Volume 236, number 2, 352–356 FEB 06233 August 1988

Cloning and complete amino acid sequences of human and murine basement membrane protein BM-40 (SPARC, osteonectin)

Brigitte Lankat-Buttgereit, Karlheinz Mann*, Rainer Deutzmann*, Rupert Timpl* and Thomas Krieg

Dermatologische Klinik der Universität München, München and **Max-Planck-Institut für Biochemie, Martinsried, FRG*

Received 1 July 1988

Amino acid sequences of 285 and 286 residues, respectively, were deduced for mouse and human BM-40 from cDNA clones isolated from expression libraries. The sequences showed 92% identity and were also essentially identical to those of bone osteonectin and of the parietal endoderm protein SPARC. About 60% of the mouse BM-40 sequence was confirmed by Edman degradation. Two of the seven disulfide bonds were localized which apparently separate two distinct domains of mouse BM-40.

Extracellular matrix; Ca^{2+}-binding protein; Disulfide bond; cDNA sequence

1. INTRODUCTION

Basement membranes represent unique extracellular matrices being composed of several distinct proteins [1,2]. Among these components is a 40 kDa glycoprotein, BM-40, which was originally purified from the mouse Engelbreth-Holm-Swarm (EHS) tumor and shown to be present in various tissues and cell culture media [3]. The protein could be released from the matrix by chelating agents, demonstrating that divalent cations are involved in its repertoire of interactions [4]. Subsequent observations then established that BM-40 binds several Ca^{2+} cooperatively, with a K_d of 1–10 μM which is accompanied by an increase in α-helical conformation [5]. How these structural changes are reflected by distinct biological functions still remains to be identified.

Partial sequence analysis of BM-40 [4] indicated identity to the mouse parietal endoderm protein SPARC, the protein sequence of which had been deduced from cDNA sequences [6]. Analysis of the distribution of mRNA for SPARC revealed that it is not restricted to basement membrane-rich regions of mouse embryonic tissues [7,8]. Recent studies demonstrated in addition a high sequence homology to bovine osteonectin [9] which was previously considered to be a bone-specific protein with binding activities for collagen, calcium and hydroxyapatite [10,11]. Apparently, the proteins referred to as BM-40, SPARC or osteonectin are distributed in quite diverse compartments of connective tissue, but represent very similar or even identical gene products. In order to substantiate this possibility we have now completed the sequence of mouse BM-40 and compared it with a similar sequence deduced from human cDNA clones. In addition, we have localized two disulfide bonds in BM-40, which lends support to previous predictions [5,9] on the domain structure of the protein.

Correspondence address: B. Lankat-Buttgereit, Dermatologische Klinik, Frauenlobstr. 9–11, D-8000 München 2, FRG

The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under the accession number Y00755

2. MATERIALS AND METHODS

A cDNA library in the λgt11 vector of 6×10^5 independent clones with an average size of 0.8 kb was prepared for us from mouse EHS tumor mRNA by Clontech Laboratories (Palo

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

Alto, CA). About 5×10^5 clones were screened with polyclonal antibodies to mouse BM-40 [3] following previously described procedures [12,13]. The clones were plaque-purified, the inserts released by *Eco*RI digestion and subcloned into pUC19 [14]. The same and another λgt11 library from human placenta (Clontech) were screened with murine BM-40 probes which were ^{32}P-nick-translated (Amersham nick-translation kit) following the supplier's protocol. The screening followed standard procedures [15]. Positive clones were plaque-purified and subcloned into pUC19. Nucleotide sequences of the 3′- and 5′-ends of cloned inserts were determined by the dideoxy chain-termination method [16] using oligonucleotide primers for the pUC19 vector (US Biochemicals, Cleveland, OH).

Total RNA was prepared from 4 M guanidine isothiocyanate extracts of the EHS tumor and cultured human fibroblasts [14], electrophoresed on 1% formaldehyde polyacrylamide gels [15] and transferred to Gene Screen membranes (New England Nuclear, Boston, MA). Northern blot hybridization with nick-translated murine and human cDNA probes followed standard protocols [15]. Runs were calibrated with λDNA digested with *Hin*dIII (Boehringer, Mannheim).

Purified, non-reduced mouse BM-40 was cleaved with SV8 protease or TPCK-trypsin [4]. Digests were separated by molecular-sieve and reversed-phase chromatography and purified peptides subjected to Edman degradation as described [4,12].

