The EMBO Journal Vol.16 No.22 pp.6659–6666, 1997

Solution structure of the transforming growth factor β-binding protein-like module, a domain associated with matrix fibrils

Xuemei Yuan[1,2], A.Kristina Downing[1,2], Vroni Knott[2,3] and Penny A.Handford[2,3,4]

[1]Department of Biochemistry, University of Oxford, Oxford OX1 3QU, [2]Oxford Centre for Molecular Sciences, New Chemistry Laboratory, University of Oxford, Oxford OX1 3QT and [3]Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, UK

[4]Corresponding author
e-mail: Penny.Handford@pathology.ox.ac.uk

Here we describe the high resolution nuclear magnetic resonance (NMR) structure of a transforming growth factor β (TGF-β)-binding protein-like (TB) domain, which comes from human fibrillin-1, the protein defective in the Marfan syndrome (MFS). This domain is found in fibrillins and latent TGF-β-binding proteins (LTBPs) which are localized to fibrillar structures in the extracellular matrix. The TB domain manifests a novel fold which is globular and comprises six antiparallel β-strands and two α-helices. An unusual cysteine triplet conserved in the sequences of TB domains is localized to the hydrophobic core, at the C-terminus of an α-helix. The structure is stabilized by four disulfide bonds which pair in a 1–3, 2–6, 4–7, 5–8 pattern, two of which are solvent exposed. Analyses of MFS-causing mutations and the fibrillin-1 cell-binding RGD site provide the first clues to the surface specificity of TB domain interactions. Modelling of a homologous TB domain from LTBP-1 (residues 1018–1080) suggests that hydrophobic contacts may play a role in its interaction with the TGF-β1 latency-associated peptide.
Keywords: 8-cysteine domain/human fibrillin-1/Marfan syndrome/NMR structure/TB domain

Introduction

The transforming growth factor β (TGFβ)-binding protein-like (TB, see Bork *et al.*, 1996) domain, also referred to as the 8-cysteine domain, is found in proteins [fibrillins 1, 2 and latent TGF-β-binding proteins (LTBPs)] localized to extracellular matrix fibrils which are involved in extracellular matrix architecture and storage of latent TGF-β (Sakai *et al.*, 1991; Dallas *et al.*, 1995; Taipale *et al.*, 1996). It is characterized by eight conserved cysteine residues, which include an unusual cysteine triplet. Functional studies have directly implicated this domain in cell adhesion of fibrillin-1 (Pfaff *et al.*, 1996; Sakamoto *et al.*, 1996) and in the association of LTBP-1 with latent TGF-β1 (Gleizes *et al.*, 1996; Saharinen *et al.*, 1996). The discovery of 11 genetic mutations so far in TB domains from human fibrillin-1, which cause the Marfan syndrome (MFS), emphasizes the biological significance of this module (reviewed in Dietz and Pyeritz, 1995; Ades *et al.*, 1996; Collod-Béroud *et al.*, 1997). However, the structural effects of these disease-causing mutations are not understood.

Here we present the structure of the sixth TB domain (TB6) from human fibrillin-1, which adopts a novel fold as assessed by a protein structural database alignment using the program STAMP (Russell and Barton, 1992). We analyse MFS-causing mutations in terms of the TB domain structure. A model for a homologous TB domain from LTBP-1 (residues 1018–1080, Kanzaki *et al.*, 1990) is also presented and evaluated in terms of covalent and non-covalent interactions required for the formation of large latent TGF-β complexes.

Results

Solution structure of TB6

The structure determination of TB6 from human fibrillin-1 (residues 2054–2125) was based on a total of 1612 nuclear Overhauser effect (NOE)-derived interproton distance constraints, 24 distance restraints for 12 backbone hydrogen bonds and 34 backbone torsion angle φ constraints (see Materials and methods). The observation of a high number of interproton interactions for TB6 is a consequence of the compactness of the fold and of the high solubility of the domain. The 21 final structures overlaid in Figure 1 were selected based on agreement with the experimental data, with no NOE violations greater than 0.4 Å and no torsion angle violations greater than 3°. The root-mean-square deviation to the unminimized average coordinates is 0.53 ± 0.14 Å for the backbone atoms and 0.97 ± 0.17 Å for the heavy atoms of residues 2058–2115, hence the coordinates for the structure are well defined. A statistical analysis of the 21 structures in terms of agreement with the experimental data and idealized covalent geometry is presented in Table I. There is evidence for *cis–trans* isomerization of two prolines in the TB domain in solution, which is associated with minor, local conformational changes (X.Yuan and A.K.Downing, in preparation).

The secondary structure of the TB6 domain is illustrated schematically in Figure 2. The six β-strands form a four-stranded β-sheet (Figure 2, strands B, C, E and F) and a two-stranded β-sheet (Figure 2, strands A and D). The four-stranded β-sheet packs very closely with helix 1 and forms the central part of the globular structure, while the two-stranded β-sheet and helix 2 pack on either side of the central core. The first two cysteines of the triplet (residues 2083–2085) adopt a helical conformation at the end of helix 1. The disulfide bonds form in a 1–3, 2–6, 4–7, 5–8 pattern to join the secondary structure elements and stabilize the overall fold. This is consistent with chemical analysis of TB6 (this study) and with previous analyses of TB4 from fibrillin-1 and a TB domain from

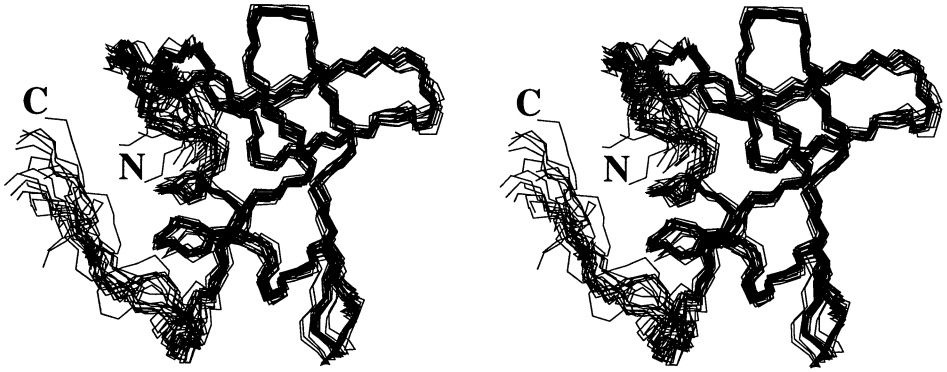

Fig. 1. Stereoview superposition of the 21 final structures calculated from the NMR data. The structures are overlaid based on the backbone (C^{α}, C, N) coordinates of residues 2058–2115 which have backbone order parameters (S^2) > 0.7. A detailed presentation of the relaxation data for this domain will form the subject of a subsequent manuscript (X.Yuan and A.K.Downing, in preparation).

