The EMBO Journal Vol.16 No.13 pp.3778–3786, 1997

Crystal structure of a pair of follistatin-like and EF-hand calcium-binding domains in BM-40

Erhard Hohenester[1], Patrik Maurer[2,3] and Rupert Timpl[2]

Department of Crystallography, Birkbeck College, London WC1E 7HX, UK and [2]Abteilung Proteinchemie, Max-Planck-Institut für Biochemie, 82152 Martinsried, Germany

[3]Present address: Institut für Biochemie der Medizinischen Fakultät, Universität Köln, 50931 Köln, Germany

[1]Corresponding author

BM-40 (also known as SPARC or osteonectin) is an anti-adhesive secreted glycoprotein involved in tissue remodelling. Apart from an acidic N-terminal segment, BM-40 consists of a follistatin-like (FS) domain and an EF-hand calcium-binding (EC) domain. Here we report the crystal structure at 3.1 Å resolution of the FS–EC domain pair of human BM-40. The two distinct domains interact through a small interface that involves the EF-hand pair of the EC domain. Residues implicated in cell binding, inhibition of cell spreading and disassembly of focal adhesions cluster on one face of BM-40, opposite the binding epitope for collagens and the *N*-linked carbohydrate. The elongated FS domain is structurally related to serine protease inhibitors of the Kazal family. Notable differences are an insertion into the inhibitory loop in BM-40 and a protruding N-terminal β-hairpin with striking similarities to epidermal growth factor. This hairpin is likely to act as a rigid spacer in proteins containing tandemly repeated FS domains, such as follistatin and agrin, and forms the heparin-binding site in follistatin.

Keywords: crystal structure/extracellular module/follistatin/glycoprotein/serine protease inhibitor

Introduction

BM-40 (also termed osteonectin or SPARC, for secreted protein, acidic and rich in cysteines) is an abundant secreted glycoprotein that modulates interactions between cells and extracellular matrix (Lane and Sage, 1994; Sage, 1997). Expression of BM-40 is generally associated with tissues undergoing remodelling, either during normal developmental processes, such as angiogenesis or mineralization, or as a response to injury. During embryogenesis, BM-40 expression is temporarily and spatially regulated, and overexpression in *Caenorhabditis elegans* leads to severe defects (Schwarzbauer and Spencer, 1993). BM-40 has been termed an anti-adhesive protein because it inhibits cell spreading, induces rounding of cultured endothelial cells and fibroblasts and disassembles focal adhesions (Lane and Sage, 1994; Sage, 1997). Significantly, BM-40 is highly expressed in different malignant tumours and it has been shown recently that BM-40 plays a crucial role in the tumorigenicity of human melanoma cells (Ledda *et al.*, 1997). Other activities of BM-40 include calcium-dependent binding to collagens and thrombospondin, binding to platelet-derived growth factor (PDGF)-AB and -BB and the regulation of cell proliferation and matrix metalloproteinase expression (Lane and Sage, 1994; Sage, 1997).

The sequence of BM-40 has been highly conserved throughout vertebrate evolution (Lane and Sage, 1994), and a homologue with 38% identity to human BM-40 has also been found in the nematode *C.elegans* (Schwarzbauer and Spencer, 1993). The human protein consists of 286 residues divided into three distinct domains: (i) an N-terminal acidic segment (residues 1–52) binding several calcium ions with low affinity and mediating the interaction with hydroxyapatite; (ii) a follistatin-like (FS; Patthy and Nikolics, 1993; Bork *et al.*, 1996) domain (residues 53–137) containing five disulfides and an *N*-linked complex carbohydrate at Asn99; and (iii) an α-helical domain containing two EF-hand calcium-binding sites (EC domain, residues 138–286; Maurer *et al.*, 1995; Hohenester *et al.*, 1996). Pairs of FS and EC domains are found in a number of extracellular proteins, including SC1/hevin (Johnston *et al.*, 1990; Girard and Springer, 1996), QR1 (Guermah *et al.*, 1991), testican (Alliel *et al.*, 1993) and tsc-36 (Shibanuma *et al.*, 1993). However, with the exception of hevin, which modulates high endothelial cell adhesion to the basement membrane (Girard and Springer, 1996), the functions of proteins related to BM-40 are unknown.

We recently have solved the crystal structure at 2.0 Å resolution of the EC domain of human BM-40 (Hohenester *et al.*, 1996). This structure revealed a canonical pair of EF-hand calcium-binding sites (Kretsinger, 1996) elaborated by three additional helices that contain the collagen-binding site of BM-40 (Sasaki *et al.*, 1997). Here we report the structure of the FS–EC domain pair of human BM-40 at 3.1 Å resolution. Apart from the presumably loosely structured N-terminal acidic segment, this domain pair encompasses the entire BM-40 molecule and thus allows a comprehensive structural rationalization of previous peptide mapping studies. In addition, we provide the first structure of a representative FS domain. We also propose a model for FS tandem repeats, which occur in several extracellular proteins, including the activin antagonist follistatin (Michel *et al.*, 1993; Hemmati-Brivanlou *et al.*, 1994; Matzuk *et al.*, 1995) and agrin (Patthy and Nikolics, 1993; Kleinman and Reichardt, 1996).

Results and discussion

Overall structure of the BM-40 FS–EC domain pair

The structure of BM-40 FS–EC is shown in Figure 1. As anticipated from the modular architecture of BM-40, the

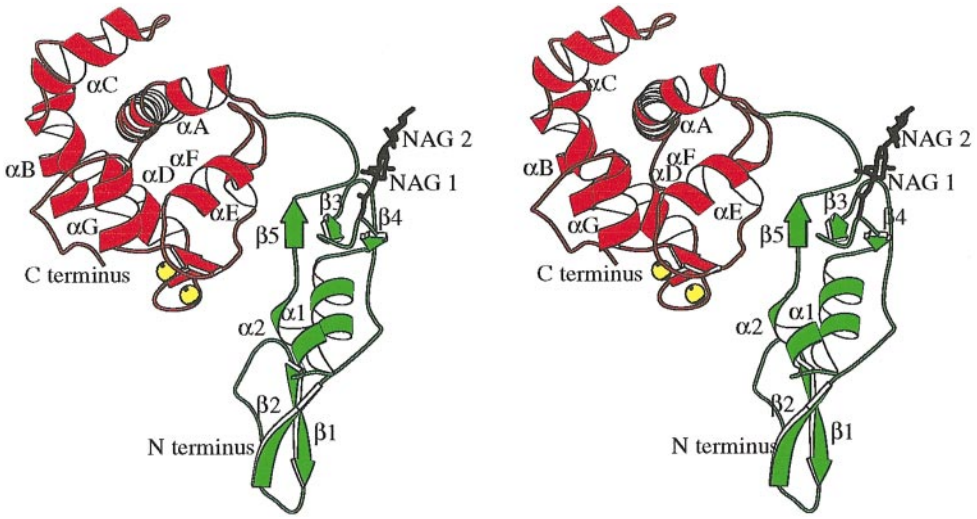

Fig. 1. Stereo view (Kraulis, 1991) of the BM-40 FS–EC structure and its secondary structure elements. The FS domain (residues 54–137) is shown in green and consists of an N-terminal β-hairpin and a small hydrophobic core of α/β structure. The EC domain (residues 138–286) is in red and consists of a pair of EF-hand calcium-binding sites and helices A, B and C. The calcium ions bound to the EF-hands are shown as yellow spheres. The FS and EC domains interact through a small interface involving mainly β5 and the preceding loop of the FS domain and αE of the EC domain. The FS domain is glycosylated at Asn99; the first two *N*-acetylglucosamine (NAG) sugar moieties are included in the crystallographic model and are shown in atomic detail.

