THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 276, No. 15, Issue of April 13, pp. 12003–12011, 2001

Printed in U.S.A.

Isolation and Characterization of EMILIN-2, a New Component of the Growing EMILINs Family and a Member of the EMI Domain-containing Superfamily*

Received for publication, December 22, 2000
Published, JBC Papers in Press, January 16, 2001, DOI 10.1074/jbc.M011591200

Roberto Doliana‡, Simonetta Bot‡, Gabriella Mungiguerra‡, Anna Canton‡, Stefano Paron Cilli‡, and Alfonso Colombatti‡§¶

From the ‡Divisione di Oncologia Sperimentale 2, Centro di Riferimento Oncologico, 33081 Aviano and §Dipartimento di Scienze e Tecnologie Biomediche, Università di Udine, 33100 Udine, Italy

EMILIN (elastin microfibril interfase located Protein) is an elastic fiber-associated glycoprotein consisting of a self-interacting globular C1q domain at the C terminus, a short collagenous stalk, an extended region of potential coiled-coil structure, and an N-terminal cysteine-rich domain (EMI domain). Using the globular C1q domain as a bait in the yeast two-hybrid system, we have isolated a cDNA encoding a novel protein. Determination of the entire primary structure demonstrated that this EMILIN-binding polypeptide is highly homologous to EMILIN. The domain organization is superimposable, one important difference being a proline-rich (41%) segment of 56 residues between the potential coiled-coil region and the collagenous domain absent in EMILIN. The entire gene (localized on chromosome 18p11.3) was isolated from a BAC clone, and it is structurally almost identical to that of EMILIN (8 exons, 7 introns with identical phases at the exon/intron boundaries) but much larger (about 40 *versus* 8 kilobases) than that of EMILIN. Given these findings we propose to name the novel protein EMILIN-2 and the prototype member of this family EMILIN-1 (formerly EMILIN). The mRNA expression of EMILIN-2 is more restricted compared with that of EMILIN-1; highest levels are present in fetal heart and adult lung, whereas, differently from EMILIN-1, adult aorta, small intestine, and appendix show very low expression, and adult uterus and fetal kidney are negative. Finally, the EMILIN-2 protein is secreted extracellularly by *in vitro*-grown cells, and in accordance with the partial coexpression in fetal and adult tissues, the two proteins shown extensive but not absolute immunocolocalization *in vitro*.

The elasticity of many tissues such as lung, dermis, and large blood vessels depends on the presence of a high content of elastic fibers in the ECM.[1] These structures are composed of two distinct morphological elements, a more abundant amorphus core of which elastin, responsible for the elastic properties, is the major constituent, and microfibrillar structures of about 10–12-nm diameter, which are located around the periphery of the amorphus component and consist primarily of fibrillin-1 and/or -2 (1, 2). Whereas the amorphous elastic core is apparently poorly organized, fibrillin-containing microfibrils are highly organized structures. Several components that contribute to the elastic fiber organization have been identified and cloned, including microfibril-associated proteins 1 to 4 (3–6), latent transforming growth factor β-binding proteins 1 to 4 (7–10), fibulins 1 and 2 (11, 12), microfibril-associated glycoprotein-2 (13), and EMILIN (14). The latter is synthesized *in vitro*, and it is deposited extracellularly as a fine network (15, 16); it is broadly expressed in connective tissues, and it is particularly abundant in blood vessels, skin, heart, lung, kidney, and cornea (17–19). EMILIN is found at the interface between amorphous elastin and microfibrils (14), and it might regulate the formation of the elastic fiber given the finding that elastin deposition *in vitro* is perturbed by the addition of anti-EMILIN antibodies (14).

EMILIN differs from all other elastin-associated proteins and has a unique multimodular organization (20); it includes a C1q-like globular domain at the C terminus, endowed with cell-adhesion-promoting functions, a short uninterrupted collagenous stalk, a long segment of about 650 residues with a high potential for forming coiled-coil α-helices, and a new cysteine-rich domain (EMI domain) at the N terminus (21). The presence of a gC1q domain and the recent identification that gC1q is structurally homologous to the tumor necrosis factor family of growth factors (22) allowed the inclusion of EMILIN in the C1q/tumor necrosis factor superfamily of proteins (21). The gC1q-like domain is shared with several other ECM constituents including type VIII and type X collagens in which it represents the equivalent of the C propeptide of fibrillar collagens (23–25). Given the tissue distribution of EMILIN, its proadhesive functions, and the characteristics of its domains, it is likely that EMILIN plays a fundamental role in the process of elastogenesis and might associate with other ECM constituents. However, their identification is difficult because of the low solubility of the tissue form (15) and by the very large size of recombinant EMILIN (26) that makes it poorly suitable for protein-protein interaction studies. To bypass these problems we have decided to isolate potential interactors of EMILIN by the yeast two-hybrid system that allows the measurement of specific protein-protein interactions *in vivo*; a vector encompassing the C-terminal gC1q-like domain that has previously been shown to interact with itself to form homotrimers (26) was constructed. In the present study this vector was used as a bait to screen a human kidney cDNA library and allowed the identification of a novel protein that interacts with EMILIN via

* This work was supported by Grants E 704 and E 1256 from Telethon, Ministero dell'Università e della Ricerca Scientifica e Tecnologica-Cofin 1998 and 1999, and Fondo Dipartimentale. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed: Divisione di Oncologia Sperimentale, Centro di Riferimento Oncologico, 33081 Aviano, Italy. Tel.: 0039-0434-659-365; Fax: 0039-0434-659-428; E-mail: acolombatti@cro.it.

[1] The abbreviations used are: ECM, extracellular matrix; gC1q, globular C1q-like domain; PCR, polymerase chain reaction; bp, base pair; EBP, EMILIN-binding protein.

their gC1q domains. This gene product, of which preliminary accounts were reported recently (21, 27) is homologous to EMILIN, it is encoded by a distinct gene, and differs in part from EMILIN in the tissue-specific expression pattern. Its mRNA was detected in a variety of human organs including fetal heart, lung placenta, and spinal cord. We propose to classify the EMILINs as a new family of extracellular proteins and to name its members as EMILIN-1 (formerly EMILIN) and EMILIN-2 (this study).