Table I. Statistics for the 21 TB domain structures

	<SA>[a]	(SA)
R.m.s. deviation from experimental distance restraints (Å)		
all (1636)	0.022 ± 0.002	0.016
interproton distances		
intraresidue (491)	0.023 ± 0.003	0.020
sequential (366)	0.015 ± 0.003	0.009
inter-residue short-range ($1<\|i–j\|\leqslant 4$) (254)	0.024 ± 0.004	0.015
inter-residue long-range ($\|i–j\|>4$) (402)	0.024 ± 0.004	0.018
ambiguous (99)	0.023 ± 0.008	0.014
H-bonds (24)	0.015 ± 0.004	0.012
R.m.s. deviations from experimental ϕ restraints (°)		0.573
(34)	0.631 ± 0.079	
Deviations from idealized covalent geometry		
bonds (Å)	0.003 ± 0.0002	0.002
angles (°)	0.462 ± 0.029	0.387
impropers (°)	0.456 ± 0.028	0.398
F_{NOE} (kJ/mol)	169.2 ± 31.2	90.1
F_{CDIH} (kJ/mol)	3.5 ± 0.9	2.8
F_{repel} (kJ/mol)	81.3 ± 19.5	41.8

[a]The notation is as follows: <SA> denotes the 21 structures calculated from the NMR data; (SA) refers to the average minimized structure.

LTBP-1 which showed that no free thiols were present in these domains (Gleizes *et al.*, 1996; Reinhardt *et al.*, 1996). NMR spectra recorded with increasing calcium concentration showed that the TB domain does not bind calcium (see Materials and methods). Two conserved carboxylate/carboxyamide residues (D2055, E2097 in TB6) map to the molecular surface of TB6 and form salt bridges with adjacent conserved basic residues (R2057, K2080, see Figure 3A). W2092, which is located in strand E, plays an important structural role in the centre of the hydrophobic core. This is consistent with the high conservation of an aromatic residue at this position in the TB modules of fibrillin-1, fibrillin-2 and LTBPs (see Figure 3). The C-terminus of the TB6 domain is flexible and the structural consequence of this for fibrillin-1 is currently under investigation.

Analysis of the disulfide bond arrangement of this domain isolated from native fibrillin has not been performed due to the difficulties of obtaining pure fibrillin and the complexity of mapping disulfides in such cysteine-rich proteins. Therefore, the evidence for TB6 adopting a native fold, although substantive, is indirect. Data suggestive of a native fold include: (i) one major form of the protein was produced by *in vitro* refolding and other disulfide-rich domains produced by these methods have adopted the native fold (Rao *et al.*, 1995; Knott *et al.*, 1996); (ii) TB6 displays a high degree of secondary structure with well-defined loop regions; (iii) the globular conformation of TB6 is consistent with the appearance, in electron micrographs, of fibrillin fragments with TB domains which are rod-shaped molecules containing globular structures (Reinhardt *et al.*, 1996); (iv) residues with functional significance (e.g. RGD motifs, glycosylation sites) map as expected to the surface of this structure; and (v) the pattern of amino acid conservation in TB domain sequences (Figure 3A) can be explained entirely in terms of the structure.

Discussion

The distribution of TB domains

The domain organization of proteins containing TB domains is illustrated schematically in Figure 4. Sequences for TB domains found in the fibrillin and LTBP families show a high level of amino acid conservation (see Materials