structure is composed of two distinct domains. The FS domain is very elongated and consists of an N-terminal β-hairpin and a small core of mixed α/β structure; the EC domain is globular and almost entirely α-helical (Hohenester *et al.*, 1996). The interface between the two domains is small (550 Å^2) and predominantly polar. In the FS domain, the interface involves strand β5 and the preceding loop, while the EC domain contributes the proline-rich loop connecting the two EF-hand calcium-binding sites and, in particular, helix E. The domain interface contains a total of four hydrogen bonds, including a potential salt bridge between Asp127 and His232. Apolar contacts are observed between Leu126 and Thr233 and between Ile135/Pro136 and His247. Interestingly, His62 in the first reverse turn of the FS domain interacts with His224, which takes part in the *cis*-peptide bond in the unusual EF-hand 1 of the EC domain (Hohenester *et al.*, 1996). Residues involved in interdomain contacts in BM-40 are highly conserved in the related proteins SC1/hevin (Johnston *et al.*, 1990; Girard and Springer, 1996) and QR1 (Guermah *et al.*, 1991), providing strong evidence that the relative domain orientation observed in the BM-40 FS–EC structure represents the biologically active conformation and is not an artefact of crystallization. The intimate involvement of the EF-hand pair in the domain interface makes it likely that the relative orientation of the FS and EC domains is dependent on the presence of calcium. The isolated EC domain has been shown to bind calcium with 10-fold reduced affinity compared with BM-40 (Maurer *et al.*, 1995), demonstrating a small but significant mutual stabilizing effect of calcium binding and interdomain contacts.

The glycosylation site of BM-40 at Asn99 is located in the reverse turn between strands β3 and β4 of the FS domain. About 10% of the total mass of BM-40 is carbohydrate, depending on the tissue from which the protein is isolated (Lane and Sage, 1994). The first two sugar moities are clearly defined in averaged $2F_{obs}–F_{calc}$ maps (not shown), and two *N*-acetylglucosamine (NAG) sugars are included in the crystallographic model. These two NAG sugars contact the linker segment connecting the FS and EC domains at Tyr134 and may contribute to the stabilization of the domain pair.

Structure of the FS domain

The FS domain folds into a non-globular structure (Figure 2) that is stabilized by a small hydrophobic core in its larger, C-terminal portion and by a total of five disulfide bonds. The domain is assembled from two rather weakly interacting substructures, a highly twisted N-terminal β-hairpin with seven interstrand hydrogen bonds (residues 54–77) and a pair of antiparallel α-helices connected to a small three-stranded antiparallel β-sheet. This subdivision is reflected in a disulfide linkage pattern of two non-overlapping sets (1–3, 2–4 and 5–9, 6–8, 7–10). The N-terminal hairpin interacts with the body of the domain mainly in the region around the disulfide bonds 60–76 and 78–113 through hydrophobic contacts between Val65 and Gln79; a further prominent contact involves His62 and Phe109 in helix α1 and a salt bridge between Lys64 and Glu116, located in the turn connecting α1 and α2. The N-terminal β-hairpin of the FS domain is rather mobile in our structure, as indicated by an average *B*-factor of 46 Å^2 compared with 26 Å^2 for the rest of the FS domain and the EC domain. Not unexpectedly, the tip of the hairpin shows the largest degree of mobility, whereas regions interacting with the main body of the FS domain are as well ordered as the rest of the structure.

The BM-40 FS domain is a representative FS domain

A sequence alignment of the FS domains of BM-40, follistatin and agrin (Figure 3) demonstrates that the structure of the BM-40 FS domain is a valid template for

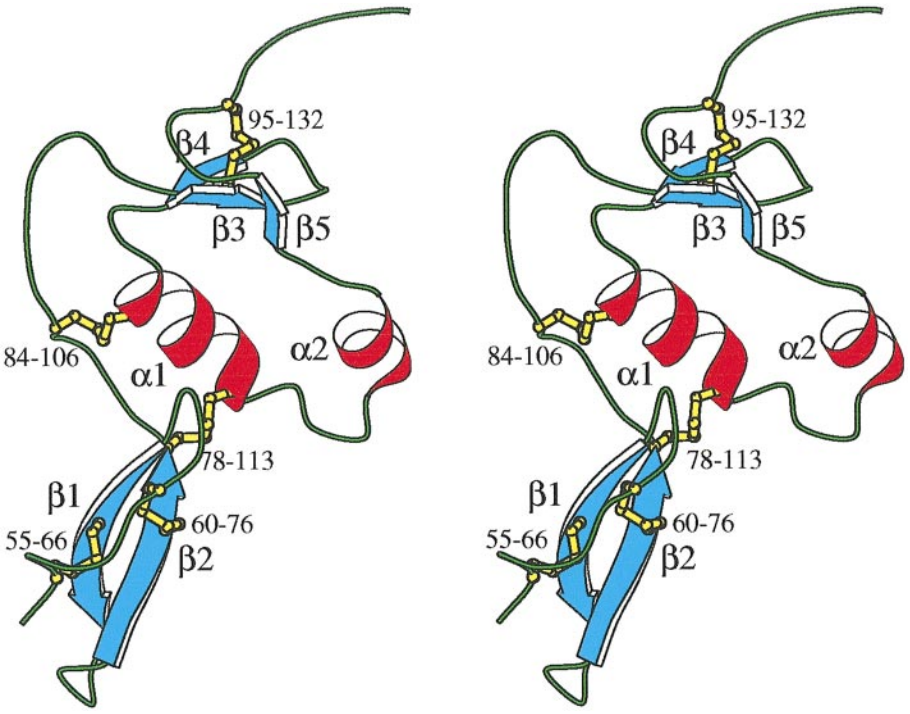

Fig. 2. Stereo view (Kraulis, 1991) of the FS domain in BM-40 and its secondary structure elements. The five disulfide bonds are shown in yellow and have been labelled. The highly twisted N-terminal β-hairpin is an independent substructure and makes only limited contact with the globular core of the FS domain. The single glycosylation site of BM-40 at Asn99 is located in the reverse turn between β3 and β4.