MATERIALS AND METHODS

Yeast Two-hybrid Library Screening—The *Saccharomyces cerevisiae* strain EGY48 (p8op-lacZ; *MAT*α,his3,trp1,ura3,LexA$_{op(x6)}$-LEU2) carrying the reporter plasmid p8pop-LacZ and YM4271 (*MAT*a,ura-3–52,his3–200,lys2–801,ade2–101,ade5,trp1–901,leu2–3,112,tyr1–501,gal4-d512,gal80-D538,ade5-hisG) was used for all assays. Yeast cultures were grown at 30 °C in either YPD (1% yeast extract, 2% peptone, and 2% glucose) or SD minimal medium (0.5% yeast nitrogen base without amino acids, 2% glucose, and 1% desired amino acid dropout solution). Growth and manipulation of yeast strains was carried out using the procedures described in the Matchmaker Two-Hybrid system user manual (CLONTECH Laboratories Inc., Palo Alto, CA). For our studies, a bait was constructed by cloning in the LexA plasmid the C-terminal domain of EMILIN-1 (gC1q-1), generated by PCR amplifications using 1 ng of pCEpu-EMILIN template (26), 10 pM each primer (see below), one unit of *Taq* polymerase (Promega Corp.), 0.2 mM each of the four deoxynucleotide triphosphates (Pharmacia Ultrapure; Amersham Pharmacia Biotech), in a final volume of 100 μl of 1× Promega PCR buffer (Promega Corp.). The primers used were as follows: sense, 5′-GGGAATTCGCACCAGCAGCCCCTGTG-3′; and antisense, 5′-CCCTCGAGCTACGCGTGTTCAAGCTCTGG-3′. The underlined bases correspond to appended *Eco*RI (sense) and *Xho*I (antisense) restriction enzyme recognition sites plus two additional protective nucleotides. The PCR fragments were digested with the appropriate restriction enzymes, ligated overnight in pLexA vector, in frame with the DNA-binding domain, and transformed in the *Escherichia coli* competent DH5α strain. Ampicillin-resistant colonies were screened for the presence of the PCR fragment by restriction analysis of their plasmids. The nucleotide sequences of plasmids carrying the insert, as determined by restriction analysis, were performed by automatic sequencing, and a selected plasmid was used as a bait in the library screening. To screen for interacting proteins the EGY48 cells were sequentially transformed with the bait and with a human kidney Matchmaker cDNA library (CLONTECH) using the LiAc method. Clones were examined for transcriptional activation of reporter genes Leu- and β-galactosidase indicating interaction between bait/binding domain and library/activation domain constructs. Only clones meeting all standard two-hybrid specificity tests were considered as positive. These tests included absence of an interaction between the target construct and p53 and pLaminC-negative control constructs and the inability of colonies containing the target construct alone, or the target and the bait vectors without any insert, to pass Leu− and β-galactosidase assay. Positive clones were sequenced.

Library Screening—The entire coding sequence of the cDNA isolated in the yeast two-hybrid system was determined by cDNA library screening. The insert from one selected yeast two-hybrid system clone (about 1000 bp) was labeled by the random primer method with a multiprime kit (Amersham Pharmacia Biotech) and utilized to screen, by the plaque hybridization method, about 300,000 clones of a human kidney cDNA library in the λgt10 vector (CLONTECH). Successive rounds of screening of a human aorta cDNA library in the λgt10 vector with the most 5′-end clones resulted in the isolation of overlapping clones comprising the full-length cDNA of EMILIN-2. The sequences were performed using the Big Dye terminator cycle sequencing kit and a model 310 DNA sequencing system (PerkinElmer Life Sciences). To correct for possible *Taq* polymerase errors all sequences were determined from both strands and were repeated on clones obtained from independent PCR products. All human cDNA sequences were confirmed by sequencing the EMILIN-2 gene (see below).

Isolation and Characterization of a Human Genomic DNA Clone—A human genomic BAC library was screened for specific clones at Genome System using a cDNA insert corresponding to the 5′-end of the EMILIN-2 (see below) cDNA. Two positive clones were identified, and one was further characterized. It was authenticated by successful PCR reamplification of insert fragments with primer pairs derived from the EMILIN-2 cDNA sequences, and it was partially characterized by restriction enzyme mapping and Southern blot analysis. Appropriate restriction fragments were gel-purified and subcloned in the pGEM 7z+ vector and then sequenced.

Dot Blot Analysis—RNA expression analysis was performed using a human multiple tissues blot from CLONTECH. A ^{32}P-labeled probe was synthesized using as template the EMILIN-binding protein-1 (EBP-1) clone and the multiprime labeling kit (Amersham Pharmacia Biotech). Hybridization was performed at 65 °C in Rapid-hyb buffer. After film exposure the blot was stripped and hybridized with a ^{32}P-labeled EMILIN-1 probe. All the other procedures were performed using standard techniques.

Production of Recombinant Prokaryotic gC1q of Human EMILIN-2 (gC1q-2) and Preparation of Monoclonal Antibodies—The sequence corresponding to the C-terminal domain of EMILIN-2 (gC1q-2) was amplified from the yeast two-hybrid system template (see above) with the following primers: sense, 5′-GGGGATCCGGGCGGGGTCTGCCGCG-3′; and antisense, 5′-GGGGTACCTTAGAGGTGGGAAAGGAAAGGAT, where the underlined nucleotides correspond to appended *Bam*HI (sense) and *Kpn*I (antisense) restriction enzyme recognition sites plus two additional protective nucleotides. The amplified gC1q-2 fragment was then ligated in frame in the 6His-tagged pQE-30 expression vector (Qiagen) and transformed in M15 cells. Positive clones were isolated, and the cloned fragment was sequenced in both directions to check for errors generated by PCR. 500 ml of liquid culture grown at 0.6 $A_{600\ nm}$ was induced with 2 mM isopropyl-1-thio-β-D-galactopyranoside for 3 h at 37 °C. The culture was then centrifuged at 4000 × *g* for 20 min, and the cell pellet was resuspended in sonication buffer (50 mM sodium phosphate, pH 8.0, 0.3 M sodium chloride) at 5 volumes per gram of wet weight. The samples were frozen in a dry ice/ethanol bath, thawed in cold water, and sonicated on ice (1 min bursts/1 min cooling/2–300 watts), and cell breakage was monitored by measuring the release of nucleic acids at $A_{260\ nm}$. The cell lysate was centrifuged at 10,000 × *g* for 20 min, the supernatant was collected, and purification of the 6His-tagged recombinant fragment was performed by affinity chromatography on nickel-nitrilotriacetic acid resin (Qiagen) under native conditions. The recombinant protein was eluted from the affinity column in sonication buffer, pH 6.0, containing 10% glycerol and 0.2 M imidazole. BALB/c mice were immunized with the recombinant gC1q-2 fragment, and hybridomas that reacted with the antigen in enzyme-linked immunosorbent assay assay were selected and subcloned twice before using.

Immunofluorescence—Indirect immunofluorescence of cells grown on tissue culture glass chamber slides (Nunc Inc., Naperville, IL) was carried out on cells fixed in 4% (v/v) paraformaldehyde in phosphate-buffered saline for 30 min before incubation with the a polyclonal rabbit anti-EMILIN-1 antiserum or a murine anti-EMILIN-2 (gC1q-2 domain) monoclonal antibody. These two antibody reagents are specific for their respective antigens and did not show any cross-reactivity (data not shown). The slides were then incubated with fluorescein-conjugated goat anti-rabbit IgG (for EMILIN-1) or with rhodamine-conjugated goat anti-mouse IgG (for EMILIN-2) and examined under the confocal laser (MRC-1024, Bio-Rad laboratories, Hercules, CA) scanning microscope (Diaphot 200, Nikon Inc., Melville, NY).

