Characterization of SMOC-2, a modular extracellular calcium-binding protein

Christian VANNAHME[1], Silke GÖSLING, Mats PAULSSON, Patrik MAURER[2] and Ursula HARTMANN[3]

Center for Biochemistry, Medical Faculty, University of Cologne, Joseph-Stelzmann Strasse 52, D-50931 Cologne, Germany

We have isolated the novel gene SMOC-2, which encodes a secreted modular protein containing an EF-hand calcium-binding domain homologous to that in BM-40. It further consists of two thyroglobulin-like domains, a follistatin-like domain and a novel domain found only in the homologous SMOC-1. Phylogenetic analysis of the calcium-binding domain sequences showed that SMOC-1 and -2 form a separate group within the BM-40 family. The human and mouse SMOC-2 sequences are coded for by genes consisting of 13 exons located on chromosomes 6 and 17, respectively. Analysis of recombinantly expressed protein showed that SMOC-2 is a glycoprotein with a calcium-dependent conformation. Results from Northern blots and reverse transcription PCR revealed a widespread expression in many tissues.

Key words: EF-hand, extracellular matrix, BM-40, osteonectin, SPARC.

INTRODUCTION

The BM-40 family of modular extracellular proteins is characterized by always containing an extracellular calcium-binding (EC) domain as well as a follistatin-like (FS) domain with ten cysteines in a typical pattern. The EC domain has two EF-hand calcium-binding motifs, each with a bound calcium ion in the X-ray structure of BM-40 [1,2]. Several members of this family have been found localized to the extracellular matrix of different tissues and at least BM-40 and SMOC-1 are, in some tissues, components of basement membranes [3,4].

BM-40 [also known as secreted protein acidic and rich in cysteines (SPARC) or osteonectin] was isolated originally from bone [5], but subsequently found in a variety of other tissues (for review see [6,7]). Its N-terminal domain, containing about 50 amino acids of which 18 are negatively charged, is followed by an FS domain and an EC domain.

BM-40 was reported to participate in the regulation of cell–matrix interactions, in particular influencing bone mineralization, wound repair and angiogenesis. *In vitro* BM-40 inhibits cell adhesion, spreading and proliferation and regulates the expression of proteins involved in matrix turnover (reviewed in [8]). It is highly expressed in some malignant tumours and was reported to play a crucial role for the tumourogenicity of human melanomas [9], although this role has been questioned [10]. The biological properties of BM-40 may be related to its reported binding of collagen IV [11] and of growth factors such as platelet-derived growth factor ('PDGF') [12,13] and vascular endothelial growth factor ('VEGF') [14]. Deletion of the BM-40 gene in mice does not alter embryogenesis. However, postnatally BM-40-deficient mice develop cataracts, which is associated with a compromised structure of the lens [15–17] and severe osteopenia [18]. These mice also display defects in wound healing [19,20] and in tissue remodelling [21,22].

Several proteins homologous to BM-40 have been identified, including SC1/hevin [23], QR1 (the quail homologue of SC1) [24], tsc36/Flik/FRP [25] and testican-1 [26], -2 [27] and -3 [28]. These are expressed in a wide variety of tissues and there is so far no clear common denominator among the proposed functions.

We recently identified a novel member of the BM-40 family with a unique domain organization [4]. This protein was termed SMOC-1, where SMOC stands for secreted modular calcium-binding protein, and could be detected in basement membranes, but also in other extracellular locations. In addition to a calcium-binding EC domain homologous to that in BM-40, it consists of two thyroglobulin-like (TY) domains, an FS domain and a novel domain. In searches of the EST sequence databases we found a second SMOC, SMOC-2, with an identical domain structure. In the present work we describe the molecular properties of this protein. In parallel to our studies, which focused mainly on the mouse SMOC-2, the human homologue of SMOC-2 was identified as a smooth-muscle-associated protein up-regulated during neointima formation and was termed smap2 [29].

EXPERIMENTAL

Isolation of mouse cDNA clones

A brain cDNA library (newborns of ICR outbred mice; Stratagene) was screened with three fragments: 1.1 and 0.45 kb derived by *Eco*RI/*Not*I restriction digestion and 0.48 kb derived by *Eco*RI/*Bgl*II restriction digestion of the mouse EST clone MM059909 (ATCC). After isolation, fragments were labelled with [^{32}P]dCTP using the random-prime labelling system (Amersham Biosciences) according to the manufacturer's protocol. Hybridization was carried out at 42 °C in 50 % formamide, 5 × Denhardt's solution (0.1 % BSA, 0.1 % Ficoll

Abbreviations used: EC, extracellular calcium-binding; FS, follistatin-like; ORF, open reading frame; PNGase F, peptide N-glycosidase F; SMOC, secreted modular calcium-binding protein; TY, thyroglobulin-like; UTR, untranslated region; RT-PCR, reverse transcription PCR.

[1] Present address: Cytos Biotechnology AG, Wagistr. 25, 8952 Zürich-Schlieren, Switzerland.

[2] Present address: Qiagen GmbH, Max-Volmer-Strasse 4, 40724 Hilden, Germany.

[3] To whom correspondence should be addressed (e-mail ursula.hartmann@uni-koeln.de).

The nucleotide sequence of SMOC-2 has been deposited in the EMBL/GenBank®/DDBJ/GSDB Nucleotide Sequence Databases under accession no. AJ249901.

400 and 0.1 % polyvinylpyrrolidone), 5 × SSPE (750 mM NaCl, 50 mM sodium phosphate, pH 7.6, and 5 mM EDTA), 1.5 % SDS and 0.2 mg/ml salmon sperm DNA. Filters were washed twice for 10–15 min in 0.1 × SSC (where 1 × SSC is 0.15 M NaCl/0.015 M sodium citrate, pH 7.5) and 0.1 % SDS and exposed to X-ray film. Positive plaques were excised and rescreened. One positive plaque from the final rescreen was *in vivo*-excised, yielding cDNA in the pBluescript vector. The plasmid was sequenced on both strands with flanking and internal primers using an ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction Kit and the products were resolved on an ABI Prism 377 Automated Sequencer (PE Biosystems).

Isolation of human cDNA clones

A human fetal brain cDNA library (Clontech) was screened with two 187 and 239 bp ^{32}P-labelled *Sty*I fragments isolated from the human synovial membrane EST clone 107131 (ATCC). Hybridization was carried out at 65 °C in aqueous solution (0.5 M sodium phosphate, pH 7.2, 7 % SDS, 1 mM EDTA, 10 % BSA and 0.2 mg/ml salmon sperm DNA). Filters were washed under low-stringency conditions (40 mM sodium phosphate, pH 7.2, and 1 % SDS) once at 55 °C for 15 min and twice at 65 °C for 15 min. Positive plaques were excised and rescreened. One positive plaque from the final rescreen was *in vivo*-excised, yielding cDNA in the pDR2 vector. The plasmid was sequenced on both strands with flanking and internal primers as described. Analysis of the nucleotide sequence and homology searches with isolated mouse and human clones in the dbEST [30] and the EMBL/GenBank®/DDJB databases were performed with the programs of the GCG package. The ClustalX program package was used to align the sequences and perform phylogenetic analysis using the neighbour-joining method and bootstrapping [31].

Northern blotting

Total RNA was extracted from various tissues of adult mice using the guanidinium thiocyanate method [32]. Poly(A)$^+$ RNA was prepared using the Oligotex mRNA Midi-Kit (Qiagen). Aliquots (5 μg) of poly(A)$^+$ RNA from each tissue were electrophoresed on a 1 % denaturing agarose gel and transferred on to Hybond-XL nylon membrane (Amersham Biosciences). A 412-bp probe was generated by *Xmn*I restriction digestion of the SMOC-2 cDNA. Hybridization with the ^{32}P-labelled fragment (Ladderman Labelling Kit; TaKaRa) was carried out in formamide mix (50 % formamide, 5 × SSPE, 1.5 % SDS and 5 × Denhardt's solution) at 42 °C overnight. The membrane was washed twice in 0.5 × SSC and 0.1 % SDS at 65 °C for 10 min and exposed overnight to a PhosphorImager photoplate (Molecular Dynamics).

