

EMILIN, a Component of the Elastic Fiber and a New Member of the C1q/Tumor Necrosis Factor Superfamily of Proteins*

(Received for publication, December 23, 1998, and in revised form, March 9, 1999)

Roberto Doliana‡, Maurizio Mongiat‡, Francesco Bucciotti‡, Emiliana Giacomello‡, Rainer Deutzmann§, Dino Volpin¶, Giorgio M. Bressan¶, and Alfonso Colombatti‡**

From the ‡Divisione di Oncologia Sperimentale 2, Centro di Riferimento Oncologico di Aviano, 33081 Aviano, Italy,

§Department of Biochemistry, Genetics and Microbiology, University of Regensburg, Regensburg, Germany D-8400,

¶Istituto di Istologia, Università di Padova, 35100 Padova, Italy, and ¶Dipartimento di Scienze e Tecnologie Biomediche, Università di Udine, 33100 Udine, Italy

EMILIN (elastin microfibril interface located protein) is an extracellular matrix glycoprotein abundantly expressed in elastin-rich tissues such as blood vessels, skin, heart, and lung. It occurs associated with elastic fibers at the interface between amorphous elastin and microfibrils. Avian EMILIN was extracted from 19-day-old embryonic chick aortas and associated blood vessels and purified by ion-exchange chromatography and gel filtration. Tryptic peptides were generated from EMILIN and sequenced, and degenerate inosine-containing oligonucleotide primers were designed from some peptides. A set of primers allowed the amplification of a 360-base pair reverse transcription polymerase chain reaction product from chick aorta mRNA. A probe based on a human homologue selected by comparison of the chick sequence with EST data base was used to select overlapping clones from both human aorta and kidney cDNA libraries. Here we present the cDNA sequence of the entire coding region of human EMILIN encompassing an open reading frame of 1016 amino acid residues. There was a high degree of homology (76% identity and 88% similarity) between the chick C terminus and the human sequence as well as between the N terminus of the mature chick protein where 10 of 12 residues, as determined by N-terminal sequencing, were identical or similar to the deduced N terminus of human EMILIN. The domain organization of human EMILIN includes a C1q-like globular domain at the C terminus, a collagenous stalk, and a longer segment in which at least four heptad repeats and a leucine zipper can be identified with a high potential for forming coiled-coil α helices. At the N terminus there is a cysteine-rich sequence stretch similar to a region of multimerin, a platelet and endothelial cell component, containing a partial epidermal growth factor-like motif. The native state of the recombinantly expressed EMILIN C1q-like domain to be used in cell adhesion was determined by CD spectra analysis, which indicated a high value of β -sheet conformation. The EMILIN C1q-like domain promoted a high cell adhesion of the leiomyosarcoma cell line SK-UT-1, whereas the fibrosarcoma cell line HT1080 was negative.

Elastic fibers are major constituents of the extracellular matrix (ECM),¹ confer to connective tissues the properties of resilience and elastic recoil, and are remarkable for the diverse range of tissues in which they are found. Elastic fibers can be identified in the ECM of many tissues as solid branching and unbranching fine and thick rod-like fibers (in elastic ligaments) or as concentric sheets of lamellae (in blood vessels) or in three-dimensional meshworks of fine fibrils (in elastic cartilage) or as a combination of these (in skin and lung) (1). Electron microscopy has provided additional insights into the structure of elastic fibers, which are composed of two morphologically distinguishable components: an amorphous core lacking any apparent regular or repeating structure (2) and a microfibrillar component (3) consisting of fibrils of 12–13 nm in diameter that are located primarily around the periphery of the amorphous core but to some extent are also interspersed within it.

Despite this simple morphology, elastic fibers are now known to be highly complex structures (4) consisting of several specific proteins, including elastin, fibrillin 1 and 2 (5, 6), microfibril-associated proteins 1 to 4 (7–10), latent-transforming growth factor β -binding protein 1 to 4 (11–14), fibulin (15), and MAGP-2 (8, 16). Elastin and fibrillins are the main components of the amorphous core and microfibrils, respectively. Fibrillin-containing microfibrils are highly organized structures, and microfibril-associated proteins and probably latent-transforming growth factor β -binding proteins are predominantly associated with them. Fibrillin-containing microfibrils are also found as elastin-free bundles in tissues where elastic fibers cannot be morphologically distinguished such as in ocular zonule, oxitalan fibers of the cornea, kidney, and spleen (17).