RESULTS

Identification of a Binding Partner for EMILIN-1 in the Yeast Two-hybrid Assay—The two hybrid system was used to screen for potential interactors with the gC1q domain of EMILIN. A segment spanning the gC1q domain of EMILIN (residues 845 to 995 of the published sequence; see Ref. 20 and GenBank™/EBI Data Bank accession number AF 088916) was fused to the LexA DNA-binding domain in pLex, and EGY48 *S. cerevisiae* cells were transformed with this plasmid. A library of human kidney cDNAs in the pB42-activating domain was then introduced to the transformants, and the colonies growing in the absence of the Leu/His/Trp/Ura markers were selected. Among the 5×10^6 cells transformed in total, 87 colonies were Leu-positive. After the β-galactosidase assay, 7 clones of 20 positive colonies were selected and identified as representing the same clone by DNA sequencing analysis. The interaction was specific, because neither the LexA DNA-binding domain-gC1q hybrid interacted with the unfused pB42-activating domain, nor did a pB42-activating domain-unrelated hybrid clone from the library with the unfused LexA-binding domain (data not shown). Fusing the original bait (gC1q-1) into the pB42-activating domain vector and the target

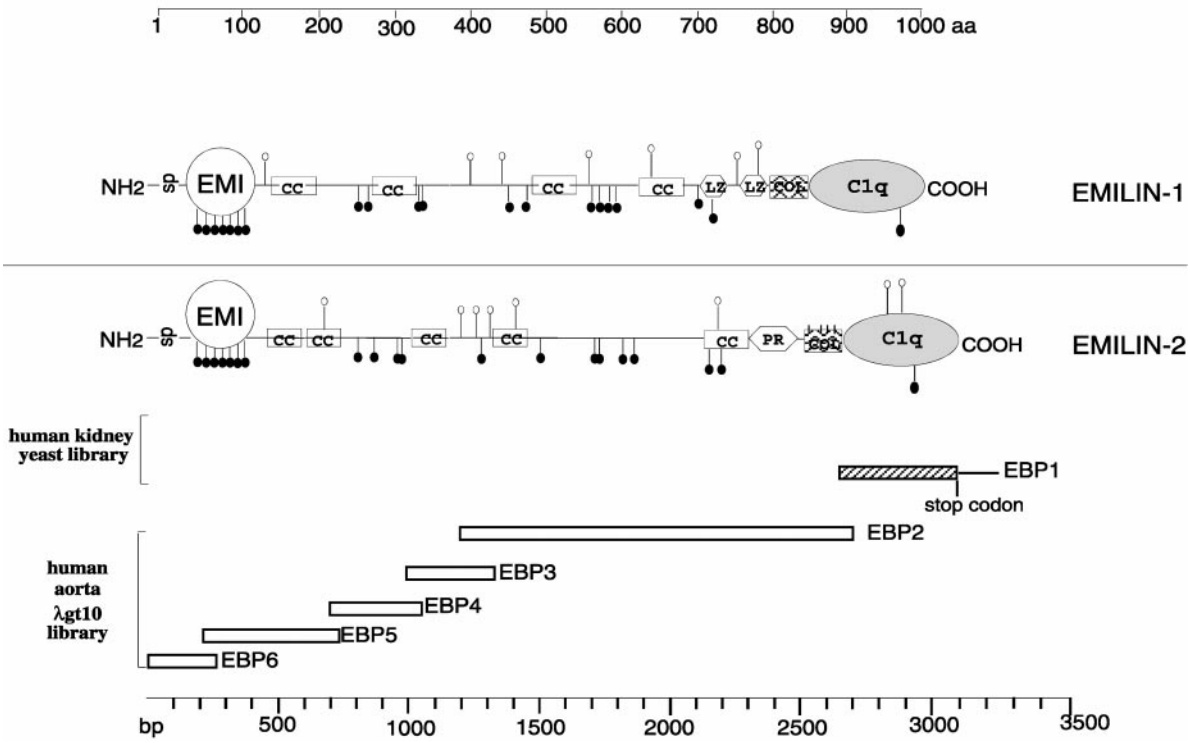


FIG. 1. **A schematic diagram with the cloning strategy and the domain structure of human EMILIN-2.** The various types of clones (*striped* and *open boxes*) used in determining the sequence are shown. *SP* indicates the signal peptide. The different domains are designated according to Bork and Koonin (48). *CC*, coiled-coil domain; *COL*, collagenous domain; *C1q*, gC1q-like domain; *LZ*, leucine zipper. *EMI* indicates the novel domain recently described in EMILIN family members (21). *PR*, proline-rich. *Bars* in the COL domain of EMILIN-2 refer to imperfections in the triple helix. Cysteines and potential glycosylation sites are indicated by *closed* and *open circles*, respectively. The diagram of EMILIN-1 is shown for direct comparison.

protein into the LexA-binding domain vector also gave a positive result. One clone, EBP-1, was then further characterized. The EBP-1 cDNA isolated contained 1000 nucleotides and encoded an in frame 192-amino acid-long open reading frame with high homology to the C-terminal end of EMILIN, including part of the collagenic region and the entire gC1q domain. The stop codon is followed by a quite long 3′-untranslated region. To obtain the full-length open reading frame, the cDNA was extended by screening a human aorta library in λgt10. Five partly overlapping cDNA clones were sequentially isolated using initially the EBP-1 fragment as the probe and then probes derived from subclones at the 5′-end of each successively isolated clone (Fig. 1, *bottom*). A composite nucleotide sequence of about 3900 base pairs was then obtained from the overlapping clones. A data bank search indicated that this novel protein (provisionally called EBP), has not been identified previously, and several partially overlapping expressed sequence tag entries showed a good match with different regions of the human EMILIN-1 transcript (20). However, expressed sequence tag clones harbored a total of 26 mismatches as compared with EBP-1, including base replacements resulting in amino acid substitutions and single base insertions or deletions. Independent sequencing of BAC clones confirmed the present sequence. The complete coding sequence and the deduced amino acid sequence and the complete 3′-untranslated region of the novel protein (GenBank™ accession number AF270513) is shown in (Fig. 2).

The characterized human cDNA spans about 3877 bp and has an open reading frame of 1053 amino acids, starting with a Met codon whose surrounding sequences fit into the eukaryotic translation start sites (28).The 3′-untranslated region of 715 nucleotides includes one putative polyadenylation signal. The predictions with the highest probabilities for the initial residue of the mature protein are between positions −1 (Ala; *X* value of 0.268) and +1 (Gly; *Y* value of 0.804) of the present sequence. Therefore, residues −30/−1 correspond most likely to the signal peptide as it agrees with the classical consensus sequence (29) and ends with a consensus signal cleavage site (30). Thus, the mature protein consists of 1023 amino acids, with a calculated molecular mass of 112 kDa, and a statistical pI of 3.8. It contains 8 potential *N*-glycosylation sites and 20 cysteines with a number of them clustered as doublets, separated by none or two residues, that could be involved in intramolecular disulfide bonding.