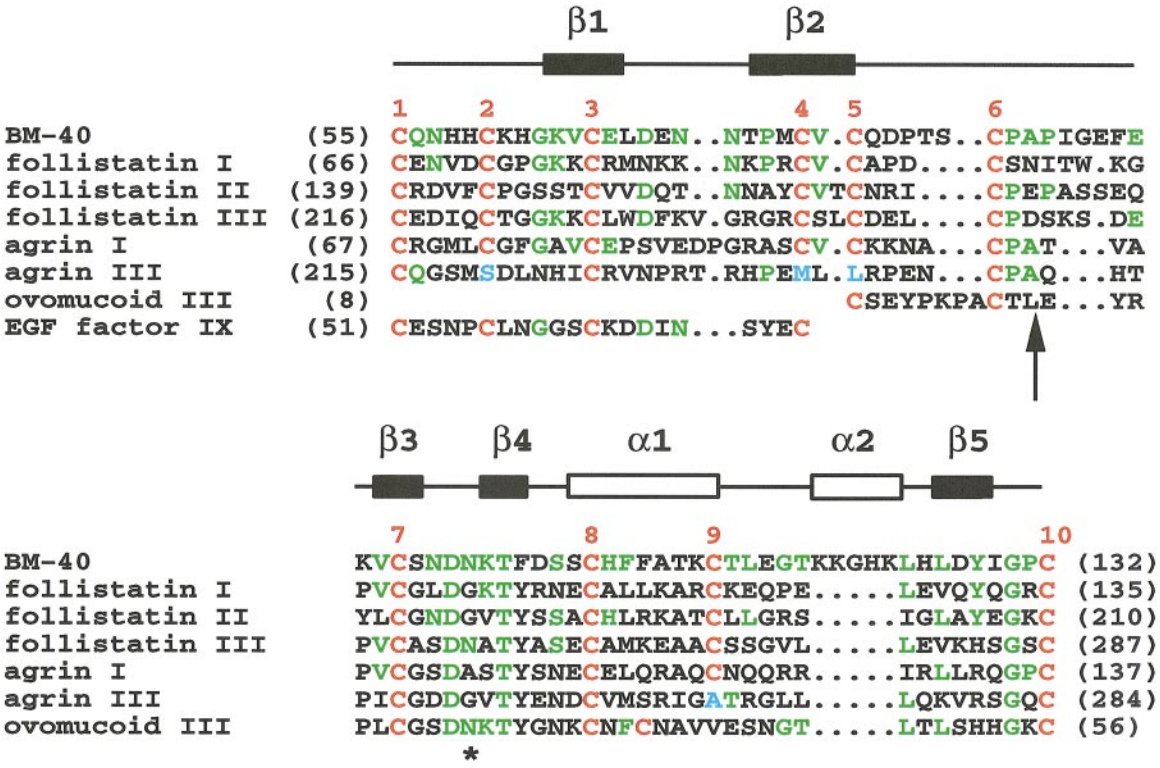

Fig. 3. Alignment of partial sequences of the FS domain of human BM-40 with FS domains I–III of human follistatin, FS domains I and III of rat agrin and domain III of turkey ovomucoid; also shown is the portion of the EGF-like module of human blood coagulation factor IX that is structurally similar to the FS domain (see text). Conserved cysteines are in red, other conserved residues are in green and residues replacing cysteines in agrin FS III are in blue. The secondary structure elements of the BM-40 FS domain are indicated above the alignment. The scissile bond of the Kazal serine protease inhibitor ovomucoid III is indicated by an arrow; the asterisk denotes the position of the glycosylation site in BM-40 (Asn99). The cysteines are linked [1–3, 2–4, 5–9, 6–8, 7–10] in the FS domain, [5–9, 6–8, 7–10] in the ovomucoid domain and [1–3, 2–4] in the EGF-like module.

FS domains in other proteins. All five disulfide bonds are conserved, with the notable exception of the third FS domain of agrin (Rupp *et al.*, 1991), where two disulfide bonds, 2–4 and 5–9, are replaced by non-covalent contacts between apolar side chains (Figure 3). Hydrophobic core residues are well conserved, as are a number of key turn residues, such as Gly63, Pro85, Ser96 and Gly130 in BM-40. With the exception of helix α2, which is unique to BM-40 and the closely related proteins SC1/hevin and QR1, secondary structure elements appear to be conserved throughout the family. The disposition of the N-terminal hairpin relative to the domain core may differ in some FS domains. In follistatin, cysteines 4 and 5 are separated by two residues in FS domains II and III, rather than by a single one as in BM-40. A more drastic insertion is seen in the FS domain of testican (Alliel *et al.*, 1993) where cysteines 4 and 5 are separated by 27 residues. This large insertion is likely to disrupt all contacts between the N-terminal hairpin and the domain core. The observation that the BM-40 deletion mutant ΔN lacking the N-terminal hairpin is properly folded and secreted (Pottgiesser *et al.*, 1994) also points to the existence of two essentially autonomous substructures in the FS domain.

The FS domain resembles Kazal protease inhibitors and EGF

It has been predicted, from sequence analysis, that the FS domain may resemble serine protease inhibitors of the Kazal family (Engel *et al.*, 1987; Esch *et al.*, 1987; Bolander *et al.*, 1988; Bode and Huber, 1991). A superposition of the BM-40 FS domain and the third domain of turkey ovomucoid (Fujinaga *et al.*, 1987) is shown in Figure 4A. The structural similarity of the C-terminal core of the FS domain and ovomucoid is evident and is reflected in a root-mean-square deviation (r.m.s.d.) of only 0.8 Å for 30 equivalent Cα atoms. Notable differences are the insertion of helix α2 in BM-40 (but not in other FS domains; see above) and a different position of the second cysteine in the central helix, resulting in an altered path of the N-terminal segment. Significantly, there is an insertion into the inhibitory loop of ovomucoid in BM-40, as indeed in all three FS domains of follistatin (Figure 3). This may explain why several studies have failed to detect a protease inhibitor activity of BM-40 (Villareal *et al.*, 1989; J.Engel, personal communication).

The N-terminal β-hairpin of the BM-40 FS domain, which has no counterpart in Kazal protease inhibitors, is stabilized by two disulfide bonds, 55–66 and 60–76. Both the separation of cysteines and their linkage pattern are reminiscent of epidermal growth factor (EGF; Campbell and Bork, 1993; Bork *et al.*, 1996). This similarity had been noted previously, but homology to EGF and Kazal protease inhibitors was considered to be mutually exclusive (Lane and Sage, 1994). Residues 54–78 of BM-40 can be superimposed on the EGF module of factor IX (Rao *et al.*, 1995) with an r.m.s.d. of 1.6 Å for 19 equivalent Cα atoms. Figure 4B shows that the two structures are very similar both in backbone conformation and in the conformation of the disulfide bridges; the FS and EGF-like modules diverge only after the fourth cysteine. Apart from the four equivalent cysteines, two other important residues are also conserved (Figure 3): a glycine in the first reverse turn (Gly63 in BM-40) and an aspartic acid stabilizing the second reverse turn (Asp69 in BM-40).

The high degree of structural correspondence of the FS domain to both ovomucoid and EGF raises some intriguing questions regarding the evolution of the FS domain. Modular proteins are thought to have evolved by gene duplication and exon shuffling, and these processes require domains with identical phases at both boundaries (Doolittle, 1995). All FS domains analysed so far indeed have domain boundaries of phase 1. Ovomucoid domains, however, are of phase 0–0, and therefore cannot participate in the assembly of class 1 mosaic proteins (Patthy and Nikolics, 1993). The structural similarity of a significant part of the FS domain to EGF complicates matters further. Only BM-40 and SC1/hevin contain an intron within the FS domain (splitting the triplet encoding residue 93 in BM-40), but the location of this intron does not coincide with the division into two substructures based on the three-dimensional structure. It is not clear, therefore, whether the structural similarity of a significant part of the FS domain to EGF is accidental or reflects an ancient fusion of exons encoding part of EGF and a Kazal protease inhibitor domain.

Functional analysis of the BM-40 EC–FS structure

Apart from an N-terminal segment of ~50 residues, which contains 15 glutamic acids and presumably adopts a defined three-dimensional structure only in the presence of a high concentration of calcium (Maurer *et al.*, 1992), the FS–EC domain pair comprises the entire BM-40 protein. The manifold activities of BM-40 have been probed by several peptide mapping studies, and these can now be analysed in the light of our structure (Figure 5).

