Solution Structure of the Disulfide-Linked Dimer of Human Intestinal Trefoil Factor (TFF3): The Intermolecular Orientation and Interactions Are Markedly Different from Those of Other Dimeric Trefoil Proteins†

Frederick W. Muskett,*,‡ Felicity E. B. May,§ Bruce R. Westley,§ and James Feeney*,‡

MRC Biomedical NMR Centre, National Institute for Medical Research, Mill Hill, London NW7 1AA, United Kingdom, and Department of Pathology, Royal Victoria Infirmary, University of Newcastle, Newcastle, NE1 4LP, United Kingdom

Received July 30, 2003; Revised Manuscript Received October 16, 2003

ABSTRACT: The trefoil protein TFF3 forms a homodimer (via a disulfide linkage) that is thought to have increased biological activity over the monomer. The solution structure of the TFF3 dimer has been determined by NMR and compared with the structure of the TFF3 monomer and with other trefoil dimer structures (TFF1 and TFF2). The most significant structural differences between the trefoil domain in the monomer and dimer TFF3 are in the orientations of the N-terminal 3_{10}-helix (residues 10−12) and in the presence in the dimer of an additional 3_{10}-helix (residues 53−55) outside of the core region. The TFF3 dimer forms a more compact structure as compared with the TFF1 dimer where the two trefoil domains are connected by a flexible region with the monomer units being at variable distances from each other and in many different orientations. Although TFF2 is also a compact structure, the dispositions of its monomer units are very different from those of TFF3. The structural differences between the dimers result in the two putative receptor/ligand binding sites that remain solvent exposed in the dimeric structures having very different dispositions in the different dimers. Such differences have significant implications for the mechanism of action and functional specificity for the TFF class of proteins.

Human TFF3[1] (previously called intestinal trefoil factor) is a small cysteine-rich secreted protein (*1*, *2*) belonging to the trefoil factor family (TFF) of proteins that share homology within a conserved trefoil domain of 42−43 amino acids (see Figure 1). The trefoil motif contains several well-conserved features including six cysteine residues with essentially conserved spacings. Three human trefoil peptides are known, namely, TFF1, TFF2, and TFF3. TFF1 (previously called pS2 or pNR-2) and TFF3 each contain one trefoil domain and are 60 and 59 amino acids long, respectively, while TFF2 (previously called spasmolytic polypeptide) contains two trefoil domains in a single chain of 106 amino acids (*3*, *4*).

The predominant site of expression of all three trefoil proteins in normal tissue is the gastrointestinal tract (*3*). TFF3 is expressed mainly by the mucus secretory goblet cells of the small and large intestine and at very low levels by the stomach (*2*, *5*). TFF1 is expressed mainly in the stomach, and TFF2 is expressed in the stomach and duodenum (*3*, *4*). TFF3 expression has also been detected in other organs including the uterus (*2*) and is found in the lower respiratory tract (*6*). All three proteins are secreted in association with mucins (*3*).

In addition to their normal distribution, trefoil peptides are expressed at exceptionally high levels at sites of ulceration throughout the human gastrointestinal tract (*3*). TFF3 expression is induced in both the adult and the embryonic rodent gastrointestinal tract following experimentally induced damage (*7*). TFF3-null mice manifest impaired defense of their intestinal mucosa and die after oral administration of dextran sulfate sodium (DSS) (*8*). These observations have led to the supposition that trefoil proteins are involved in tissue repair after damage.

It is thought that the normal function of trefoil proteins is to protect and maintain the gastrointestinal mucosa both by stabilizing the mucus barrier and by stimulating restitution. Evidence for the former is provided by the observation that gastric antral mucus is absent in TFF1-null mice (*9*) and by our demonstration that TFF1 is intimately associated with human gastric antral mucus and that the TFF1 homo-dimer interacts specifically with gastric mucin (*10*). All three-trefoil proteins promote wound closure in an in vitro wounding assay, which would be consistent with a role in restitution (*11*−*13*). Work with animal models has also demonstrated the importance of TFF3 in the maintenance and protection of the gastrointestinal mucosa (*8*, *14*).

† Funded by Cancer Research UK, the Wellcome Trust, and the Medical Research Council.

* Corresponding authors. Telphone: 020 8959 3666. Fax: 020 8906 4477. E-mail: (J.F.) jfeeney@nimr.mrc.ac.uk; (F.W.M.) fwm1@leicester.ac.uk.

‡ National Institute for Medical Research.

§ University of Newcastle.

[1] Abbreviations: COSY, correlated spectroscopy; DQF, double quantum filtered; HNHA, 3-D experiment correlating amide NH and α-CH protons; HSQC, heteronuclear single quantum coherence spectroscopy; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; rms, root-mean-square; TFF1, trefoil factor family 1 protein (previously called pS2 or pNR-2); TFF2, trefoil factor family 2 protein (previously called spasmolytic polypeptide); TFF3, trefoil factor family 3 protein (previously called intestinal trefoil factor); TOCSY, total correlation time spectroscopy; 3-D, three-dimensional; 2-D, two-dimensional; WATERGATE, pulse sequence for supressing water signal.