The Novel Protein Belongs to the EMILINs Family—We had previously reported on the isolation and characterization of EMILIN (20), a multimodular protein composed by a C-terminal gC1q domain, a short collagenic domain, a long region with high propensity to form coiled-coil structures and, at the N terminus, a characteristic cysteine-rich domain (EMI domain; see Ref. 21). The primary structure of the novel protein reported here, including the lack of a transmembrane segment, the presence of a putative secretory signal peptide, and the overall sequence composition and domain organization suggests that this gene product is an EMILIN-related ECM protein that may also form oligomers via its C-terminal gC1q domain and the potential coiled-coil domain (see Fig. 1, *top*). Based on the primary sequence, on the domain composition, and on the gene structure organization (see below), EMILIN and EBP were renamed EMILIN-1 and EMILIN-2, respectively. The novel protein includes an N-terminal EMI domain and a C-terminal gC1q domain, which both have up to 70% sequence similarity with the corresponding human EMILIN-1 domains. Pairwise alignment of the EMILINs further emphasizes the close kinship between these two members. In the C1q domain, for example, both protein present an insertion of 10 amino acids in comparison to all the other members of the C1q-containing proteins family, one of which, ACRP30/Adipo Q, is shown for comparison in Fig. 3*B*. The EMI domain at the

FIG. 2. **Nucleotide and predicted amino acid sequence of human EMILIN-2.** *First line*, nucleotide sequence; *second line*, deduced amino acid sequence. *Plain* and *bold numbers* on the *right* indicate nucleotides and amino acids, respectively. Amino acids are numbered starting at the predicted beginning of the putative mature sequence. The presumed N terminus of the mature protein is marked by a *closed arrow*, and the UAA stop codon is indicated by an *asterisk*. The polyadenylation signal is *bold* and *underlined*. Potential N attachment sites for oligosaccharides are *boxed*, and cysteine residues are *circled*. Several structural features are highlighted. The coiled-coil sequences with the residues in the *a* and *d* position are marked by a *dot* and by *Greek letters*; the glycines (*G*) of the collagenous domain are shown within *triangles*; the prolines of the proline-rich region are indicated by *reverse type*; and the C1q-like C-terminal domain is *boxed*.

N terminus is much more conserved between the two proteins in comparison to the other EMI domain-containing gene products (Fig. 3*A*). The short collagenic stretch of 17 triplets present in EMILIN-1 is conserved in EMILIN-2, although in the latter there are four imperfections. Although lacking any detectable sequence homology in the extended central region, EMILIN-1 and EMILIN-2 have structural motifs, consisting of heptad repeats in which positions 1 and 4 are preferentially occupied by aliphatic moieties and positions 5 and 7 are filled with polar residues, in common. The presence of these repeats, which are characteristic of coiled-coil α helices (Fig. 4, *top*), also suggest that EMILIN-2 might form extended homoassociations, as was determined to be the case for EMILIN-1 (26). Finally, one striking difference between the two members was the finding that in EMILIN-2 the collagenic region is preceded by a sequence, absent in EMILIN-1, characterized by an unusually high proline content. This novel proline-rich 55-residue-long sequence, in which the proline content exceeds 41% (compared with an 8–12% of proline content in the EMI and gC1q domains and a 2–4% in the coiled-coil regions), might be implicated in additional protein-protein interactions.

Chromosomal Localization and Analysis of the EMILIN-2 Gene—The exon/intron boundaries of the protein-coding region of the EMILIN-2 gene were identified by comparison between a human BAC isolated with an EMILIN-2 probe carrying the entire EMLIN gene and the cDNA sequence. As for the EMILIN-1 gene, each intron is in phase 1 with the exception of the first two introns, which are in phase 2, and all the sequences at the exon/intron boundaries are in full agreement with the consensus rules established for the splice sites of vertebrate genes (32). The gene consists of 8 exons and 7 introns (Fig. 5) as in the EMILIN-1 gene (31). However, whereas the exon structure is remarkably similar between the two genes, several introns of EMI-2 are much larger than those of the EMI-1 gene resulting in an overall

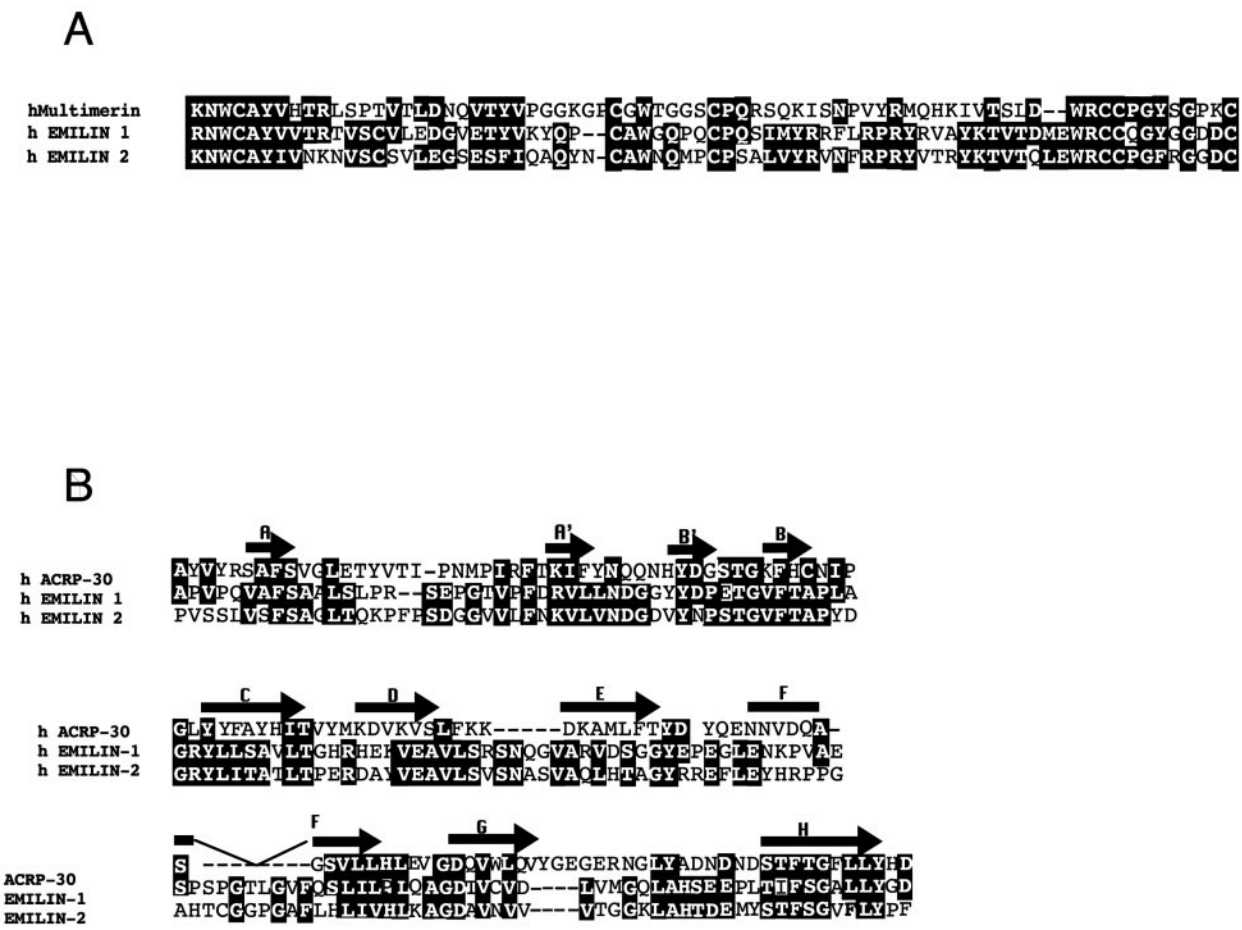


FIG. 3. **Sequence comparison.** *A*, EMI domain. *B*, gC1q domain. Identical/similar residues are indicated in *reverse type*. The locations of the β strands according to the crystal structure of the gC1q-like domain of ACRO-30/Adipo Q (22) are indicated *above* the sequence by *arrows* and *capital letters*. The insertion of 10 residues shared between EMILIN-1 and EMILIN-2 is *double-underlined*. ACRP-30/Adipo Q is shown here as representing a prototype gC1q-containing member of the superfamily.