Synthetic peptides from two regions in the BM-40 FS–EC domain pair have been shown to have biological effects mirroring those of intact BM-40 (Lane and Sage, 1994; Sage, 1997): residues 55–74 inhibit the proliferation of aortic endothelial cells (Funk and Sage, 1991, 1993) and mediate disassembly of focal adhesions of endothelial cells (Murphy-Ullrich *et al.*, 1995); residues 255–274 modulate cell shape (Lane and Sage, 1990), bind to endothelial cells and inhibit their proliferation (Yost and Sage, 1993; Sage *et al.*, 1995) and also participate in the disassembly of focal adhesions (Murphy-Ullrich *et al.*, 1995). Residues 114–130 and shorter peptides containing the copper-binding GHK sequence stimulate proliferation and angiogenesis, but these effects are not observed with intact BM-40 (Funk and Sage, 1993; Lane *et al.*, 1994).

In our structure, residues 55–74 map to the N-terminal hairpin of the FS domain; residues 114–130 map to helix α2 and the preceding and following loops, also in the FS domain, and residues 255–274 map to the second, disulfide-bonded EF-hand in the EC domain (Hohenester *et al.*, 1996). Figure 5 shows that all three regions are located on one side of BM-40 FS–EC, roughly opposite to the collagen-binding site located on the solvent-accessible face of helix A in the EC domain (Sasaki *et al.*, 1997). The wide range of biological effects elicited by peptides 55–74 and 255–274 has been taken as evidence that BM-40 is recognized at these two regions by a specific receptor (or several receptors) on endothelial cells (Yost and Sage, 1993). We have shown that a disulfide-bonded peptide spanning residues 250–275 dimerizes in the pres-

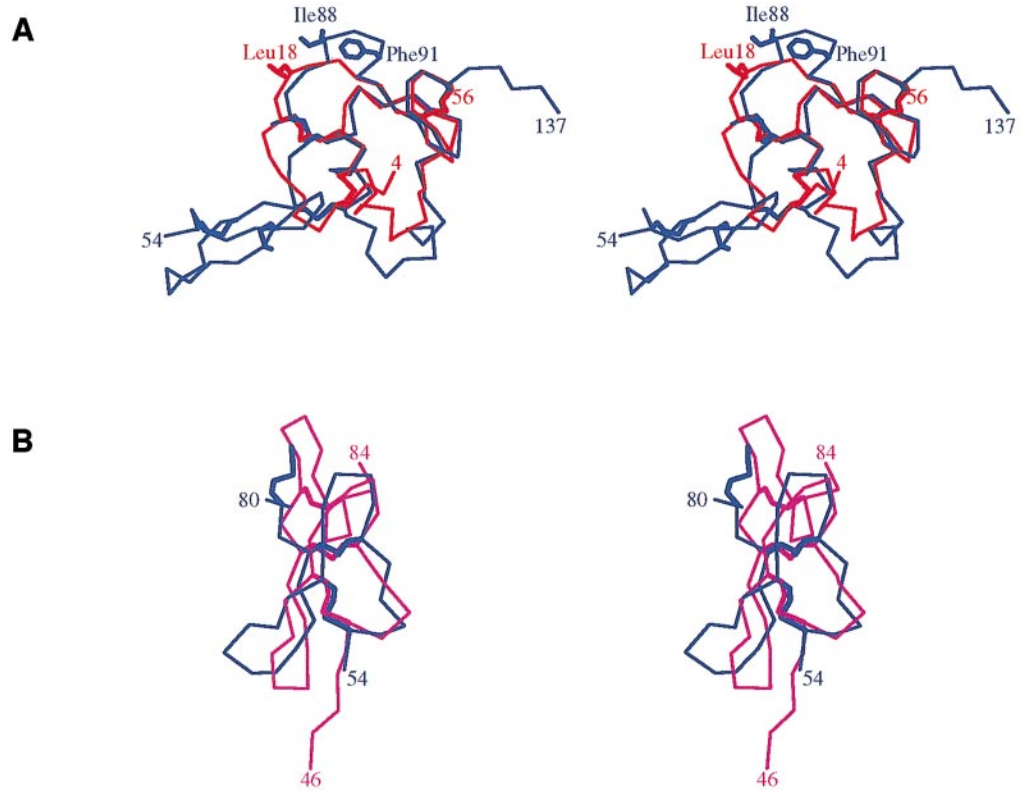

Fig. 4. (**A**) Stereo view (Kraulis, 1991) of a superposition of the BM-40 FS domain (in blue) and the third domain of turkey ovomucoid (in red; Fujinaga *et al.*, 1987). Thirty equivalent Cα atoms can be superimposed with an r.m.s.d. of 0.8 Å. The N- and C-termini are marked in both structures. The P_1 residue of ovomucoid III, Leu18, is shown as a stick model. In BM-40, two large hydrophobic residues, Ile88 and Phe91, are exposed on the tip of the loop corresponding to the inhibitory segment of ovomucoid; these two residues may be involved in ligand binding to the FS domain (see text). (**B**) Stereo view (Kraulis, 1991) of a superposition of the BM-40 FS domain (in blue) and the EGF-like module of factor IX (in purple; Rao *et al.*, 1995). Nineteen equivalent Cα atoms can be superimposed with an r.m.s.d. of 1.6 Å; the disulfide bonds were not used for the superposition. The two structures diverge after the fourth cysteine: in EGF, the polypeptide chain folds back underneath the hairpin and forms a third disulfide bridge between cysteines 5 and 6. In BM-40, the chain leaves the hairpin in an extended conformation; Cys78, corresponding to cysteine 5 of the EGF-like module, forms a disulfide bond with Cys113 (cysteine 9) at the C-terminus of helix α1.

ence of calcium. Thus, the biological effects ascribed to peptide 255–274 are likely to represent those of a symmetric peptide homodimer (Maurer *et al.*, 1995). Like other canonical EF-hand pairs, the EF-hands in the BM-40 EC domain possess 2-fold pseudosymmetry, violated only by the one-residue insertion into the first EF hand (Hohenester *et al.*, 1996). We therefore believe that the putative receptor-binding epitope of BM-40 involves the accessible surface of the entire EF-hand pair, rather than just EF-hand 2. This proposed epitope is contiguous with, and may extend to, the β-hairpin in the FS domain. This would explain why peptides from two regions distant in sequence (55–74 and 255–274) display silmilar effects in a range of biological assays (Murphy-Ullrich *et al.*, 1995; Sage *et al.*, 1995). However, it has to be borne in mind that epitope mapping using synthetic peptides is inherently limited, as it is always arguable whether isolated peptides adopt the same conformation as the corresponding stretches of polypeptide in the folded protein. For this reason, it will be important to establish the proposed epitopes more firmly by site-directed mutagenesis.

The region around helix α2 is very basic, with several lysine and histidine side chains projecting into the solvent. It is close in space to the N-terminus of the FS domain, which is also basic. This conspicuous concentration of positive charges may indicate that the FS domain of BM-40 contains a binding site for glycosaminoglycans similar to follistatin (Nakamura *et al.*, 1991; see below). Preliminary experiments indicate that BM-40 binds moderately to heparin through its FS domain, a property which is lost by deletion of a large portion of the N-terminal hairpin (unpublished data). The binding of proteoglycans may be functionally important in the recruitement of BM-40 to its cell surface receptor(s).