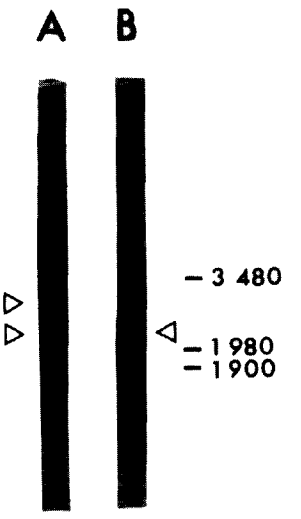

Fig.1. Northern blot hybridization of human fibroblast (A) and mouse EHS tumor (B) RNA to human or mouse BM-40 cDNA probes, respectively. Each lane was loaded with 1 μg total RNA. The nick-translated probes corresponded to the 1300 nucleotide human or 600 nucleotide mouse inserts. Open triangles show the mRNA bands identified. Numbers refer to the sizes in nucleotides of marker DNA.

3. RESULTS

Screening of an expression library prepared from the mouse EHS tumor with rabbit antiserum against mouse BM-40 yielded eight positive clones. All clones were identical as shown by sequencing and *Eco*RI digestion which resulted in inserts of about 600 nucleotides from the 3′-end and of 300 nucleotides from the 5′-end. The same internal *Eco*RI recognition site is also present in mouse SPARC cDNA [6]. Screening with the 300 nucleotide probe yielded another overlapping 300 nucleotide clone which covered the 5′-end of the coding region of mouse BM-40. Several more cDNA clones were obtained from a human placenta library by using the various mouse probes for screening. They included a single 500 nucleotide clone from the 5′-end, and several 1300 nucleotide clones from the 3′-end of human BM-40 mRNA.

Northern blot hybridization of mouse EHS tumor RNA with two different mouse probes demonstrated a single 2.2 kb band (fig.1B) as previously found for mouse SPARC mRNA [6]. Examination of human fibroblast RNA with the corresponding human probes identified the same major mRNA band but in addition a weak reaction with a 3.0 kb band (fig.1A). The nature of the latter band was not examined further.

Nucleotide sequences were determined from both inserts which overlapped at the *Eco*RI site (fig.2). The human sequence showed an open reading frame of 909 nucleotides encoding 17 amino acid residues of a signal peptide and 286 residues of the BM-40 protein. The N-terminal 36 amino acid residues of human BM-40 were identical to those reported for human osteonectin [9] except for two, Ser-Val in position 25–26 (fig.2), which had been determined as Pro-Thr in osteonectin. The mouse sequence coding for BM-40 showed 89% identity at the nucleotide and 92% identity at the amino acid level compared to the human sequence. The mouse BM-40 sequence was also identical to that of mouse SPARC [6] with two exceptions (fig.2).

About 60% of the deduced mouse sequence was confirmed by Edman degradation of peptides generated from BM-40 by SV8 protease and trypsin (fig.2). The only difference was found at amino acid position 271 where a Gly instead of Ser was found by Edman degradation which was also the position at variance between BM-40 and SPARC