gene size of the EMILIN-2 gene around 40 kilobases as compared with the highly compact EMILIN-1 gene (8 kilobases) (31). The C1q-like domain is split in exons 7 and 8, the latter containing also the 3′-untranslated sequence; the collagenic region is encoded by exon 6 and part of 5, the latter also encoding the EMILIN-2-specific proline-rich region. Finally, characteristic is the presence in both genes of a very uncommon large exon of about 1900 bp, in which the coiled-coil regions potentially involved in interchain interaction are clustered. The present amino acid sequences of EMILIN-1 and EMILIN-2 and the strong similarities of exon organization indicates that they are the products of closely related but distinct genes likely to be derived from a common ancestor. While this study was under way, a GenBank™ search using sequences originated from the BAC clone retrieved a cluster of partially characterized human genomic clones localized on chromosome 18p11.3 between the markers D18S476 and D18S481. One of the clones, corresponding to GenBankTM entry AC015958, allowed the complete characterization of the 3′-end of the gene, from exon 5 to the 3′-untranslated region. A partial overlap exists between the very end of the EMILIN-2 gene (AF270513) and GenBank™ entry AL117592, corresponding to a cDNA of about 2100 bp. Moreover, a predicted gene (KIAA0249) with a transcript of about 6000 bp lays very close to the 3′-end of the EMILIN-2 gene on the opposite strand.

EMILIN-1 and EMILIN-2 Are Differentially Expressed in a Variety of Tissues—Distribution of EMILIN-1 and EMILIN-2 mRNAs in various adult and fetal human tissues was studied by RNA blot hybridization on a multiple tissue blot containing 50 different tissues and developmental stages. Many tissues express EMILIN-1 mRNA in different amounts, with the highest levels in the adult small intestine, aorta, lung, uterus, and appendix and in the fetal spleen, kidney, lung, and heart; intermediate expression was detected in adult liver, ovary, colon, stomach, lymph node, and spleen; adult heart, bladder, prostate, adrenal gland, mammary gland, placenta, and kidney showed low expression whereas a series of other adult tissues, including skeletal muscle and different regions of adult brain, did not express EMILIN-1 mRNA at all. (Fig. 6). The mRNA expression for EMILIN-2 resulted much more restricted, with a relatively high expression in fetal heart and adult lung, intermediate levels in peripheral leukocytes, placenta, and spinal cord and low expression in fetal brain, spleen, thymus, and lung and in adult heart, aorta, testis, bone marrow, small intestine, thymus, lymph node, and appendix. Although RNA spotted amounts are accurately normalized allowing a semiquantitative comparative analysis among the tissue mRNAs hybridized with the same probe, a direct quantitative comparison between EMILIN-1 and EMILN-2 mRNA expression is not feasible because of possible differences in hybridization efficiency between the two probes. Nevertheless, the conclusion can be reached that (i) EMILIN-1 is more widely distributed in both fetal and adult tissues; (ii) EMILIN-1 is expressed at higher levels in fetal heart and fetal lung compared with adult tissues (or any other tissue); (iii) EMILIN-2 is much more expressed in fetal than in adult heart; (iv) conversely, adult lung shows the highest expression for EMILIN-2 as compared with fetal lung and all the other tissues; and (v) finally, in uterus only EMILIN-1 is expressed.

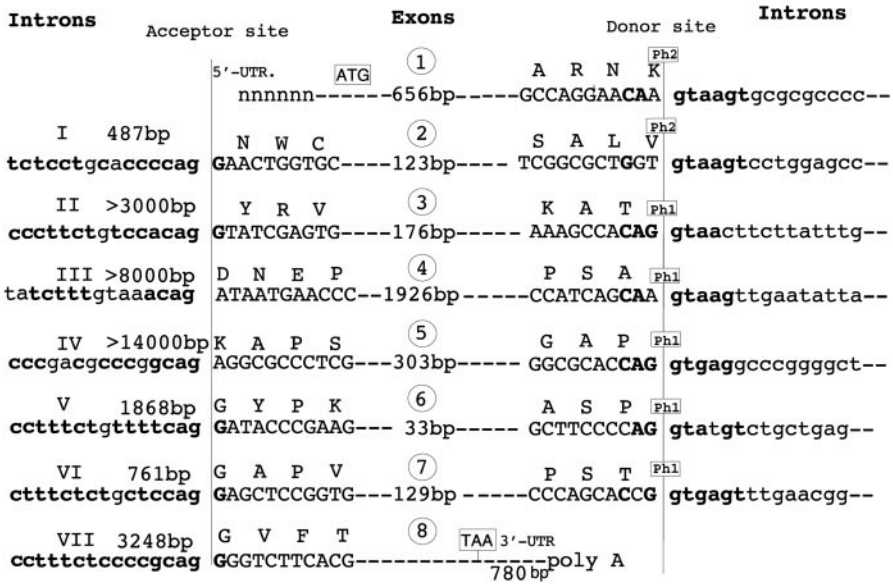


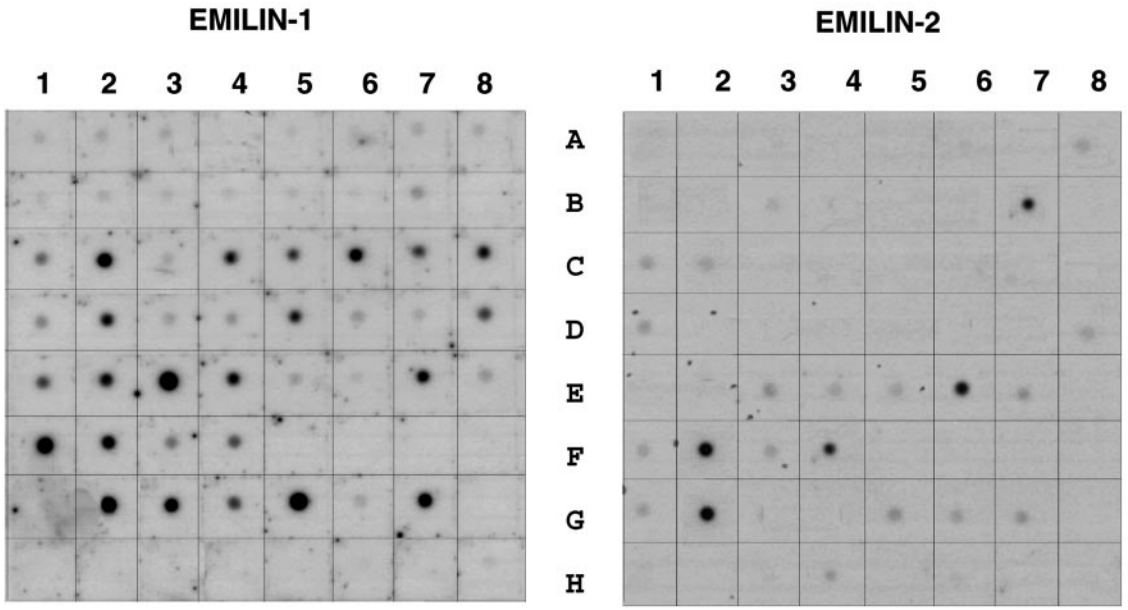
FIG. 4. **Sequences of the intron:exon junctions of the human EMILIN-2 gene.** Translated sequences are given in *uppercase letters*; intron sequences are given in *lowercase letters*. Amino acids encoded near and at splice junctions are indicated in *one-letter code* above their codons. Exon and intron sizes are also shown. Consensus sequences of the splice acceptor and donor sites are in *bold*. Splice site consensus sequences are shown at the *bottom*. *y*, pyrimidine; *n*, any nucleotide. *Ph1* and *Ph2* indicate introns that interrupt a codon triplet after the first or after the second nucleotide, respectively. *UTR*, untranslated region.