Reverse transcription PCR (RT-PCR) analysis

Reverse transcription was performed with 500 ng of total RNA/tissue and reaction. In the first incubation RNA together with 13 pmol of oligo(dT) primer 5′-AAGGATCCGTCGACATCGA-TAATACGACTCACTATAGGGATTTTTTTTTTTTTTTTN-3′ in a total volume of 6 μl was denatured at 70 °C for 10 min and then chilled on ice. In a modification of the manufacturer's protocol first-strand buffer (2 μl), dithiothreitol (0.1 M, 0.1 μl) and dNTPs (10 mM each, 0.5 μl) were added. Transcription was started by adding 0.5 μl of reverse transcriptase (Superscript II; Invitrogen) per reaction and incubated at 42 °C for 50 min. Finally the reaction was inactivated by heating samples to 70 °C for 15 min. Negative controls for each tissue were treated in the same way without adding enzyme. Forward primer 5′-GCTCACGTT-CTTGAGAGTCG-3′ and reverse primer 5′-TGTAGCTGTGA-CACTGGACC-3′ were used in the 35 cycles of PCR performed with Amplitaq polymerase (PerkinElmer) under the following conditions: 0.5 min at 93 °C, 1 min at 57 °C and 1.5 min at 72 °C. Authenticity of the PCR products was confirmed by direct sequencing of the purified PCR product. Aliquots (20 μl) of each PCR reaction were electrophoresed on a 2 % agarose gel using 0.5 × TBE (45 mM Tris/borate, pH 8.5, and 1 mM EDTA) and stained with ethidium bromide. The agarose gel was incubated in denaturation solution (0.5 M NaOH and 1.5 M NaCl) for 30 min followed by renaturation (0.5 M Tris/borate, pH 8.0, and 1.5 M NaCl) for 7 min and transferred overnight on to a Hybond-N nylon membrane (Amersham Biosciences). Hybridization was carried out in 1 M NaCl with a PCR-derived 304-bp probe which was labelled with [^{32}P]dCTP as described above. The membrane was washed twice with 0.1 × SSC for 20 min at 65 °C and exposed to an X-ray film (New RX; Fuji) for 10–15 min.

Protein expression and purification of mouse SMOC-2

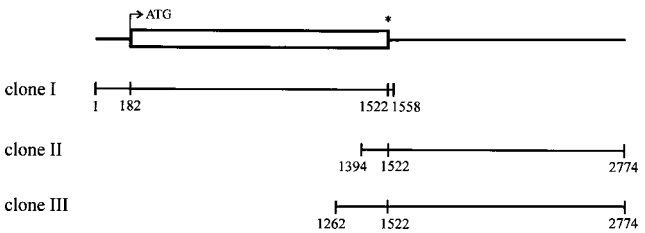
Primers were designed to amplify cDNA fragments spanning the full-length cDNA of murine SMOC-2. Forward primer 5′-GCCCGCTAGCCCAGAAGTTCTCAGCGCTC-3′ introduced an *Nhe*I restriction site at the 5′ side and the reverse primer 5′-CA-ATGACTGCGGCCGCTCATCCTTGTTTCCTGGG-3′ intro-duced a stop codon together with a *Xho*I restriction site. PCR-amplified cDNA was cloned in the pCRII vector (Invitrogen) and sequenced on both strands by cycle sequencing. *Nhe*I/*Xho*I restriction fragments of the pCRII-SMOC-2 plasmid were purified and cloned in the eukaryotic expression vector pCEP-Pu-His-Myc [33]. Correct insertion of the construct in the pCEP-Pu-SMOC-2 was verified by sequencing. Plasmids were transfected into the human embryonal kidney cell line 293-EBNA using an electroporator (Bio-Rad) according to the instructions of the manufacturer. Growth and selection of transfected cells was carried out as described [34]. Conditioned serum-free media of 293-EBNA cells were collected and were passed over a column of Talon Matrix (Clontech). Proteins were eluted in a linear gradient of 0–250 mM imidazol, in 50 mM sodium phosphate, pH 7.4, containing 300 mM NaCl. The SMOC-2-containing fractions were pooled and submitted to ion-exchange chromatography on a Resource S column (Amersham Biosciences) eluted with a gradient of NaCl (0–1 M) in 50 mM Tris/HCl, pH 7.4, to separate the full-length protein from a 30-kDa degradation fragment.

SDS/PAGE and analysis of N-linked glycosylation

Samples of purified recombinant SMOC-2 (1 μg) were electrophoresed on an SDS/12 %-polyacrylamide gel with or without prior reduction of disulphide bonds and stained with Commassie Brilliant Blue. Recombinant SMOC-2 (1 μg) was digested in 50 mM Tris/HCl, pH 7.4, with 1 μl of peptide N-glycosidase F (PNGase F; Roche Diagnostics) at 37 °C for 2 h. A control sample was treated the same way without adding enzyme. Both these samples were submitted to SDS/PAGE after reduction.

