High-Resolution Solution Structure of Human Intestinal Trefoil Factor and Functional Insights from Detailed Structural Comparisons with the Other Members of the Trefoil Family of Mammalian Cell Motility Factors[†,‡]

Xavier Lemercinier,[§] Frederick W. Muskett,[∥] Bruce Cheeseman,[§] Pauline B. McIntosh,[§] Lars Thim,[⊥] and Mark D. Carr*[,∥,#]

Laboratory of Molecular Structure, National Institute for Biological Standards and Control, Potters Bar, Hertfordshire EN6 3QG, U.K., Department of Biosciences, University of Kent, Canterbury, Kent CT2 7NJ, U.K., Protein Chemistry, Novo Nordisk A/S, DK-2880 Bagsvaerd, Denmark, and Department of Biochemistry, University of Leicester, Leicester LE1 7RH, U.K.

Received January 29, 2001; Revised Manuscript Received May 4, 2001

ABSTRACT: The secreted proteins intestinal trefoil factor (ITF, 59 residues), pS2 (60 residues), and spasmolytic polypeptide (SP, 106 residues) form a small family of trefoil domain-containing mammalian cell motility factors, which are essential for the maintenance of all mucous-coated epithelial surfaces. We have used ^{1}H NMR spectroscopy to determine the high-resolution structure of human ITF, which has allowed detailed structural comparisons with the other trefoil cell motility factors. The conformation of residues 10−53 of hITF is determined to high precision, but the structure of the N- and C-terminal residues is poorly defined by the NMR data, which is probably indicative of significant mobility. The core of the trefoil domain in hITF consists of a two-stranded antiparallel β-sheet (Cys 36 to Asp 39 and Trp 47 to Lys 50), which is capped by an irregular loop and forms a central hairpin (loop 3). The β-sheet is preceded by a short α-helix (Lys 29 to Arg 34), with the majority of the remainder of the domain contained in two loops formed from His 25 to Pro 28 (loop 2) and Ala 12 to Arg 18 (loop 1), which lie on either side of the central hairpin. The region formed by the surface of loop 2, the cleft between loop 2 and loop 3, and the adjacent face of loop 3 has previously been proposed to form the functional site of trefoil domains. Detailed comparisons of the backbone conformations and surface features of the family of trefoil cell motility factors (porcine SP, pS2, and hITF) have identified significant structural and electrostatic differences in the loop 2/loop 3 regions, which suggest that each trefoil protein has a specific target or group of target molecules.

The secreted proteins pS2 (TFF1, 60 residues), spasmolytic polypeptide (SP or TFF2, 106 residues), and intestinal trefoil factor (ITF or TFF3, 59 residues) constitute a small but very important family of mammalian cell motility factors, which are characterized by the presence of either one (pS2 and ITF) or two (SP) copies of a fairly highly conserved (44−58% identity), 41- or 42-residue functional motif, commonly referred to as a trefoil domain (Figure 1, *1*− *3*). Expression of the three proteins is mainly restricted to mucin-producing epithelial cells, and they all play essential roles in the maintenance and repair of mucous membranes, such as the lining of the gut and respiratory tract (*3*). This was perhaps most dramatically demonstrated by the severe pathological effects seen in knockout mice lacking either pS2 or ITF (*4*, *5*). The three mammalian trefoil proteins have all been shown to be upregulated at sites of mucosal injury and to act as powerful motogens, which are able to stimulate the movement of epithelial cells without promoting cell division (*3*, *6*, *7*). The proteins clearly participate in the repair of mucous membranes by stimulating the migration of surviving cells from the edges of damaged regions to cover denuded areas, a process known as epithelial restitution. The evidence available to date suggests that this cell motility activity is

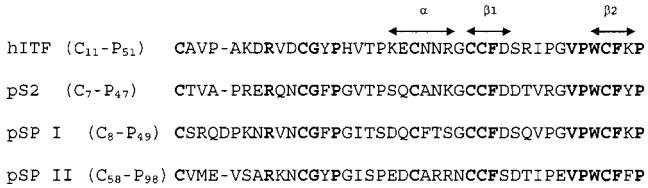

FIGURE 1: Comparison of the amino acid sequences of the trefoil domains found in hITF (Cys 11 to Pro 51), pS2 (Cys 7 to Pro 47), and pSP (domain I, Cys 8 to Pro 49, and domain II, Cys 58 to Pro 98). Residues conserved in all four trefoil domains are highlighted in bold, and the locations of the conserved α-helix (α) and two-stranded β-sheet (β1 and β2) are indicated above the sequences.

[†] This work was supported by the Department of Health (U.K.), the Medical Research Council (U.K.), and the Wellcome Trust (U.K.).

[‡] The atomic coordinates of the family of converged human intestinal trefoil factor structures, together with the NMR constraints, have been deposited in the Protein Data Bank (accession codes 1e9t and r1e9tmr, respectively).

* Address correspondence to this author at the University of Leicester. Tel: +44 (0)116 252 3054. Fax: +44 (0)116 223 1503. E-mail: mdc12@le.ac.uk.

[§] National Institute for Biological Standards and Control.

[∥] University of Kent.

[⊥] Novo Nordisk A/S.

[#] University of Leicester.

10.1021/bi010184+ CCC: $20.00
Published on Web 07/14/2001

mediated via the trefoil proteins' binding to high-affinity, cell surface receptors (*8*, *9*). More recently, the trefoil proteins have also been proposed to play a role in stabilizing the mucous layer that coats and protects all endodermal surfaces by cross-linking mucins; however, the evidence for this second role is currently fairly weak and is largely based on the coexpression and apparent cosecretion of specific trefoil proteins and mucins (*3*). Trefoil motifs have also been identified in a number of large vertebrate proteins, including several frog mucins (*10*, *11*) and a major zona pellucida protein (*12*), which clearly indicates a more widespread role as a shuffled functional module, probably involved in protein−protein or protein−sugar interactions (*12*, *13*).

