide. Similarities to porcine spasmolytic peptide and evidence for a monomeric structure. *Eur. J. Biochem.* **233**, 847–855.

Prud'homme, J. F., Fridlansky, F., Le Cunff, M., Atger, M., Mercier-Bodard, C., Pichon, M. F. & Milgrom, E. (1985). Cloning of a gene expressed in human breast cancer and regulated by estrogen in MCF-7 cells. *DNA,* **4**, 11–21.

Rio, M., Bellocq, J. P., Daniel, J. Y., Tomasetto, C., Lathe, R., Chenard, M. P., Batzenschlager, A. & Chambon, P. (1988). Breast cancer-associated pS2 protein: synthesis and secretion by normal stomach mucosa. *Science,* **241**, 705–708.

Ryckaert, J.-P., Ciccutti, G. & Berendsen, H. J. C. (1977). Numerical-integration of cartesian equations of motion of a system with constraints: molecular dynamics of N-alkanes. *J. Comput. Phys.* **23**, 327–351.

Rye, P. D., Keyte, J., Sutcliffe, M. J., Bailey, K. & Walker, R. A. (1994). Solid phase synthesis of pS2 peptide: suggestion of an alternative trefoil structure. *Protein Pept. Letters,* **1**, 54–59.

Sands, B. E. & Podolsky, D. K. (1996). The trefoil peptide family. *Annu. Rev. Physiol.* **58**, 253–273.

Schwartz, L. H., Koerner, F. C., Edgerton, S. M., Sawicka, J. M., Rio, M., Bellocq, J. P., Chambon, P. & Thor, A. D. (1991). pS2 expression and response to hormonal therapy in patients with advanced breast cancer. *Cancer Res.* **51**, 624–628.

Sklenar, V., Piotto, M., Leppik, R. & Saudek, V. (1993). Gradient-tailored water suppression for ^{1}H-^{15}N HSQC experiments optimised to retain full sensitivity. *J. Magn. Reson. ser. A,* **102**, 241–245.

Stone, M. J., Fairbrother, W. J., Palmer, A. G., III, Reizer, J., Saier, M. H. & Wright, P. E. (1992). Backbone dynamics of the *Bacillus subtilis* glucose permease IIA domain determined from ^{15}N NMR relaxation measurements. *Biochemistry,* **31**, 4394–4406.

Stonehouse, J., Shaw, G. L., Keeler, J. & Laue, E. D. (1994). Minimizing sensitivity losses in gradient-selected ^{15}N-^{1}H HSQC spectra of proteins. *J. Magn. Reson. ser. A,* **107**, 178–184.

Thim, A. (1989). A new family of growth factor-like peptides. ''Trefoil'' disulphide loop structures as a common feature in breast cancer associated peptide (pS2), pancreatic spasmolytic polypeptide (PSP), and frog skin peptides (spasmolysins). *FEBS Letters,* **250**, 85–90.

Tomasetto, C., Rio, M.-C., Gautier, C., Wolf, C., Hareuveni, M., Chambon, P. & Lathe, R. (1990). hSP, the domain-duplicated homolog of pS2 protein is co-expressed with pS2 in stomach but not in breast carcinoma. *EMBO J.* **9**, 407–414.

Van Zijl, P. C. M. & Moonen, C. T. W. (1990). Complete water suppression for solutions of large molecules based on diffusional differences between solute and solvent (DRYCLEAN). *J. Magn. Reson.* **87**, 18–25.

Vuister, G. W. & Bax, A. (1993). Quantitative J correlation: a new approach for measuring homonuclear three-bond $3(H^{N}H^{\alpha})$ coupling constants in ^{15}N-enriched proteins. *J. Am. Chem. Soc.* **115**, 7772–7777.

Wider, G. & Wüthrich, K. (1993). A simple experimental scheme using pulse field gradients for coherence-pathway rejection and solvent suppression in phase-sensitive heternocular correlation spectra. *J. Magn. Reson. ser. B,* **102**, 239–241.

Williams, R., Stamp, G. H. W., Gilbert, C., Pignatelli, M. & Lalani, E.-N. (1996). pS2 transfection of murine adenocarcinoma cell line 410.4 enhances dispersed growth pattern in a 3-D collagen gel. *J. Cell Sci.* **109**, 63–71.

Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L. & Sykes, B. D. (1995). ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR,* **6**, 135–140.

Wright, N. A., Poulsom, R., Stamp, G. W., Hall, P. A., Jeffery, R. E., Longcroft, J. M., Rio, M., Tomasetto, C. & Chambon, P. (1990). Epidermal growth factor (EGF/URO) induces expression of regulatory peptides in damaged human gastrointestinal tissues. *J. Pathol.* **162**, 279–284.

Wright, N. A., Poulsom, R., Stamp, G. W., van Noorden, S., Sarraf, C., Elia, G., Ahenen, D., Jeffery, R., Longcroft, J., Pile, C., Rio, M. & Chambon, P. (1993). Trefoil peptide gene expression in gastrointestinal epithelial cells in inflammatory bowel disease. *Gastroenterology*, **104**, 12–20.

Edited by P. E. Wright

(Received 2 October 1996; received in revised form 30 December 1996; accepted 7 January 1997)

http://www.hbuk.co.uk/jmb

Supplementary material comprising one Table is available from JMB Online