CD spectroscopy

CD spectra were recorded in a Jasco model 715 CD spectropolarimeter at 25 °C in thermostat-controlled quartz cells of optical pathlength 1 mm. The molar ellipticity (θ; expressed in deg · cm^2 · dmol^{-1}) was calculated on the basis of a mean residue molecular mass of 110 Da. The Ca^{2+} dependence of the CD




```
  1-GCGGCAGTTCGGGGAGCGCCGGCCAGAGCGCACGGAGGGGCCGTGCGGTCTCCACCAGCG
 61-GCCATAGGACCGCGAAGCAGTTCTAGCCCGCTCGCCCGCACGTTCGCACACCGGATCTTC
121-GCCGAGTGCCAGGGCGCAGCGCGTGGGGCGTCTGCCTCGCTTGGTCCCCTCCAGCGTCAC
181-CATGCTGCCGCCACAGCTGTGCTGGCTGCCGCTGCTCGCTGCCTTGCTGCCGCCGGTGCC
     M  L  P  P  Q  L  C  W  L  P  L  L  A  A  L  L  P  P  V  P  -

241-CGCGCAGAAGTTCTCAGCGCTCACGTTCTTGAGAGTCGATCAAGACAAAGACAGAGACTG
     A  Q  K  F  S  A  L  T  F  L  R  V  D  Q  D  K  D  R  D  C  -

301-CAGCCTGGACTGCCCAAGCTCCCCTCAGAAGCCACTCTGTGCCTCGGACGGGAGGACCTT
     S  L  D  C  P  S  S  P  Q  K  P  L  C  A  S  D  G  R  T  F  -

361-CCTGTCCCGATGTGAGTTCCAGCGGGCCAAGTGCAAAGATCCACAGCTGGAGATCGCTCA
     L  S  R  C  E  F  Q  R  A  K  C  K  D  P  Q  L  E  I  A  H  -

421-CCGTGGGAATTGCAAAGATGTGTCCAGGTGTGTGGCTGAGAGGAAGTATACCCAGGAGCA
     R  G  N  C  K  D  V  S  R  C  V  A  E  R  K  Y  T  Q  E  Q  -

481-GGCCCGGAAGGAGTTCCAGCAAGTGTTCATTCCAGAATGCAATGATGACGGCACCTACAG
     A  R  K  E  F  Q  Q  V  F  I  P  E  C  N  D  D  G  T  Y  S  -

541-TCAGGTCCAGTGTCACAGCTACACAGGATACTGTTGGTGTGTTACACCAAATGGAAGACC
     Q  V  Q  C  H  S  Y  T  G  Y  C  W  C  V  T  P  N  G  R  P  -

601-CATCAGTGGCACTGCTGTGGCCCACAAGACACCCAGGTGCCCCGGTTCAATAAATGAAAA
     I  S  G  T  A  V  A  H  K  T  P  R  C  P  G  S  I  N  E  K  -

661-GGTGCCGCAGCGGGAAGGAGCAGGGAAAGCAGATGATGCTGCAGCCCCAGCATTGGAGAC
     V  P  Q  R  E  G  A  G  K  A  D  D  A  A  A  P  A  L  E  T  -

721-TCAGCCCCAAGGAGATGAAGAAGATATTGCCTCACGCTACCCTACACTCTGGACCGAGCA
     Q  P  Q  G  D  E  E  D  I  A  S  R  Y  P  T  L  W  T  E  Q  -

781-AGTTAAGAGTCGGCAGAACAAGACCAATAAAAATTCAGCATCCTCCTGTGATCAGGAGCA
     V  K  S  R  Q  N  K  T  N  K  N  S  A  S  S  C  D  Q  E  H  -

841-TCAGTCAGCTCTTGAGGAAGCCAAGCAGCCCAAGAATGACAATGTAGTGATCCCCGAGTG
     Q  S  A  L  E  E  A  K  Q  P  K  N  D  N  V  V  I  P  E  C  -

901-TGCACATGGTGGTCTCTACAAGCCAGTGCAATGCCATCCATCCACCGGATACTGCTGGTG
     A  H  G  G  L  Y  K  P  V  Q  C  H  P  S  T  G  Y  C  W  C  -

961-TGTGCTAGTGGACACTGGACGGCCCATTCCTGGGACCTCCACAAGGTATGAGCAACCTAA
     V  L  V  D  T  G  R  P  I  P  G  T  S  T  R  Y  E  Q  P  K  -

1021-GTGTGACAACACAGCCCGAGCTCACCCAGCGAAGGCCCGGGACCTGTACAAGAACAGGCC
     C  D  N  T  A  R  A  H  P  A  K  A  R  D  L  Y  K  N  R  P  -

1081-ACTGCAGGGTTGTCCTGGTGCCAAAAAGCACGAGTTTCTGACAAGTGTCCTGGATGCGCT
     L  Q  G  C  P  G  A  K  K  H  E  F  L  T  S  V  L  D  A  L  -

1141-CTCCACAGACATGGTTCATGCTGTCTCTGACCCCTCTTCCTCTTCTGGCAGGCTGTCAGA
     S  T  D  M  V  H  A  V  S  D  P  S  S  S  S  G  R  L  S  E  -

1201-GCCAGACCCCAGCCACACCCTGGAGGAGAGGGTTGTACATTGGTACTTCAAGCTGCTTGA
     P  D  P  S  H  T  L  E  E  R  V  V  H  W  Y  F  K  L  L  D  -

1261-TAAGAACTCTAGTGGAGACATTGGCAAGAAGGAGATCAAACCCTTTAAGAGGTTCCTGCG
     K  N  S  S  G  D  I  G  K  K  E  I  K  P  F  K  R  F  L  R  -

1321-AAAGAAATCCAAGCCCAAAAAGTGTGTGAAGAAGTTTGTGGAGTACTGCGACATGAACAA
     K  K  S  K  P  K  K  C  V  K  K  F  V  E  Y  C  D  M  N  N  -

1381-TGACAAGTCCATCACCGTGCAGGAGCTCATGGGCTGCTTGGGTGTCACCAGAGAGGAGGG
     D  K  S  I  T  V  Q  E  L  M  G  C  L  G  V  T  R  E  E  G  -

1441-TAAAGCCAACACCAGGAAGCGCCACACCCCCAGAGGAAATGCTGAAAGTACTTCTAATAG
     K  A  N  T  R  K  R  H  T  P  R  G  N  A  E  S  T  S  N  R  -

1501-ACAGCCCAGGAAACAAGGATGAACGGCTGACTACTCAAGACAGTTCCTAGACATGTGGGA
     Q  P  R  K  Q  G  *

1561-ATTTTCCCTCACCAAAGAGCAATTAAAAACAAAAAACATATAGTATTTGCACTTTGTACT
1621-TTAAATGTAAATTCACTTTGTAGAAATGAGATATTTAAACGGACTGTTGTGATCTGTGAA
1681-AACGGAAGGCTGGCTTCGGGAAATTAATCACATACAATGTATGTGTCCTTGTTTGACCTT
1741-CAAGATCTGTGCTGGTGGGGTGGGTTTTAAATGCATTTCAACTTCACTTCCTCGTCCCTC
1801-TGTGGAGGGCCTGGTTTATTCCCTTGCCCTTTCCTCTGCCTGCAGTCTCCGTTGGCTTGC
1861-TCTCCGTGCTCTTGATGCCAGCCAGAGCCTACCGTATCCTTAGAAGCTTGCAGTGGTTGC
1921-CTGTTTCATAAAGTTTTGCTCTCTAGAGATGTCACTCACACCATGGGTCACACAGGCACT
1981-AGCTCAGGCATAGCCTTAACTTCCAGTAGCCCCTGCTTGAGACTATGCTGTTAGGCTCCA
2041-GTTTTAGCTTCTGCTTGAGCCAACCCTACTGTTAAATCCCCGGGGAGCTGAATGGACAAG
2101-CAGCCAGAGAGTCCTCTTTCAGAGAGCACATCTAACTCTGGCCTGCCGACTCCATCTCAT
2161-TGGCCCAGGAAGGATCCAGTCTACCTGCCTCACTTCTGAGTGCCTGGGCCATCCTGCTAG
2221-GGTCCACTGGAAAGGAAAAATGCCTTCTACAGAGATCATGGACCACATCCTGAGGAAGGC
2281-ACAGCTAGACTACATCGTCCCAAGAGTGCTGAACAACAGGCAGCTGAGAATGTCAATTTC
2341-ATTTTGTGGAGGCATCGGGTTTGAGGGAATTCACCCAAGGCCACTATGGGATGAAGGTTA
2401-TGAACTGGATGACCATACTGTGAGAGACTGAACTAGATTGGGTGTATGGCTGGCTGTCAC
2461-TTTCTGACCTAGGTCAGTCTGTACTCTCTGTTCTCAAGGAACAGTGGCAGATGGGAGGAG
2521-AGCTGGGGCAATTCTTTCAGATTGTGGTTTATATGGGAAATTGTTCATGGTCCCCAGCCT
2581-CTCAATGTATGTCTGTAGATGCATTGTGAACGCATGCTGCAGAAGGCCTTTGCTCATTTG
2641-AACTGTGTAGAGGCAAGAAGTGACCGGCTCATGTCAATGCTGTCTACAAACGAGTATGAT
2701-CCTAACTGTTTTGGATAATCTTTTATATTTCTGAACTCTGAATTTAATCATTTTATTAGA
2761-TTAAAATATGCGAT
```

spectrum was measured by addition of 2 mM $CaCl_2$. Reversibility of the conformational change was tested by subsequent addition of 4 mM EDTA. A baseline with buffer (5 mM Tris/HCl, pH 7.4) was recorded separately and subtracted from each spectrum.

RESULTS

Cloning and characterization of SMOC-2

Scrutiny of the dbEST database [30] revealed a mouse EST derived from fetal tissues. The 503 bp of this EST contained an open reading frame (ORF) of 82 amino acids and a potential stop codon. The sequence showed about 75 % similarity to mouse SMOC-1 on the amino acid level and therefore encoded a homologue. We screened a mouse cDNA library and isolated one positive clone after two rounds of rescreening. This clone encompassed 1558 bp which contained an ORF of 1341 bp. The EST was also sequenced completely and covered the 3′-part of the isolated clone, but contained an additional 2116 bp. The sequence of a second EST (accession number MM117770) confirmed this 3′-untranslated region (UTR) sequence (Figure 1, top panel). The composite mouse cDNA thus encompasses 2774 bp (Figure 1, bottom panel). A conserved translational start site [35] surrounds the putative ATG initiation codon. A putative polyadenylation signal (ATTAAA) is found near the ending of the 3′-UTR. Screening of the human fetal brain cDNA library revealed a clone of 469 bp which could be completed by the exons of genomic DNA (GenBank accession number NT_007302). The translated protein sequences of the mouse and human cDNA were 94 % identical to each other, whereas they were 55 % identical to the SMOC-1 proteins. Thus the new sequences comprise paralogues of SMOC-1 and were termed SMOC-2.

Domain organization of SMOC-2

The isolated mouse cDNA encodes a putative protein sequence of 446 amino acids (Figure 2). A stretch of 21 amino acids at the N-terminus of mouse SMOC-2 conforms well to the signal peptide consensus and ends with a signal peptidase cleavage site [36]. As no potential transmembrane-spanning hydrophobic region is present in the sequence, SMOC-2 will presumably be secreted from the cells. Mature SMOC-2 consists of 425 amino acids. A putative polymorphism was detected between cDNAs derived from ICR outbred mice and the C57/Bl6 strain, which differed by one triplet encoding a phenylalanine. Comparison with SMOC-1 revealed the same domain organization of five modules: an FS domain is followed by a TY domain, a domain unique to SMOC proteins, a second TY domain and an EC domain (Figures 3 and 4). As seen for SMOC-1 (Figure 4) all essential structural features of FS, TY and EC domains are conserved in SMOC-2. In contrast to the EC domain of BM-40, which contained a variant EF-hand motif in addition to a canoncial one, both EF-hand motifs of SMOC-1 and SMOC-2 are of the canonical type and each is predicted to bind a calcium ion. The SMOC-specific

Figure 1 Schematic overview of the *in vivo*-excised clones (top panel) and complete cDNA and deduced amino acid sequences (bottom panel) of mouse SMOC-2

Top panel: the clone isolated from the cDNA library (clone I) encompassed the complete ORF, whereas two EST clones, MM117770 (clone II, 1381 bp) and MM059909 (clone III, 1513 bp), spanned only parts of the ORF and consisted mainly of 3′-UTR. Bottom panel: the putative signal peptide sequence is underlined.