High-resolution solution and crystal structures have been determined for porcine SP (pSP, *13*−*16*), and more recently, a high-resolution solution structure was also obtained for pS2 (*17*, *18*). The topology of the trefoil domains in pSP and pS2 is pretty similar, with the majority of the protein backbone forming three looplike regions, which viewed edge-on resemble a clover leaf. The core of the domain is formed by a short, C-terminal antiparallel β-sheet, with the two strands linked by a well-defined but irregular loop and which together form a central hairpin (loop 3). The β-sheet is preceded by a 6-residue α-helix that packs against the N-terminal β-strand, with the remainder of the protein backbone taken up by two loops (loop 2 and loop 1) that lie on either side of the central hairpin and are linked by an extended region that wraps around the C-terminal β-strand. Independent consideration of the surface features of pSP and pS2, including the location of conserved residues, has highlighted the region formed by the surface of loop 2, the cleft between loop 2 and loop 3, and the adjacent face of loop 3 as the probable functional site of the trefoil domains (*13*, *15*, *18*).

In this paper we report the determination of the high-resolution solution structure of the monomeric form of human ITF (hITF),[1] which has allowed us to carry out detailed comparisons of the topologies and surface features of the complete family of trefoil cell motility factors (pSP, pS2, and hITF). This type of structural analysis of a family of functionally related proteins, together with a consideration of the locations of patches of conserved surface residues, can clearly be very informative about the location of functional sites on the proteins by allowing the identification of real trends and differences across the family. In the case of the trefoil proteins, this approach has identified significant structural and electrostatic potential differences in the loop 2/loop 3 regions of the proteins, which have been proposed to form the functional sites, and suggests that each trefoil protein has a specific target or group of target molecules, which fits with the lack of functional degeneracy revealed by the mice knockout experiments (*4*, *5*). The detailed structural comparisons also provide a clearer picture of the conserved residues probably involved in the binding of target molecules, which can hopefully be confirmed in the future by a focused program of site-directed mutagenesis.

[1] Abbreviations: hITF, human intestinal trefoil factor; pSP, porcine spasmolytic polypeptide; rmsd, root mean squared deviation; DQF-COSY, double-quantum-filtered correlation spectroscopy; TOCSY, total correlation spectroscopy; NOESY, nuclear Overhauser effect spectroscopy; ROESY, rotating frame nuclear Overhauser effect spectroscopy.

EXPERIMENTAL PROCEDURES

Sample Preparation. The expression and purification of recombinant hITF were carried out as described previously using a *Saccharomyces cerevisiae* secretion system, with the monomeric form of the protein used for the NMR studies (*19*). The NMR experiments were performed on 0.6 mL samples of 2.8 mM hITF dissolved in a 25 mM potassium phosphate and 100 mM potassium chloride buffer at pH 6.0. The samples were prepared in either 100% D_2O or 90% H_2O/10% D_2O, as appropriate.

NMR Spectroscopy. All the NMR experiments were carried on Varian Unity and UnityPlus spectrometers operating at 500 MHz, with the data collected in the phase-sensitive mode (*20*). The following spectra were collected from both D_2O and H_2O samples of hITF at 35 °C: DQF-COSY (*21*), TOCSY using isotropic mixing times of 30−50 ms (*22*, *23*), NOESY with NOE buildup periods of 250 ms (*24*, *25*), and ROESY using spin-lock times of 60 ms (*26*, *27*). In addition, TOCSY and NOESY spectra with mixing times of 50 and 150 ms, respectively, were acquired from a H_2O sample of hITF at 15 °C. The 2D spectra were all acquired over about 40 h, with typical acquisition times of 42−48 ms in F_1 and 340−380 ms in F_2.

In the TOCSY experiments isotropic mixing was achieved using an MLEV17 pulse sequence at a radio-frequency field strength of about 10.3 kHz (*28*), while the ROESY spectra were acquired using a continuous wave spin-lock field of about 3 kHz. The D_2O solutions of hITF were prepared from protein repeatedly freeze-dried from D_2O, and thus only minimal presaturation was required to suppress the residual HOD signal. Water suppression in 90% H_2O/10% D_2O solutions of hITF was obtained either by selective presaturation of the H_2O signal or by the pulsed-field gradient-based WATERGATE method (*29*).

The 2D NMR data were processed using the Varian VNMR software package on SUN workstations. The original data were usually zero-filled twice in F_1 and once in F_2 prior to Fourier transformation, and mild resolution enhancement was achieved by applying a $\pi/3$-shifted sine-squared apodization function in both dimensions. All of the spectra were analyzed on-screen using the program XEASY (*30*).

Structural Calculations. Structurally significant intraresidue and interresidue NOEs were identified in 2D NOESY spectra of hITF recorded with mixing times of 250 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 hITF, and on this basis the NOEs were converted to upper distance constraints using the program CALIBA (*31*) with the maximum upper distance limit set to 6 Å. Where appropriate, standard distance corrections were applied to constraints involving methyl and aromatic ring protons (*32*, *33*). In addition, where spectral overlap prevented reliable determination of volumes for NOE cross-peaks, the distance constraints were set to the upper limit of 6 Å.

The TOCSY spectrum recorded from an H_2O sample of hITF at 35 °C allowed reliable $^3J_{HN-H\alpha}$ coupling constants to be determined for 51 residues using the inverse Fourier transformation method (*34*). These coupling constant data, together with intraresidue and sequential NOEs, were used