We had originally isolated from chick aorta a novel glycoprotein component associated with the ECM of blood vessels, gp115 (18), later christened EMILIN (19). The major characteristics of this protein were the following. EMILIN was preferentially extracted from tissues using buffers containing guanidine HCl and reducing agents; it formed a fibrillar network in the ECM of aorta; the amino acid composition was characterized by a high content of glutamic acid and arginine (18). Subsequent studies have established that (i) EMILIN is broadly expressed in connective tissues and is particularly abundant in blood vessels, skin, heart, lung, kidney, and cornea, whereas it is undetectable in the serum (20–22); (ii) the protein is synthesized by aortic smooth muscle cells and by

* This work was supported by grants from Associazione Italiana per la Ricerca sul Cancro, Telethon, MURST-Cofin 1997 (to A. C.), and Ricerca Sanitaria Finalizzata 793/03/97 (D.V.). 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.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF088916.

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

¹ The abbreviations used are: ECM, extracellular matrix; gC1q-like, globular C1q-like domain; TNF, tumor necrosis factor; PCR, polymerase chain reaction; EST, Expressed Sequence Tag; bp, base pair(s); RACE, rapid amplification of cDNA ends; CAFCA, centrifugal assay for fluorescence-based cell adhesion.

tendon fibroblasts, and it is deposited extracellularly as a fine network (18, 23); (iii) soon after secretion, EMILIN undergoes intermolecular cross-linking by disulfide bonds, giving rise to high molecular weight aggregates (23); (iv) EMILIN is a component of elastic fibers and is localized mainly at the interface between amorphous elastin and microfibrils (19); (v) finally and more important for the functional significance of EMILIN, the process of elastin deposition *in vitro* is perturbed by the addition of anti-EMILIN antibodies in the culture medium (19). Therefore, given the close co-distribution of elastin and EMILIN, the fine localization at the interface between elastin and microfibrils, and the interference with the deposition of elastin *in vitro*, it is likely that EMILIN plays a fundamental role in the process of elastogenesis also *in vivo*. To initiate addressing the question of the functional role of EMILIN directly, we have sought to clone its gene.

We report here the cDNA sequence and the analysis of the deduced amino acid sequence of human EMILIN. This glycoprotein, a new member of the C1q/TNF superfamily of proteins, is characterized by a gC1q-like C-terminal domain, a short collagenous domain, two leucine zippers, and an extended coiled-coil region that is uniquely shared with another member of this superfamily, multimerin (24). At the N terminus of these two members of the superfamily there is a short region of homology including a partial epidermal growth factor-like motif. In addition, the isolated recombinantly produced EMILIN gC1q-like C-terminal domain is able to support cell adhesion.

EXPERIMENTAL PROCEDURES

Purification of Avian EMILIN and Peptide Sequences—EMILIN was purified as described previously (18). Briefly, aortas and associated blood vessels were excised from 19-day-old chick embryos and dropped into ice-cold 20 mM sodium phosphate, pH 7.4, 150 mM NaCl containing the following protease inhibitors: 25 mM EDTA, 2 mM phenylmethylsulfonyl fluoride, 5 mM *N*-ethylmaleimide, and 1 mM *p*-aminobenzamide hydrochloride. All subsequent procedures were carried out at 4 °C. The tissues were homogenized, and the pellet obtained after centrifugation at 12,000 × *g* for 45 min was re-extracted twice in the same buffer and twice in 0.1 M Tris-HCl, pH 7.5, containing 6 M guanidine HCl and protease inhibitors (guanidine buffer). The last pellet was extracted twice in guanidine buffer containing 25 mM dithioerythritol. The last two supernatants were extensively dialyzed against distilled water and lyophilized. The lyophilized extract was reduced and alkylated, and it was then fractionated by DEAE-cellulose chromatography followed by agarose gel filtration. The EMILIN-containing fractions, as assessed by an enzyme-linked immunosorbent assay and by immunoblotting with a specific monoclonal antibody, 147H11 (21), were pooled, dialyzed against distilled water, and lyophilized. An aliquot of the pooled fractions was analyzed by SDS-polyacrylamide gel electrophoresis to check for the presence of contaminating proteins.