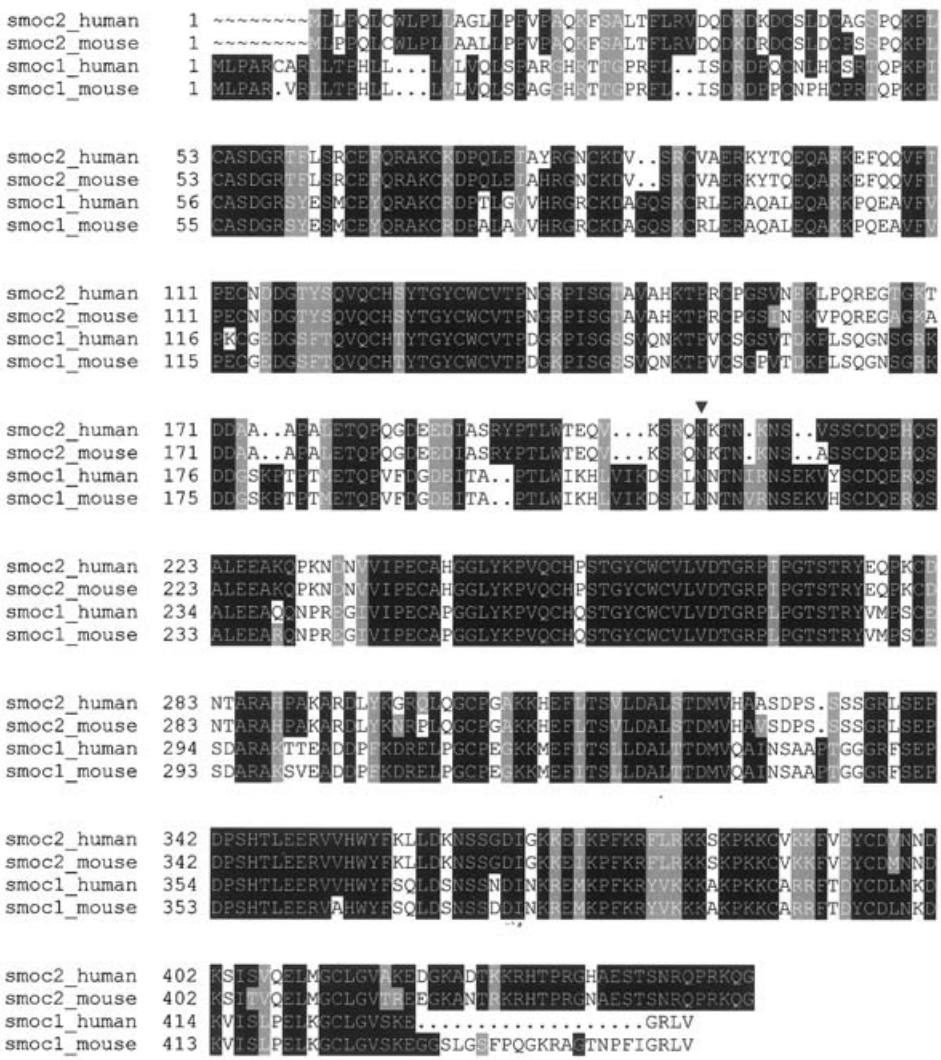


Figure 2 Alignment of the amino acid sequences of human and mouse SMOC-1 and SMOC-2

The sequence for human SMOC-2 is derived from our partial cDNA clone (EMBL/GenBank/DDBJ accession no. AJ249902) and the genomic sequence as described. The mouse SMOC-1 sequence is derived from the EST MM000223. The sequence of human SMOC-1 is available from EMBL/GenBank/DDBJ under accession no. AJ249900. Identical amino acids are shown in white on a black background, similar amino acids in white on a grey background. The triangle marks the putative N-glysosylation site.

domain between the two TY domains shows no similarity to known modules. A putative Asn-glycosylation site is found at Asn-206 and is conserved between SMOC-1 and SMOC-2. A second NXS sequence is found in the EC domain; however, the Asn and the Ser are predicted to be involved in calcium co-ordination and it remains to be elucidated whether this site is glycosylated *in vivo*.

Evolutionary relationship of SMOC-2 to other members of the BM-40 family

A phylogenetic tree was constructed for the EC domain that is shared between all members of the BM-40 family (Figure 5). Four major groups can be distinguished. SMOC-2 together with SMOC-1 forms a distinct group with its closest relatives being the group of tsc36 orthologues. The testicans form a distinct, farther-removed subfamily, as do the BM-40 orthologues together with SC1. Inclusion of the sequences of the pufferfish *Fugu rubripes* orthologues (P. Maurer, unpublished work) showed clearly that the diversification within the four groups occurred before the appearance of the vertebrates. The neighbour-joining method used to calculate the phylogenetic tree allows for different branch lengths. The fact that the branch lengths for the four groups are quite similar indicates that the rate of mutations within the EC domains has remained constant.

Structure of the human and mouse SMOC-2 genes

The structure of the human SMOC-2 gene was elucidated by analysis of a genomic contig (NT_007302) originating from the sequencing of chromosome 6. The SMOC-2 gene was mapped to chromosome 6q27. The gene spans about 226 kb. The coding region of the SMOC-2 gene consists of 13 exons (Figure 6, Table 1). Each domain of SMOC-2 is encoded by one or more exons and the domain borders coincide with splice sites. The mouse SMOC-2 gene is located on chromosome 17 and its intron–exon structure is highly conserved compared with the human gene (Figure 6, Table 2).

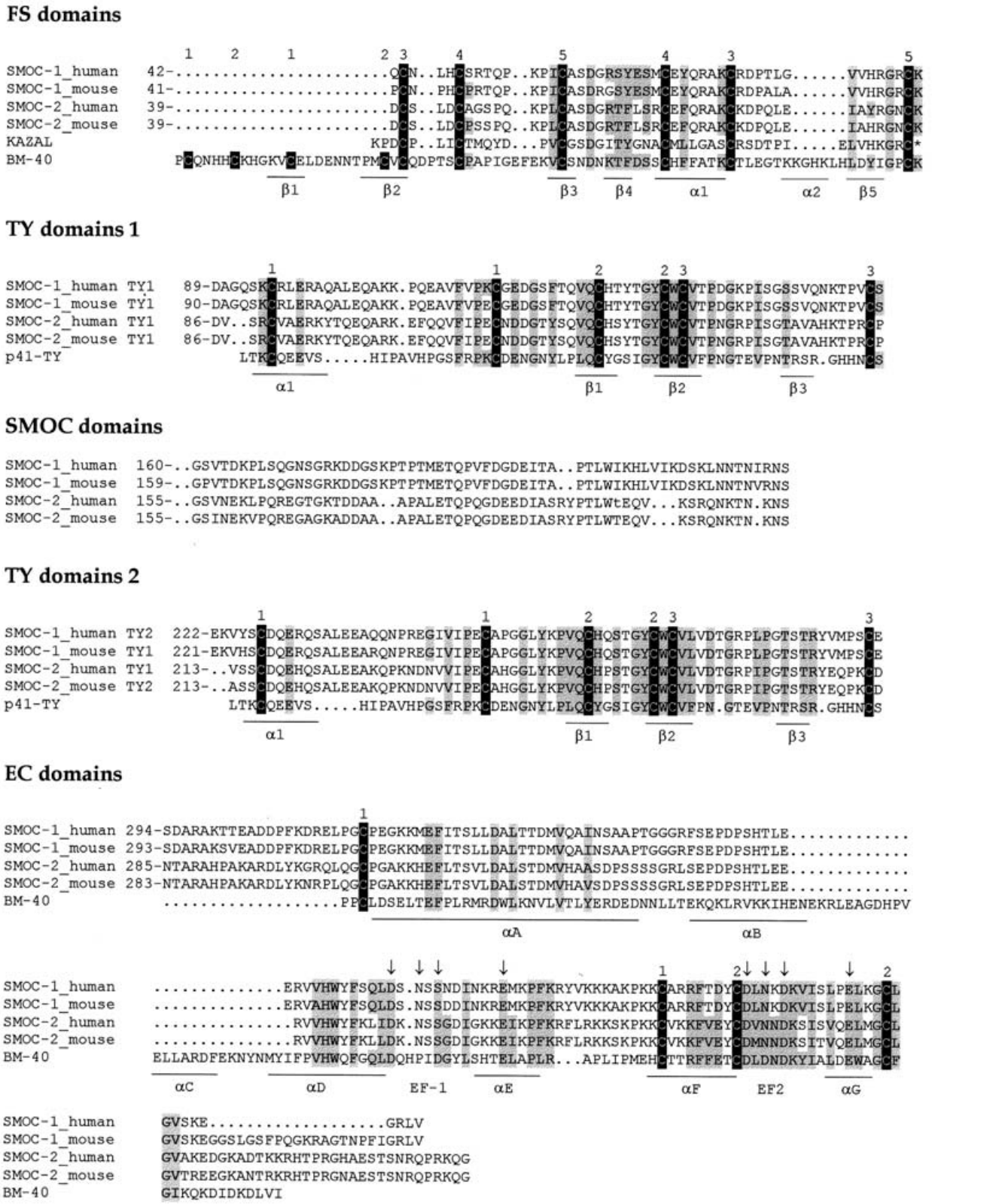


Figure 3 Structure-based alignment of SMOC-2 and related proteins

The sequences were aligned to the sequence of the FS and EC domains of BM-40, the Kazal domain of elastase inhibitor [49] and the TY domain of p41. Elements of secondary structure as revealed by the structures of BM-40 [37] and p41 [50] are given below the consensus with β-sheets and α-helices marked. Pairs of numbers above the sequence correspond to cysteines forming disulphide bonds. Calcium-co-ordinating residues are marked by arrows. Conserved amino acids are indicated with a grey background; cysteines are shown with a black background.