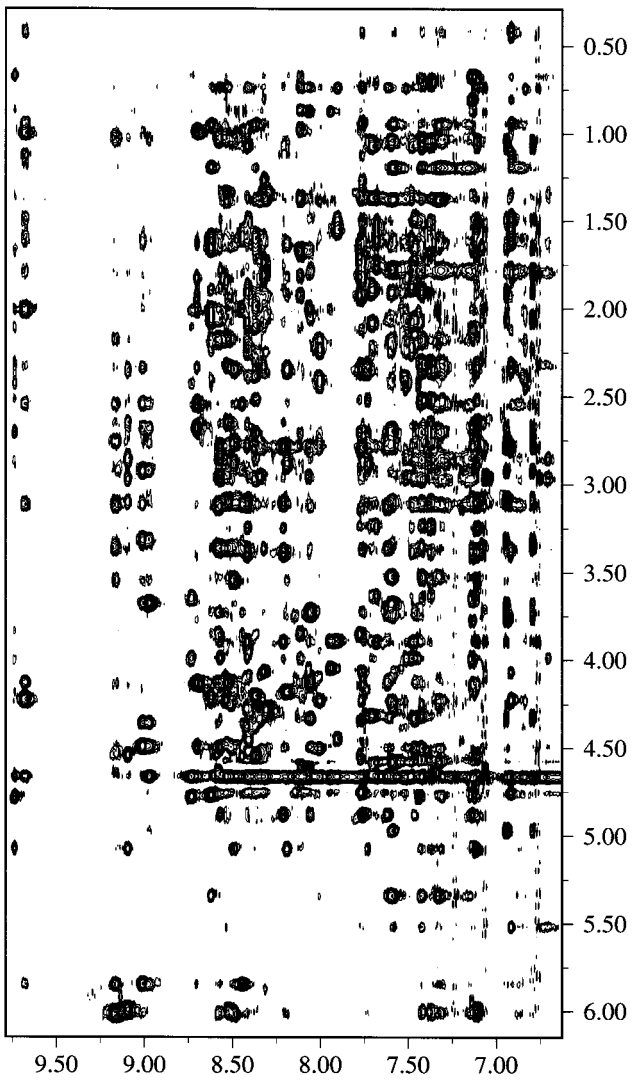

FIGURE 2: Selected region from a 2D NOESY spectrum of hITF acquired with an NOE mixing time of 250 ms. The region shown contains cross-peaks corresponding to NOEs between aliphatic protons and backbone amide or aromatic ring protons.

as input for the program HABAS (*31*), which produced initial constraints for 29 ϕ, 29 ψ, and 11 χ^1 torsion angles in hITF.

The high-resolution solution structure of hITF was calculated using the program DYANA (*35*), which uses simulated annealing combined with torsion angle dynamics. The initial structures were calculated using 936 upper distance limits and 69 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 hITF calculations was the result of several cycles of an interactive process that involved using converged hITF structures from the previous generation of calculations to assign NOEs that were ambiguous with chemical shift information alone. In addition, preliminary hITF structures, together with estimates of $^3J_{H\alpha-H\beta}$ coupling constants from a short mixing time TOCSY spectrum and rotating frame

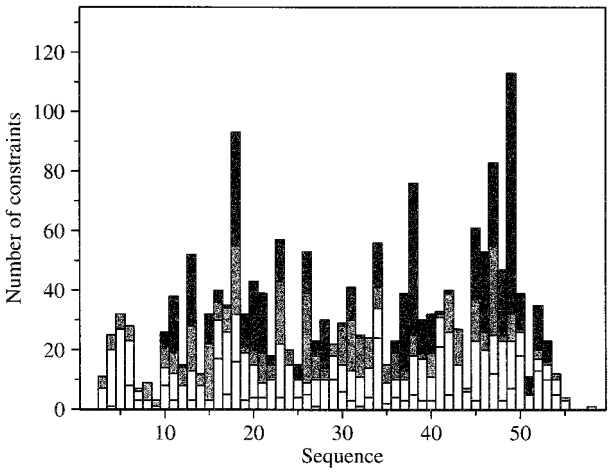

FIGURE 3: Summary of the number and distribution of the intraresidue (white), sequential (i, $i+1$, light gray), medium-range (i, $i \leq 4$, dark gray), and long-range (i, $i \geq 5$, black) NOEs used to determine the solution structure of hITF.

Table 1: NMR Constraints and Structural Statistics for hITF

(a) No. of Constraints Used in Final Structural Calculations		
intraresidue NOEs	255	
sequential NOEs (i, $i+1$)	278	
medium-range NOEs (i, $i \leq 4$)	176	
long-range NOEs (i, $i \geq 5$)	305	
torsion angles	69	(28 ϕ, 29 ψ, and 12 χ^1)
hydrogen bonding	8	
disulfide bonding	24	
(b) Maximum and Total Constraint Violations in 85 Converged hITF Structures		
upper distance limits (Å)	0.41 ± 0.06	16.7 ± 0.5
lower distance limits (Å)	0.14 ± 0.02	0.4 ± 0.1
van der Waals contacts (Å)	0.25 ± 0.04	7.3 ± 0.6
torsion angle ranges (deg)	2.5 ± 0.9	6.9 ± 2.3
av DYANA target function (Å^2)	4.33 ± 0.34	
(c) Structural Statistics for Family of Converged hITF Structures		
no. of residues within allowed region of Ramachandran plot (%)	86	
backbone atom rmsd for structured region (residues 10−53) (Å)	0.18 ± 0.04	
heavy atom rmsd for structured region (residues 10−53) (Å)	0.66 ± 0.08	
backbone atom rmsd for trefoil domain (residues 11−51) (Å)	0.17 ± 0.04	
heavy atom rmsd for trefoil domain (residues 11−51) (Å)	0.62 ± 0.09	

Overhauser effect data, were used to determine stereospecific assignments and χ^1 torsion angle constraints for progressively more residues. At each stage in the calculation cycle new generations of hITF structures were calculated from 100 random starting coordinates, with the additional NMR-derived structural constraints included. In the final stages of refinement of the hITF structure, additional distance constraints were incorporated in the calculations corresponding to NMR-determined disulfide and backbone hydrogen bonds. The constraints for the four 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 with an amide proton detectable after several weeks in D_2O and where the distance between the hydrogen bond acceptor and donor atoms was less than 2.5 Å and the NH−O bond angle greater than 135°.

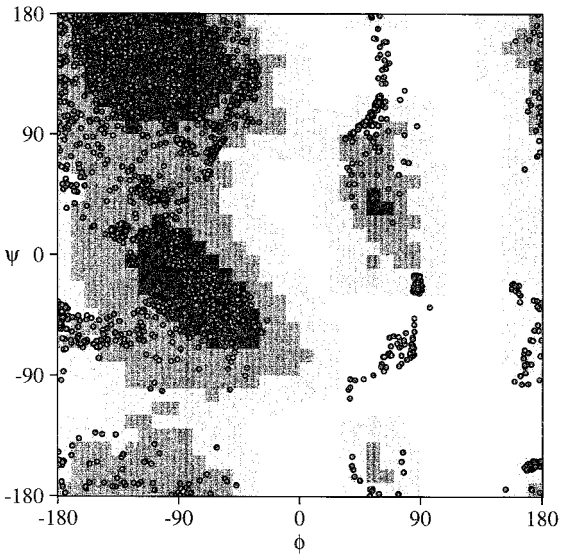

FIGURE 4: Ramachandran plot showing the distribution of the backbone ϕ and ψ dihedral angles for all non-glycine and non-proline residues in the 85 converged hITF structures.

Structural Analysis. Analysis of the family of hITF structures (root mean squared deviation values, Ramachandran plots, angular order parameters, etc.) and comparisons with the structures of other trefoil proteins were carried out using the programs MOLMOL (*36*) and DYANA (*35*). In addition, MOLMOL was used to prepare all of the protein structure figures.