EMILIN was then resolved on a 4–10% SDS gradient gel. The Coomassie-stained band corresponding to EMILIN was excised, cut into small pieces, and washed with 0.2 M NH_4HCO_3 followed by 0.2 M NH_4HCO_3 , acetonitrile (1:1). This procedure was repeated twice. The gel pieces were lyophilized for 2 h and then rehydrated by adding three portions of 0.2 M NH_4HCO_3 containing 0.002% Tween 20 at 5-min intervals. The first aliquot added contained 1 mg of trypsin/100- μl gel pieces, and only as much liquid was used as was necessary to restore the original gel volume. Protease digestion was performed overnight at 37 °C, and the peptides were extracted twice with 5% trifluoroacetic acid and once with 2.5% trifluoroacetic acid in 50% aqueous acetonitrile. Alternatively, the gel pieces were repeatedly washed with 50 mM Tris buffer, pH 9.0, and 50 mM Tris buffer, acetonitrile (1:1), and cleavage was carried out with a lysine-specific protease from *Achromobacter* (Wako) at pH 9.0 and 30 °C. The peptides were then separated by high performance liquid chromatography on a reversed phase Nucleosil-120 C18 column using a linear gradient of 0–70% aqueous acetonitrile in 20 mM ammonium acetate buffer. Sequence analysis was done on a Procise protein sequencing system (Applied Biosystems) according to the manufacturer's instructions. The amino acid sequences obtained were used to search the Swiss-Prot protein sequence data base (Geneva University Hospital and University of Geneva, Geneva, Switzerland).

Primers Design—Four of the nine chick EMILIN peptides sequenced

appeared suitable for the purpose of degenerate primers design. A set of degenerate inosine-containing oligodeoxynucleotides based on peptides 4, 6, 8 and 9 were synthesized on an ABI-381A synthesizer (Applied Biosystem) in both the sense and the antisense orientations (Fig. 1). The primers were purified by native polyacrylamide gel electrophoresis (20% gel) using a standard sequencing apparatus.

PCR-based Cloning Strategy and Sequencing of Chick and Human EMILIN—Total RNA from aortas of 19-day-old chick embryos was isolated using RNA fast (Molecular System, San Diego, CA). The poly(A)+ RNA fraction was purified from total RNA with the use of Oligotex kit (Qiagen GmbH, Germany). The first-strand cDNA was synthesized starting from 1 μg of poly(A)+ RNA primed with hexanucleotides and reverse-transcribed with 20 units of AMV-RT (Promega Corp., Madison, WI). PCR was performed on a Robocycler Gradient Apparatus (Stratagene, La Jolla, CA) with the degenerate inosine-containing oligonucleotides used as primers in all the possible different combinations and at different annealing temperatures ranging from 48 to 72 °C with 2 °C steps. The amplification products were cloned into the pGEM-T vector (Promega Corp.) and sequenced by the dideoxynucleotide chain termination method using the modified T7 polymerase (Sequenase, Amersham Pharmacia Biotech).

To clone the human EMILIN cDNA, the FASTA computer program (25) located at the EBI server (26) was used to search the GenBank™ data base for expressed sequence tags (ESTs) containing sequences homologous to those of chick EMILIN. Several human and mouse entries showed a significant homology to the 3' end of the chick EMILIN. Two primers, ESTH-N (5'-ATTATGATCCAGAGACAGGC-3') and ESTH-R (5'-CCGAGTGCAGCCAGCTGCC-3'), were designed based on the entry HSAA1823 and used in PCR to obtain a human EMILIN-specific probe. The reaction was performed on a template constituted by total kidney RNA primed with hexanucleotides and reverse-transcribed with 20 units of avian myeloblastosis virus reverse transcriptase. The amplification product was cloned into the pGEM-T vector (EST-H clone) and sequenced to confirm its identity. The EST-H insert (290 bp) was labeled by the random primer method with the multiprimer 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 $\lambda\text{gt}10$ vector (CLONTECH Laboratories Inc., Palo Alto, CA). The longest insert of the five positive plaques identified, K1, was used to rescreen the library. The second screening yielded only one positive clone, K2, constituted by a fusion between a specific EMILIN cDNA and an unrelated cDNA. Three additional rounds of screening of a human aorta cDNA library resulted in the isolation of clones A1, A2, and A3. Also in this library several clones appeared to be cloning artifacts carrying short EMILIN-specific sequences fused to unrelated cDNAs. The human sequences were performed using the Big Dye terminator cycle sequencing kit and a model 310 DNA sequencing system (Perkin-Elmer Applied Biosystem). 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 gene by an independent analysis of a BAC clone, used to characterize the EMILIN gene.²

Rapid Amplification of cDNA Ends (RACE)—To determine the sequence extending toward the 3' end, the RACE method, using the 5'-3' RACE kit (Roche Molecular Biochemicals) was applied. Reverse transcriptase of chick aorta poly(A)+ RNA was performed with the use of exanucleotides, and the product was subjected to PCR using the EMILIN specific sense primer 5'-GGAGCCGCTCACCATCTTCAGCGGGGCCC-3' in combination with the anchor-poly(dT) from the kit.