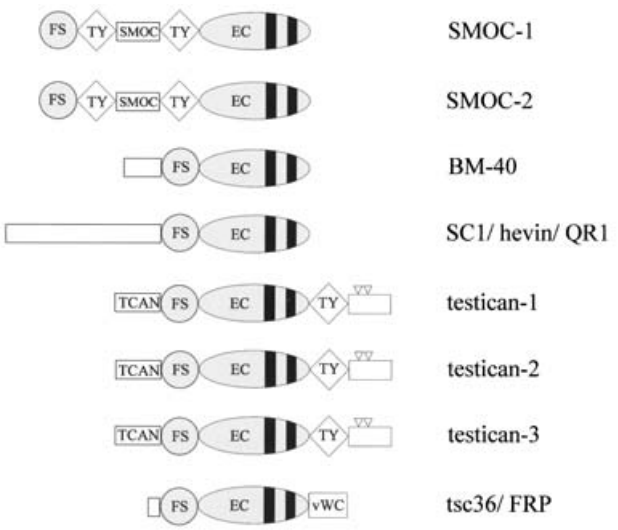


Figure 4 Domain organization of SMOC-2 and related proteins

Black bars within the EC domain refers to the EF-hand motifs. vWC, a domain with partial similarity to von Willebrand factor-type-C domains; TCAN and SMOC, domains only found in testicans and SMOC-1 and -2, respectively. Triangles indicate putative glycosaminoglycan-attachment sites in testicans. Domains with no homology with other proteins are shown as empty boxes. Signal peptides are not shown. FRP, follistatin-related protein, the name for the human and rat homologues of tsc36.

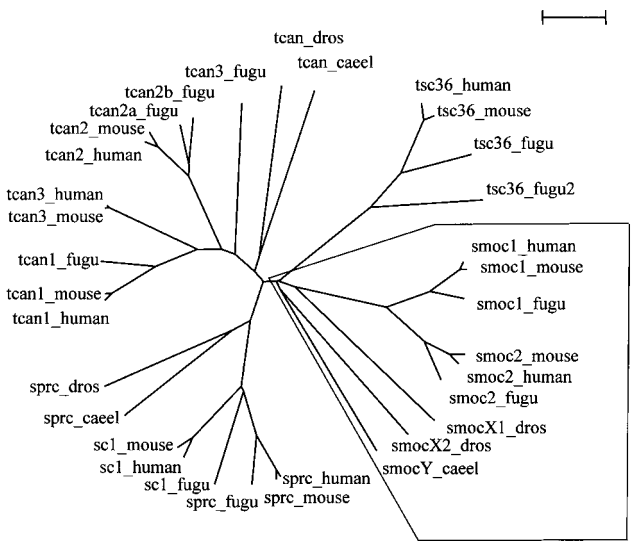


Figure 5 Phylogenetic relationships between the EC domains in the BM-40 protein family

The EC domains were aligned using ClustalX software and phylogenetic trees calculated with the neighbour-joining method. The bar indicates a branch length of 0.2 substitutions. Four subgroups can be discerned within the EC domain family; the testicans, the tsc36 orthologues, the SMOCs and the group of BM-40 (sprc) orthologues together with the rather closely related SC1. Bootstrap replicates (1000) were calculated and supported the separation of the four major groups in more than 79 % of the replicates. Nodes leading to the vertebrate orthologues were usually supported in more than 90 % of the bootstraps.

Recombinant expression of SMOC-2

A SMOC-2 expression vector was constructed based on the pCEP-Pu plasmid [34] modified to code for an N-terminal hexahistidine peptide followed by a Myc tag and a cleavage site for factor X (pCEP-Pu-His-Myc-FX [33]). The His–Myc-tagged SMOC-2 was expressed in the human embryonal kidney cell line 293-EBNA and could be purified from the cell culture supernatant by immobilized metal-ion-affinity chromatography on a cobalt column. The SMOC-2 fraction obtained from this column consisted of the full-length protein together with a 30-kDa degradation fragment which was apparently derived from the N-terminal part of SMOC-2 as it retained the tag. Recombinant His–Myc-tagged SMOC-2 migrated as a double band at 56–64 kDa in reducing and 50–60 kDa in non-reducing SDS/PAGE (Figure 7). The theoretical mass for mature His–Myc-tagged SMOC-2, calculated from the amino acid sequence, is 52909.3 Da, and the discrepancy with the estimated mass indicates the presence of post-translational modifications and/or abnormal behaviour of SMOC-2 on SDS/PAGE.

Digestion with PNGase F resulted in a shift in electrophoretic mobility corresponding to a slight loss of about 2500 Da (Figure 7). Accordingly, SMOC-2 produced in 293-EBNA cells is a glycoprotein presumably carrying one N-linked oligosaccharide. Treatment with PNGase F did not reduce the apparent mass determined by SDS/PAGE to the theoretical level, indicating a remaining abnormal electrophoretic mobility. Further, the PNGase F-digested SMOC-2 retained both bands seen in SDS/PAGE, indicating a residual heterogeneity of unknown cause.

Table 1 Exon–intron structure of the human SMOC-2 gene

Intron sequences are in lower-case letters; exon sequences are in capital letters. The bold letters indicate the consensus sequence of the splice-donor (gt) and the splice-acceptor sites (ag). SP, signal peptide.

Exon number	Domain	Exon size (bp)	Splice donor	Intron size (bp)	Splice acceptor	Codon phase
1	UTR + SP	> 273	CTC ACG **gt**aagc	68460	ttta**ag** TTT TTG	0
2	FS	172	GCA AAG **gt**aagc	16259	ctac**ag** ACG TGT	1
3	TY_1	107	AGT CAG **gt**cagg	1175	tcgc**ag** GTC CAG	0
4	TY_1	100	GCC CGG **gt**aggt	15897	tttt**ag** GTT CCG	1
5	SMOC	48	AAA CAG **gt**aact	3413	attc**ag** ATG ATG	1
6	SMOC	84*	AAG AAG **gt**gagc	1959	tcat**ag** ATA TTG	1
7	SMOC	75	ATT CAG **gt**aaga	48690	ccgc**ag** TGT CAT	1
8	TY_2	187	CAC AAG **gt**aaat	9152	ccat**ag** GTA CGA	2
9	TY_2	83	TAC AAG **gt**gagc	42441	tttc**ag** GTT GTC	1
10	EC	103	AGG CAG **gt**acgc	2170	ttcc**ag** GCT CTC	2
11	EC	275	GCC ACA **gt**aaga	10845	tgct**ag** CCC CCA	1
12	EC	38	AGA CAG **gt**aagt	2309	tttc**ag** CCA AGG	0
13	EC + UTR	1571				

* Use of a different splice-acceptor site: exon 6 size, 51 bp.

Table 2 Exon–intron structure of the mouse SMOC-2 gene

Details are as for Table 1.