RESULTS AND DISCUSSION

Sequence-Specific Assignments for hITF. hITF is a small, stable protein, which gives rise to well-resolved 2D ^{1}H NMR spectra at pH 6.0 and 35 °C, as illustrated by the backbone amide to aliphatic region from the 2D NOESY spectrum shown in Figure 2. The good dispersion and sensitivity observed in DQF-COSY, TOCSY, and NOESY spectra of hITF allowed essentially complete sequence-specific ^{1}H resonance assignments to be obtained for the majority of the protein (residues 3−55) in a straightforward manner, using the well-proven sequential assignment procedure (*33*). Signals were not identified for the backbone amide protons of several residues at the N- and C-termini of hITF (Glu 2, Glu 56, Cys 57, and Phe 59) and for the single cysteine residue (Cys 60) linked by a disulfide bond to Cys 57. This probably reflects rapid exchange of these backbone amide protons with water. The comprehensive sequence-specific ^{1}H resonance assignments obtained for hITF have been deposited in the BioMagResBank database.

High-Resolution Solution Structure of hITF. The solution structure of hITF was determined using a total of 1115 NMR-derived structural constraints (an average of 18.6 per residue, Table 1), including 1014 NOE-based upper distance limits [255 intraresidue, 278 sequential (i, $i+1$), 176 medium range (i, $i \leq 4$), and 305 long range (i, $i \geq 5$), Figure 3] and 69 torsion angle constraints (28 ϕ, 29 ψ, and 12 χ^1). The final structural calculations also included 8 distance constraints corresponding to two backbone hydrogen bonds detected between Cys 37 and Phe 49 and 24 distance constraints required to impose the four disulfide bonds determined from the structural data (Cys 11−Cys 37, Cys 21−Cys 36, Cys 31−Cys 48, and Cys 57−Cys 60). Stereospecific assignment

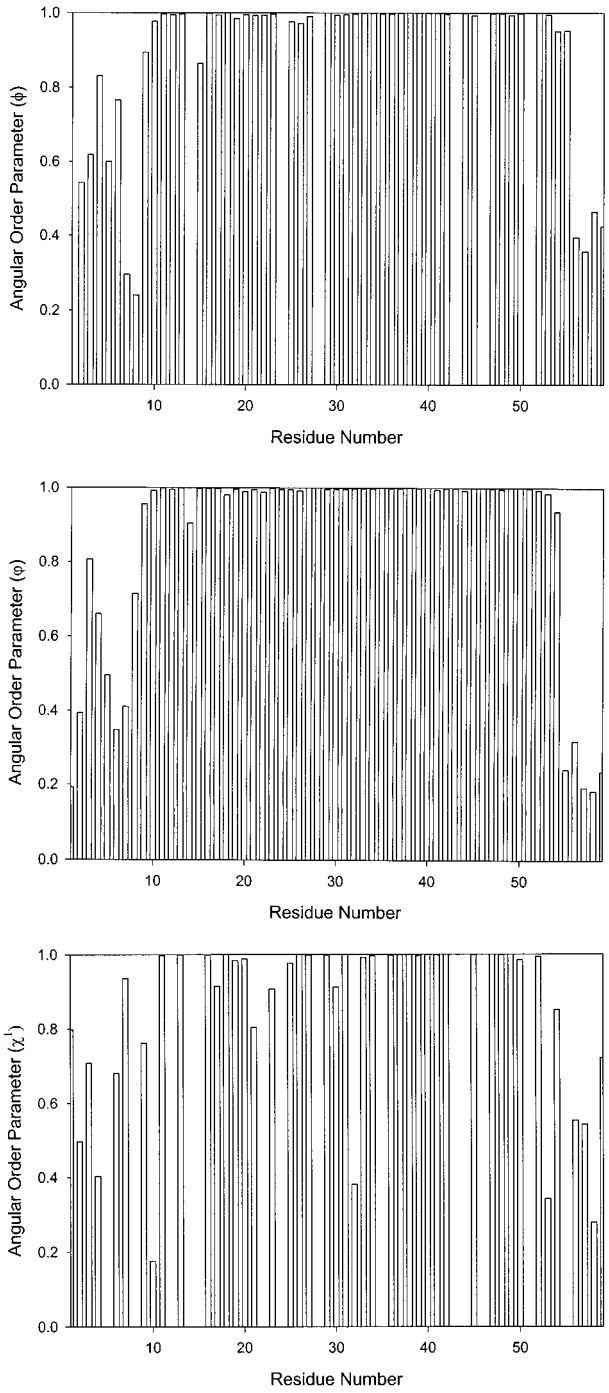

FIGURE 5: Summary of the precision to which dihedral angles have been determined in the family of 85 converged hITF structures. The histograms show the ϕ, ψ, and χ^1 angular order parameters for appropriate residues of hITF, and as a guide values of 0.95, 0.9, and 0.8 correspond to uncertainties in the angles of about ±17°, ±24°, and ±36°, respectively.

information was incorporated for 15 β, 4 γ, and 4 δ-methylene groups, 2 γ-methyl groups, and the side chain amide group of Gln 10.

After the final round of DYANA calculations, 85 satisfactorily converged hITF structures were obtained from 100 random starting conformations. The converged structures contain no distance constraint or van der Waals violations greater than 0.5 Å and no dihedral angle violations greater than 5°, with an average value for the final DYANA target function of 4.33 ± 0.34 Å^2 (Table 1). The family of