10.1021/bi030182k CCC: $25.00
Published on Web 12/02/2003

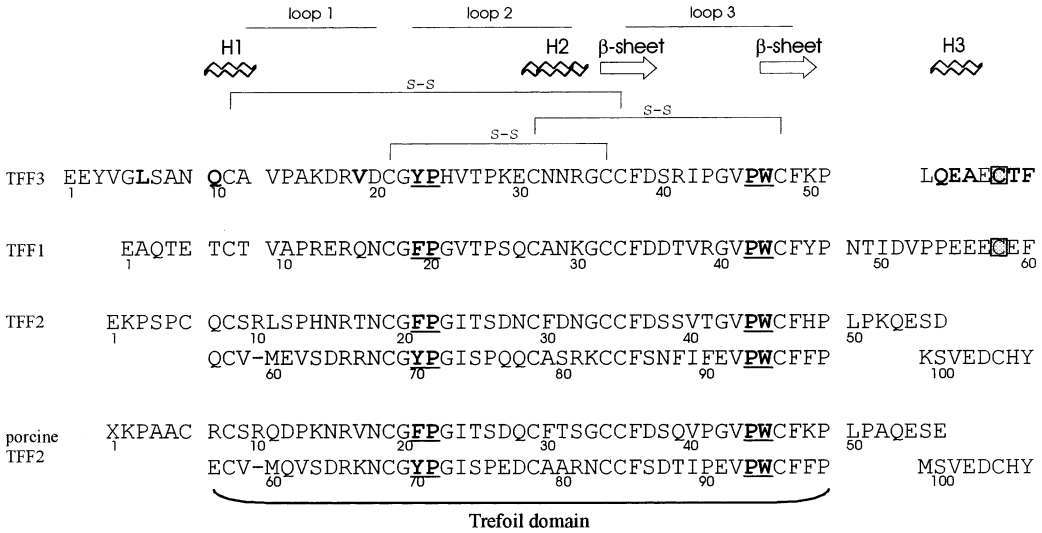

FIGURE 1: Amino acid sequences for human TFF3, TFF1, TFF2, and porcine TFF2 showing the trefoil domains aligned (residues 10−51 in TFF3). The positions of the three disulfide bonds in the trefoil domains are indicated. In the symmetrical TFF3 homodimer, an interchain disulfide bond connects the Cys57 residues (□), and in the TFF1 dimer, the corresponding bond involves the Cys58 residues (□). For TFF2, there is an additional disulfide bond between Cys6 and Cys104 in its two domains. The positions of the secondary structural elements (β-sheets; helices H1, H2, and H3; and loops) present in the TFF3 dimer are indicated above the sequences. The residues that are involved in intermolecular interactions are shown in bold, and the conserved hydrophobic residues from loops 2 and 3 that form part of the exposed cleft are shown in bold and underlined.

Trefoil proteins have been shown to be expressed by a number of tumors. In neoplastic human colonic mucosa, TFF3 expression is associated with more differentiated, less advanced tumors and with the expression of mucin (*15*, *16*). TFF3 expression in gastric adenocarcinoma is associated with more aggressive pathological features in men and is correlated significantly with reduced overall survival of male patients (*17*). TFF3 mRNA expression has been detected in breast tumor cells (*4*, *18*) and is induced by oestrogens in breast cancer cells in culture (*5*). The close association between oestrogen receptor and TFF3 expression in breast cancer has been confirmed by two recent microarray based studies of primary breast tumors. TFF3 was ranked as the first and third most important gene other than the oestrogen receptor for predicting the oestrogen receptor status of the tumor (*19*, *20*). Trefoil proteins have also been implicated in the dispersal of cancer cells (*4*). Consistent with such a role, TFF1 has been shown to stimulate breast cancer cell migration (*21*), and all three trefoil proteins have been shown to promote invasion of transformed cells into collagen gel (*22*).

The extra-trefoil domain cysteine residues of trefoil proteins have attracted interest as possible mediators of covalent intermolecular interactions (*4*). Both mammalian single-domain trefoil proteins contain one conserved extra-trefoil domain cysteine residue, three amino acids away from the carboxy-terminus (see Figure 1). The conservation of this free cysteine residue implies that it is of functional significance. Previous studies with recombinant TFF1 and TFF3 have shown that the C-terminal cysteine is capable of forming intermolecular disulfide bonds to produce TFF1 and TFF3 homodimers (*23*, *24*).

We have shown that in normal human gastric mucosa, TFF1 is present in three distinct molecular forms: as a TFF1 monomer, TFF1 homodimer, and as a ~25 kDa complex formed with another as yet unidentified protein that is stabilized by a disulfide bond (*10*). TFF1 secreted from breast cancer cells also forms homodimers and interacts via an intermolecular disulfide bond with a second larger protein to form a heterodimer (*23*). It has been reported that TFF3 dimers are present in an intestinal sample from a patient with Crohn's disease (*25*).

The relative activities of the recombinant TFF1 monomer and dimer have been compared in colonic cells using an in vitro wounding assay. Notably, the TFF1 dimer was motogenic, whereas the TFF1 monomer was inactive (*11*). Similarly, the TFF1 dimer was considerably more protective than the TFF1 monomer in a rat model of indomethacin-induced gastric damage (*11*). In breast cancer cells, the TFF1 monomer and dimer both stimulate cell migration in vitro, but the TFF1 dimer is 8-fold more potent than the monomer (*21*). Although the activity of the TFF3 monomers and dimers has not been compared, the activities of 21 kDa fusion proteins of TFF3 in which individual cysteine residues in TFF3 have been replaced with serine have been studied (*26*). The analogue Cys57Ser, which would be unable to form the intermolecular disulfide bond present in TFF3 dimers, had either reduced activity or was inactive in biological assays (*26*).

The three-dimensional structures of the human TFF1 monomer and dimer (*27*, *28*), porcine TFF2 (*29*−*31*), and human TFF3 monomer (*32*) have already been determined. For the TFF1 monomer, the nondimerizing analogue TFF1 Ser58 was studied, and for the TFF3 monomer, a recombinant protein in which Cys57 was covalently bound to an eighth cysteine residue was analyzed (*27*, *32*). Within the trefoil domains, the cysteine residues are paired 1−5, 2−4, and 3−6 (see Figure 1) to form the compact trefoil motif, which comprises three closely packed loops with the third loop positioned between the first and the second. In all the trefoil domains, there is a short α-helix in loop 2 that is packed against two antiparallel β-strands from loop 3. Several conserved surface residues are juxtaposed to form a small hydrophobic patch comprising adjacent parts of the second and third loops. It has been proposed that this area could represent a receptor or ligand binding site (*27*−*30*).

Analysis of the structure of the TFF1 homodimer using NMR techniques indicates that the trefoil domains have very similar structures in the monomer and dimer with no evidence for any contact between the two domains in the dimer (*28*). The two monomer units have no fixed orientation with respect to each other and are held apart by a flexible linker as a mixture of many conformations. This is in sharp contrast to the fixed orientation and relatively compact structure of the two covalently linked trefoil domains in porcine TFF2 (*31*). We have speculated previously that differences in function or receptor specificity of different trefoil proteins may derive in part from the orientation and separation of conserved surface residues in the different molecules (*28*).

TFF3 has eight amino acids C-terminal to the 42 amino acid trefoil domain, whereas the equivalent amino acid chain of TFF1 is 13 residues long (see Figure 1). The TFF3 dimer will therefore have a shorter linker sequence than the TFF1 dimer, and one might expect its structure to be intermediate between those of the TFF1 dimer and TFF2. We now report the structure of dimeric TFF3.

EXPERIMENTAL PROCEDURES

Production of Recombinant TFF3 Dimer. The human TFF3 expression plasmid contains TFF3 cDNA immediately downstream of the codons for an in-frame consensus recognition sequence for Factor Xa in the multiple cloning site of the pEZZ18 vector. Factor Xa cleaves immediately 3′ of its four-amino acid recognition site. BL21 *Escherichia coli* cells that contained the expression plasmid were grown overnight to stationary phase in 2 × TY medium at 37 °C, retrieved by centrifugation, and washed in 20% (w/v) sucrose, 1 mM EDTA, 0.3 M Tris/HCl, pH 8.0. The fusion protein was released from the periplasm by subjecting the cells to an osmotic shock in 0.5 mM $MgCl_2$ for 20 min at 4 °C and purified by affinity chromatography on an IgG-Sepharose (Pharmacia) column equilibrated in 150 mM NaCl, 50 mM Tris/HCl, pH 7.6 (*33*). Nonspecifically bound material was removed with 5 mM ammonium acetate, pH 5.0, and the bound fusion protein eluted with 0.5 M acetic acid, pH 3.4. The fusion protein was transferred to 100 mM NaCl, 1 mM $CaCl_2$, 50 mM Tris/HCl, pH 8.0 and digested for 48−72 h at 37 °C with Factor Xa. The ZZ fusion partner was removed from the mature TFF3 protein by passage over IgG-Sepharose.