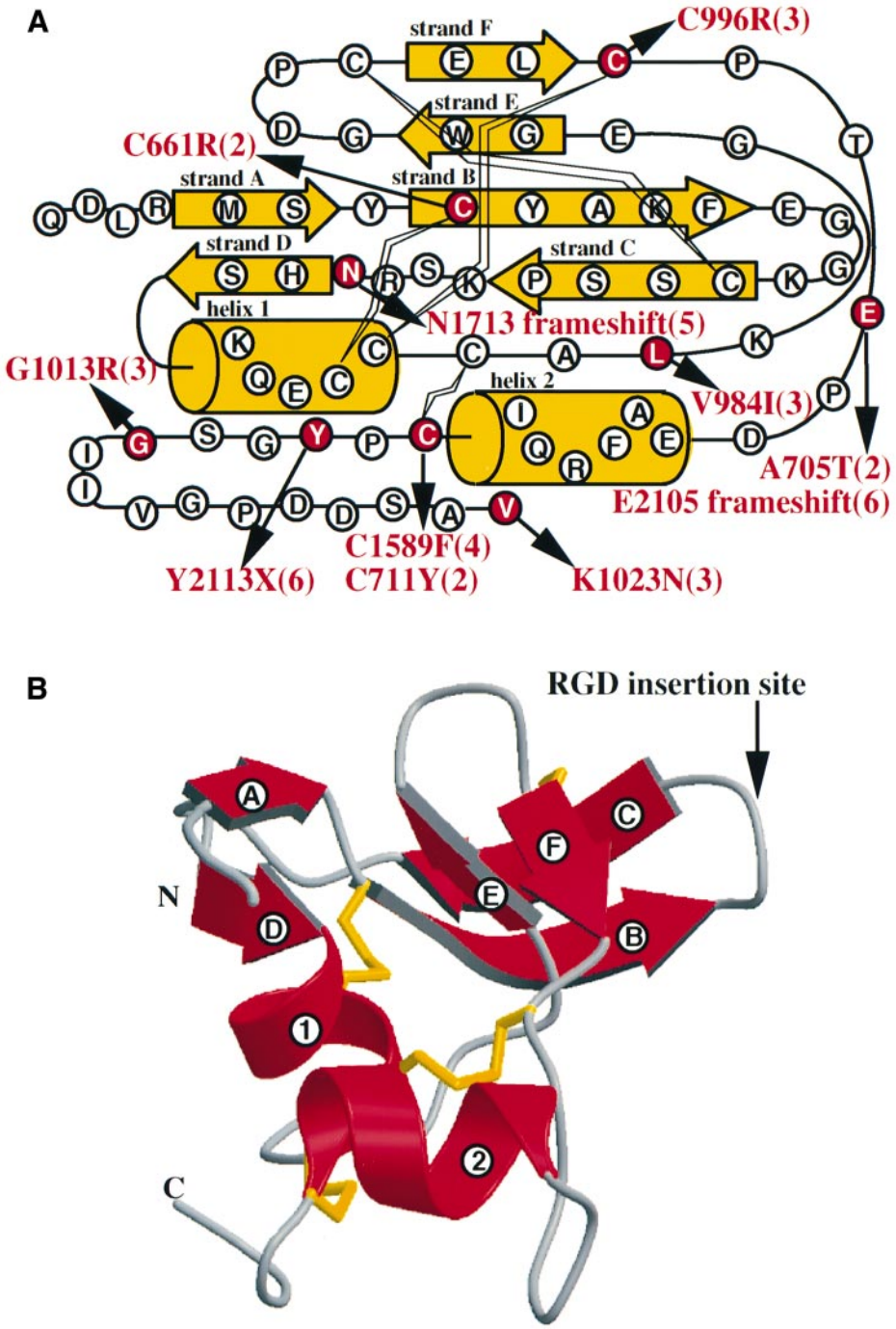

Fig. 2. Illustrations of the (**A**) secondary structure and (**B**) tertiary structure of the TB module. β-strands are shown as arrows and α-helices are depicted as cylinders (A) or coils (B). In (A) the positions of the 11 MFS-causing mutations reported for TB domains in human fibrillin-1 are highlighted; the number in brackets refers to the TB domain in which the mutation is located. In (B) disulfide bonds are shown in yellow and the site of the RGD sequence in fibrillin TB4 is highlighted. Part (B) was rendered (Merritt and Murphy, 1994) from MOLSCRIPT (Kraulis, 1991) input.

and methods), and one would predict that these would adopt similar folds (see Figure 3).

Other proteins exist which contain less similar domains of at least eight cysteine residues including an unusual Cys triplet. These include a homologue of NADH dehydrogenase subunit 7 (Koslowsky *et al.*, 1990), a high sulfur keratin (Swart and Haylett, 1973), interleukin-14 precursor (Ambrus *et al.*, 1993) and bovine immunodeficiency virus transactivating transcriptional regulatory proteins (BIV Tats, Garvey *et al.*, 1990). However, many of the secondary structure elements seen in TB6 encoded by residues between the cysteines are absent in these proteins, so it is uncertain whether these homologies are of structural significance. It will be particularly interesting to see if the cysteine triplet in these proteins is at the end of a helix as in TB6.

Role of the TB domain in microfibril assembly

The domain organization of fibrillin is striking in that the TB modules interrupt multiple tandem stretches of

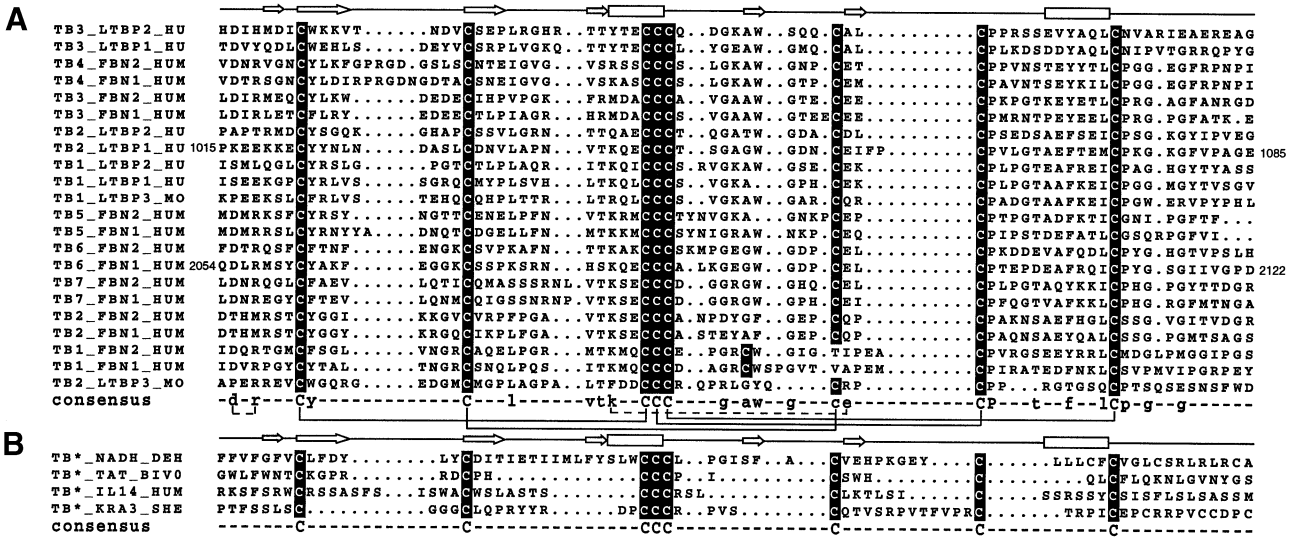

Fig. 3. Multiple sequence alignments and consensus sequences for (**A**) human fibrillin and LTBP TB domains (the first and last residue numbers of TB6_FBN1_HUM and TB2_LTBP1_HU are highlighted), and (**B**) other proteins containing the TB domain pattern of eight cysteine residues. The secondary structure of TB6 is illustrated above each alignment, with arrows shown for β-strands and boxes for α-helices; the sequence order displayed is according to the program MULTALIGN (Barton, 1990). In (A) the mouse sequences for LTBP-3 TB domains are shown because the human sequence for this protein has not been reported. The disulfide connections (solid lines) and salt bridges (dashed lines) are shown under the alignment. In the LTBP proteins an additional N-terminal TB-like domain may exist which contains seven rather than eight cysteines (Gleizes *et al.*, 1996; Saharinen *et al.*, 1996). In (B) TB* represents a module which contains the TB pattern of eight cysteines but lacks the additional conserved residues seen in the fibrillin and LTBP proteins (see text). This figure has been shaded using BOXSHADE written by K.Hofmann and M.Baron.