BM-40 binds to collagens I, III, IV and V. We have suggested that the binding site resides on helix A in the EC domain (Sasaki *et al.*, 1997) and this recently has been corroborated by site-directed mutagenesis (unpublished data). We have also shown that glycosylation at Asn99 has no effect on collagen binding, in contrast to another study on collagen V binding (Xie and Long, 1995). While the structure of BM-40 FS–EC cannot provide a conclusive answer, we note that Asn99 is at >30 Å distance from helix C, whose cleavage between residues 197 and 198 results in an increased affinity of BM-40 for collagens (Sasaki *et al.*, 1997).

BM-40 binds with high affinity to PDGF-AB and -BB (Raines *et al.*, 1992). The significance of this interaction is not clear and the region(s) involved in PDGF binding have not been mapped, although some evidence appears to implicate the FS domain (Patthy and Nikolics, 1993; see below).

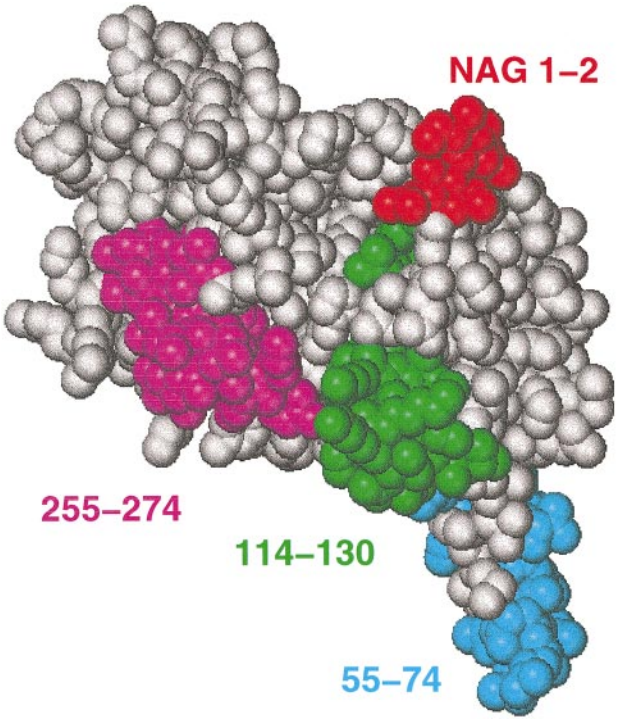

Fig. 5. Regions proposed to be important for the biological functions of BM-40 mapped onto a van der Waals representation of the FS–EC domain pair. The view direction is similar to that in Figure 1; the FS domain is on the right, with the N-terminus in the lower right corner. Residues 55–74 and 255–274 may be involved in binding to endothelial cells and the disassembly of focal adhesions; peptides spanning residues 114–130 contain the mitogenic GHK sequence and stimulate angiogenesis (Lane and Sage, 1994; Sage, 1997). The *N*-linked carbohydrate of BM-40 is shown in red. Residues 55–74 and 255–274 may be part of a contiguous receptor-binding surface (see text).

FS domains in follistatin and agrin

Several extracellular mosaic proteins contain FS domains whose structures can now be predicted reliably based on the high homology to the BM-40 FS domain (Figure 3; Bork *et al.*, 1996). Follistatin itself consists of a small N-terminal domain followed by three tandemly repeated FS domains. Follistatin binds activin, a member of the transforming growth factor-β superfamily, and this interaction is directly responsible for the inhibition of follicle-stimulating hormone release (Michel *et al.*, 1993). Overexpression of follistatin in *Xenopus* induces neuralization (Hemmati-Brivanlou *et al.*, 1994), and transgenic mice lacking follistatin have multiple defects and die shortly after birth, indicating an interaction with several growth factors during development (Matzuk *et al.*, 1995). The activin-binding epitope of follistatin has not been mapped to a specific domain. Preliminary data point to an involvement of the N-terminal, non-FS domain and residues 75–86 in the first FS domain (Inouye *et al.*, 1991; Sumitomo *et al.*, 1995). These latter residues form the N-terminal hairpin of the first FS domain of follistatin (Figure 3), and homology modelling predicts that six basic side chains emanate from one face of the hairpin (pointing downwards in the view of Figure 2). This region is, therefore, a prime candidate for the heparin-binding site of follistatin (Esch *et al.*, 1987; Nakamura *et al.*, 1991) and is unlikely to participate directly in activin binding.

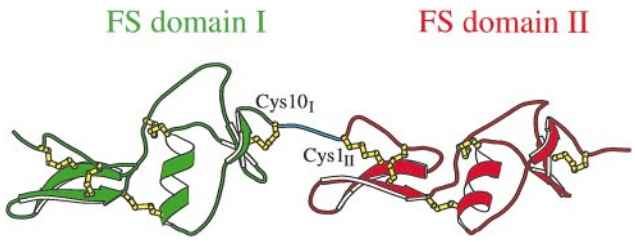

Fig. 6. Model of a tandem of FS domains in follistatin and agrin. The N-terminal domain (in green) and the C-terminal domain (in red) are linked by an extended segment of three residues (in blue); the relative rotation between the two domains is ~0°. The tip of the N-terminal β-hairpin of the C-terminal domain contacts the N-terminal domain at the small three-stranded β-sheet and stabilizes the linear arrangement of domains. In this model, helix α2, which is unique to BM-40, has been replaced with the corresponding segment of ovomucoid (compare Figures 3 and 4A).

Table I. Data collection and refinement statistics

Data collection (25–3.1 Å)	
Observations	47 563
Unique reflections	14 328
Completeness (%)	90.7 (85.1)
R_{merge}[a]	0.086 (0.267)
Refinement (8–3.1 Å)	
Reflections (no amplitude cut-off)	13 378
Completeness (%)	90.4 (85.7)
Non-H atoms	2×1860
R-factor	0.262
R_{free}[b]	0.322
Bond lengths (Å)	0.007
Bond angles (°)	1.8
r.m.s.d. of *B*-factors for bonded atoms (Å^2)	1.5
NCS for all atoms (Å)	0.03

[a]$R_{merge} = \Sigma_h\Sigma_i|I_{h,i}-<I_h>|/\Sigma_h\Sigma_i I_{h,i}$. For R_{merge} and completeness, the numbers in parentheses refer to data in the highest resolution shell.
[b]R_{free} was calculated for 5% of the data not included in the refinement.

Interestingly, the corresponding region of the BM-40 FS domain is also basic and may be involved in heparin binding (see above).