Purification of the TFF3 Dimer. After the second passage over IgG Sepharose, 1 mM DTT was added to the TFF3 protein, which was then incubated at 4 °C for 20 h. The recombinant protein was bound to a Mono Q anion exchange column in Tris HCl, pH 8.5 and eluted with a gradient of NaCl. The purified TFF3 was concentrated using a vivaspin concentrator (Vivascience) with a 5000 molecular weight cutoff filter. The TFF3 dimer was purified by gel filtration on Superdex 75 (Pharmacia) in 50 mM sodium phosphate buffer, pH 6.5 and concentrated to >15 mg/mL with a vivaspin concentrator.

Production of ^{15}N-Labeled TFF3 Dimer. The TFF3 expression plasmid was electroporated into the *E. coli* strain BL21-CodonPlus-RP (Stratagene), which compensates for the presence of codons for tRNAs of arginine and proline that are present in low abundance in *E. coli*. The bacteria were grown in modified M9 medium (42.3 mM Na_2PO_4, 22.1 mM KH_2PO_4, 8.6 mM NaCl, 1 mM $MgCl_2$, 0.1 mM $CaCl_2$, 0.006 mM thiamine, 0.4% glucose) containing 0.025 mg/mL ampicillin and 18.7 mM $(^{15}NH_4)_2SO_4$ as the sole nitrogen source. The bacteria were grown to mid log phase and harvested. The purification procedure was as for unlabeled TFF3 dimer.

NMR Spectroscopy. The NMR experiments were performed on 0.6 mL samples of 2.0 mM TFF3 dissolved in 50 mM potassium phosphate buffer pH 6.5, prepared in either 100% D_2O or 90% H_2O/10% D_2O as appropriate. The NMR experiments were carried out on a Varian Inova spectrometer operating at 600 MHz, with data collected in phase-sensitive mode using the method of States and co-workers (*34*), at a temperature of 35 °C. The following spectra were acquired with both D_2O and H_2O samples: DQF−COSY (*35*), TOCSY using an isotropic mixing time of 60 ms (*36*, *37*), and NOESY with an NOE build-up period of 100 ms (*38*, *39*). The 2-D spectra were acquired over about 36 h, with typical acquisition times of 40−50 ms in F_1 and 300−400 ms in F_2.

With the ^{15}N-labeled TFF3 dimer the following experiments were conducted: 2-D ^{15}N HSQC (*40*−*42*) and ^{15}N-(CLEANEX-PM)-FHSQC (*43*), heteronuclear {^{1}H}-^{15}N NOE data measured on the TFF3 dimer at 25 °C as described previously (*27*), 3-D ^{15}N NOESY−HSQC (*44*) with an NOE build-up period of 100 ms, ^{15}N TOCSY−HSQC with a mixing time of 70 ms, and a ^{15}N HNHA spectrum (*45*). Typically, the three-dimensional ^{15}N-edited spectra were recorded over about 65 h with acquisition times in the indirect dimensions of 10−16 ms for ^{15}N, 16−22 ms for ^{1}H, and for 100−150 ms in the real time domain. Water suppression was achieved either by selective presaturation of the H_2O signal or by the pulsed-field gradient-based WATERGATE method (*46*). All NMR data were processed using the software package NMRPipe (*47*) and analyzed on-screen using the program SPARKY (T. D. Goddard and D. F. Kneller, SPARKY 3, University of California, San Francisco, CA).

Structure Calculations. Structurally significant intra- and interresidue NOEs were identified in two-dimensional NOESY and three-dimensional ^{15}N NOESY-HSQC spectra recorded with mixing times of 100 ms. The relationship between NOE intensity and interproton distance was calibrated using NOEs corresponding to known distances in regular α-helical and β-sheet regions of TFF3. On this basis, the NOEs were converted into upper distance constraints using the CALIBA macro of the program DYANA (*48*) with the maximum upper distance limit set to 6 Å. Where appropriate, standard distance corrections were applied to constraints involving nonstereospecifically assigned methylene, methyl, and aromatic ring protons (*49*, *50*). In addition, where spectral overlap prevented the reliable determination of NOE cross-peak volumes, the distance constraints were set to the upper limit of 6 Å. The ratio of diagonal to cross-peak volumes in the ^{15}N HNHA spectrum allowed reliable $^3J_{HN-H\alpha}$ coupling constants (±1 Hz) to be determined for 28 residues that were then converted to ϕ-angle constraints.

The high-resolution solution structure of the TFF3 dimer was calculated using the program DYANA (version 1.5, Güntert et al. (*48*)), which uses simulated annealing combined with torsion angle dynamics. The TFF3 structure was

calculated as a monomer using 1068 upper distance limits and 28 torsion angle constraints. The simulated annealing protocol used consisted of a high-temperature conformational search stage of 2000 steps, followed by slow cooling over 8000 steps, with conjugate gradient minimization at the end. To maximize the number of converged structures obtained from an initial set of 100 random starting coordinates, each stage of the calculations included five cycles of the redundant dihedral angle (REDAC) procedure. The set of NOE-based upper distance limits used in the final TFF3 calculations was the result of several cycles of an interactive process that involved using converged TFF3 structures from the previous generation of calculations to assign NOEs that were ambiguous on the basis of chemical shift information alone. At each stage in the calculation cycle, new generations of TFF3 structures were calculated from 100 random starting coordinates, with the additional NMR-derived structural constraints included. In the final stages of refinement of the TFF3 monomer, additional distance constraints were included in the calculations corresponding to NMR-determined disulfide and backbone hydrogen bonds. Constraints for the three disulfide bonds were progressively included, as it became clear that only one possible linkage of a pair of cysteine residues was consistent with the structures obtained. Backbone hydrogen bond constraints were only included for residues whose amide proton was observed to be hydrogen bonded via the ^{15}N-(CLEANEX-PM)−FHSQC (*43*) experiment and where the distance between the hydrogen bond acceptor and the hydrogen bond donor atoms was less than 2.5 Å and the HN−O bond angle greater than 135°.

Once a family of converged TFF3 monomer structures was obtained, the structure of the dimer was calculated. The intermolecular disulfide bond (Cys57−Cys57) connecting the monomers was included in the structure calculation, and intermolecular NOEs were identified and included only if they were unambiguously assigned in the NOESY spectra and were inconsistent with the structure of the monomer. A total of 18 interdomain NOEs were included in the calculation. The 47 structures with the lowest restraint violation energies were selected to represent the solution structure of the TFF3 dimer.