Exon number	Domain	Exon size (bp)	Splice donor	Intron size (bp)	Splice acceptor	Codon phase
1	UTR + SP	> 265	CTC ACG **gt**aagt	50019	ccca**ag** TTC TTG	0
2	FS	172	GCA AAG **gt**gagc	10840	ctgc**ag** ATG TGT	1
3	TY_1	107	AGT CAG **gt**gact	1008	ctgc**ag** GTC CAG	0
4	TY_1	100	GCC CCG **gt**aggt	6367	tctt**ag** GTT CAA	1
5	SMOC	48	AAA CAG **gt**aact	3142	attc**ag** ATG ATG	1
6	SMOC	51	AAG AAG **gt**gagc	1446	ctct**ag** ATA TTG	1
7	SMOC	75	ATT CAG **gt**aagt	20134	ctgc**ag** CAT CCT	1
8	TY_2	187	CAC AAG **gt**aaat	?	tcct**ag** GTA TGA	2
9	TY_2	83	TGC AGG **gt**gata	?	ttt**ag** GTT GTC	1
10	EC	103	TGG CAG **gt**atac	1747	ctct**ag** GCT GTC	2
11	EC	275	GCC ACA **gt**aaga	1956	ttcc**ag** CCC CCA	1
12	EC	41	AGA CAG **gt**gagt	1912	ctcc**ag** CCC AGG	0
13	EC + UTR	1270				

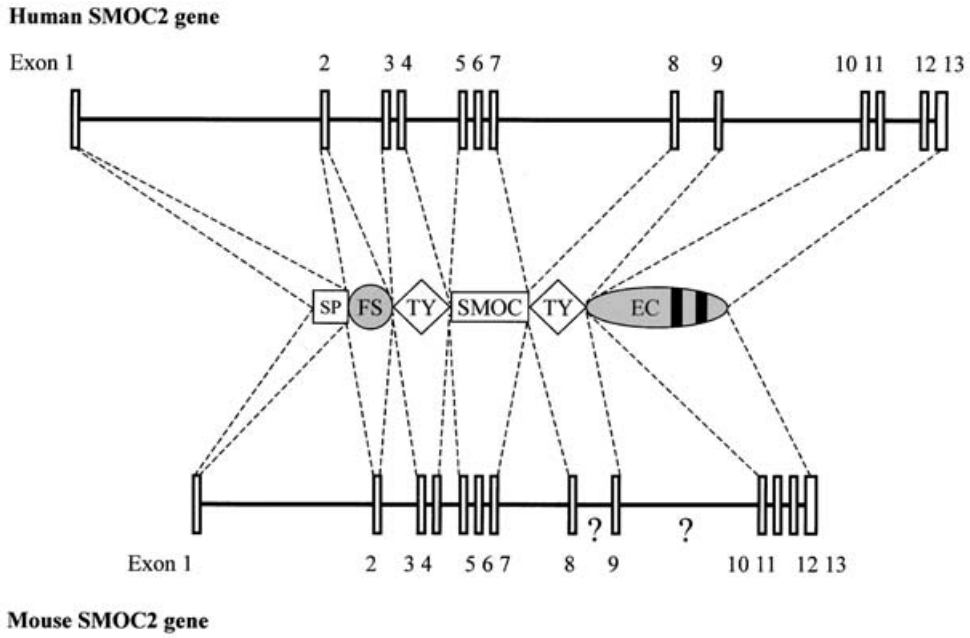


Figure 6 Exon–intron structure of the human and mouse SMOC-2 genes

All exons are numbered and span the full coding region of the SMOC-2 genes. Question marks indicate unknown intron lengths in the mouse gene. SP, signal peptide.

Conformation and calcium-binding of SMOC-2

CD spectroscopy of the tagged SMOC-2 was performed in both the presence and absence of calcium (Figure 8). The spectra showed a distinct folding with low proportions of α-helix ($\approx$10%) and predominant β-sheet structure ($\approx$40%). Calcium induced a conformational change reminiscent of that seen for BM-40, indicating that the conserved EC domain is functional.

Tissue distribution of SMOC-2 mRNA

Northern-blot analysis of mRNA from different tissues of adult mice indicates a broad expression of SMOC-2. A prominent signal of 2.7 kb appears in ovary, but clear signals are also detected in heart, muscle and spleen (Figure 9, top panel). RT-PCR amplification of mRNAs combined with hybridization with a SMOC-2 probe and autoradiographic detection demonstrated a much broader expression with signals found in brain, thymus, lung, heart, liver, kidney, spleen, testis, ovary and skeletal muscle (Figure 9, bottom panel). Thus SMOC-2 was present in all tissues analysed, and shows a somewhat broader tissue distribution than SMOC-1, which was not detected in spleen [4].

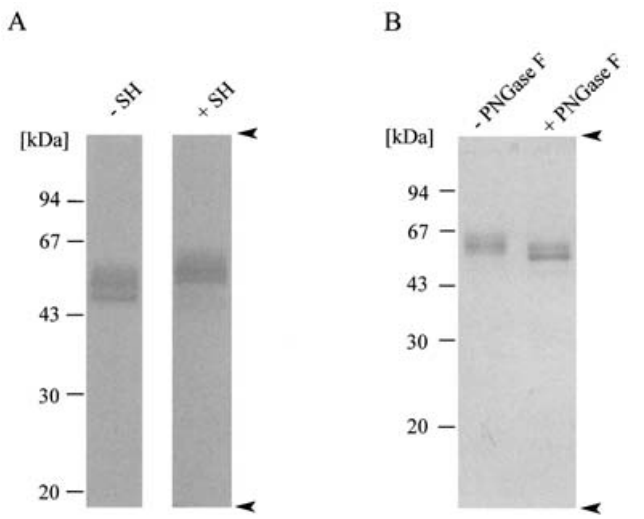


Figure 7 SDS/PAGE of recombinant SMOC-2 after reduction and after treatment with PNGase F

Gel electrophoresis of SMOC-2 was performed on a 12% polyacrylamide gel with (+SH) or without (−SH) prior reduction. SMOC-2 (1 μg) was incubated in 50 mM Tris/HCl, pH 7.4, without (−PNGase F) or with (+PNGase F) 1 μl of PNGase F at 37 °C for 2 h and electrophoresed under reducing conditions. The gels were stained with Commassie Brilliant Blue. The migration positions of calibrating proteins are given on the left. Arrows indicate the top and bottom of the separating gels.

DISCUSSION

We have isolated a novel gene that encodes a modular secreted protein. The protein is a paralogue of the recently decribed SMOC-1 [4] and was thus termed SMOC-2. SMOC-1 and SMOC-2 share a common domain organization built from one FS domain, one EC domain, two TY domains and a novel domain without known homologues. SMOC-1 and SMOC-2 are members of the BM-40 family of proteins as defined by containing a pair of a FS domains and an EC domain. The modular structure of this family is expanded in testicans and tsc36 where additional domains such as the TY domain or a domain with partial similarity to von Willebrand factor-type-C domains have been inserted during evolution. The domain organization of SMOC-1 and SMOC-2 shows a further variation of this modular setup: in all other members of the BM-40 protein family the FS domain is immediately followed by the EC domain, and both domains interact via a small surface [37]. Although the interaction of the FS

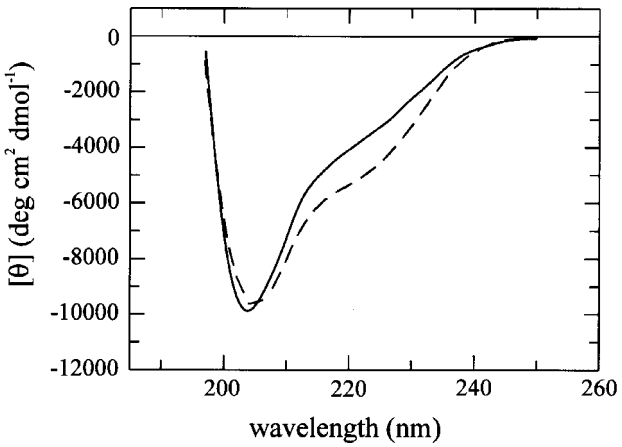


Figure 8 CD spectra of SMOC-2

Far-UV spectra were recorded at a protein concentration of 0.12 mg/ml in the presence of 2 mM $CaCl_2$ (broken line) and 4 mM EDTA (solid line) in 5 mM Tris/HCl, pH 7.4.