FIG. 5. **Schematic representation of the human EMILIN-2 gene.** A schematic diagram of the EMILIN-1 gene, as reported in Ref. 31, is shown at the *top*. *Broken lines* connect the exons of EMILIN-1 with the corresponding exons of EMILIN-2 that are numbered from 1 to 8. *Numbers above* the various exons and introns refer to their length in base pairs. The various domains corresponding to the exons are indicated as in Fig. 1. *PR*, proline-rich; *SP*, signal peptide; *col*, collagenous domain.

Codistribution of EMILIN-1 and EMILIN-2 in Vitro—Confocal microscopy analysis was performed on several tumor cell lines, and in a number of them a positive expression of EMILIN-2 was detected. As shown in Fig. 7 for the leiomyosarcoma cell line SK-LMS-1, EMILIN-2-specific immunofluorescence was extracellular with a diffuse meshwork pattern. In colabeling experiments it partially colocalized with EMILIN-1, although in some areas a predominant deposition of EMILIN-2 or EMILIN-1 could also be detected.

DISCUSSION

Candidate interactors for human EMILIN-1 were investigated by the yeast two-hybrid system. One ligand, EMILIN-2, that is secreted extracellularly and is deposited *in vitro* in the ECM with a meshwork pattern, was identified using as a bait the gC1q-1 domain of EMILIN-1; its cDNA and gene and a preliminary mRNA tissue distribution pattern are reported. The structural characteristics and the predicted domain organization of EMILIN-2 replicate closely those recently established for EMILIN-1 (20). As a result, the structural/functional criteria defining the EMILIN members are beginning to emerge; accordingly, they are expected to be constituted of four structurally distinct regions preceded by a signal peptide. In fact, both EMILIN-1 (20) and EMILIN-2 display the newly identified EMI domain at their N terminus. This domain is consistently found at the N terminus downstream of the signal peptide in all EMI domain-containing proteins, except for multimerin, which also has a propeptide upstream of the EMI domain that is cleaved before secretion of the mature protein (33). Using both qualitative and quantitative yeast two-hybrid systems, the EMI-1 domain was recently found to interact with the gC1q-1 domain and even more strongly with the gC1q-2 domain (21). This finding suggests that, in addition to the gC1q-1/gC1q-2 interaction that was instrumental in isolating the first EMILIN-2 clone from the library, the heterotypic EMI-1/gC1q-2 interaction detected *in vivo* in the two-hybrid system might be related to the macroassembly and tissue organization of EMILINs. The EMI-2 domain is followed by an extended discontinuous sequence with the potential of forming amphipathic coiled-coil α-helices. Although the sequence homology between the EMILINs is negligible in this domain as is



1	2	3	4	5	6	7	8	
whole brain	amygdala	caudate nucleus	cere-bellum	cerebral cortex	frontal lobe	hippo-campus	medulla oblongata	A
occipital lobe	putamen	sub-stantia grigia	temporal lobe	thalamus	sub-talamic nucleus	spinal cord		B
heart	aorta	skeletal muscle	colon	bladder	uterus	prostate	stomach	C
testis	ovary	pancreas	pituitary gland	adrenal gland	thyroid gland	salivary gland	mammary gland	D
kidney	liver	small intestine	spleen	thymus	peri-pheral leukocyte	lymph node	bone marrow	E
appendix	lung	trachea	placenta					F
fetal brain	fetal heart	fetal kidney	fetal liver	fetal spleen	fetal thymus	fetal lung		G
yeast total RNA 100ng	yeast tRNA 100ng	E.coli rRNA 100ng	E.coli DNA 100ng	Poly r(A) 100ng	human Cot DNA 100ng	human DNA 100ng	human DNA 500ng	H

FIG. 6. **mRNA expression.** Expression of EMILIN-1 and EMILIN-2 using a human multiple tissue blot is shown at the *top*. Corresponding tissues are indicated at the *bottom*. The two blots were exposed for the same length of time.

the relative position of the heptad repeats, the overall propensity to form coiled-coil structures (34, 35) is similarly high in both molecules (20, 27).

EMILIN-2 is slightly larger than EMILIN-1 and harbors, right upstream of the collagenous domain, a unique proline-rich motif of 53 residues. The proline-rich region is also of potential interest for EMILIN-2 interactions and assembly. For instance and by analogy to proteins such as dystrophin, which has an overall extended conformation interrupted by proline-rich sequences representing sites of protein-protein interaction (36) or may allow bending of the protein, the proline-rich region of EMILIN-2 could provide some flexibility that is not present in EMILIN-1. Among the elastic/microfibril-associated glycoproteins fibrillin-1 has one proline-rich region of equivalent length. However, different from the proline-rich region of fibrillin-1 that is largely hydrophobic (30%) and thus unlikely to form a surface loop (37, 38), in EMILIN-2 this region is highly hydrophylic and potentially exposed to the solvent and thus available for interactions with other ligands.

While displaying four interruptions of the Gly-X-Y triplets not detected in EMILIN-1 (20), the collagenous domain of EMILIN-2 could still form a trimeric collagen-like region, as shown for instance in type IV collagen (39). This sequence could participate in trimerization providing additional binding strength to the trimers. At variance from the rigid stalk that the collagenous domain would form in EMILIN-1, these imperfections could confer to the collagenous domain of EMILIN-2 more flexibility or bending capability. Considering that upstream of this domain there is the hydrophylic proline-rich domain, a flexible rod could confer a higher probability of protein-protein interaction with potential ligands.

The close identity between the EMILIN-1 and EMILIN-2 cDNAs is further emphasized by their gene organization, which is almost identical. The exon size pattern and location of introns in the coding sequence are very well conserved between the two genes, and the two genes have probably evolved from a common ancestor. However, the intron sequences have diverged, because the intron sizes in EMILIN-2 are much larger,

and the overall gene size is around 40 kilobases, five times larger than the EMILIN-1 gene (20). The divergent evolution of the two genes probably reflects random loss from and/or uptake of intervening sequences into the noncoding regions of the genes after they duplicated. Interestingly, the EMILIN-2 gene is located on chromosome 18p11.3, centromerically positioned but close to the LAMA1 gene coding for the laminin α1 chain (GenBank™). The precise chromosomal mapping of EMILIN-2 is not possible yet, because that chromosomal region is still ill defined, but the EMILIN-2 gene is very likely between the markers D18S476 and D18S481 right upstream of the KIAA0249 gene.