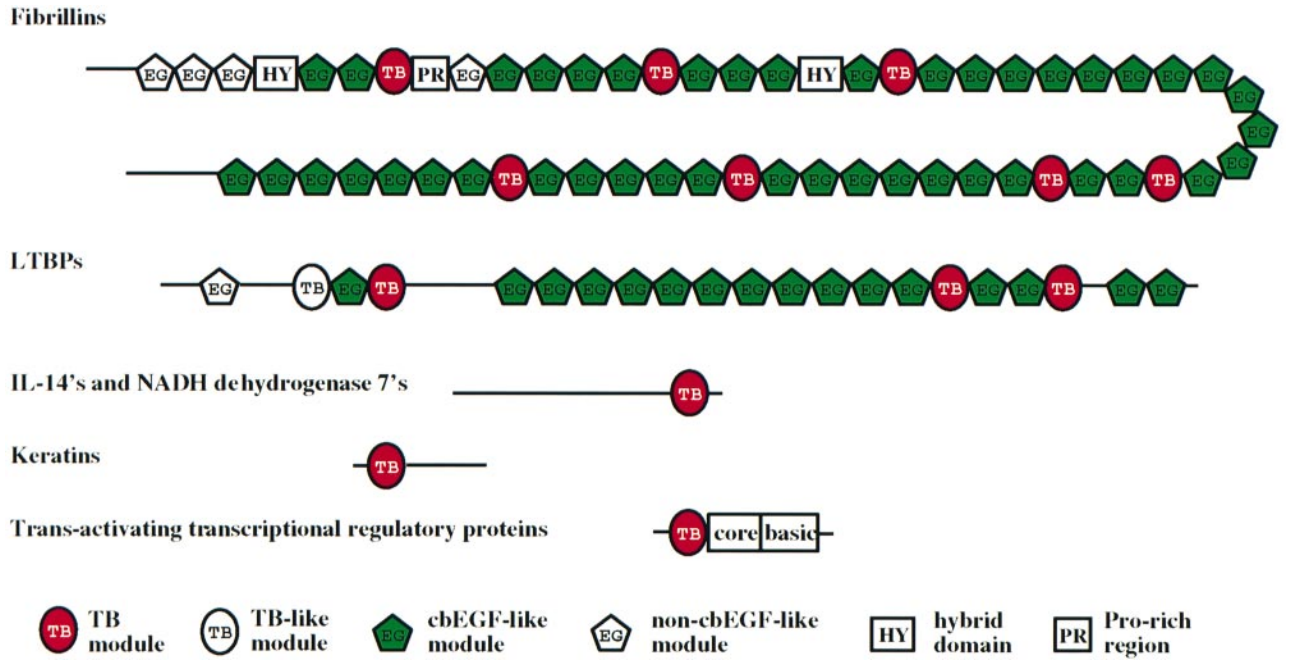

Fig. 4. Domain organization of proteins containing the TB domain consensus pattern of eight cysteine residues (see text).

calcium-binding epidermal growth factor-like (cbEGF) domains which are thought to form rod-like structures (Downing *et al.*, 1996). Do all TB domains in the fibrillins play the same structural role? In the fibrillins, the sequence of TB1 differs from those for TB2–TB7. In Figure 3, the sixth cysteine of TB1 is not aligned. However, if TB1 is modelled based on the alignment shown in Figure 3, (i) the aromatic residue in strand E which plays a key structural role (see above) is conserved and (ii) cysteines 2 and 6 are still in close enough spatial proximity to form a disulfide bond. Hence it is likely that all fibrillin TB domains adopt the same fold.

What role might the TB module play in stabilizing fibrillin interactions within the microfibril? The TB module has been shown to mediate covalent protein interactions via disulfide exchange in the LTBP–latent TGF-β complex during secretion (see below); however, monomeric fibrillin fragments containing TB domains have been expressed in mammalian cells (Reinhardt *et al.*, 1996), and no evidence of dimerization was observed in the structural analysis of

TB6, where protein concentrations up to 5–6 mM were used. Moreover, the ability of the TB module to reshuffle disufide bonds (Gleizes *et al.*, 1996) does not seem likely in the extracellular matrix since the redox potential of the extracellular matrix is favourable to the formation of stable disulfide bonds.

Mutations in TB domains implicated in the Marfan syndrome

Fibrillin-1 is a major structural component of 10 nm connective tissue microfibrils in the extracellular matrix (Dietz and Pyeritz, 1995). Mutations in the gene encoding human fibrillin-1 cause the Marfan syndrome (MFS), an autosomal dominant disease of connective tissue which occurs at a frequency of at least 1 in 5000 in the population (Sakai and Keene, 1994; Dietz and Pyeritz, 1995). Human fibrillin-1 is mainly comprised of 47 tandemly clustered EGF-like domains separated by seven TB modules (Figure 4). These two families of modules are thought to play an important role in the organization and/or function of fibrillin within the connective tissue microfibril, and mutations in both module types cause MFS.

It is now possible to evaluate the consequences of MFS-causing mutations in terms of the TB module structure. There are altogether 11 reported mutations to the TB modules in human fibrillin-1 which cause MFS (see Figure 2A). Those which involve cysteines (C661R, C711Y, C996R and C1589F) are likely to cause domain misfolding, since one of the four disulfide bonds would be disrupted. Three mutations related to either a frameshift (N1713 and E2105) or to the formation of a premature stop codon (Y2113X) will cause disruption of the primary sequence of this domain. Point mutations to TB domains include A705T, V984I, G1013R and K1023N. We are unable to interpret K1023N, since this residue lies in the flexible C-terminus of the structure. Residues A705 and V984 correspond to E2105 and L2087 in TB6 respectively, whose side chains are both completely exposed. Mutations of these surface residues could therefore interrupt intra- or intermolecular domain interactions. The G1013R mutation is located five to six residues after the eighth cysteine residue; this glycine is strictly conserved in the TB domains which are followed by a cbEGF domain, and it is interesting to speculate that a glycine may be required at this position to achieve a specific orientation of the TB domain with respect to the adjacent cbEGF domain in the microfibril.

TB domains involved in cell adhesion and TGF-β1 targeting

An RGD sequence located in the fourth TB module of both fibrillin-1 and fibrillin-2 (see Figure 3) defines a major cell-binding epitope in the specific interaction between fibrillins and cell surface receptor integrin αvβ3 (Pfaff *et al.*, 1996; Sakamoto *et al.*, 1996). TB4 is the only TB module in human fibrillin-1 that has an RGD sequence. The position of the RGD insertion relative to the TB6 structure is at the end of the β-hairpin formed by strands B and C which points into solution (Figure 2), and thus would be accessible for integrin binding. Interestingly, a second RGD site is present in the third TB module of fibrillin-2 which is not functional in adhesion studies (Sakamoto *et al.*, 1996). This is predicted by homology to be located in the flexible C-terminal region of the TB structure. Structural studies of the TB6–cbEGF32 pair of domains (currently in progress) will elucidate whether or not this region is involved in inter-domain packing interactions and therefore inaccessible for protein binding. The identification of functional regions of TB domains which map to the surface will facilitate future modelling studies of fibrillin, since they specify the orientation of these domains within the microfibril.