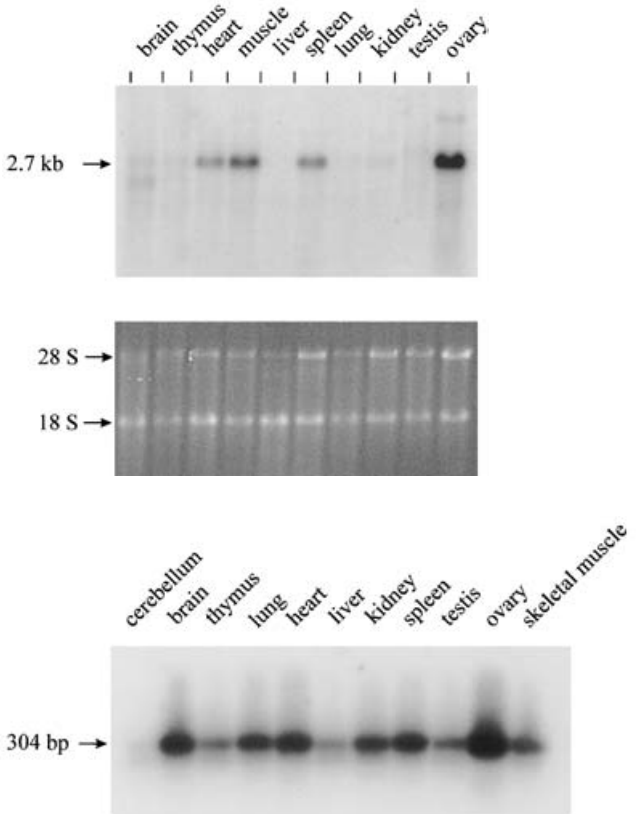


Figure 9 Tissue distribution of SMOC-2 mRNA

Top panel: 5 μg of poly(A)$^+$ RNA of various tissues from adult mice was electrophoresed on a 1% denaturing agarose gel and transferred on to a nylon membrane (uppermost panel). The membrane was hybridized with a 412-bp ^{32}P-labelled DNA fragment generated by restriction digestion of SMOC-2 cDNA. A specific signal is seen at 2.7 kb, and additional signals are due to prior hybridizations with unrelated probes (lower panel). Ethidium bromide staining of the agarose gel was performed prior to blotting to document equal loading of all lanes and the integrity of the RNAs. Bottom panel: RT-PCR was performed with 500 ng of total RNA of various tissues from adult mice using PCR primers producing a 304-bp PCR product. Aliqouts of the reaction mixture were electrophoresed on a 2% agarose gel and transferred on to a nylon membrane. After hybridization with a SMOC-2-specific probe, the membrane was exposed to X-ray film for 10–15 min.

domain with the EC domain influences calcium binding to the latter [38], the EC domain is functional and assumes the same structure when expressed separately [2]. In SMOC-1 and SMOC-2, the FS and the EC domain are separated by the two TY domains which are themselves split by the novel SMOC-specific domain.

Phylogenetic analysis of the EC domain sequences of all members of the BM-40 family clearly distinguished four groups which are in agreement with the classification of the proteins according to their domain organizations (Figure 5). BM-40 and SC1 form one group, testican-1, -2 and -3 a second, tsc36 the third, and SMOC-1 and SMOC-2 form the fourth group. Presumbably, a gene duplication in an ancestor gene containing an EC domain led to the separation of the BM-40/SC1/testican and the tsc36/SMOC-1/SMOC-2 gene precusors. A second duplication then led to the four main groups. Within these groups further duplications produced the eight different EC domain paralogues found in present-day vertebrates. In the genome of the puffer fish *Fugu rubripes* [39] orthologues of all eight paralogues are present (P. Maurer, unpublished work). The presence of *Fugu* orthologues of SMOC-1 and SMOC-2 shows that the duplication leading to SMOC-1 and SMOC-2 took place at least 450 million years ago when the last common ancestor of humans and teleost fishes lived. It can thus be speculated that the functions of SMOC-1 and -2 are conserved in vertebrates. In the invertebrates *Caenorhabditis elegans* and *Drosophila melanogaster* no orthologues of SMOC-2 and SMOC-1 with a conserved domain organization could be detected. However, genes encoding proteins with TY and EC domains are present (termed SMOCX1,2 and SMOCY in Figure 5; P. Maurer, unpublished work). Bootstrap analysis of the phylogenetic tree did not unequivocally place these proteins in the SMOC branch, thus their classification as orthologues of SMOC-1 and -2 remains unclear.

Analysis of the gene structure shows that an intron of phase I is present at each domain border in both human and mouse SMOC-2. Domains that have become mobile during evolution are characterized by introns of the same phase at their domain borders [40], ensuring that the reading frame is maintained when the domain is inserted into an intron of another gene. The intron–exon structure and the position of the introns is conserved strictly in the genes of mice and humans. For human SMOC-2 the use of an alternate splice acceptor site is responsible for the expression of two isoforms. One of the isoforms contains 11 additional amino acids within the SMOC domain. So far alternative splice products have not been found for mouse SMOC-2, although an alternative splice site is also present in the genome sequence and it remains to be determined to what extent this exon influences the function of SMOC-2.

SMOC-2 mRNA was found in a wide variety of tissues and was most prominently in ovary, heart, skeletal muscle and spleen. SMOC-1 is also strongly transcribed in ovary, although expression in spleen was not detected and weak expression was found in heart.

Predicting the functions of paralogues separated by many millions of years of evolution is very difficult, especially in the case of multidomain proteins with domain insertions and deletions. Predictions are even more complicated because binding sites within multidomain proteins are often not yet clearly assigned to individual domains and experimental verification is needed. However, the three-dimensional structure of the EC domain of BM-40 allows a structure-based prediction of functional calcium-binding sites in the EC domain of SMOC-2. The crucial acidic residues at positions 1, 3, 5, 9 and 12 and the signatures for helices encompassing the calcium-binding loops are fully conserved for both EF-hand motifs. This prediction was confirmed experimentally, with a conformational change observed

for SMOC-2 when CD spectra were recorded in the presence and absence of calcium. Calcium binding is presumably important for the structure of SMOC-2 as seen for BM-40, which is in the calcium-bound form when present in the extracellular environment [38].

The FS domain is a widespread module found not only in follistatin and members of the BM-40 family but also present in the follistatin-related gene, in the complement proteins C6, C7 and factor I, in agrin, and in the transmembrane receptors tomoregulin-1/transmembrane protein with EGF-like and two FS domains (TMEFF) 1 and TMEFF2 [41–44]. Whereas both follistatin and BM-40 bind growth factors, neither uses the FS domain for this [13,45]. The crystal structure of the FS domain revealed two subdomains, an N-terminal one with similarity to epidermal growth factor and a C-terminal subdomain that is homologous to the Kazal-type protease inhibitor domain [37]. The FS domain of SMOC-1 and -2 lack the N-terminal subdomain and the positively charged α2 helix (Figure 3). Based on the alignments it is not possible to predict whether the FS domain of SMOC-2 has protease-inhibiting activity, particularly as the sequence of the active loop between Cys-8 and Cys-16 of the elastase inhibitor, the most closely related Kazal domain, is not conserved in SMOC-2.

Eleven copies of the TY domain are present in thyroglobulin, but the function of these domains is unknown. TY modules have spread into additional modular proteins. A subgroup of TY-containing proteins including equistatin, the cysteine protease inhibitor ECI, saxiphilin and the major histocompatibility class II-associated invariant chain p41 (CD74) have proven protease inhibitory function [46]. However, nidogens and the tumour-associated antigens GA-733-1 and -2 have no such activity. Whereas testican-3 is able to inhibit membrane-type-1 matrix metalloproteinase this appears not to be due to the TY domain [47].

SMOC-1 is localized in basement membranes [4]; however, the interaction partner and the domain of SMOC-1 responsible for binding has not yet been identified. BM-40 is localized to basement membranes by the interaction of its EC domain with collagen IV [1,48]. The collagen-binding epitope on the EC domain was mapped to five crucial residues located on the opposite site of the EF-hand motifs on the N-terminal-helix and the loop that connects the EF-hands [11]. However, the residues used for collagen binding in BM-40 are not conserved in SMOC-1. In particular, the linker region between the EF-hands is longer and contains six lysine residues. Binding of SMOC-1 to collagen IV in a similar manner as BM-40 is thus questionable. The residues important for collagen binding in BM-40 are also not conserved in SMOC-2 and future studies will ascertain the importance of these domains in SMOC-2.

This work was supported by grants from the Deutsche Forschungsgemeinschaft (Ma 1932/2 and Pa 660/6) and from the Köln Fortune programme of the Medical Faculty of the University of Cologne.