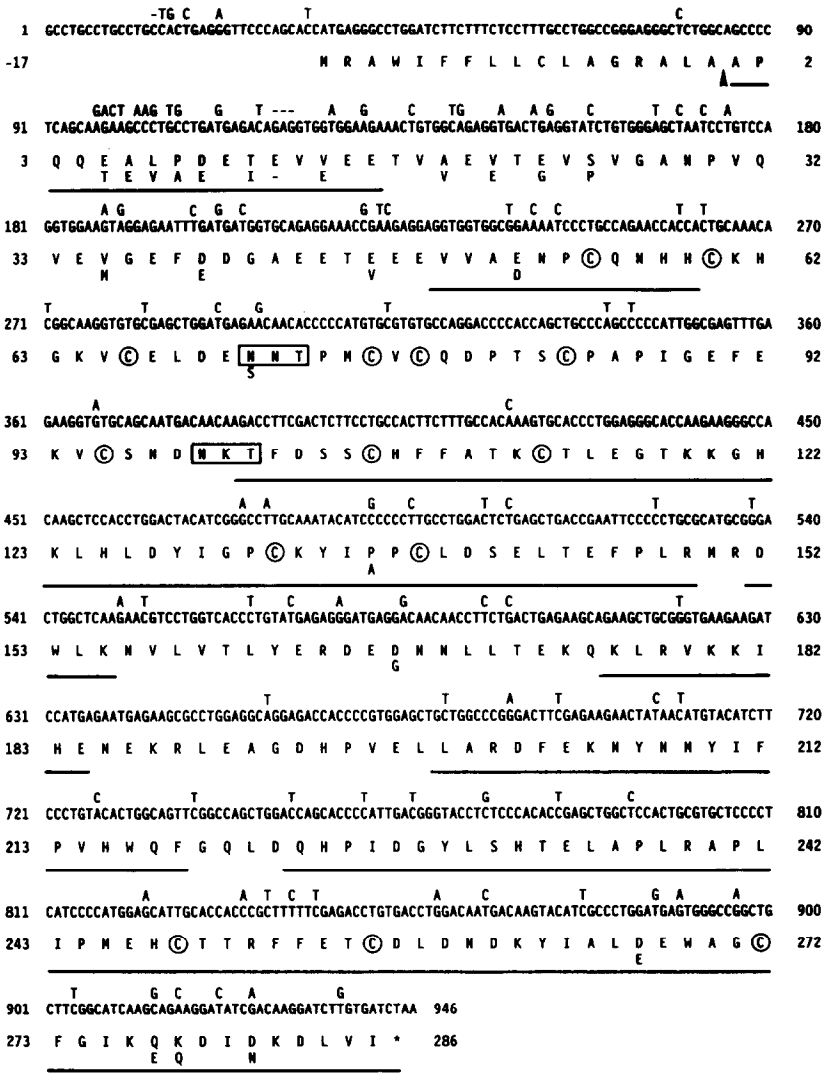

Fig.2. Nucleotide sequence and deduced amino acid sequence of cDNA clones encoding human BM-40. Nucleotide and amino acid substitutions within the corresponding murine sequence are shown above and below the human sequence. Sequences confirmed by Edman degradation of mouse BM-40 peptides are underlined. Cysteines are encircled and NXT putative carbohydrate acceptor sites are shown in boxes. An arrowhead denotes the cleavage site for the signal peptide. Numbering is for the human sequence; the mouse amino acid sequence is one residue less after position 11. Two nucleotide positions (199,896) of mouse BM-40 are at variance with SPARC cDNA [6]. Two more nucleotides (nos 200,201) were ambiguous and were taken from the SPARC sequence.

cDNA sequences (see above). Trypsin cleavage was also used to identify several disulfide-bonded peptides. A large fragment containing 8–10 of the 14 cysteine residues could be separated from smaller peptides and showed multiple N-terminal sequences. Two more peptides contained cysteines and allowed the localization of disulfide bonds between Cys-13 and -14 (peptide T5) and between Cys-11 and -12 (peptide T4). Peptide T5 occurred either with a single sequence FFETC-DLDNDKYIALEEXA(X) or as a variant with an additional cleavage of the KY bond. Peptide T4 showed only a double sequence interpreted as YIAPCLDSELTEFPLR and APLIPMEHCTTR according to the cDNA sequences (fig.2). Since only single cysteine residues were present in the

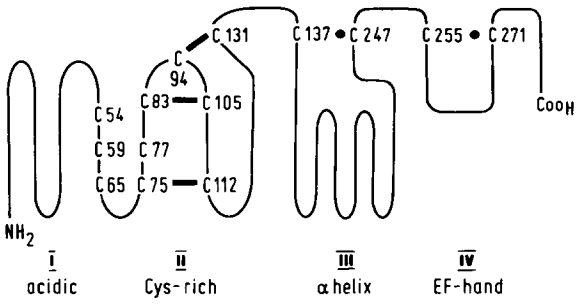

Fig.3. Schematic domain structure of BM-40 as proposed for SPARC/osteonectin [5,9] and positions of cysteines (C) and some disulfide bonds. Those determined by Edman degradation are denoted by black circles. Bonds predicted from the trypsin inhibitor homology ([9,18]; see text) are connected by black bars. Numbering of cysteines follows the mouse sequence (fig.2).

predicted tryptic peptides there was no ambiguity in assigning the disulfide bonds.

4. DISCUSSION

Complete amino acid sequences deduced mainly from cDNA clones demonstrated a high degree of identity (92%) between human and mouse BM-40. Major differences were noted close to the N-terminus including a Glu in position 12 of human BM-40 which is deleted in the mouse sequence, making both proteins slightly different in size (286 vs 285 residues, respectively). The human sequence also contains two Asn residues in typical carbohydrate acceptor sequences (positions 71 and 99, fig.2) while only the latter is conserved in the mouse. A further Asp/Glu substitution (position 267) is located within the loop of the EF-hand domain (fig.3), a presumed Ca^{2+}-binding site, but is not in conflict with the consensus sequence motif [5]. All 14 Cys residues are, however, invariant in the sequences, indicating identical folding patterns for the two proteins.