As illustrated in Figure 4, LTBPs (LTBP-1, -2 and -3) contain multiple cbEGF domains and TB modules (Kanzaki *et al.*, 1990; Moren *et al.*, 1994; Gibson *et al.*, 1995; Li *et al.*, 1995; Yin *et al.*, 1995). LTBPs mediate the secretion and subsequent localization of latent TGF-β complexes (comprising TGF-β and its latency-associated propeptide) to the extracellular matrix where, upon activation, TGF-β plays an important role in many biological processes including the maintenance of tissue homeostasis (reviewed in McPherron and Lee, 1996). In the covalent association of LTBP-1 and TGF-β1 latency-associated propeptide (TGF-β1 LAP), a TB module of LTBP-1 exchanges disulfide bonds with TGF-β1 LAP during secretion (Gleizes *et al.*, 1996; Saharinen *et al.*, 1996). In TB6, two of the disulfide bonds joining cysteines 2–6 and 4–7 (C2070–C2096 and C2084–C2099) are localized to the surface of the domain and therefore have the potential to disulfide exchange under appropriate conditions. However, additional specificity must be required for this to occur because only one of the TB modules from LTBP-1 is involved in this interaction.

The TB domain from LTBP-1 which mediates the interaction with TGF-β1 LAP contains two additional residues (F and P) at the end of strand F (see Figures 2 and 4), which is part of a four-stranded β-sheet. It is possible that a conformational change associated with the insertion of these residues contributes to the interaction of TGF-β1 LAP with LTBP-1. Modelling of this TB domain from LTBP-1 reveals that the domain has markedly different surface hydrophobicity to TB6. As shown in Figure 5, the site of the FP insertion is in the middle of a hydrophobic patch adjacent to the seventh cysteine, which suggests that hydrophobic contacts may play a role in recognition of LTBP-1 by TGF-β1 LAP.

From studies of the LTBPs and fibrillins, it is clear that the TB module plays an important role in protein–protein interactions and can mediate binding to other proteins either via covalent or non-covalent interactions. However, this disulfide-rich domain may also contribute to the biomechanical properties of fibrillar structures in the matrix. Recent investigations of fibrillin-containing microfibrils in the sea cucumber *Cucumaria frondosa* have shown that intact disulfide bonds are important for the linear elasticity of these structures (Thurmond and Trotter, 1996).

Materials and methods

Protein expression and purification

A DNA fragment encompassing nucleotides 6293–6508 of the human fibrillin-1 cDNA, corresponding to amino acids 2054–2125 (numbering according to Pereira *et al.*, 1993), was amplified by PCR using *Pfu* polymerase (Stratagene). The primers used for amplification of the TB domain were: 5′TAGTAGGGATCCATAGAAGGACGATCAGCACAAGATTTGCGAATGAGCTAC (forward primer) and 5′TAGTAGAA-

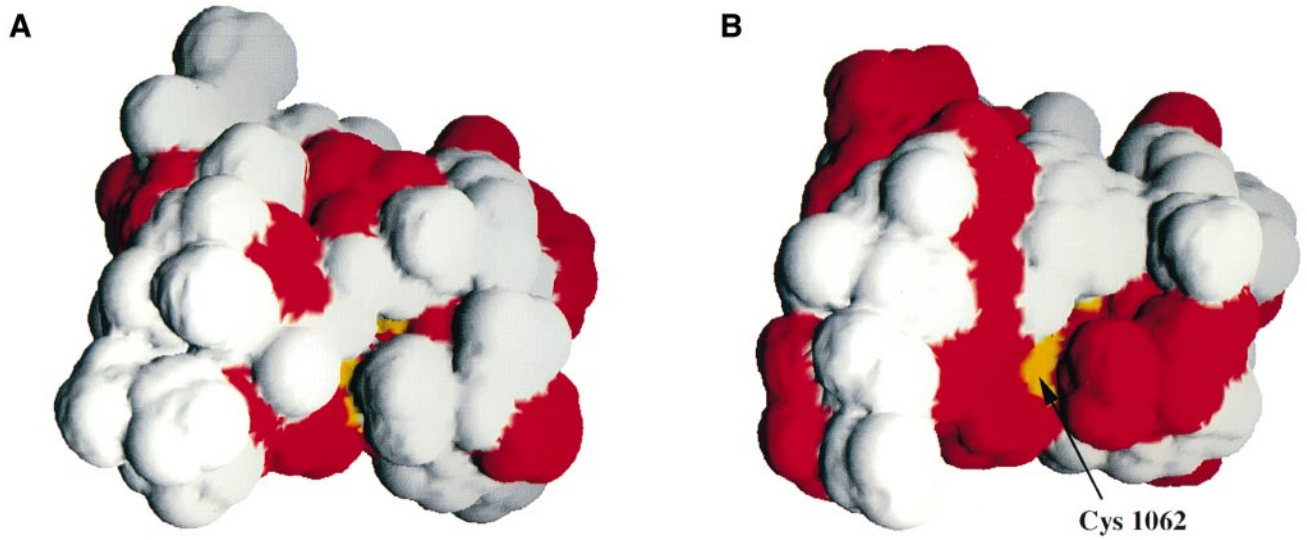

Fig. 5. Comparison of the molecular surfaces of (**A**) fibrillin-1 TB6 and (**B**) TB domain (residues 1018–1080) from LTBP-1. Hydrophobic residues (Ala, Leu, Val, Ile, Met, Pro, Phe, Tyr, Trp) are shown in red. The seventh cysteine in each domain is shaded in yellow. Cys1062 is adjacent to the FP insertion in the TB domain from LTBP-1.

GCTTTTATGCTGAATCATCAGGTCCC (reverse primer), containing restriction sites for cloning purposes (*Bam*HI for the forward primer and *Hin*dIII for the reverse primer). The forward primer also contained a sequence encoding a factor Xa cleavage site. The PCR fragment was cloned into plasmid pQE30 (Qiagen) downstream of a sequence encoding a six-histidine tag. The resulting plasmid was grown in *Escherichia coli* NM554 (Raleigh *et al.*,1988) containing a *lac* repressor plasmid (pRep4, Qiagen). The positive clone selected for protein expression was sequenced using an automated DNA sequencer (Applied Biosystems).