Production of Recombinant Prokaryotic gC1q-like Domain of Human EMILIN—The EMILIN gC1q-like domain was amplified by PCR from Clone K1, ligated in-frame in the 6-His-tagged pQE-30 expression vector (Qiagen GmbH), and grown in M15 cells. To check for errors generated by PCR, all the cloned fragments were sequenced in both directions. M15 cells were 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 NaCl) at 2–5 volumes/g of wet weight. The sample was 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 EMILIN C1q-like domain was performed by affinity chromatography on nickel nitrilotriacetic acid resin

² R. Doliana, A. Canton, F. Buccioti, M. Mongiat, P. Bonaldo, and A. Colombatti, manuscript in preparation.

(Qiagen GmbH) under native conditions (sonication buffer). The recombinant protein was eluted from the affinity column in sonication buffer, pH 6.0, containing 10% glycerol and 0.2 M imidazole. After dialysis against cold phosphate-buffered saline, the C1q-like domain was used for CD spectra analysis and cell adhesion assays.

Circular Dichroism Spectroscopy—The purified polyhistidine EMILIN gC1q-like peptide was used for CD spectroscopy analysis. A Jasco J-600 CD/ORD spectrophotometer interfaced to an Olidata computer for data collection was used for all the measurements. Calibration of the instruments was performed with D(+)-10-camphorsulfonic acid at 290 nm. Standard conditions were 25 mM Na₃PO₄, 150 mM NaCl, pH 6.0, 10 °C, using a 0.2-cm path-length cuvette. The temperature was controlled by a water bath. UV circular dichroism spectra are presented as millidegrees of ellipticity. The reported results are the smoothed average over 10 measurements. To calculate the secondary structural content, data were transformed in terms of mean residue molecular ellipticity (θ) ($\text{deg} \times \text{cm}^2 \times \text{dmol}^{-1}$), based on a mean residue weight of 104.3. Then, the spectra were analyzed using the Menendez-Arias program (27).

Cell Adhesion—The capability of isolated recombinant human EMILIN gC1q-like domain to support cell attachment was evaluated using several cell lines and the centrifugal assay for fluorescence-based cell adhesion (CAFCA) (28, 29). Briefly, specifically devised six-well strips of flexible polyvinyl chloride (CAFCA miniplates; TECAN Polyfiltronics, Inc., Boston, MA) covered with double-sided tape (bottom miniplates) were coated overnight with recombinant EMILIN gC1q-like fragment and subsequently incubated with 1% heat-denatured bovine serum albumin for 2–4 h at room temperature to block uncovered areas of the plastic. Two human smooth muscle (SK-UT-1 and SK-LMS-1 leiomyosarcomas) and a human fibroblastic (HT-1080 fibrosarcoma) cell lines were fluorescently labeled by incubation with the vital fluorochrome calcein AM (Molecular Probes Europe BV, Leiden, The Netherlands), rinsed extensively in Ca²⁺- and Mg²⁺-free phosphate-buffered saline, and then aliquoted into the bottom CAFCA miniplates at a density of $1\text{--}3 \times 10^4$ cells/well. Cell adhesion to substrates was assayed in phosphate-buffered saline containing 0.1% bovine serum albumin, 1 mM MgCl₂ and CaCl₂, and 2% India ink as a fluorescence quencher. CAFCA miniplates were placed on specifically devised hard plastic holders (TECAN Polyfiltronics, Inc.) and centrifuged at $142 \times g$ for 5 min at 37 °C to synchronize contact of the cells with the substrate. The miniplates were incubated for 30 min at 37 °C and then mounted together with a similar CAFCA miniplate lacking double-sided tape (top miniplate) such as to create communicating chambers to be reverse-centrifuged. The relative number of cells bound to the substrate (*i.e.* remaining bound to the bottom miniplates) and unbound cells (in wells of the top miniplates) was estimated by top/bottom fluorescence detection in a computer-interfaced SPECTRAFluor microplate fluorometer (TECAN Polyfiltronics, Inc.). Fluorescence values were analyzed by custom CAFCA software (TECAN Polyfiltronics, Inc.) to determine the percentage adherent cells of the total cell population analyzed according to a previously published formula (28, 29).