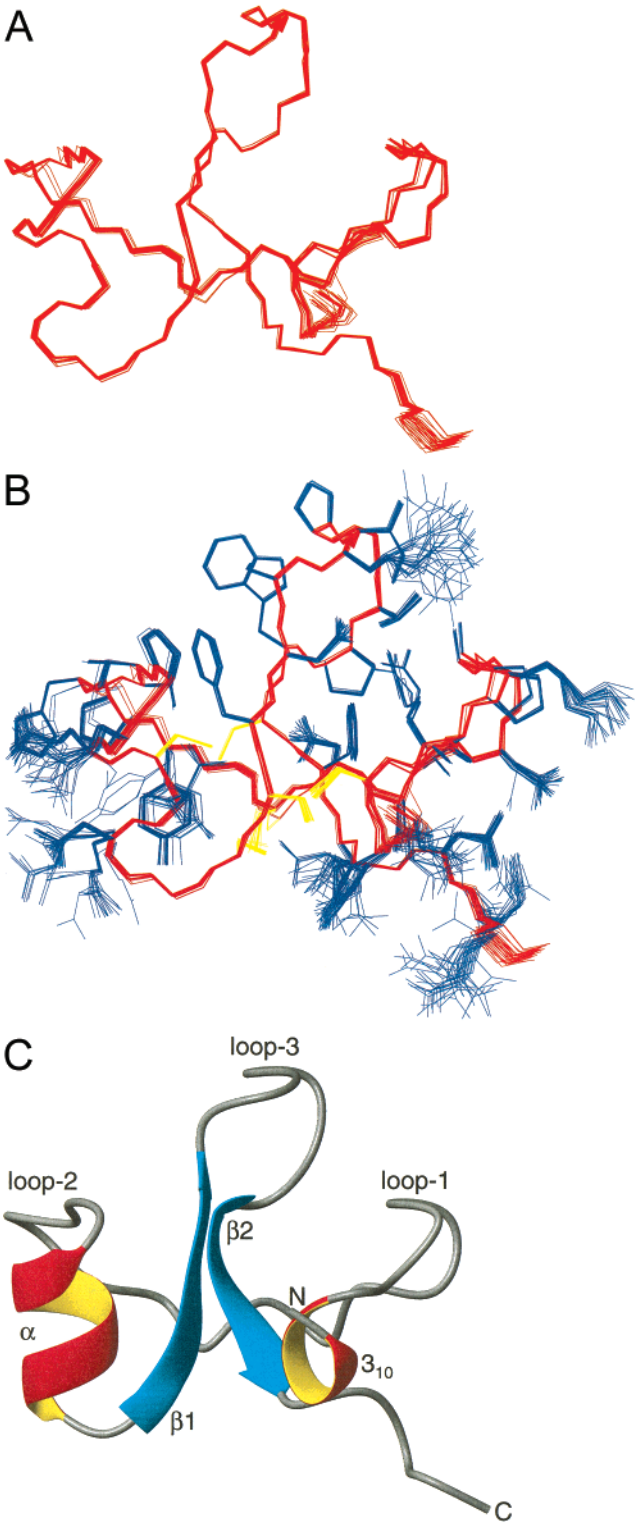

FIGURE 6: Panel A shows a best-fit superposition of the protein backbone for residues 9−54 from the 30 converged hITF structures with the lowest DYANA target functions. In panel B the side chains of the residues have also been included, with the backbone still shown in red, conserved cysteines in yellow, and all other side chains in blue. Panel C contains a ribbon representation of the backbone topology of hITF, in which the main structural features of the protein have been labeled including the α- and 3_{10}-helices (α and 3_{10}), the two-stranded β-sheet (β1 and β2), the three loop regions (loop 1, loop 2, and loop 3), and the N- and C-termini (N and C). In the view shown the central hairpin loop (loop 3) is clearly visible, with the short α-helix and loop 2 located to the left, and the N-terminal loop 1 to the right.

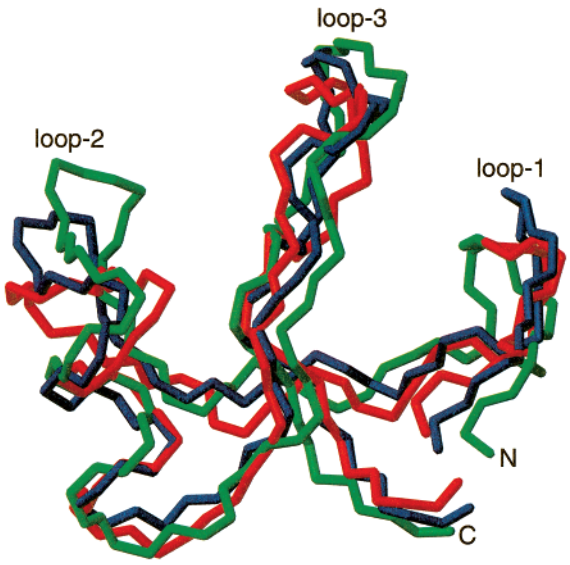

FIGURE 7: Comparison of the backbone topologies of the single trefoil domains present in hITF (red) and pS2 (green, *18*) and of the second trefoil motif found in pSP (blue, *13*). The trefoil domains in hITF (Cys 11 to Pro 51), pS2 (Cys 7 to Pro 47), and pSP (Cys 58 to Pro 98) were placed in the same orientation by superimposing the backbone atoms of the conserved α-helical and β-sheet regions. The orientation of the trefoil domains is very similar to that shown in Figure 6, with the central loop 3 clearly visible and the preceding α-helix and loop 2 located to the left and loop 1 to the right. The very significant differences in the conformation of the loop 2 region are clearly visible.

converged hITF structures, together with the NMR constraints, has been deposited in the Protein Data Bank (accession codes 1e9t and r1e9tmr, respectively).

The hITF structures all show good nonbonded contacts, as illustrated by the Ramachandran plot shown in Figure 4, in which 56% of the non-glycine and non-proline residues lie within the most favored regions, 30% in additional allowed regions, 12% in generously allowed regions, and only 2% in disallowed regions (*37*). On average, there is only one residue per structure (0.95 ± 0.93) with a combination of ϕ and ψ that lies outside of the favored regions and no residues that consistently adopt unfavorable backbone conformations. In addition, all of the residues that occasionally show nonpreferred combinations of ϕ and ψ (Glu 2, Tyr 3, Ser 7, Ala 8, Asn 9, Glu 56, Cys 57, and Thr 58) are poorly constrained by the NOE data (Figure 3) and consequently have low angular order parameters (less than 0.9, which corresponds to an uncertainty of over ±24° in the angles; *38*) for one or both of these backbone dihedral angles (Figure 5).