The recombinant clone was grown at 37°C in 2× TY medium containing 25 μg/ml kanamycin and 100 μg/ml ampicillin. Expression of the recombinant protein was induced by adding IPTG to a concentration of 2 mM at the end of log phase, and growth continued for a further 3 h. Cells were lysed in 6 M guanidine–HCl, 50 mM sodium phosphate, 5 mM 2-mercaptoethanol (pH 6) at room temperature for 1–2 h (50 ml buffer/3 l culture), sonicated to reduce viscosity and the solution clarified by centrifugation at 40 000 r.p.m. in a Beckman type 60 fixed angle rotor for 45 min. The cleared lysate was loaded onto a column containing Ni^{2+} chelating Sepharose (Pharmacia) at 0.5 ml/min, washed with the same buffer and eluted in loading buffer containing 20 mM EDTA. The pooled eluate was adjusted to 0.1 M DTT and 0.1 M Tris–HCl (pH 8.3) and reduced for 1–2 h at room temperature, acidified to pH 3 and dialysed against 0.1% TFA prior to purification by reverse phase HPLC (Handford *et al.*, 1990). The purified protein was lyophilized to concentrate and digested with bovine factor Xa (Denzyme) in 50 mM Tris–HCl, 0.1 M NaCl, 1 mM $CaCl_2$ (pH 7.5) at 37°C for 16 h (1:1000 enzyme:protein) to remove the histidine tag. The cleavage product was reduced in 6 M guanidine–HCl, 0.1 M DTT, 0.1 M Tris–HCl (pH 8.3) for 1–2 h at room temperature before refolding by dialysis against 50 mM Tris–HCl, 3 mM cysteine, 0.3 mM cystine (pH 8.3) with three changes. The refolded material was purified by ion-exchange chromatography (MonoQ-Pharmacia) using a salt gradient from 0 to 0.5 M NaCl in 50 mM Tris–HCl pH 7.5, followed by reverse phase HPLC and lyophilization.

Ellman assay for thiols

Protein samples were assayed for free thiols by the addition of 50 μl of 3 mM 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB) to 1 ml protein solution (containing >2 nmol purified protein) in 6 M guanidinium chloride (GdmCl) in 0.1 M phosphate buffer (pH 7.3)/0.001 M EDTA. The absorbance of the solution was measured at 412 nm and free thiols calculated from the molar absorbance of the TNB anion (ε_{412} = 13 700/Mcm in GdmCl).

NMR spectroscopy

Protein samples for homonuclear experiments contained approximately 2.7 mM purified protein at pH 5.6 in 0.55 ml 90% H_2O/10% ^2H_2O or in 0.55 ml 99.9% ^2H_2O. The concentrations for the labelled samples were ~5.7 mM and 0.8 mM for the ^{15}N-labelled sample and the ^{15}N,^{13}C-labelled sample respectively, at pH 5.6 in 0.55 ml 90% H_2O/10% ^2H_2O. NMR spectra were recorded at 500, 600 and 750 MHz proton frequencies and at temperatures of 25°C and 30°C. NOE intensity calibrations and distance restraints derived for structure calculations were based on the NMR spectra acquired at 25°C. Felix 2.3 (Biosym, Inc.) was used to process the NMR data. Initial sequential assignments were made based on comparison of homonuclear two-dimensional NOESY (Jeener *et al.*, 1979; Macura *et al.*, 1981) at mixing times of 150 and 200 ms, HOHAHA (Aue *et al.*, 1976; Brown *et al.*, 1988) with t_m = 62 ms and COSY (Braunschweiler and Ernst, 1983; Bax and Davis, 1985) spectra. These assignments were confirmed and completed using comparison of strips extracted from within the amide region of a gradient-enhanced ^{15}N-separated NOESY spectrum (Kay *et al.*, 1989, 1992) and an HBHA(CBCACO)NH spectrum (Grzesiek and Bax, 1993). These spectra were recorded on a home-built/GE Omega spectrometer operating at 600 MHz, fitted with a triple resonance probe with self-shielded pulsed field gradients. The three-dimensional NOESY experiment (t_m = 150 ms) was recorded with acquisition times of 114 ms in the direct ^1H (F_3) dimension, 11 ms in the ^{15}N (F_2) dimension and 48 ms in the indirect ^1H (F_1) dimension, in ~3 days. Linear prediction was used to double the number of points in the F_2 dimension. The three-dimensional HBHA(CBCACO)NH spectrum was acquired with acquisition times of 41 ms in the direct ^1H (F_3) dimension, 14.8 ms in the ^{15}N (F_2) dimension and 3.7 ms in the indirect ^1H (F_1) dimension, in ~3 days. $^3J_{HNH\alpha}$ coupling constants were measured by line-shape fitting F_1 traces from a HMQC-J dataset (Kay and Bax, 1990), which was recorded at acquisition times of 114 ms in the direct ^1H (F_2) dimension and 110 ms in the ^{15}N (F_1) dimension. Slowly exchanging amide protons were identified in a series of HSQC spectra (Kay *et al.*, 1992) recorded after dissolving the sample in ^2H_2O. To assess calcium binding of the TB module, the ^{15}N-labelled NMR sample dissolved in ^2H_2O was adjusted to pH = 6.5. Three aliquots of calcium chloride (10 μl of 200 mM stock solution each) were added to the NMR sample and a HSQC spectrum was recorded after each addition.

Derivation of experimental restraints

A two-dimensional NOESY spectrum, recorded on the unlabelled NMR sample in 99.9% ^2H_2O, and a three-dimensional NOESY spectrum, acquired on the ^{15}N-labelled sample in 90% H_2O/10% ^2H_2O, were used for deriving the distance restraints. The mixing times for these two experiments were both 150 ms. The program NMRView (Merck and Co., Inc.) was used to assist with NOE assignment, based on proton chemical shifts and quantitation of crosspeak intensities. These intensities were converted to four categories of distance constraint: 2.8, 3.5, 5.0 and 7.0 Å. The first three categories were calibrated according to known

secondary structure distances (Wüthrich, 1986). Very weak NOEs in the fourth category (containing 260 distance restraints) were assigned in the 150 ms mixing time three-dimensional NOESY spectrum based on the preliminary structures calculated in their absence.

Backbone torsion angle ϕ restraints were incorporated for residues with $^3J_{HNH\alpha} < 6.0$ Hz and $^3J_{HNH\alpha} > 8.0$ Hz with a minimum range of $\pm$ 30°. Slowly exchanging backbone amide protons were identified and those that could be assigned to regions of α-helix and β-sheet secondary structures on the basis of initial structure calculations were each constrained to form regular HN–CO bonds by two distance restraints, $d_{O-N} = 3.0 \pm 0.3$ Å and $d_{O-HN} = 2.0 \pm 0.3$ Å. All torsion angle and hydrogen bond restraints were consistent with initial structures calculated in their absence.