The gC1q domains of both EMILINs have a high sequence homology including a unique stretch of 10 residues absent in all other members of the C1q/tumor necrosis factor superfamily identified so far (20). The gC1q-1 domain has been shown experimentally to promote homotrimerization of EMILIN-1 (26); similarly, gC1q domains of other members of the superfamily can form homo- or heterotrimers (22, 40–45). Thus, it is very likely that gC1q-2 will also form trimers. The following question then arises. Given the finding that the gC1q-1 bait interacted with a gC1q-2 cDNA clone of the library, is there the possibility that EMILIN-1 and EMILIN-2 can form heterotrimeric assemblies, or are they compatible only with the formation of homotrimers (Fig. 8)? Although the gC1q domains are highly similar, the exclusive presence of the proline-rich region in EMILIN-2, the detection of four imperfections in its collagenous domain, and the fact that the *in vitro* ECM-deposited EMILIN molecules display only a partial colocalization favor the hypothesis that distinct homotrimers are formed.

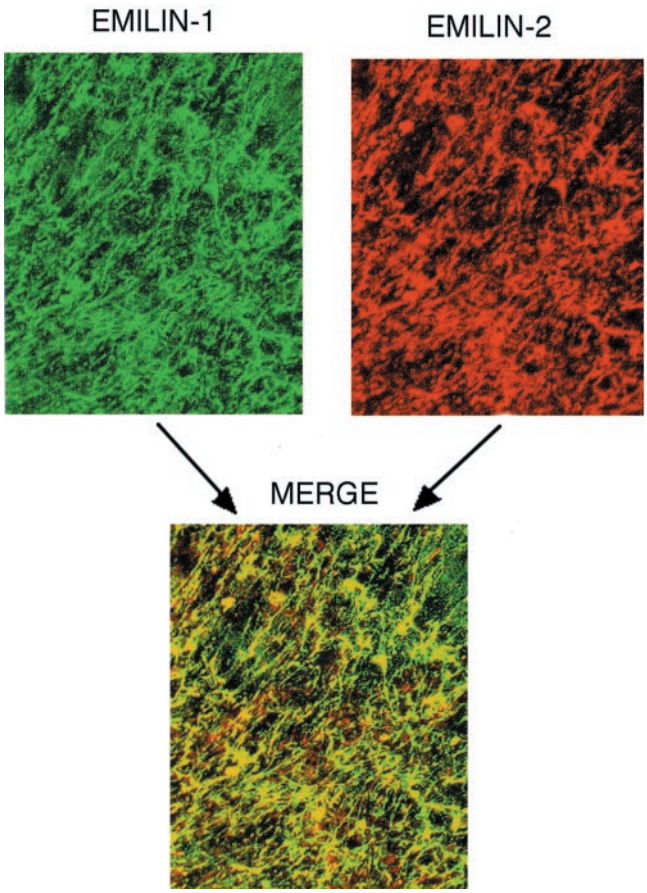


FIG. 7. **Double-labeling immunostaining of EMILIN-1 and EMILIN-2.** Human leiomyosarcoma cells were grown *in vitro* for four days, fixed, and incubated with anti EMILIN-1 (*left*) or anti EMILIN-2 (*right*) antibodies. On the *center panel* with the merged images the partial colocalization is indicated by the *yellow* staining.

The gC1q domain-containing molecules assemble to quaternary structures composed of multimers of several polypeptides reaching sizes of several millions of daltons. Both *in vivo* (15) and in EMILIN-1-transfected 293-EBNA cells (26) EMILIN-1 is present as large molecular aggregates. The closely related multimerin platelet protein similarly forms large multimers (46). These polymers are apparently because of intermolecular S-S bonds, because both EMILIN-1 and multimerin migrate as a trimeric protomer of about 500 kDa under reducing conditions in SDS gels. EMILIN-2 also has a number of cysteines that might potentially be involved in intermolecular S-S bonds.

Recombinant EMILIN-1 promoted cell adhesion of a number of hematopoietic and nonhematopoietic cell lines (27). In addition, a proadhesive function was also associated with the isolated recombinant and native gC1q-1 domain (20) suggesting that at least part of the cell binding activity could reside in this domain. Among the numerous members of the C1q/tumor necrosis factor superfamily, a cell adhesive function had been reported previously for the gC1q domain of the complement C1q, as well (47). It will be a matter of further studies to investigate whether EMILIN-2 and/or its gC1q-2 domain are endowed with a similar proadhesive function.

The data on tissue and developmental expression of EMILIN-2, although still very preliminary, allow some considerations to be drawn. The prominent expression in the fetal heart and the drastic reduction in the adult heart suggest that EMILIN-2 might be involved in or promote the development of heart chambers. On the contrary, the striking reverse pattern observed in the lung, *i.e.* low expression in the fetus and high expression in the adult, indicates a potential role of EMILIN-2 in the physiology of respiration. More in depth studies and a comparative analysis of EMILIN expression in the developing mouse are necessary and should help elucidate the role played by these molecules. Finally, although formal ultrastructural proof that also EMILIN-2 is located at the elastin-microfibril interface is still lacking, the finding that the tissue distribu-

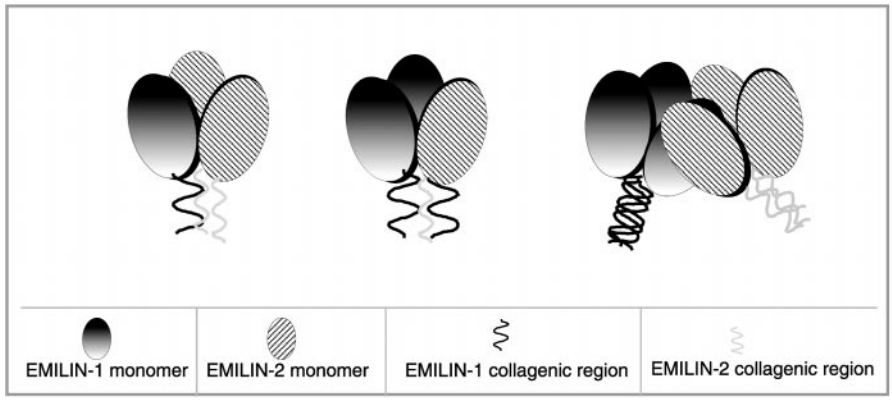


FIG. 8. **Schematic representation of the protomers of EMILIN-1 and EMILIN-2: potential interaction models between these two members.**

tion of EMILIN-1 and EMILIN-2 is only partly overlapping supports the notion that EMILINs contribute to the compositional and maybe functional heterogeneity of ECM structures.

Acknowledgments—We thank Francesco Bucciotti for excellent technical assistance and Dr. Paola Spessotto for performing immunofluorescence staining.