Figure 3 clearly shows that the conformation of residues 10−53 from hITF is determined by a large number of NOE-based distance constraints (an average of 21.8 per residue) and the structure of this region of the protein is determined to very high precision. This is reflected in low root mean squared deviations (rmsds) to the mean structure for both the backbone and all heavy atom coordinates of these residues of 0.18 ± 0.04 Å and 0.66 ± 0.08 Å, respectively, for the family of 85 converged hITF structures (Table 1). In addition, the angular order parameters for the backbone ϕ and ψ dihedral angles of residues 10−53 are all over 0.95, except for the ψ of Pro 14 and ϕ of Ala 15 (Figure 5), which indicates an uncertainty of less than ±17° (*38*). The side

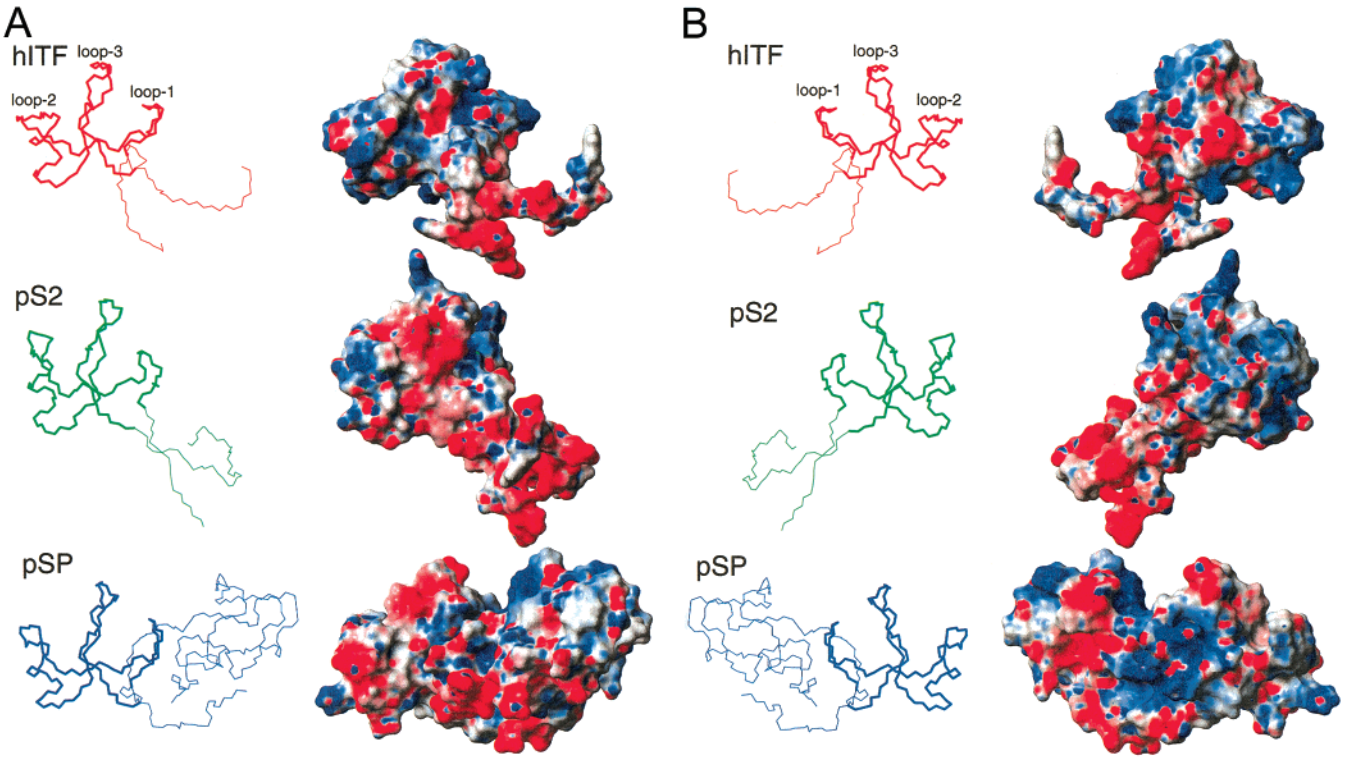

FIGURE 8: Comparison of the molecular surfaces of the solution structures of hITF (top), pS2 (middle), and pSP (bottom). In panel A the proteins are shown in the same orientation as in Figure 7, while in panel B the molecules are rotated by 180° about the vertical axis to show the opposite face. For clarity, the backbone topology of each of the proteins is shown to the left of the surface view, with the relevant trefoil domain highlighted by thicker bonds for the polypeptide backbone. The molecular surface is colored according to electrostatic potential, with areas of significant negative charge shown in red, significant positive charge in blue, and neutral as white. The electrostatic potential surface was calculated using MOLMOL (*36*) with the threshold for depicting significant areas of charge set to ±15, which is about twice that required to obtain a neutral representation for the fully solvent exposed aromatic ring of Phe 59 in hITF.

chain conformations of many residues in this region of the protein are also very well defined with 29 of the 33 non-glycine, proline, or alanine residues having angular order parameters of over 0.9 for χ^1, corresponding to a variation of less than ±24° (Figure 5). The N- (Glu 1 to Ala 8) and in particular C-terminal residues (Ala 55 to Phe 59) of hITF are poorly constrained by the NMR data (Figure 3), which is reflected in low angular order parameters (Figure 5) and in average local rmsds for the backbone atoms of 0.67 ± 0.21 and 2.03 ± 1.52 Å respectively, as compared to 0.05 ± 0.03 Å for the structured core of the protein (residues 10−53). The poorly defined conformation for the N- and C-termini of hITF is probably indicative of significant mobility in solution, which is consistent with fast exchange of a number of these backbone amide protons with water.

Figure 6A shows a best fit of the protein backbone for residues 9−54 from the best 30 converged hITF structures, with the side chains added in Figure 6B, and clearly illustrates the high precision to which the structure of hITF has been determined. In common with the structures determined for the trefoil domains in pSP (*13*−*16*) and pS2 (*17*, *18*), the trefoil domain in hITF (Cys 11 to Pro 51) is clearly composed of three looplike regions (Figure 6C). The central core of the domain consists of a short antiparallel β-sheet formed from residues Cys 36 to Asp 39 and Trp 47 to Lys 50, which are joined by an irregular but well-defined loop and together form the central hairpin loop of the trefoil domain (loop 3). The β-sheet is preceded by a short 6-residue α-helix formed from residues Lys 29 to Arg 34, which packs against the N-terminal strand of the sheet. The majority of the remainder of the domain is taken up by two loops formed from residues His 25 to Pro 28 (loop 2) and Ala 12 to Arg 18 (loop 1), which lie on either side of the central hairpin and are joined by an extended region formed from residues Val 19 to Pro 24 (Figure 6C). The other notable feature is that residues Asn 9 to Cys 11 form a single turn of a 3_{10}-helix. The N- and C-termini of the trefoil domain are located adjacent to each other on the lower right-hand side of the hITF structure as shown in Figure 6. The disulfide bonding pattern in the trefoil domain of hITF (Cys 11−Cys 37, Cys 21−Cys 36, and Cys 31−Cys 48) was unequivocally determined from preliminary structures calculated in the absence of any disulfide bond constraints and corresponds to that predicted from the known structures of pSP and pS2 (1−5, 2−4 and 3−6; *13*, *15*, *16*, *18*).