Agrin is a large multidomain proteoglycan that plays a key role in the formation of the neuromuscular junction (Patthy and Nikolics, 1993; Kleinman and Reichardt, 1996). It contains nine FS domains, eight of which are arranged in an uninterrupted tandem repeat (Rupp *et al.*, 1991). Electron microscopy of agrin shows that the N-terminus, which contains the large FS repeat, adopts a rigid rod-like structure, indicating substantial interactions between adjacent FS domains (M.Rüegg, personal communication). This prompted us to construct a model for an FS domain pair incorporating interdomain contacts (Figure 6). Two FS domains were connected by an extended three-residue linker, based on the separation of domains in agrin and follistatin, which ranges from two to six residues (between Cys10$_i$ and Cys1$_{i+1}$). Assuming a linear arrangement of FS domains, we found essentially two orientations that allow for interdomain contacts, one with a relative rotation of ~0° and one with a rotation of ~180°. The former alternative, which is the one shown in Figure 6, results in contacts between the β1–β2 hairpin and the small three-stranded β-sheet of the preceding FS domain, whereas a rotation of 180° would place the tip

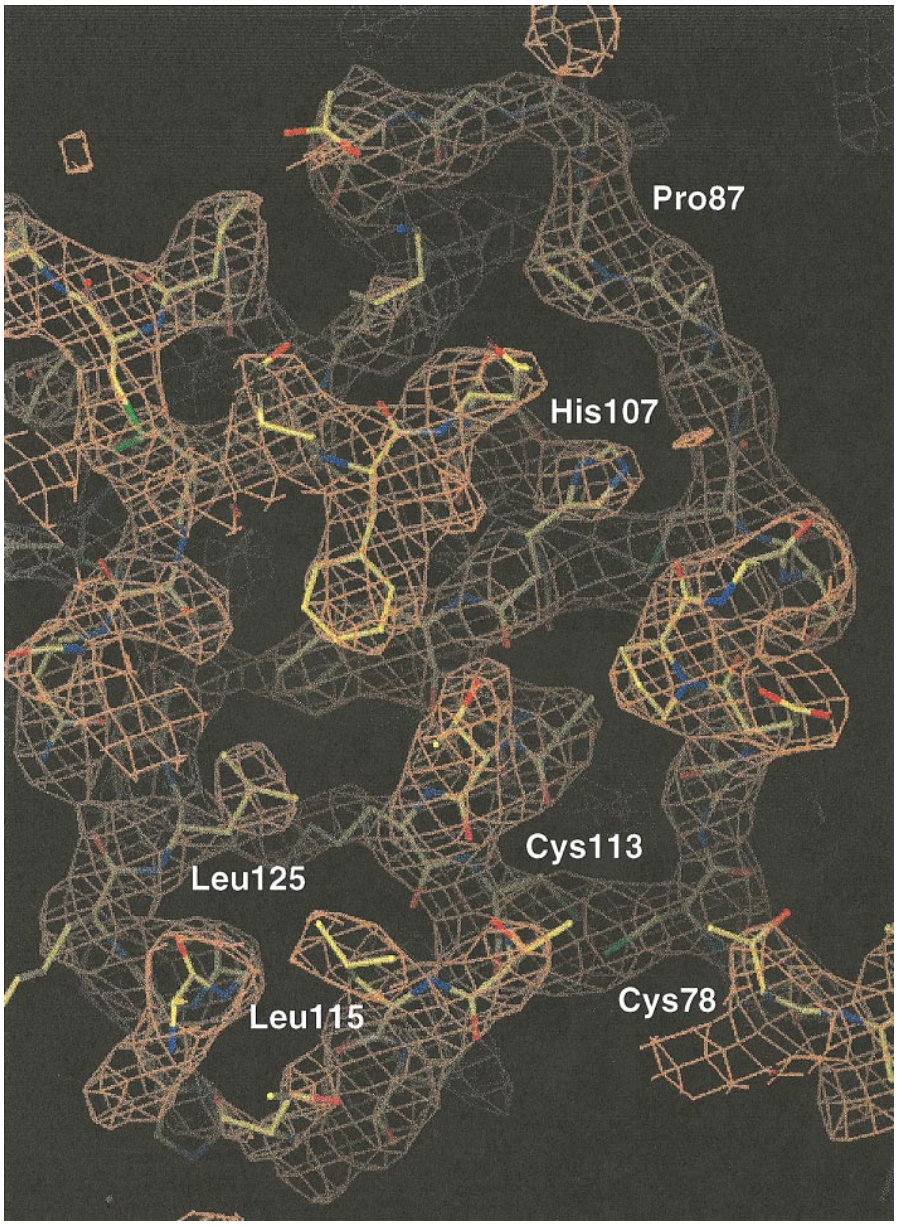

Fig. 7. Initial, unbiased electron density of the FS domain in the BM-40 FS–EC structure. Phases calculated from the two EC domains were greatly improved by 2-fold non-crystallographic symmetry averaging and solvent flattening (70% solvent). The electron density shown is calculated using data between 15 and 3.1 Å resolution and is contoured at 1.0σ. Superimposed on the map is the current refined model. The central helix α1 runs from top to bottom and is linked to an extended segment at the right by disulfide bonds 78–113 and 84–106 (partly obscured by the side chain of His107).

of the hairpin close to the segment corresponding to the inhibitory loop in ovomucoid. In our model, the N-terminal β-hairpins of the FS domains act as rigid spacers between the globular cores, and the resulting structure is that of a rod rigidified by contacts across domain boundaries. This arrangement is rather similar to that of multiple repeats of calcium-binding EGF-like domains (Bork *et al.*, 1996; Downing *et al.*, 1996); however, the functional significance of FS tandem repeats is unclear at present.

Although some modules in extracellular proteins may have a universal function, it is generally thought that modules acquire different functions depending on their structural setting (Bork *et al.*, 1996). Binding of growth factors has been suggested to be a common function of FS domains, based on the observation that BM-40 and follistatin bind PDGF and activin, respectively (Patthy and Nikolics, 1993). The BM-40 FS domain and, by homology, the first FS domain of follistatin present large hydrophobic side chains at the tip of their potential inhibitory loops (Figures 3 and 4A). These residues could be (part of) a ligand-binding site, and it is tempting to speculate that an ancestral protease inhibitor could have become a cytokine-binding domain by suitable modifications to its binding segment. By additionally containing distinct binding sites for glycosaminoglycans, FS domains could then function in the recruitment of growth factors to the cell surface and thus play an important role in cytokine signalling. Alternatively, some FS domains may indeed be protease inhibitors, as suggested by the close structural relationship to Kazal protease inhibitors.

Materials and methods

Protein production, crystallization and X-ray data collection

The recombinant fragment BM-40 (53–286) was expressed in 293 cells and purified as described (Maurer *et al.*, 1995). For large-scale production, cells were grown in a hollow-fibre system (Tillet *et al.*, 1994). Crystals were obtained with difficulty by the hanging drop method at room temperature: 2 μl of protein solution (8–10 mg/ml in 10 mM Tris pH 7.5, 2 mM $CaCl_2$) were mixed with 2 μl of 14% (w/v) PEG4000, 0.15 M sodium acetate, 0.1 M Na-HEPES pH 7.5 and equilibrated against the latter solution. Irregular plates measuring up to 80 μm in thickness grew after several days in some of the drops.

The crystals diffracted X-rays very weakly and suffered from persistent twinning. Eventually, one batch yielded a small number of sufficiently large single crystals. Diffraction data to 3.1 Å resolution were collected at 100 K from two cryoprotected crystals (10 s soaks in mother liquor with 10, 20 and 30% glycerol added) using an MAR Research image plate detector mounted on an Elliott GX21 rotating anode X-ray generator. The majority of the data (85%) were collected from one good crystal; these data were merged with previously collected low-resolution data to obtain >98% completeness to 4 Å resolution. The data were processed with MOSFLM (Leslie, 1994) and programs of the CCP4 suite (Collaborative Computing Project No. 4, 1994). The crystals belong to space group $P2_1$ with a = 70.77 Å, b = 56.35 Å, c = 111.60 Å, β = 104.38°. There are two molecules of BM-40 (53–286) in the asymmetric unit related by a 2-fold axis at polar angles ω = 48°, φ = 0°. Neglecting the unknown fraction of carbohydrate, the V_m is 4.0 $Å^3$/Da, corresponding to a solvent content of the crystals of 70%. Data collection and reduction statistics are summarized in Table I.