Structure calculations

A total number of 1670 constraints, comprised of 491 intraresidue ($|i-j| = 0$), 366 sequential ($|i-j| = 1$), 254 short-range ($|i-j| \leq 4$), 402 long-range ($|i-j| > 4$) and 99 ambiguous interproton distance restraints, 24 distance restraints for 12 backbone hydrogen bonds and 34 backbone torsion angle ϕ restraints were used in the structure calculations. Structures were calculated from the experimental restraints by *ab initio* simulated annealing from an extended template structure with ideal covalent geometry (Nilges *et al.*, 1988, 1991), using the program X-PLOR v3.1 (Brünger, 1992) and v3.851 topology and parameter files. NOE-derived constraints were implemented using SUM averaging, so no pseudoatom corrections were applied. Ambiguous crosspeaks were treated as described by Nilges (1995). The energy minimization routine utilized a 'floating chirality' protocol, in which non-stereospecifically assigned valine and leucine methyl carbons and methylene hydrogens are allowed to swap positions over the course of the initial calculations. Disulfide bond constraints were treated ambiguously in initial calculations, and the pairing of the cysteines was clearly defined by the NMR data. Structures were refined with two additional cycles of simulated annealing. Twenty-one final structures were selected based on no distance violations greater than 0.4 Å, no torsion angle violations greater than 3°, and $F_{NOE} < 225$ kJ/mol. An average coordinate structure was calculated and energy minimized in X-PLOR (Brünger, 1992).

TB module identification, alignment and conservation analysis

The program MOTIFS (Genetics Computing Group, 1994) was used to search the Owl protein sequence database (release 24.0) using PROSITE (Bairoch, 1992) patterns. TB domains were identified using the pattern: ×7C~C{2,16}C~C{2,16}CCC~C{2,16}C~C{2,12}C~C{2,16}C×12. Multiple sequence alignments were performed using the program MULTALIGN (Barton, 1990). The criteria for amino acid conservation in Figure 3A were satisfied if at any one position, one, two or three residues of similar size and charge/hydrophobicity were observed, with a net conservation of >60%.

LTBP-1 TB domain modelling

The structure of the TB domain from LTBP-1 (residues 1018–1080, numbering according to Kanzaki *et al.*, 1990) which mediates the interaction with TGF-β1 LAP was modelled by homology using the program Insight 2.3 (Biosym, Inc.). The FP insertion (residues 1060–1061) was not included. The structure was energy minimized in X-PLOR (Brünger, 1992) to alleviate unfavourable non-bonded interactions.

Acknowledgements

The authors are grateful to Iain Campbell for critical reading of the manuscript and helpful discussions, and to Jonathan Boyd, Nick Soffe and Jörn Werner for technical advice. X.Y. is supported by an ORS grant and the OCMS. A.K.D. and V.K. thank the support of the Wellcome Trust. P.A.H. is a Royal Society University Research Fellow.

References

Ades,L.C., Haan,E.A., Colley,A.F. and Richards,R.I. (1996) Characterization of 4 novel fibrillin-1 (FBN1) mutations in Marfan-syndrome. *J. Med. Genet.*, **33**, 665–671.

Ambrus,J.L. *et al.* (1993) Identification of a cDNA for a human high-molecular-weight B-cell growth-factor. *Proc. Natl Acad. Sci. USA*, **90**, 6330–6334.

Aue,W.P., Bartholdi,E. and Ernst,R.R. (1976) Two-dimensional spectroscopy: application to nuclear magnetic resonance. *J. Chem. Phys.*, **64**, 2229–2246.

Bairoch,A. (1992) Prosite—a dictionary of sites and patterns in proteins. *Nucleic Acids Res.*, **20**, 2013–2018.

Barton,G.J. (1990) Protein multiple sequence alignment and flexible pattern-matching. *Methods Enzymol.*, **183**, 403–428.

Bax,A. and Davis,D.G. (1985) MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Resonance*, **65**, 355–360.

Bork,P., Downing,A.K., Kieffer,B. and Campbell,I.D. (1996) Structure and distribution of modules in extracellular proteins. *Q. Rev. Biophys.*, **29**, 119–167.

Braunschweiler,L. and Ernst,R.R. (1983) Coherence transfer by isotropic mixing—application to proton correlation spectroscopy. *J. Magn. Resonance*, **53**, 521–528.

Brown,S.C., Weber,P.L. and Mueller,L. (1988) Toward complete ^{1}H NMR spectra in proteins. *J. Magn. Resonance*, **77**, 166–169.

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

Collod-Béroud,G. *et al.* (1997) Marfan database (second edition): software and database for the analysis of mutations in the human FBN1 gene. *Nucleic Acids Res.*, **25**, 147–150.

Dallas,S.L., Miyazono,K., Skerry,T.M., Mundy,G.R. and Bonewald,L.F. (1995) Dual role for the latent transforming growth factor-β binding protein in storage of latent TGF-β in the extracellular matrix and as a structural matrix protein. *J. Cell Biol.*, **131**, 539–549.

Dietz,H.C. and Pyeritz,R.E. (1995) Mutations in the human gene for fibrillin-1 (FBN1) in the Marfan syndrome and related disorders. *Hum. Mol. Genet.*, **4**, 1799–1809.

Downing,A.K., Knott,V., Werner,J.M., Cardy,C.M., Campbell,I.D. and Handford,P.A. (1996) Solution structure of a pair of calcium-binding epidermal growth factor-like domains: Implications for the Marfan syndrome and other genetic disorders. *Cell*, **85**, 597–605.

Garvey,K.J., Oberste,M.S., Elser,J.E., Braun,M.J. and Gonda,M.A. (1990) Nucleotide-sequence and genome organization of biologically-active proviruses of the bovine immunodeficiency-like virus. *Virology*, **175**, 391–409.

Genetics Computing Group (1994) *Program manual for the GCG package, version 8*. University of Wisconsin, Madison, WI, USA.

Gibson,M.A., Hatzinikolas,G., Davis,E.C., Baker,E., Sutherland,G.R. and Mecham,R.P. (1995) Bovine latent transforming growth factor β1-binding protein 2: Molecular cloning, identification of tissue isoforms and immunolocalization to elastin-associated microfibrils. *Mol. Cell. Biol.*, **15**, 6932–6942.

Gleizes,P.-E., Beavis,R.C., Mazzieri,R., Shen,B. and Rifkin,D.B. (1996) Identification and characterization of an eight-cysteine repeat of the latent transforming growth factor-β binding protein-1 that mediates bonding to the latent transforming growth factor-β1. *J. Biol. Chem.*, **271**, 29891–29896.

Grzesiek,S. and Bax,A. (1993) Amino-acid type determination in the sequential assignment procedure of uniformly C-13/N-15-enriched proteins. *J. Biomol. NMR*, **3**, 185–204.

Handford,P.A., Baron,M., Mayhew,M., Willis,A., Beesley,T., Brownlee,G.G. and Campbell,I.D. (1990) The first EGF-like domain from human factor IX contains a high-affinity calcium binding site. *EMBO J.*, **9**, 475–480.