Structural Analysis. Analysis of the family of TFF3 dimer structures, preparation of protein structure figures, and comparisons with the structures of other trefoil proteins was carried out using the program MOLMOL (*51*).

Table 1: Structural Statistics

Number of Restraints Used in Final Structure Calculation[a]	
intraresidue NOEs	418
sequential NOEs ($i,i+1$)	554
medium-range NOEs ($i,i=4$)	694
long-range NOEs ($i,i=5$)	470
intersubunit NOEs	18
torsion angles	56
hydrogen bonding	80
disulfide bonds	7
Maximum Constraint Violations in 47 Converged TFF3 Dimer Structures	
upper distance limits (Å)	0.42 ± 0.04
lower distance limits (Å)	0.34 ± 0.03
van der Waals contacts (Å)	0.41 ± 0.04
torsion angles (deg)	4.5 ± 0.4
average DYANA target function (Å^2)	13.52 ± 1.10
Structure Quality and Coordinate Precision	
Ramachandran plot analysis[b]	
most favored region	61.3%
additionally allowed region	33.5%
generously allowed region	5.0%
disallowed region	0.2%
coordinate precision (Å)	
(i) single domain	
backbone atoms[c]	0.35 ± 0.1
heavy atoms[c]	0.77 ± 0.10
(ii) dimer	
backbone atoms[c]	0.68 ± 0.29
heavy atoms[c]	1.01 ± 0.25

[a] The number of unique restraints is half this number since the protein is a symmetric homodimer (excluding intersubunit NOEs). None of the structures exhibited distance restraint violations >0.5 Å or torsion angle violations >5°. [b] The overall quality of the protein was assessed using the program PROCHECK (*55*). [c] The precision of the coordinates is defined as the average atomic rmsd between the 47 individual structures and the mean coordinates. The values refer to residues 5−55 of both subunits.

RESULTS AND DISCUSSION

Sequence-Specific Assignments for the TFF3 Dimer. Sequence-specific ^{1}H resonance assignments for the majority of the TFF3 monomer residues (residues 3−55) have already been determined (*32*). However, the possibility of assignment changes following structural rearrangements as a consequence of forming the dimer led us to perform a complete reassignment of the ^{1}H resonances for the TFF3 dimer. The TFF3 dimer gives rise to well-resolved 2-D ^{1}H and ^{1}H -^{15}N NMR spectra: Figure 1S (Supporting Information) shows the excellent dispersion obtained in a 2-D ^{1}H -^{15}N HSQC spectrum. The good dispersion and sensitivity obtained in the three-dimensional ^{15}N NOESY− and TCOSY−HSQC spectra allowed the near complete sequence-specific assignment of backbone and side-chain ^{1}H and ^{15}N resonances. In the ^{1}H -^{15}N HSQC spectrum of the dimer, each pair of corresponding amino acids is represented by a single amide resonance indicating that the homodimer is completely symmetric in its time averaged form on the NMR time scale (ms). Resonance assignments agree very well with those obtained for the monomer (*32*) but with some minor differences. The backbone amide resonances for residues Glu2, Glu56, Cys57, and Phe59 that were unassigned in the monomer have all been identified in this study (although the backbone amide proton resonance for Gly5 could not be assigned). The sequence specific ^{1}H and ^{15}N resonance assignments obtained for the TFF3 dimer have been deposited in the BioMagResBank database (BMRB 5771).

High-Resolution Structure of the TFF3 Dimer. The solution structure of the TFF3 dimer was determined using a total of 1149 NMR derived structural restraints per monomer as summarized in Table 1. From 100 starting structures, 47 satisfactorily converged TFF3 dimer structures were obtained. The family of converged TFF3 dimer structures, together with the NMR constraints, have been deposited in the Protein Data Bank (accession code 1PE3).

The TFF3 dimer structure was determined from a relatively large number of NMR derived constraints (an average of 19.5 per residue), which resulted in a relatively low value for the overall root-mean-squared deviations (rmsds) to the mean for the backbone atom coordinates (0.35 ± 0.09 Å for the individual trefoil domains and 0.68 ± 0.29 Å for the

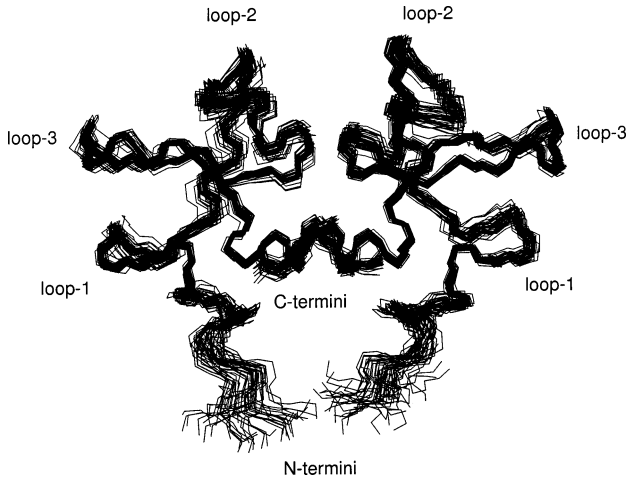

FIGURE 2: Best-fit superposition of the family of converged structures showing the full protein backbone superimposed on residues 5−55 of the TFF3 dimer.

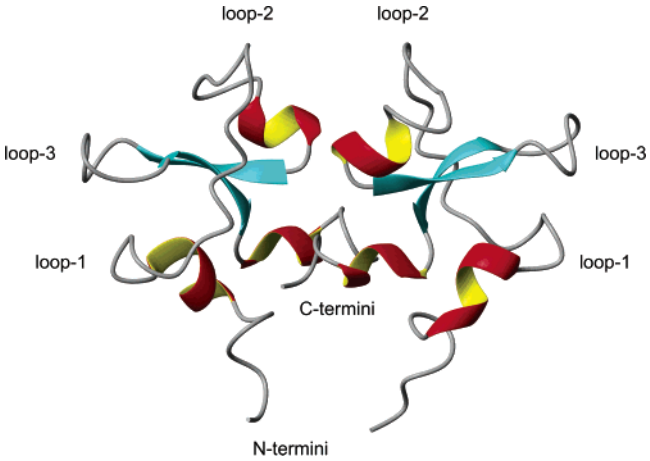

FIGURE 3: MOLMOL ribbon representation of the TFF3 dimer showing the secondary structural elements of the two domains and the disposition of the two monomers in the symmetrical dimer.

complete dimer structure) for the family of structures (see Table 1). Figure 2 shows the family of structures of the full protein (best-fit superposition of the protein backbone for residues 5−55), clearly showing a well-defined dimer structure with a 2-fold axis of rotational symmetry. The secondary structural features in the dimer are represented in the ribbon diagram of the dimer shown in Figure 3. The high-resolution solution structure of the TFF3 monomer has been determined previously by Lemercinier and co-workers (*32*) (PDB accession code 1E9T). The best-fit superposition of the core residues 14−50 between the monomer and the dimer structures shows reasonable agreement, with an rmsd of 1.07 ± 0.08 Å for backbone atoms and 1.87 ± 0.05 Å for all heavy atoms. Figure 4 shows the structure of the TFF3 monomer (Figure 4B) and the corresponding structure of the monomer unit in the TFF3 dimer (Figure 4A). The most significant differences between the two structures are in the orientations of the N-terminal 3_{10}-helix (residues 10−12) and in the presence of the second 3_{10}-helix (helix 3) involving residues 53−55 in the TFF3 dimer that was not found in the monomer.