Structure solution and refinement

The structure of BM-40 (53–286) was solved by molecular replacement and 2-fold non-crystallographic symmetry (NCS) averaging. The search model was the EC domain of human BM-40 (residues 138–286) refined at 2.0 Å resolution (Hohenester *et al.*, 1996). Cross-rotation and translation functions were calculated with AMoRe (Navaza, 1994) using data between 15 and 3.5 Å resolution. The cross-rotation function gave two solutions consistent with the self-rotation function, with peak heights of 7.9σ and 6.7σ (highest noise peak 3.9σ). The corresponding translation functions were unambiguous, with *R*-factors of 0.501 and 0.514, respectively, and rigid-body refinement of the two independent EC domains resulted in an *R*-factor of 0.458 (8–3.1 Å). A SIGMAA-weighted (Read, 1986) map with phases calculated from the EC domains showed electron density that could be attributed to the missing FS domains but was not interpretable. Six cycles of NCS averaging with RAVE (Kleywegt and Jones, 1994) using data between 15 and 3.1 Å resolution and separate masks and NCS operators for the FS and EC domains improved the map correlation from 0.56 to 0.82 and from 0.82 to 0.88, respectively. The *R*-factor between the experimental structure factor amplitudes and those calculated from the averaged map (R_{aver}) was 0.215. This map was of good quality (Figure 7) and the model could be extended readily by 60 residues with O (Jones *et al.*, 1991). The remaining N-terminal 24 residues of the FS domain were added after a round of slow-cooling simulated annealing refinement with X-PLOR (Brünger, 1992), followed by another round of averaging with improved masks (final R_{aver} = 0.201).

The structure was refined with very tight NCS restraints against all data between 8 and 3.1 Å resolution. A bulk solvent correction was used (r = 0.25 Å, k = 0.33, B = 50 $Å^2$). The final model comprises two essentially identical molecules of BM-40 FS–EC including residues 54–286, the first two NAG sugars attached to Asn99 and two calcium ions; the *R*-factor is 0.262 (R_{free} = 0.322). The Ramachandran plot shows no residues in disallowed or generously allowed regions as defined by PROCHECK (Laskowski *et al.*, 1993), and the final model scores well above average for a structure at 3.1 Å resolution with an overall G-factor of 0.28. The average *B*-factors are 46 $Å^2$ (residues 54–77), 26 $Å^2$ (78–137) and 26 $Å^2$ (138–286). Simulated annealing OMIT maps, both unaveraged and averaged, were used to verify the correctness of the structure. This proved to be essential for the mobile N-terminal hairpin, where the unaveraged electron density is weak in several places. The *N*-linked carboydrate is close to several crystal packing contacts, and we observe strong but patchy electron density extending from NAG2. The presence of partially ordered carbohydrate is only poorly accounted for by the bulk solvent correction, and this may explain the relatively high final *R*-factor. In any case, values around 0.25 are typical of conservative models at modest resolution (Kleywegt and Jones, 1995). Refinement statistics are summarized in Table I. The atomic coordinates and experimental structure factor amplitudes have been deposited with the Brookhaven Protein Data Bank (accession codes 1BMO and R1BMOSF, respectively).

Acknowledgements

E.H. would like to thank his colleagues at Birkbeck College for many helpful discussions and Dr Steve Wood (University of Southampton) for encouragement. E.H. is a fellow of the Human Frontier Science Program.

References

Alliel,P.M., Perin,J.-P., Jolles,P. and Bonnet,F.J. (1993) Testican, a multidomain testicular proteoglycan resembling modulators of cell social behaviour. *Eur. J. Biochem.*, **214**, 347–350.

Bode,W. and Huber,R. (1991) Natural protein proteinase inhibitors and their interactions with proteinases. *Eur. J. Biochem.*, **204**, 433–451.

Bolander,M.E., Young,M.F., Fisher,L.W., Yamada,Y. and Termine,J.D. (1988) Osteonectin cDNA sequence reveals potential binding regions for calcium and hydroxyapatite and shows homologies with both a basement membrane protein (SPARC) and a serine proteinase inhibitor (ovomucoid). *Proc. Natl Acad. Sci. USA*, **85**, 2919–2923.

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.

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

Campbell,I.D. and Bork,P. (1993) Epidermal growth factor-like modules. *Curr. Opin. Struct. Biol.*, **3**, 385–392.

Collaborative Computing Project No. 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.

Doolittle,R.F. (1995) The multiplicity of domains in proteins. *Annu. Rev. Biochem.*, **64**, 287–314.

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 amd other genetic disorders. *Cell*, **85**, 597–605.

Engel,J., Taylor,W., Paulsson,M., Sage,H. and Hogan,B. (1987) Calcium binding domains and calcium-induced conformational transition of SPARC/BM-40/osteonectin, an extracellular glycoprotein expressed in mineralized and nonmineralized tissues. *Biochemistry*, **26**, 6958–6965.

Esch,F.S., Shimasaki,S., Mercado,M., Cooksey,K., Ling,N., Ying,S., Ueno,N. and Guillemin,R. (1987) Structural characterization of follistatin: a novel follicle-stimulating hormone release-inhibiting polypeptide from gonad. *Mol. Endocrinol.*, **1**, 849–855.

Fujinaga,M., Sielecki,A.R., Read,R.J., Ardelt,W., Laskowski,M. and James,M.N.G. (1987) Crystal and molecular structures of the complex of α-chymotrypsin with its inhibitor turkey ovomucoid third domain at 1.8 Å resolution. *J. Mol. Biol.*, **195**, 397–418.

Funk,S.E. and Sage,E.H. (1991) The Ca^{2+}-binding glycoprotein SPARC modulates cell cycle progression in bovine aortic endothelial cells. *Proc. Natl Acad. Sci. USA*, **88**, 2648–2652.

Funk,S.E. and Sage,E.H. (1993) Differential effects of SPARC and cationic SPARC peptides on DNA synthesis by endothelial cells and fibroblasts. *J. Cell. Physiol.*, **154**, 53–63.

Girard,J.-P. and Springer,T.A. (1996) Modulation of endothelial cell adhesion by hevin, an acidic protein associated with high endothelial venules. *J. Biol. Chem.*, **271**, 4511–4517.

Guermah,M., Crisanti,P., Laugier,D., Dezelee,P., Bidou,L., Pessac,B. and Calothy,G. (1991) Transcription of a quail gene expressed in embryonic retinal cells is shut off sharply at hatching. *Proc. Natl Acad. Sci. USA*, **88**, 4503–4507.

Hemmati-Brivanlou,A., Kelly,O.G. and Melton,D.A. (1994) Follistatin, an antagonist of activin is expressed in the Spemann organizer and displays direct neuralizing activity. *Cell*, **77**, 283–295.

Hohenester,E., Maurer,P., Hohenadl,C., Timpl,R., Jansonius,J.N. and Engel,J. (1996) Structure of a novel extracellular Ca^{2+}-binding module in BM-40. *Nature Struct. Biol.*, **3**, 67–73.

Inouye,S., Ling,N. and Shimasaki,S. (1991) Site-specific mutagenesis of human follistatin. *Biochem. Biophys. Res. Commun.*, **179**, 352–358.