Comparisons of the Structures of Trefoil Domains in hITF, pSP, and pS2. Figure 7 provides a visual comparison of the backbone conformations of the single trefoil domains present in hITF (Cys 11 to Pro 51) and pS2 (Cys 7 to Pro 47) and of the second trefoil motif found in pSP (Cys 58 to Pro 98). The overall topology of the trefoil domains is clearly pretty similar in all three proteins and is confirmed by the pairwise rmsd values for the backbone atoms, which for the solution structures range from 1.44 ± 0.13 Å for hITF, compared to pSP, to 1.98 ± 0.18 Å for hITF, compared to pS2. In addition, the locations of the core β-sheet and preceding α-helix are conserved in all three trefoil proteins (Figure 1). There is also a striking similarity in the positions and

interactions made by the side chains of many highly conserved residues, such as those corresponding to Cys 11, Arg 18, Cys 21, Cys 31, Cys 36, Cys 37, Phe 38, Asp 39, Pro 46, Trp 47, Cys 48, and Phe 49 in hITF (Figure 1). A number of long-range side chain to side chain interactions that are characteristic of the trefoil domains in pSP and pS2 are also observed in hITF, including the arrangement of the disulfide bonds and the salt bridge between the conserved residues corresponding to Arg 18 and Asp 39 in hITF (*15*, *18*).

The major differences between the trefoil domains in hITF, pS2, and pSP are in the conformation of the loop 2 region. In hITF the backbone of residues Tyr 23 to Pro 28 forms an essentially flat, extended loop; however, the equivalent region in pS2 and in both domains of pSP adopts an uplifted, turn-like conformation, which gives the appearance of a pronounced cleft between loop 2 and loop 3 (Figures 6 and 7). This region contains a conserved aromatic residue, corresponding to Tyr 23 in hITF, Phe 19 in pS2, and Phe 21/Tyr 70 in pSP (Figure 1), which is located on the exposed surface of loop 2 in the pS2 and pSP trefoil structures (*13*, *18*). The position of this aromatic side chain is significantly shifted in hITF where it is found between the base of loop 2 and the central hairpin and appears to have a more structural role (Figure 6). The side chain of Pro 24 in hITF occupies a very similar position on the surface of loop 2 to that of Phe 19 in pS2 and Phe21/Tyr 70 in pSP, with the imidazole ring of His 25 adjacent to it and also solvent exposed. The substitution of a histidine at position 25 in hITF compared to a glycine in the trefoil domains of pS2 and pSP is clearly not structurally conservative and is the only significant difference in the sequences of this loop 2 region, which strongly suggests that it is responsible for the distinct conformation of loop 2 in hITF.

Comparisons of the Proposed Functional Sites on hITF, pSP, and pS2. The presence of a cluster of conserved residues on the surface of loop 2 and the adjacent loop 3 in pS2 and pSP, together with the pronounced cleft between the tips of loop 2 and loop 3 in pSP, has previously led to proposals that this region of the trefoil domain forms the functional site for protein−protein or protein−-sugar interactions (*13*, *15*, *18*). Panels A and B of Figure 8 provide views of both faces of the molecular surfaces of hITF, pSP, and pS2, with the single trefoil domains in hITF and pS2 shown in the same orientation as domain II of pSP. The shape of the loop 2/loop 3 region is strikingly different in the three proteins; in particular, the obvious cleft between the loops in domains I and II of pSP is completely closed in pS2, while in hITF it has become a fully open, shelf-like feature. In the case of hITF and pSP, the shelf or cleft and the facing surface of loop 3 are predicted to be predominantly nonpolar in nature at neutral pH (Figure 8). Similarly, electrostatic calculations suggest that the external surface of loop 2 is primarily positively charged in hITF and pS2, while in both domains of pSP it is predicted to have a mainly positively charged face (Figure 8B) but with the edge formed by the α-helix carrying a significant net negative charge (Figure 8A). This is probably due to the presence of several acidic residues at the N-terminus of the helix in the trefoil domains of pSP.

CONCLUSIONS

Comparisons of the high-resolution structures determined for trefoil domains in hITF, pS2, and pSP clearly identify the conformation and nature of the loop 2 region and of the cleft between loop 2 and loop 3 as the most variable features of the trefoil motifs. The striking differences in both shape and charge for the proposed functional site on the trefoil domains of hITF, pS2, and pSP strongly suggest that this family of cell motility factors will act via binding to distinct target molecules, most probably a group of related membrane-bound receptors. This would certainly explain the lack of functional redundancy observed in vivo, which was dramatically demonstrated by the severe pathological effects seen in transgenic mice lacking either pS2 or ITF, where the normal functions of ITF or pS2 in maintaining the integrity of the lining of the gastrointestinal tract could clearly not be compensated for by the two remaining trefoil proteins (*4*, *5*). The detailed structural comparisons have also provided a clearer idea of which conserved surface residues are probably involved in the binding of target molecules, which can hopefully be confirmed in the future by a focused program of site-directed mutagenesis.

ACKNOWLEDGMENT

We thank Dr. Richard Williamson for many useful discussions.