REFERENCES

1. Sakai, L. Y., Keene, D. R., and Engvall, E. (1986) *J. Cell Biol.* **103,** 2499–2509
2. Zhang, H., Apfelroth, S. D., Hu, W., Davis, E. E., Sanguineti, C., Bonadio, J., Mecham, R. P., and Ramirez, F. (1994) *J. Cell Biol.* **124,** 855–863
3. Henderson, M., Polewski, R., Fanning, J. C., and Gibson, M. A. (1996) *J. Histochem. Cytochem.* **44,** 1389–1397
4. Gibson, M. A., Hatzinikolas, G., Kumaratilake, J. S., Sandberg, L. B., Nicholl, J. K., Sutherland, G. R., and Cleary, E. G. (1996) *J. Biol. Chem.* **271,** 1096–1103
5. Abrams, W. R., Ma, R. I., Kucich, U., Bashir, M. M., Decker, S., Tsipouras, P., McPherson, J. D., Wasmuth, J. J., and Rosenbloom, J. (1995) *Genomics* **26,** 47–54
6. Zhao, Z., Lee, C.-C., Jiralerspong, S., Juyal, R. C., Lu, F., Baldini, A., Greenberg, F., Caskey, C. T., and Patel, P. I. (1995) *Hum. Mol. Gen.* **4,** 589–597
7. Kanzaki, T., Olofsson, A., Moren, A., Werntedt, C., Hellman, U. K., Claesson-Welsh, L., and Heldin, C. H. (1990) *Cell* **6,** 1051–1061
8. Gibson, M. A., Hatzinikolas, G., Davis, E. C., Baker, E., Sutherland, G. R., and Mecham, R. P. (1995) *Mol. Cell. Biol.* **15,** 6932–6942
9. Yin, W., Smiley, E., Flanders, K. C., and Sporn, M. B. (1995) *J. Biol. Chem.* **270,** 10147–10160
10. Saharinen, J., Taipale, J., Monni, O., and Keski-Oja, J. (1998) *J. Biol. Chem.* **273,** 18459–18469
11. Roak, E. F., Keene, D. R., Haudenschild, C. C., Godyna, S., Little, C. D., and Argraves, W. S. (1995) *J. Histochem. Cytochem.* **43,** 401–411
12. Reinhardt, D. P. (1996) *J. Biol. Chem.* **271,** 19489–19496
13. Raghunath, M., Tschodrich-Rotter, M., Sasaki, T., Meuli, M., Chu, M.-L., and Timpl, R. (1999) *J. Invest. Dermatol.* **112,** 97–101
14. Bressan, G. M., Daga-Gordini, D., Colombatti, A., Castellani, I., Marigo, V., and Volpin, D. (1993) *J. Cell Biol.* **121,** 201–212
15. Bressan, G. M., Castellani, I., Colombatti, A., and Volpin, D. (1983) *J. Biol. Chem.* **258,** 13262–13267
16. Colombatti, A., Bonaldo, P., Volpin, D., and Bressan, G. M. (1988) *J. Biol. Chem.* **263,** 17534–17540
17. Colombatti, A., Bressan, G. M., Castellani, I., and Volpin, D. (1985) *J. Cell Biol.* **100,** 18–26
18. Colombatti, A., Bressan, G. M., Volpin, D., and Castellani, I. (1985) *Collagen Relat. Res.* **5,** 181–191
19. Colombatti, A., Poletti, A., Bressan, G. M., Carbone, A., and Volpin, D. (1987) *Collagen Relat. Res.* **7,** 259–275
20. Doliana, R., Mongiat, M., Bucciotti, F., Giacomello, E., Deutzmann, R., Volpin, D., Bressan, G. M., and Colombatti, A. (1999) *J. Biol. Chem.* **274,** 16773–16781
21. Doliana, R., Bot, S., Bonaldo, P., and Colombatti, A. (2000) *FEBS Lett.* **484,** 164–168
22. Shapiro, L., and Scherer, P. E. (1998) *Curr. Biol.* **8,** 335–338
23. Sage, H., Pritzl, P., and Bornstein, P. (1980) *Biochemistry* **19,** 5747–5755
24. Kittleberger, R., Davis, P. F., and Greenhill, N. S. (1989) *Biochem. Biophys. Res. Commun.* **159,** 414–419
25. Barber, R. E., and Kwan, A. P. (1996) *Biochem. J.* **320,** 479–485
26. Mongiat, M., Mungiguerra, G., Bot, S., Mucignat, M.-T., Giacomello, E., Doliana, R., and Colombatti, A. (2000) *J. Biol. Chem.* **275,** 25471–25480
27. Colombatti, A., Doliana, R., Bot, S., Canton, A., Mongiat, M., Mungiguerra, G., Paron-Cilli, S., and Spessotto, P. (2000) *Matrix Biol.* **19,** 289–301
28. Kozak, M. (1989) *J. Cell Biol.* **108,** 229–241
29. Nielsen, H., Engelbrecht, J., Brunak, S., and von Hejne, G. (1997) *Protein Eng.* **10,** 1–6
30. Perlman, D., and Halvorson, H. O. (1983) *J. Mol. Biol.* **167,** 391–409
31. Doliana, R., Canton, A., Bucciotti, F., Mongiat, M., Bonaldo, P., and Colombatti, A. (2000) *J. Biol. Chem.* **275,** 785–792
32. Mount, S. M. (1982) *Nucleic Acids Res.* **10,** 459–472
33. Polgar, J., Magnenat, E., Wells, T. N., and Clemetson, K. J. (1998) *Thromb. Haemost.* **80,** 645–648
34. Berger, B., Wilson, D. B., Wolf, E., Tonchev, T., Milla, M., and Kim, P. S. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 8259–8263
35. Wolf, E., Kim, P. S., and Berger, B. (1997) *Protein Sci.* **6,** 1179–1189
36. Ervasti, J. M., and Campbell, K. P. (1991) *Cell* **66,** 1121–1131
37. Pereira, L., D'Alessio, M., Ramirez, F., Lynch, J. R., Sykes, B., Pangilinan, T., and Bonadio, J. (1993) *Hum. Mol. Genet.* **2,** 961–968
38. Corson, G. M., Chalberg, S. C., Dietz, H. C., Charbonneau, N. L., and Sakai, L. Y. (1993) *Genomics* **17,** 476–484
39. Schuppan, D., Timpl, R., and Glanville, R. W. (1980) *FEBS Lett.* **115,** 297–300
40. Chan, D., Weng, Y. M., Hocking, A. M., Golub, S., McQuillan, D. J., and Bateman, J. F. (1996) *J. Biol. Chem.* **271,** 13566–13572
41. Frischholtz, S., Beier, F., Girkontaite, I., Wagner, K., Poschl, E., Turnay, J., Mayer, U., and von der Mark, K. (1998) *J. Biol. Chem.* **273,** 4547–4555
42. Rosenblum, N. D. (1996) *Biochem. Biophys. Res. Commun.* **227,** 205–210
43. Illidge, C., Kielty, C., and Shuttleworth, A. (1998) *J. Biol. Chem.* **273,** 22091–22095
44. Reid, K. B. (1989) *Behring Inst. Mitt.* 8–19
45. Smith, K. F., Haris, P. I., Chapman, D., Reid, K. B. M., and Perkins, S. J. (1994) *Biochem. J.* **301,** 249–256
46. Hayward, C. P. M., Warkentin, T. E., Horsewood, P., and Kelton, J. G. (1991) *Blood* **77,** 2556–2560
47. Nicholson-Weller, A., and Klickstein, L. B. *Curr. Opin. Immunol.* **11,** 42–46
48. Bork, P., and Koonin, E. V. (1996) *Curr. Opin. Struct. Biol.* **6,** 366–375

Isolation and Characterization of EMILIN-2, a New Component of the Growing EMILINs Family and a Member of the EMI Domain-containing Superfamily

Roberto Doliana, Simonetta Bot, Gabriella Mungiguerra, Anna Canton, Stefano Paron Cilli and Alfonso Colombatti

J. Biol. Chem. 2001, 276:12003-12011.
doi: 10.1074/jbc.M011591200 originally published online January 16, 2001