Our data also demonstrate identity of the mouse proteins BM-40 and SPARC [6] which was predicted from previous partial sequence analysis [4]. Single nucleotide differences between both sequences are presumably due to a mismatch and were not confirmed by Edman degradation. A high level of homology was also found for the bovine osteonectin sequence which was recently deduced from bone cDNA [9]. Human BM-40 differs by two successive non-conservative amino substitutions (see above) from the N-terminal sequence of human osteonectin as was determined by Edman degradation [9]. This difference is located beyond the region of high interspecies sequence variability which is encoded within a single exon of the osteonectin gene [17]. This might indicate that the BM-40 and osteonectin sequences are polymorphic variants or encoded by different, yet closely related genes.

Analyses of the SPARC [5] and osteonectin [9] sequences have led to the prediction of four protein domains (fig.3), with domains I and IV containing potential Ca^{2+}-binding sites. Our data on disulfide bond localization show in fact that the last two pairs of cysteine residues in mouse BM-40 stabilize both the α-helical domain III and the EF-hand domain IV. The latter bond was predicted from a comparison with other Ca^{2+}-binding proteins [5]. Six further Cys residues show homology to pancreatic trypsin inhibitor, ovomucoid and other protease inhibitors [9,18] and are likely to stabilize a similar disulfide-bonded structure (fig.3). These six cysteines together with the remaining four endow domain II with a compact structure which can be only incompletely cleaved by trypsin. Osteonectin contains an additional Cys located within the α-helical domain III [9] which is missing in BM-40/SPARC. This residue is unlikely to participate in intramolecular disulfide bonds but may be involved in other functions of the protein. Further sequence analyses may show whether this particular residue is unique to bone osteonectins.

Acknowledgements: We thank Dr M. Bolander for providing the osteonectin sequence data prior to publication, Mr Wilfried Schwarz and Mrs Vera van Delden for expert technical assistance, and the Deutsche Forschungsgemeinschaft and Wilhelm Sander Stiftung for financial support.

REFERENCES

[1] Martin, G.R. and Timpl, R. (1987) Annu. Rev. Cell Biol. 3, 57–85.
[2] Timpl, R. and Dziadek, M. (1986) Int. Rev. Exp. Pathol. 29, 1–22.
[3] Dziadek, M., Paulsson, M., Aumailley, M. and Timpl, R. (1986) Eur. J. Biochem. 161, 455–464.
[4] Mann, K., Deutzmann, R., Paulsson, M. and Timpl, R. (1987) FEBS Lett. 218, 167–172.

[5] Engel, J., Taylor, W., Paulsson, M., Sage, H. and Hogan, B. (1987) Biochemistry 26, 6958–6965.
[6] Mason, I.J., Taylor, A., Williams, J.G., Sage, H. and Hogan, B.L.M. (1986) EMBO J. 5, 1465–1472.
[7] Mason, I.J., Murphy, D., Münke, M., Francke, U., Elliot, R.W. and Hogan, B.L.M. (1986) EMBO J. 5, 1831–1837.
[8] Holland, P.W.H., Harper, S.J., McVey, J.H. and Hogan, B.L.M. (1987) J. Cell Biol. 105, 473–482.
[9] Bolander, M.E., Young, M.F., Fisher, L.W., Yamada, Y. and Termine, J.D. (1988) Proc. Natl. Acad. Sci. USA 85, 2919–2923.
[10] Termine, J.D., Kleinman, H.K., Whitson, S.W., Conn, K.M., McGarvey, M.L. and Martin, G.R. (1981) Cell 26, 99–105.
[11] Romberg, R.W., Werness, P.G., Lollar, P., Riggs, B.L. and Mann, K.G. (1985) J. Biol. Chem. 260, 2728–2736.
[12] Chu, M.-L., Mann, K., Deutzmann, R., Pribula-Conway, D., Hsu-Chen, C.C., Bernard, M.P. and Timpl, R. (1987) Eur. J. Biochem. 168, 309–317.
[13] Young, R.A. and Davis, R.W. (1983) Proc. Natl. Acad. Sci. USA 80, 1194–1198.
[14] Davis, L.G., Dibner, M.D. and Battey, J.F. (1986) Basic Methods in Molecular Biology, Elsevier, Amsterdam, New York.
[15] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[16] Chen, E.Y. and Seeburg, P.H. (1985) DNA 4, 165–170.
[17] Findlay, D.M., Fisher, L.W., McQuillan, C.I., Termine, J.D. and Young, M.F. (1988) Biochemistry 27, 1483–1489.
[18] Guy, O., Shapanica, R. and Greene, L.J. (1971) J. Biol. Chem. 246, 7740–7747.