Jeener,J., Meier,B.H., Bachmann,P. and Ernst,R.R. (1979) Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.*, **71**, 4546–4553.

Kanzaki,T.A., Olofsson,A., Moren,A., Wernstedt,C., Hellman,U., Miyazono,K., Claesson-Welsh,L. and Heldin,C.-H. (1990) TGF-β1 binding protein: a component of the large latent complex of TGF-β1 with multiple repeat sequences. *Cell*, **61**, 1051–1061.

Kay,L.E. and Bax,A. (1990) New methods for the measurement of NH-C-alpha-H coupling-constants in N-15-labeled proteins. *J. Magn. Resonance*, **86**, 110–126.

Kay,L.E., Marion,D. and Bax,A. (1989) Practical aspects of 3D heteronuclear NMR of proteins. *J. Magn. Resonance*, **84**, 72–84.

Kay,L.E., Keifer,P. and Saarinen,T. (1992) Pure absorption gradient-enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.*, **114**, 10663–10665.

Knott,V., Downing,A.K., Cardy,C.M. and Handford,P. (1996) Calcium binding properties of an epidermal growth factor-like domain pair. *J. Mol. Biol.*, **255**, 22–27.

Koslowsky,D.J., Bhat,G.J., Perrollaz,A.L., Feagin,J.E. and Stuart,K. (1990) The murf3 gene of *Trypanosoma brucei* contains multiple domains of extensive editing and is homologous to a subunit of NADH dehydrogenase. *Cell*, **62**, 901–911.

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

Li,X., Yin,W., Perez-Jurado,L., Bonadio,J. and Francke,U. (1995) Mapping of human and murine genes for latent TGF-β binding protein-2 (LTBP-2). *Mamm. Genome*, **6**, 42–45.

Macura,S., Huang,Y., Suter,D. and Ernst,R.R. (1981) 2-dimensional chemical-exchange and cross-relaxation spectroscopy of coupled nuclear spins. *J. Magn. Resonance*, **43**, 259–281.

McPherron,A.C. and Lee,S.-J. (1996) In LeRoith,D. and Bondy,C. (eds), *Growth Factors and Cytokines in Health and Disease. Vol. 1B*. JAI, Greenwich, pp. 357–393.

Merritt,E.A. and Murphy,M.E.P. (1994) Raster3D, version 2.0: a program for photorealistic molecular graphics. *Acta Crystallogr.*, **50**, 869–873.

Moren,A. *et al.* (1994) Identification and characterization of LTBP-2, a novel latent transforming growth factor-β binding protein. *J. Biol. Chem.*, **269**, 32469–32478.

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

Nilges,M., Clore,G.M. and Gronenborn,A.M. (1988) Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. *FEBS Lett.*, **239**, 129–136.

Nilges,M., Kuszewski,J. and Brunger,A.T. (1991) Sampling properties of simulated annealing and distance geometry. In Hoch,J.C. *et al.* (eds), *Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy.* Plenum Press, New York, pp. 451–455.

Pereira,L., Dalessio,M., Ramirez,F., Lynch,J.R., Sykes,B., Pangilinan,T. and Bonadio,J. (1993) Genomic organization of the sequence coding for fibrillin, the defective gene-product in Marfan-syndrome. *Hum. Mol. Genet.*, **2**, 961–968.

Pfaff,M., Reinhardt,D.P., Sakai,L.Y. and Timpl,R. (1996) Cell-adhesion and integrin binding to recombinant human fibrillin-1. *FEBS Lett.*, **384**, 247–250.

Raleigh,E.A. *et al*. (1988) *Mcra* and *mcrb* restriction phenotypes of some *Escherichia coli* strains and implications for gene cloning. *Nucleic Acids Res.*, **16**, 1563–1575.

Rao,Z., Handford,P., Mayhew,M., Knott,V., Brownlee,G.G. and Stuart,D. (1995) The structure of a Ca^{2+}-binding epidermal growth factor-like domain: Its role in protein–protein interactions. *Cell*, **82**, 131–141.

Reinhardt,D.P., Keene,D.R., Corson,G.M., Poschl,E., Bachinger,H.P., Gambee,J.E. and Sakai,L.Y. (1996) Fibrillin-1: Organization in microfibrils and structural properties. *J. Mol. Biol.*, **258**, 104–116.

Russell,R.B. and Barton,G.J. (1992) Multiple protein-sequence alignment from tertiary structure comparison—assignment of global and residue confidence levels. *Proteins*, **14**, 309–323.

Saharinen,J., Taipale,J. and Keski-Oja,J. (1996) Association of the small latent transforming growth-factor-β with an 8-cysteine repeat of its binding-protein LTBP-1. *EMBO J.*, **15**, 245–253.

Sakai,L.Y. and Keene,D.R. (1994) Fibrillin: Monomers and microfibrils. *Methods Enzymol.*, **245**, 29–52.

Sakai,L.Y., Keene,D.R., Glanville,R.W. and Bachinger,H.P. (1991) Purification and partial characterization of fibrillin, a cysteine-rich structural component of connective-tissue microfibrils. *J. Biol. Chem.*, **266**, 14763–14770.

Sakamoto,H., Broekelmann,T., Cheresh,D.A., Ramirez,F., Rosenbloom,J. and Mecham,R.P. (1996) Cell-type specific recognition of RGD- and non-RGD-containing cell binding domains in fibrillin-1. *J. Biol. Chem.*, **271**, 4916–4922.

Swart,L.S. and Haylett,T. (1973) Studies on the high-sulphur proteins of reduced merino wool: amino acid sequence of protein SCMKB-IIIA3. *Biochem. J.*, **133**, 641–654.

Taipale,J., Saharinen,J., Hedman,K. and Keski-Oja,J. (1996) Latent transforming growth factor-β1 and its binding protein are components of extracellular matrix microfibrils. *J. Histochem. Cytochem.*, **44**, 875–889.

Thurmond,F.A. and Trotter,J.A. (1996) Morphology and biomechanics of the microfibrillar network of sea cucumber dermis. *J. Exp. Biol.*, **199**, 1817–1828.

Wüthrich,K. (1986) *NMR of Proteins and Nucleic Acids*. John Wiley and Sons, Inc., New York.

Yin,W., Smiley,E., Germiller,J., Mecham,R.P., Florer,J.B., Wenstrup,R.J. and Bonadio,J. (1995) Isolation of a novel latent transforming growth factor-β binding protein gene (LTBP-3). *J. Biol. Chem.*, **270**, 10147–10160.

Received on August 1, 1997; revised on September 4, 1997