In common with all trefoil domains studied so far, each monomer of the TFF3 dimer is composed of three well-defined looplike regions, the central core of which consists of a short, irregular β-sheet (residues 36−39 and 47−50). Between loop 2 and the β-sheet is a short α-helix (helix 2—residues 31−34), which packs against the N-terminal strand of the β-sheet. Residues 10−12 (helix 1) and 53−55 (helix 3) form single turn 3_{10}-helices. In the TFF3 dimer, the two loop-3 regions point almost directly away from each other on opposite sides of the dimer, and the N-termini of the monomers are located at the base of the structure (Figure 3). A single disulfide bond involving residues Cys57 joins the two monomers, and the C-terminal 3_{10}-helices (helix 3—residues 53−55) lie antiparallel to each other at the interface between the monomer units. The N-terminal portion, a poorly defined region in the TFF3 monomer (*32*), is positioned further underneath the trefoil domain in the dimer to accommodate the interaction of the two domains. The C-terminal region is different in the dimer where it loops back on itself to position the additional 3_{10}-helix (residues 53−55) beneath the α-helix (residues 31−34) and at the tip of the β-sheet. These structural features can be seen in Figure 3.

Interface between the TFF3 Monomer Units. Most of the NOEs observed between the two monomers involve residues in helix 1 (Gln10), helix 3 (Gln53, Glu54, and Ala55), and nearby residues (Leu6, Val19, Cys57, Thr58, and Phe59). Helix 3 (residues 53−55) forms part of the interface between the two monomer units and is an integral part of the dimeric form of TFF3 (see Table 1S in the Supporting Information). This probably contributes to the stabilization of this 3_{10}-helix and offers an explanation for its presence only in the dimeric form of TFF3. It is interesting to note that the interface between the monomer units in the well-defined porcine TFF2 structure contains a single helical region (residues 55−59) (*29*, *30*) that corresponds well with the 3_{10}-helical (helix 3) regions found at the interface in the TFF3 dimer. The interacting residues in the case of TFF2 (residues 59−61 in domain 2) are close to the N-terminus of domain 2. For TFF3, the interactions with helix 3 of each monomer unit are with residues from the C-terminal end of the other monomer unit.

Structural Comparisons of the TFF3 Dimer with Other Trefoil Dimers. Although the TFF1 and TFF3 dimers are both formed in the same way (via a disulfide bond linkage involving a Cys residue near the C-terminus of each monomer unit), the two symmetrical dimers have very different overall structures. For the TFF1 dimer, no interactions were detected between the two domains, whereas the TFF3 dimer is shown here to have a well-defined interface between the monomer units. In the TFF1 dimer, the region containing the disulfide-bond connected C-termini has an unstructured, extended conformation that positions the two trefoil domains of the TFF1 dimer at variable distances from each other in a plethora of different orientations. It appears that the ability of the C-terminal tail of TFF3 to form a 3_{10}-helix (helix 3) plays an important role in the formation of the dimer interface: such a helix does not form in the TFF1 dimer (possibly because of the presence of Pro53 and Pro54), and this could be a contributory factor in the nonformation of a well-defined dimer structure for TFF1.

In the case of TFF2, the dimeric structure is of necessity an asymmetric structure formed from a continuous stretch of peptide and containing seven disulfide bonds. Although there are general similarities in the secondary structural

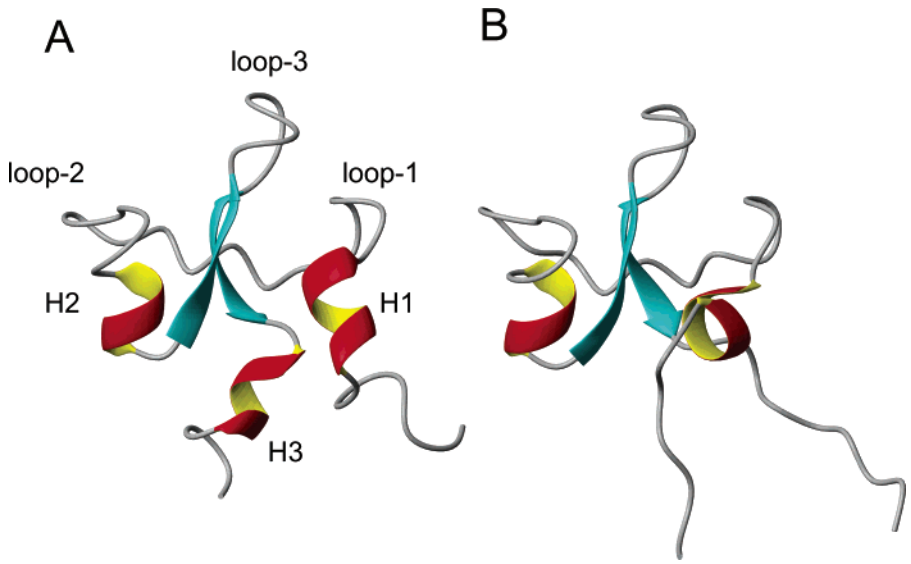

FIGURE 4: MOLMOL ribbon representation of the monomer domain in (A) the TFF3 dimer and (B) the TFF3 monomer. The additional 3_{10}-helix is indicated as helix 3 in Figure 4A.