Johnston,I.G., Paladino,T., Gurd,J.W. and Brown,I.R. (1990) Molecular cloning of SC1: a putative brain extracellular matrix glycoprotein showing partial similarity to osteonectin/BM40/SPARC. *Neuron*, **2**, 165–176.

Jones,T.A., Zou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.

Kleinman,R.J. and Reichardt,L.F. (1996) Testing the agrin hypothesis. *Cell*, **85**, 461–464.

Kleywegt,G.J. and Jones,T.A. (1994) Halloween ... masks and bones. In Bailey,S., Hubbard,R. and Waller,D. (eds), *From First Map to Final Model*. SERC Daresbury Laboratory, Daresbury, UK, pp. 59–65.

Kleywegt,G.J. and Jones,T.A. (1995) Where freedom is given, liberties are taken. *Structure*, **3**, 535–540.

Kraulis,J.P. (1994) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Kretsinger,R.H. (1996) EF-hands reach out. *Nature Struct. Biol.*, **3**, 12–15.

Lane,T.F. and Sage,E.H. (1994) The biology of SPARC, a protein that modulates cell–matrix interactions. *FASEB J.*, **8**, 163–173.

Lane,T.F., Iruela-Arispe,M., Johnson,R.S. and Sage,E.H. (1994) SPARC is a source of copper-binding peptides that stimulate angiogenesis. *J. Cell Biol.*, **125**, 929–943.

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.

Ledda,M.F., Adris,R., Bravo,A.I., Kairiyama,C., Bover,L., Chernajovsky,Y., Mordo,J. and Podhajcer,O.L. (1997) Suppression of SPARC expression by antisense RNA abrogates the tumorigenicity of human melanoma cells. *Nature Med.*, **3**, 171–176.

Leslie,A.G.W. (1994) *MOSFLM Users Guide*. MRC-LMB Cambridge, UK.

Matzuk,M.M., Lu,N., Vogel,H., Sellheyer,K., Roop,D.R. and Bradley,A. (1995) Multiple defects and perinatal death in mice deficient in follistatin. *Nature*, **374**, 360–363.

Maurer,P., Mayer,U., Bruch,M., Jenö,P., Mann,K., Landwehr,R., Engel,J. and Timpl,R. (1992) High-affinity and low-affinity calcium binding and stability of the multidomain extracellular 40-kDa basement membrane glycoprotein (BM-40/SPARC/osteonectin). *Eur. J. Biochem.*, **205**, 233–240.

Maurer,P., Hohenadl,C., Hohenester,E., Göhring,W., Timpl,R. and Engel,J. (1995) The C-terminal portion of BM-40 (SPARC/osteonectin) is an autonomously folding and crystallisable domain that binds calcium and collagen IV. *J. Mol. Biol.*, **253**, 347–357.

Michel,U., Farnworth,P. and Findlay,J.K. (1993) Follistatins: more than follicle-stimulating hormone suppressing proteins. *Mol. Cell. Endocrinol.*, **91**, 1–11.

Murphy-Ullrich,J.E., Lane,T.F., Pallero,M.A. and Sage,E.H. (1995) SPARC mediates focal adhesion disassembly in endothelial cells through a follistatin-like region and the Ca^{2+}-binding EF-hand. *J. Cell. Biochem.*, **57**, 341–350.

Nakamura,T., Sugino,K., Titani,K. and Sugino,H. (1991) Follistatin, an activin-binding protein, associates with heparan sulfate chains of proteoglycans on follicular granulosa cells. *J. Biol. Chem.*, **266**, 19432–19437.

Navaza,J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr.*, **A50**, 157–163.

Patthy,L. and Nikolics,K. (1993) Functions of agrin and agrin-related proteins. *Trends Neurosci.*, **16**, 76–81.

Pottgiesser,J., Maurer,P., Mayer,U., Nischt,R., Mann,K., Timpl,R., Krieg,T. and Engel,J. (1994) Changes in calcium and collagen IV binding caused by mutations in the EF-hand and other domains of extracellular matrix protein BM-40 (SPARC, osteonectin). *J. Mol. Biol.*, **238**, 563–574.

Raines,E.W., Lane,T.F., Iruela-Arispe,M.L., Ross,R. and Sage,E.H. (1992) The extracellular glycoprotein SPARC interacts with platelet-derived growth factor (PDGF)-AB and -BB and inhibits the binding of PDGF to its receptors. *Proc. Natl Acad. Sci. USA*, **89**, 1281–1285.

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.

Read,R.J. (1986) Improved Fourier coefficients for maps using phases from partial structure factors with errors. *Acta Crystallogr.*, **A42**, 140–149.

Rupp,F., Payan,D.G., Magill-Solc,C., Cowan,D.M. and Scheller,R.H. (1991) Structure and expression of rat agrin. *Neuron*, **6**, 811–823.

Sage,E.H. (1997) Terms of attachment: SPARC and tumorigenesis. *Nature Med.*, **3**, 144–146.

Sage,E.H., Bassuk,J.A., Yost,J.C., Folkman,M.J. and Lane,T.F. (1995) Inhibition of endothelial cell proliferation is mediated through a Ca^{2+}-binding EF-hand sequence. *J. Cell. Biochem.*, **57**, 127–140.

Sasaki,T., Göhring,W., Mann,K., Maurer,P., Hohenester,E., Knäuper,V., Murphy,G. and Timpl,R. (1997) Limited cleavage of extracellular matrix protein BM-40 by matrix metalloproteinases increases its affinity for collagens. *J. Biol. Chem.*, **272**, 9237–9243.

Schwarzbauer,J.E. and Spencer,C.S. (1993) The *Caenorhabditis elegans* homologue of the extracellular calcium binding protein SPARC/osteonectin affects nematode body morphology and mobility. *Mol. Biol. Cell*, **4**, 941–952.

Shibanuma,M., Mashimo,J., Mita,A., Kuroki,T. and Nose,K. (1993) Cloning from a mouse osteoblastic cell line of a set of transforming-growth-factor-β1-regulated genes, one of which seems to encode a follistatin-related polypeptide. *Eur. J. Biochem.*, **217**, 13–19.

Sumitomo,S., Inouye,S., Liu,X.-J., Ling,N. and Shimasaki,S. (1995) The heparin binding site of follistatin is involved in its interaction with activin. *Biochem. Biophys. Res. Commun.*, **208**, 1–9.

Tillet,E., Wiedemann,H., Golbik,R., Pan,T.-C., Zhang,R.-Z., Mann,K., Chu,M.-L. and Timpl,R. (1994) Recombinant expression and structural and binding properties of α1(VI) and α2(VI) chains of human collagen type VI. *Eur. J. Biochem.*, **221**, 177–185.

Villareal,X.C., Mann,K.G. and Long,G.L. (1989) Structure of human osteonectin based upon analysis of cDNA and genomic sequences. *Biochemistry*, **28**, 6483–6491.

Xie,R.-L. and Long,G.L. (1995) Role of *N*-linked glycosylation in human osteonectin. *J. Biol. Chem.*, **270**, 23212–23217.

Yost,J.C. and Sage,E.H. (1993) Specific interaction of SPARC with endothelial cells is mediated through a carboxyl-terminal sequence containing a calcium-binding EF-hand. *J. Biol. Chem.*, **268**, 25790–25796.

Received on March 26, 1997; revised on April 22, 1997