REFERENCES

1 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

2 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. Nat. Struct. Biol. **3**, 67–73

3 Dziadek, M., Paulsson, M., Aumailley, M. and Timpl, R. (1986) Purification and tissue distribution of a small protein (BM-40) extracted from a basement membrane tumor. Eur. J. Biochem. **161**, 455–464

4 Vannahme, C., Smyth, N., Miosge, N., Gösling, S., Frie, C., Paulsson, M., Maurer, P. and Hartmann, U. (2002) Characterization of SMOC-1, a novel modular calcium-binding protein in basement membranes. J. Biol. Chem. **277**, 37977–37986

5 Termine, J. D., Kleinman, H. K., Whitson, S. W., Conn, K. M., McGarvey, M. L. and Martin, G. R. (1981) Osteonectin, a bone specific protein linking mineral to collagen. Cell **26**, 99–105

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

7 Maurer, P. (1996) SPARC/osteonectin/BM-40. In Guidebook to the Calcium Binding-Proteins (Celio, M. R., ed.), pp. 169–171, Oxford University Press, Oxford

8 Brekken, R. A. and Sage, E. H. (2001) SPARC, a matricellular protein: at the crossroads of cell-matrix communication. Matrix Biol. **19**, 816–827

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

10 Nischt, R., Wallich, M., Reibetanz, M., Baumann, P., Krieg, T. and Mauch, C. (2001) BM-40 and MMP-2 expression are not coregulated in human melanoma cell lines. Cancer Lett. **162**, 223–230

11 Sasaki, T., Hohenester, E., Göhring, W. and Timpl, R. (1998) Crystal structure and mapping by site-directed mutagenesis of the collagen-binding epitope of an activated form of BM-40/SPARC/osteonectin. EMBO J. **17**, 1625–1634

12 Raines, W. E., Lane, T. F., Iruela-Arispe, M. L., Ross, R. and Sage, 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. U.S.A. **89**, 1281–1285

13 Göhring, W., Sasaki, T., Heldin, C.-H. and Timpl, R. (1998) Mapping of the binding of platelet-derived growth factor to distinct domains of the basement membrane proteins BM-40 and perlecan and distinction from the BM-40 collagen-binding epitope. Eur. J. Biochem. **255**, 60–66

14 Kupprion, C., Motamed, K. and Sage, E. H. (1998) SPARC (BM-40, osteonectin) inhibits the mitogenic effect of vascular endothelial growth factor on microvascular endothelial cells. J. Biol. Chem. **273**, 29635–29640

15 Gilmour, D. T., Lyon, G. J., Carlton, M. B., Sanes, J. R., Cunningham, J. M., Anderson, J. R., Hogan, B. L., Evans, M. J. and Colledge, W. H. (1998) Mice deficient for the secreted glycoprotein SPARC/osteonectin/BM40 develop normally but show severe age-onset cataract formation and disruption of the lens. EMBO J. **17**, 1860–1870

16 Norose, K., Lo, W. K., Clark, J. I., Sage, E. H. and Howe, C. C. (2000) Lenses of SPARC-null mice exhibit an abnormal cell surface-basement membrane interface. Exp. Eye Res. **71**, 295–307

17 Yan, Q., Clark, J. I., Wight, T. N. and Sage, E. H. (2002) Alterations in the lens capsule contribute to cataractogenesis in SPARC-null mice. J. Cell Sci. **115**, 2747–2756

18 Delany, A. M., Amling, M., Priemel, M., Howe, C., Baron, R. and Canalis, E. (2000) Osteopenia and decreased bone formation in osteonectin-deficient mice. J. Clin. Invest. **105**, 915–923

19 Basu, A., Kligman, L. H., Samulewicz, S. J. and Howe, C. (2001) Impaired wound healing in mice deficient in a matricellular protein SPARC (osteonectin, BM-40). BMC Cell Biol. **2**, 15

20 Bradshaw, A. D., Reed, M. J. and Sage, E. H. (2002) SPARC-null mice exhibit accelerated cutaneous wound closure. J. Histochem. Cytochem. **50**, 1–10

21 Strandjord, T. P., Madtes, D. K., Weiss, D. J. and Sage, H. (1999) Collagen accumulation is decreased in SPARC-null mice with bleomycin-induced pulmonary fibrosis. Am. J. Physiol. **277**, L628–L635

22 Savani, R. C., Zhou, Z., Arguiri, E., Wang, S., Vu, D., Howe, C. C. and DeLisser, H. M. (2000) Bleomycin-induced pulmonary injury in mice deficient in SPARC. Am. J. Physiol. Lung Cell. Mol. Physiol. **279**, L743–L750

23 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

24 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. U.S.A. **88**, 4503–4507

25 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-beta 1-regulated genes, one of which seems to encode a follistatin-related polypeptide. Eur. J. Biochem. **217**, 13–19

26 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

27 Vannahme, C., Schübel, S., Herud, M., Gösling, S., Hülsmann, H., Paulsson, M., Hartmann, U. and Maurer, P. (1999) Molecular cloning of testican-2: defining a novel calcium-binding proteoglycan family expressed in brain. J. Neurochem. **73**, 12–20

28 Hartmann, U. and Maurer, P. (2001) Proteoglycans in the nervous system – the quest for functional roles *in vivo*. Matrix Biol. **20**, 23–35
29 Nishimoto, S., Hamajima, Y., Toda, Y., Toyoda, H., Kitamura, K. and Komurasaki, T. (2002) Identification of a novel smooth muscle associated protein, smap2, upregulated during neointima formation in the rat carotid endarterectomy model. Biochim. Biophys. Acta **1576**, 225–230
30 Boguski, M. S. (1995) The turning point in genome research. Trends Biochem. Sci. **20**, 295–296
31 Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **24**, 4876–4882
32 Chomzynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162**, 156–159
33 Wuttke, M., Müller, S., Nitsche, P., Paulsson, M., Hanisch, F.-G. and Maurer, P. (2001) Structural characterization of human recombinant and bone-derived bone sialoprotein (BSP). Functional implications for cell attachment and hydroxyapatite binding. J. Biol. Chem. **276**, 36839–36848
34 Kohfeldt, E., Maurer, P., Vannahme, C. and Timpl, R. (1997) Properties of the extracellular calcium-binding module of the proteoglycan testican. FEBS Lett. **414**, 557–561
35 Kozak, M. (1996) Interpreting cDNA sequences: some insights from studies on translation. Mamm. Genome **7**, 563–574
36 Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. **10**, 1–6
37 Hohenester, E., Maurer, P. and Timpl, R. (1997) Crystal structure of a pair of follistatin-like and EF-hand calcium-binding domains in BM-40. EMBO J. **16**, 3778–3786
38 Busch, E., Hohenester, E., Timpl, R., Paulsson, M. and Maurer, P. (2000) Calcium affinity, cooperativity, and domain interactions of extracellular EF-hands present in BM-40. J. Biol. Chem. **275**, 25508–25515
39 Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J. M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A. et al. (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science **297**, 1301–1310
40 Patthy, L. (1999) Genome evolution and the evolution of exon-shuffling — a review. Gene **238**, 103–114
41 Ullman, C. G. and Perkins, S. J. (1997) The Factor I and follistatin domain families: the return of a prodigal son. Biochem. J. **326**, 939–941
42 Schneyer, A., Tortoriello, D., Sidis, Y., Keutmann, H., Matsuzaki, T. and Holmes, W. (2001) Follistatin-related protein (FSRP): a new member of the follistatin gene family. Mol. Cell. Endocrinol. **180**, 33–38
43 Eib, D. W. and Martens, G. J. (1996) A novel transmembrane protein with epidermal growth factor and follistatin domains expressed in the hypothalamo-hypophysial axis of *Xenopus laevis*. J. Neurochem. **67**, 1047–1055
44 Horie, M., Mitsumoto, Y., Kyushiki, H., Kanemoto, N., Watanabe, A., Taniguchi, Y., Nishino, N., Okamoto, T., Kondo, M., Mori, T. et al. (2000) Identification and characterization of TMEFF2, a novel survival factor for hippocampal and mesencephalic neurons. Genomics **67**, 146–152
45 Sidis, Y., Schneyer, A. L., Sluss, P. M. Johnson, L. N. and Keutmann, H. T. (2001) Follistatin: essential role for the N-terminal domain in activin binding and neutralization. J. Biol. Chem. **276**, 17718–17726
46 Lenarcic, B. and Bevec, T. (1988) Thyropins – new structurally related proteinase inhibitors. Biol. Chem. **379**, 105–111
47 Nakada, M., Yamada, A., Takino, T., Miyamori, H., Takahashi, T., Yamashita, J. and Sato, H. (2001) Suppression of membrane-type 1 matrix metalloproteinase (MMP)-mediated MMP-2 activation and tumor invasion by testican 3 and its splicing variant gene product, N-Tes. Cancer Res. **61**, 8896–8902
48 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
49 Tschesche, H., Kolkenbrock, H. and Bode, W. (1987). The covalent structure of the elastase inhibitor from *Anemonia sulcata* – a "non-classical" Kazal-type protein. Biol. Chem. Hoppe-Seyler **368**, 1297–1304
50 Guncar, G., Pungercic, G., Klemencic, I., Turk, V. and Turk, D. (1999) Crystal structure of MHC class II-associated p41 Ii fragment bound to cathepsin L reveals the structural basis for differentiation between cathepsins L and S. EMBO J. **18**, 793–803

Received 8 April 2003; accepted 13 May 2003
Published as BJ Immediate Publication 13 May 2003, DOI 10.1042/BJ20030532