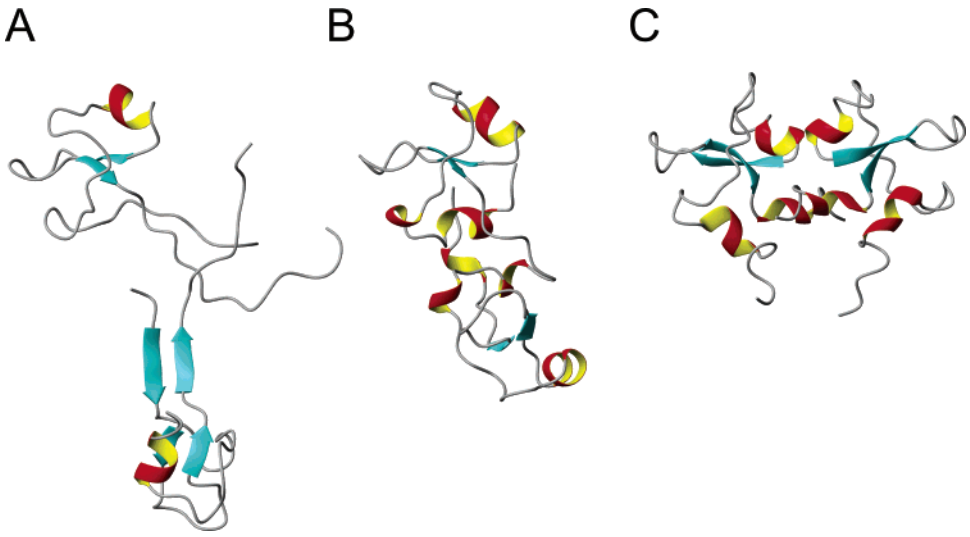

FIGURE 5: MOLMOL ribbon representation of (A) TFF1, (B) TFF2, and (C) TFF3 dimers. For TFF1, one of the family of 10 structures was used, for TFF2 the crystal structure was used (PDB accession code 2PSP), and for TFF3 the structure closest to the mean was used.

features of the trefoil domains in TFF2 and TFF3, the dispositions of the monomer units in the two dimers are very different. The distances and relative angles between the corresponding secondary structural elements in each dimer are very different. This can be seen from the structures shown in Figure 5. It is worth examining how these structural differences manifest themselves in terms of the disposition of the conserved hydrophobic residues on the surface of the proteins.

Differences in Conserved Clusters of Hydrophobic Surface Residues in the Different Dimers. Structural studies of trefoil proteins have identified a cleft formed between loops 2 and 3 on the protein surface consisting largely of hydrophobic residues, and this has been suggested as a potential binding site capable of accommodating either an oligosaccharide (*29*) or a protein aromatic side chain (*27*). In TFF1, the cleft is narrow (6 Å as measured between the Cα atoms of Pro20 and Trp43) (*27*), whereas in TFF2 the clefts are somewhat wider (10 and 8 Å for domains 1 and 2, respectively) (*29*), and in TFF3, where the relevant residues are Tyr23, Pro24, Pro46, and Trp47, the cleft is even more open having a shelf-like structure (*32*) (11−12 Å between Cα of Pro 24 and Trp 47). Formation of the TFF3 dimer structure does not change the solvent accessibility of the conserved hydrophobic cluster; thus, the hydrophobic sites in the dimeric form are still available for any required interactions with receptors.

Earlier, it had been conjectured that disulfide-bonded dimer formation of single trefoil domain proteins such as TFF1 might result in the clusters of surface conserved hydrophobic residues being similarly located to those in the two domains of the rigid TFF2 structure (*25*). However, the positioning and orientations of the two patches of conserved hydrophobic residues are very different in the different dimers. For TFF3, the conserved residues are on the same face of the dimer with the distance between the α-carbons of the Pro24 residues being 13−19 Å. For the TFF1 dimer (*28*), where the linkage between the two domains is very flexible, the distance between the hydrophobic clusters and their relative orientations are fairly random with the Pro−Pro distances being in the range of ∼36−73 Å (with the separation in the majority of structures being greater than 50 Å). In the case of the rigid TFF2 structure, the hydrophobic clusters are

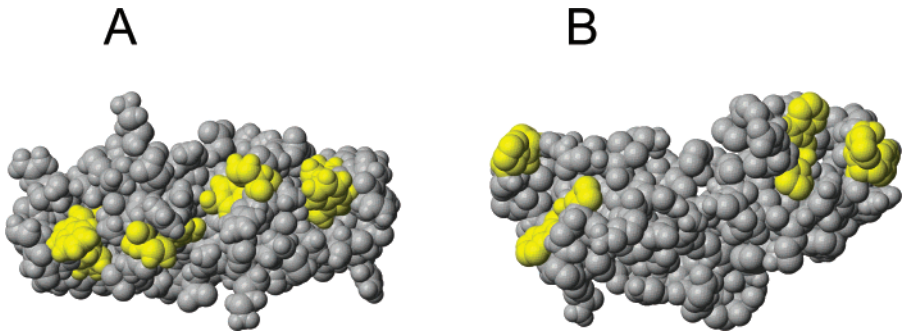

FIGURE 6: Space-filling models of the structures of (A) the TFF3 dimer and (B) TFF2. In each case, the conserved patches of hydrophobic residues are indicated in yellow. The two clefts are closer together in the TFF3 dimer with the distance between the α-carbons of the two Pro24 residues in TFF3 being 13−19 Å and that between those of Pro22 and Pro71 in TFF2 being 44 Å.

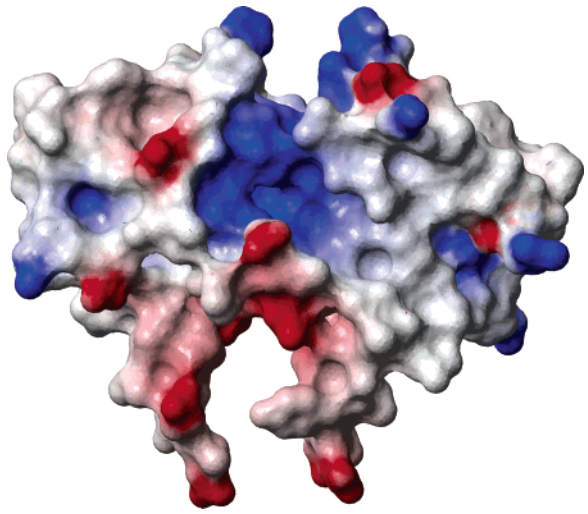

FIGURE 7: Electrostatic surface potential of the TFF3 dimer as calculated by MOLMOL. The molecule is in essentially the same configuration as the ribbon structure of the dimer shown in Figure 3. Red indicates negative charge, blue indicates positive charge, and white is neutral.

positioned at the opposite ends of the elongated molecule (see Figure 5B) resulting in a 44 Å separation between the Cα-carbons of Pro22 and Pro71, again very different from TFF3.

Figure 6 shows space filling models of the structures of the TFF3 dimer and TFF2 with the corresponding conserved surface hydrophobic residues marked in yellow in both cases. In TFF2, the two clusters of conserved surface hydrophobic residues are seen at the two opposite ends of the molecule, whereas in the TFF3 dimer, the two clusters are positioned much closer together and are located on the same face of the molecule. Such an arrangement might be required to make efficient bidentate interactions with a receptor or adapter molecule specific for TFF3. The large differences in the distances between the binding sites and their different orientations in the different trefoil two domain structures could thus be contributing to different binding specificities.