REFERENCES

1. Thim, L. (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 Lett. 250*, 85−90.
2. Hoffman, W., and Hauser, F. (1993) The P-domain or trefoil motif: a role in renewal and pathology of mucous epithelia, *Trends Biochem. Sci. 18*, 239−243.
3. Wong, W. M., Poulsom, R., and Wright, N. A. (1999) Trefoil peptides, *Gut 44*, 890−895.
4. Lefebvre, O., Chenard, M.-P., Masson, R., Linares, J., Dierich, A., LeMeur, M., Wendling, C., Tomasetto, C., Chambon, P., and Rio, M.-C. (1996) Gastric mucosa abnormalities and tumorigenesis in mice lacking the pS2 trefoil protein, *Science 274*, 259−262.
5. Mashimo, H., Wu, D.-C., Podolsky, D. K., and Fishman, M. C. (1996) Impaired defence of intestinal mucosa in mice lacking intestinal trefoil factor, *Science 274*, 262−265.
6. Dignass, A., Lynch-Devaney, K., Kindon, H., Thim, L., and Podolsky, D. K. (1994) Trefoil peptides promote epithelial migration through a transforming growth factor beta-independent pathway, *J. Clin. Invest. 94*, 376−383.
7. Playford, R. J., Marchbank, T., Chinery, R., Evison, R., Pignatelli, M., Boulton, R. A., Thim, L., and Hanby, A. M. (1995) Human spasmolytic polypeptide is a cytoprotective agent that stimulates cell migration, *Gastroenterology 108*, 108−116.
8. Frandsen, E. K., Jogensen, K. H., and Thim, L. (1986) Receptor binding of pancreatic spasmolytic polypeptide (PSP) in rat intestinal mucosal cell membranes inhibits the adenylate cyclase activity, *Regul. Pept. 16*, 291−297.
9. Chinery, R., and Cox, H. (1995) Immunoprecipitation and characterisation of a binding protein specific for the peptide, intestinal trefoil factor, *Peptides 16*, 749−755.
10. Hoffman, W. (1988) A new repetitive protein from *Xenopus laevis* skin highly homologous to pancreatic spasmolytic polypeptide, *J. Biol. Chem. 263*, 7686−7690.

11. Hauser, F., and Hoffman, W. (1992) P-domains as shuffled cysteine-rich modules in integumentary mucin C.1 (FIM-C.1) from *Xenopus laevis*, *J. Biol. Chem. 267*, 24620−24624.
12. Bork, P. (1993) A trefoil domain in the major rabbit zona pellucida protein, *Protein Sci. 2*, 669−670.
13. Carr, M. D., Bauer, C. J., Gradwell, M. J., and Feeney, J. (1994) Solution structure of a trefoil-motif-containing cell growth factor, porcine spasmolytic protein, *Proc. Natl. Acad. Sci. U.S.A. 91*, 2206−2210.
14. Carr, M. D. (1992) ^{1}H NMR-based determination of the secondary structure of porcine pancreatic spasmolytic polypeptide: one of a new family of trefoil motif containing cell growth factors, *Biochemistry 31*, 1998−2004.
15. Gajhede, M., Petersen, T. N., Henriksen, A., Petersen, J. F. W., Dauter, Z., Wilson, K. S., and Thim, L. (1993) Pancreatic spasmolytic polypeptide: first three-dimensional structure of a member of the mammalian trefoil family of peptides, *Structure 1*, 253−262.
16. De, A., Brown, D. G., Gorman, M. A., Carr, M. D., Sanderson, M. R., and 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. U.S.A. 91*, 1084−1088.
17. Polshakov, V. I., Frenkiel, T. A., Westley, B., Chadwick, M., May, F., Carr, M. D., and Feeney, J. (1995) 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.
18. 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) High-resolution solution structure of human pNR-2/pS2: a single trefoil motif protein, *J. Mol. Biol. 267*, 418−432.
19. Thim, L., Wöldike, H. F., Nielsen, P. F., Christensen, M., Lynch-Devaney, K., and Podolsky, D. K. (1995) Characterisation of human and rat intestinal trefoil factor produced in yeast, *Biochemistry 34*, 4757−4764.
20. States, D. J., Haberkorn, R. A., and Ruben, D. J. (1982) A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants, *J. Magn. Reson. 48*, 286−292.
21. Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., and Wüthrich, K. (1983) Improved spectral resolution in COSY ^{1}H NMR spectra of proteins via double quantum filtering, *Biochem. Biophys. Res. Commun. 117*, 479−485.
22. Braunschweiler, L., and Ernst, R. R. (1983) Coherence transfer by isotropic mixing: application to proton correlation spectroscopy, *J. Magn. Reson. 53*, 521−528.
23. Davis, D. G., and Bax, A. (1985) Assignment of complex ^{1}H NMR spectra via two-dimensional homonuclear Hartmann−Hahn spectroscopy, *J. Am. Chem. Soc. 107*, 2820−2821.
24. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) Investigation of exchange processes by two-dimensional NMR, *J. Chem. Phys. 71*, 4546−4553.
25. Macura, S., and Ernst, R. R. (1980) Elucidation of cross relaxation in liquids by two-dimensional NMR, *Mol. Phys. 41*, 95−117.
26. Bothner-By, A. A., Stephens, R. L., Lee, J., Warren, C. D., and Jeanloz, R. W. (1984) Structure determination of a tetrasaccharide: transient nuclear Overhauser effects in the rotating frame, *J. Am. Chem. Soc. 106*, 811−813.
27. Bax, A., and Davis, D. G. (1985) Practical aspects of two-dimensional transverse NOE spectroscopy, *J. Magn. Reson. 65*, 207−213.
28. Bax, A., and Davis, D. G. (1985) MLEV-17-based two-dimensional homonuclear magnetisation transfer spectroscopy, *J. Magn. Reson. 65*, 355−360.
29. Sklenar, V., Piotto, M., Leppik, R., and Saudek, V. (1993) Gradient-tailored water suppression for ^{1}H-^{15}N HSQC experiments optimised to retain full sensitivity, *J. Magn. Reson. 102*, 241−245.
30. Bartels, C., Xia, T., Billeter, M., Güntert, P., and Wüthrich, K. (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules, *J. Biomol. NMR 5*, 1−10.
31. Güntert, P., Braun, W., and Wüthrich, K. (1991) Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA, *J. Mol. Biol. 217*, 517−530.
32. Wüthrich, K., Billeter, M., and Braun, W. (1993) Pseudo-structures for the 20 common amino acids for use in studies of protein conformation by measurement of intramolecular proton−proton distance constraints with nuclear magnetic resonance, *J. Mol. Biol. 169*, 949−961.
33. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
34. Szyperski, T., Güntert, P., Otting, G., and Wüthrich, K. (1992) Determination of scalar coupling constants by inverse Fourier transformation of in-phase multiplets, *J. Magn. Reson. 99*, 552−560.
35. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA, *J. Mol. Biol. 273*, 283−298.
36. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures, *J. Mol. Graphics 14*, 51−55.
37. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
38. Hyberts, S. G., Goldberg, M. S., Havel, T. F., and Wagner, G. (1992) The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures, *Protein Sci. 1*, 736−751.

BI010184+