The conserved hydrophobic surface residues are located on the rim of a large positively charged patch on the surface of TFF3 (Figure 7 shows the electrostatic surface potential of the TFF3 dimer as calculated by MOLMOL (*51*)). It is possible that nearby hydrophilic residues would also contribute to the specificity of receptor binding. For example, His25, a residue not conserved in the other trefoil structures, is adjacent to the surface hydrophobic cluster in TFF3 and remains solvent exposed in the dimeric structure.

Differences in the Compactness and Mobilities of Dimeric Trefoil Proteins. Comparison of the sequences (Figure 1) indicates that the extra-trefoil domain C-terminus is significantly longer in TFF1 than in TFF3. We have speculated that the shorter C-terminus of TFF3 would reduce the distance between the two trefoil domains in the TFF3 dimer and result in its overall size and mobility being intermediate between those of the TFF1 dimer and TFF2 (*28*). The results presented here show that the TFF3 dimer is in fact much more compact than the TFF1 dimer and somewhat less mobile. The juxtaposition of helices 1 and 3, together with the intramolecular interactions around helix 2 at the base of loop 2, give the TFF3 dimer a compact structure that is almost as constrained as that of TFF2 with the former having a more globular discoid structure (26 × 19 × 10 $Å^3$) as compared with the somewhat elongated TFF2 (50 × 20 × 20 $Å^3$) (*29*).

Some indication of the relative mobilities of the TFF3 and TFF1 dimers can be obained from their {^{1}H}-^{15}N heteronuclear NOE data shown in Figure 8. The amide NHs signals of residues in the core of the proteins have steady-state NOEs in the range of 0.6−0.7, which is in good agreement with the values expected for the rigid part of proteins of this size (*52*). The average NOE for the core residues of the TFF3 dimer (0.73) is larger than that for the TFF1 dimer (0.65), indicating that there is less mobility even in the core region of the TFF3 dimer. The positive NOE values in the C-terminal linker regions (51−59 of TFF3 and 48−60 of TFF1) indicate that in both cases there is restricted rather than free mobility in these regions. However, the NOE values for the linker residues close to the Cys57−Cys57 bond are larger for the TFF3 dimer, indicating lower mobility in its linker region as compared with that of the TFF1 dimer. This is fully consistent with the increased compactness of the TFF3 dimer structure as compared with TFF1 found in the structural studies. Earlier ultracentrifugation and gel filtration experiments had also indicated that the TFF3 dimer is more compact than the TFF1 dimer (*53*).

CONCLUSION

The present structural work on the TFF3 dimer will provide a basis for guiding mutagenesis work aimed at identifying residues important for its structure and homodimerization. For example, the work has identified interface residues that could be investigated for their effects on dimerization. The availability of the structural data will also allow aspects of the biological activity of the TFF3

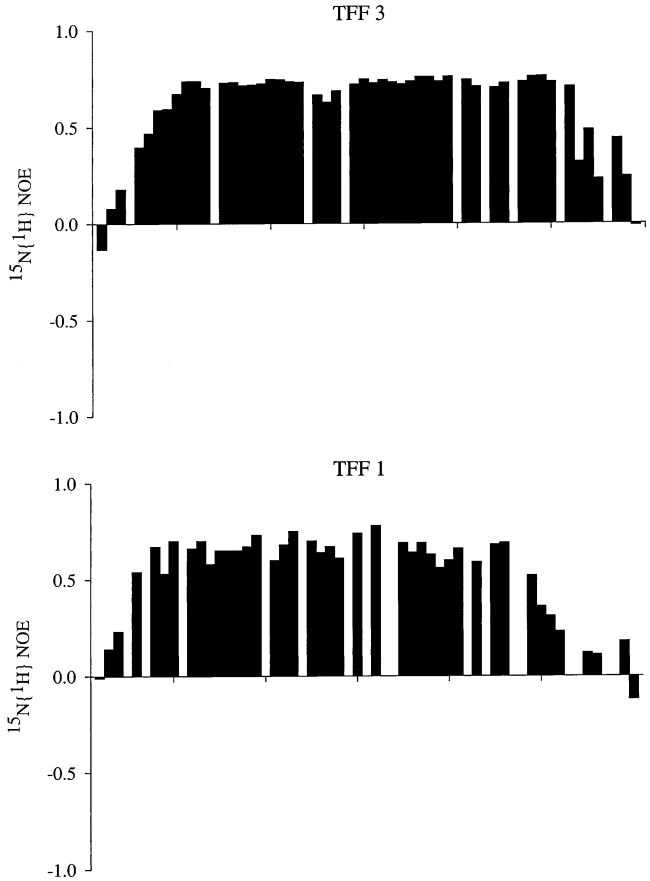

FIGURE 8: Measured ^{15}N{^{1}H} heteronuclear NOE values for the amide backbone groups of the TFF3 dimer (A) and the TFF1 dimer (B) at 25 °C.

dimer (e.g., its ability to induce phosphorylation of EGF receptors in intestinal cells (*54*)) to be studied by site directed mutagenesis in a more systematic manner.

There are clearly substantial differences in structure, mobility, and flexibility in the various dimeric trefoil proteins. These differences could be particularly important for the clusters of conserved surface hydrophobic residues thought to be important for receptor interactions. The observed structural differences in the relative positions and orientations of the clusters and in the shapes of their clefts implies that these putative binding sites are disposed very differently in the different molecules. Clearly, such structural differences have the potential of conferring specificity to interactions with particular receptors, and knowing the structures of these trefoil proteins will be important for the eventual understanding of such interactions.

SUPPORTING INFORMATION AVAILABLE

A ^{1}H-^{15}N HSQC spectrum and a table of interdomain NOEs. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Podolsky, D. K., Lynch-Devaney, K., Stow J. L., Oates, P., Murgue, B., DeBeaumont, M., Sands, B. E., and Mahida, Y. R. (1993) *J. Biol. Chem. 268*, 6694−6702.
2. Hauser, F., Poulsom, R., Chinery, R., Rogers, L. A., Hanby, A. M., Wright, N. A., and Hoffmann, W. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 6961−6965.
3. Poulsom, R., and Wright, N. A. (1993) *Am. J. Physiol. 265*, G205−G213.
4. May, F. E. B., and Westley, B. R. (1997) *J. Pathol. 183*, 4−7.
5. May, F. E. B., and Westley, B. R. (1997) *J. Pathol. 182*, 404−413.
6. Wiede, A., Jagla, W., Welte, T., Köhnlein, T., Busk, H., and Hoffman, W. (1999) *Am. J. Respir. Crit. Care Med. 159*, 1330−1335.
7. Otto, W. R., and Patel, K. (1999) *Anat. Embryol. 199*, 499−508.
8. Mashimo, H., Wu, D.-C., Podolsky, D. K., and Fishman, M. C. (1996) *Science 274*, 262−265.
9. Lefebvre, O., Chenard, M. P., Masson, R., Linares, J., Dierich, A., LeMeur, M., Wendling, C., Tomasetto, C., Chambon, P., and Rio, M. C. (1996) *Science* 274, 259−262.
10. Newton, J. L., Allen, A., Westley, B. R., and May, F. E. B. (2000) *Gut 46*, 312−320.
11. Marchbank, T., Westley, B. R., May, F. E. B., Calnan, D. P., and Playford R. J. (1998) *J. Pathol. 185*, 153−158.
12. Playford, R. J., Marchbank, T., Chinery, R., Evison, R., Pignatelli, M., Boulton, R. A., Thim, L., and Hanby, A. M. (1995) *Gastroenterology 108*, 108−116.
13. Dignass, A., Lynch-Devaney, K., Kindon, H., Thim, L., and Podolsky, D. K. (1994) *J. Clin. Invest. 94*, 376−383.
14. Babyatsky, M. W., Rossiter, G., and Podolsky, D. K. (1996) *Gastroenterology 110*, 975−984.
15. Taupin, D., Ooi, K., Yeomans, N. D., and Giraud, A. (1996) *Lab. Invest. 75*, 25−32.
16. Uchino, H., Kataoka, H., Itoh, H., and Koono, M. (1997) *J. Clin. Pathol. 50*, 932−934.
17. Yamachika, T., Werther, J. L., Bodian, C., Babyatsky, M., Tatematsu, M., Yamamura, Y., Chen, A., and Itzkowitz, S. (2002) *Clin. Cancer Res. 8*, 1092−1099.
18. Poulsom, R., Hanby, A. M., Lalani, E. N., Hauser, F., Hoffmann, W., and Stamp, G. W. (1997) *J. Pathol. 183*, 30−38.
19. Gruvberger, S., Ringner, M., Chen, Y., Panavally, S., Saal, L. H., Borg, A., Ferno, M., Peterson, C., and Meltzer, P. S. (2001) *Cancer Res. 61*, 5979−5984.
20. West, M., Blanchette, C., Dressman, H., Huang, E., Ishida, S., Spang, R., Zuzan, H., Olson, J. A., Jr., Marks, J. R., and Nevins, J. R. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 11462−11467.
21. Prest, S. J., May, F. E. B., and Westley, B. R. (2002) *FASEB J. 16*, 592−594.
22. Emami, S., Le Floch, N., Bruyneel, E., Thim, L., May, F., Westley, B., Rio, M., Mareel, M., and Gespach, C. (2001) *FASEB J. 15*, 351−361.
23. Chadwick, M. P., Westley, B. R., and May, F. E. B (1997) *Biochem. J. 327*, 117−123.
24. Thim, L., Wöldike, H. F., Nielsen, P. F., Christensen, M., Lynch-Devaney, K., and Podolsky, D. K. (1995) *Biochemistry 34*, 4757−4764.
25. Chinery, R., Bates, P. A., De, A., and Freemont, P. S. (1995) *FEBS Lett. 357*, 50−54.
26. Kinoshita, K., Taupin, D. R., Itoh, H., and Podolsky, D. K. (2000) *Mol. Cell Biol. 20*, 4680−4690.
27. Polshakov, V. I., Williams, M. A., Gargaro, A. R, Frenkiel, T. A., Westley, B. R., Chadwick, M. P., May, F. E. B., and Feeney, J. (1997) *J. Mol. Biol. 267*, 418−432.
28. Williams, M. A., Westley, B. R., May, F. E. B., and Feeney, J. (2001) *FEBS Lett. 493*, 70−74.
29. Gajhede, M., Petersen, T. N., Henriksen, A., Petersen, J. F., Dauter, Z., Wilson, K. S., and Thim, L. (1993) *Structure* 253−262.
30. De, A., Brown, D. G., Gorman, M. A., Carr, M. D., Sanderson, M. R., and Freemont, P. S. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 1084−1088.
31. Carr, M. D., Bauer, C. J., Gradwell, M. J., and Feeney, J. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 2206−2210.
32. Lemercinier, X., Muskett, F. W., Cheeseman, B., McIntosh, P. B., Thim, L., and Carr, M. D. (2001) *Biochemistry 40*, 9552−9559.
33. Chadwick, M. P., May, F. E. B., and Westley, B. R. (1995) *Biochem. J. 308*, 1001−1007.
34. States, D. J., Haberkorn, R. A., and Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286−292.
35. Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., and Wüthrich, K. (1993) *Biochem. Biophys. Res. Commun. 117*, 479−485.
36. Braunschweiler, L., and Ernst, R. L. (1983) *J. Magn. Reson. 53*, 521−528.

37. Davis, D. G., and Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2820−2821.
38. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R R. (1979) *J. Chem. Phys. 71*, 4546−4553.
39. Macura, S., and Ernst, R. R. (1980) *Mol. Phys. 41*, 95−117.
40. Bodenhausen, G., and Ruben, D. J. (1980) *Phys. Lett. 69*, 185−189.
41. Kay, L. E., Marion, D., and Bax, A. (1989) *Magn. Reson. 84*, 72−84.
42. Norwood, T. J., Boyd, J., Heritage, J. E., Soffe, N., and Campbell, I. D. (1990) *Magn. Reson. 87*, 488−501.
43. Hwang, T. L., van Zijl, P. C. M., and Mori, S. (1998) *J. Biomol. NMR 11* (2), 221−226.
44. Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., and Bax, A. (1989) *J. Am. Chem. Soc. 111*, 1515−1517.
45. Vuister, G., and Bax, A. (1993) *J. Am. Chem. Soc. 115*, 7772−7777.
46. Sklenar, V., Piotto, M., Leppik, R., and Suadek, V. (1993) *J. Magn. Reson. 102*, 241−245.
47. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR 6*, 277−293.
48. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) *J. Mol. Biol. 273*, 283−298.
49. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
50. Wüthrich, K., Billeter, M., and Braun, W. (1993) *J. Mol. Biol. 169*, 949−961.
51. Korida, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graphics 14*, 51−55.
52. Polshakov, V. I., Frenkiel, T. A., Westley, B., Chadwick, M., May, F., Carr, M. D., and Feeney, J. (1995) *Eur. J. Biochem. 233*, 847−855.
53. May, F. E. B., Church, S. T., Major, S., and Westley, B. R. (2003) *Biochemistry 42*, 8250−8259.
54. Liu, D., El-Hariry, I., Karayiannakis, A. J., Wilding, J., Chinery, R., Kmiot, W., McCrea, P. D., Gullick, W. J., and Pignatelli, M. (1997) *Lab. Invest. 77*, 557−563.
55. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1966) *J. Biomol. NMR 8*, 477−486.

BI030182K