RESULTS

Purification and Peptide Sequences of Chick EMILIN—A chick aorta EMILIN preparation obtained following differential extraction procedures, DEAE ion-exchange, and size exclusion chromatography was analyzed by SDS-polyacrylamide gel electrophoresis under reducing conditions and found to be composed of a single major band with an estimated M_r of about 115,000. This finding was in accord with previous results (18), and the few low M_r contaminants represented less than 10% of the stained material as judged by quantitative scanning of the gel (data not shown). The protein was cleaved by trypsin and by a lysine-specific protease, the peptides were separated by reverse phase chromatography on a Nucleosil-120 C18 column, and several components were selected for sequencing (Fig. 1).

Cloning Strategy—Based on the peptide sequences, several degenerate oligonucleotides in the sense and antisense orientations were synthesized and used to amplify fragments from poly(A)⁺ RNA of chick aortas. Several amplification products of different lengths were obtained using the various combinations of oligonucleotides. The sequence of the cloned products revealed that all but one contained stop codons in all the possible frames. On the contrary, the 384-bp-long amplification

Peptide Number	PEPTIDE SEQUENCE
#1	V L L N D G G A Y D P E T G T F
#3	A D F Q P A X A T F
#4	L A P P A P P D E L L R <u>ccicccgncicccgavga</u>
#5	L A H H P D
#6	P V G E X G S V P H I A F S A A <u>ccicayathgcittywsigcngc</u>
#7	H V A G L Q M A V
#8	E P A D L G D E E E P <u>gayytigggigaygargargarc</u>
#9	A X A P D E P L T V F S A <u>gciccgaygarcciytiacngt</u>
#10	E A G G E G V P G G V

FIG. 1. Sequences of chick EMILIN peptides. Peptides were generated as described under "Experimental Procedures." Amino acid sequences are indicated in 1-letter code, with X representing unidentified residues; the sequences of degenerate and inosine-containing oligonucleotides derived from peptides 4, 6, 8, and 9 are in *small letters* and *underlined*; y, c/t; h, a/c/t; w, t/a; n, a/g/c/t; r, a/g; s, g/c; i, inosine.

product obtained with the use of the primers corresponding to peptides 9 (sense orientation) and 6 (antisense orientation) encompassed an open reading frame (clone D1) that contained peptide 1, thus suggesting that the amplified sequence corresponded to that of genuine EMILIN protein (Fig. 1).

After the initial successful search of the GenBank™ for EST-containing homologous sequences to those of clone D1 of chick EMILIN, five partly overlapping cDNA clones were sequentially isolated from λ gt10 human kidney and aorta libraries using a human specific PCR-derived probe based on the EST entries and probes derived from subclones at the 5' end of each successively isolated clone (Fig. 2). Remarkably, walking toward the 5' end of the human cDNA was hampered by the surprisingly high number of fused clones isolated from both libraries.

Nucleotide and Predicted Amino Acid Sequences—The partial chick cDNA corresponds to an open reading frame of 128 amino acids (data not shown). The coding sequence and the deduced amino acid sequence of the human EMILIN cDNA (GenBank™ accession number AF 088916) is shown in Fig. 3. Several partially overlapping EST entries (HSAA23367, HSAA1823, HS1269888) showed an almost perfect match with the 3' end of the human EMILIN transcript, including the last 465 bp of the coding sequence and the entire 3'-untranslated region. However, numerous substitutions and gaps have also been detected: C2772-GAP; C2823G; C2826G; G2827T; C2857A; C2877T; C2883-GAP; G2895-GAP. Independent sequencing of BAC clones confirmed the present sequence.²

The open reading frame of the human EMILIN begins with a Met codon whose surrounding sequences fit into the eukaryotic translation start sites (30) and is preceded by an in-frame stop codon (data not shown). The human cDNA spans about 3400 bp and has an open reading frame of 1016 amino acids. The predictions with the highest probabilities for the initial residue of the mature protein are between position 2 (Ser, Y value of 0.508) and 3 (Tyr, Y value of 0.580) of the present sequence. Although the best prediction for the cleavage site of the signal

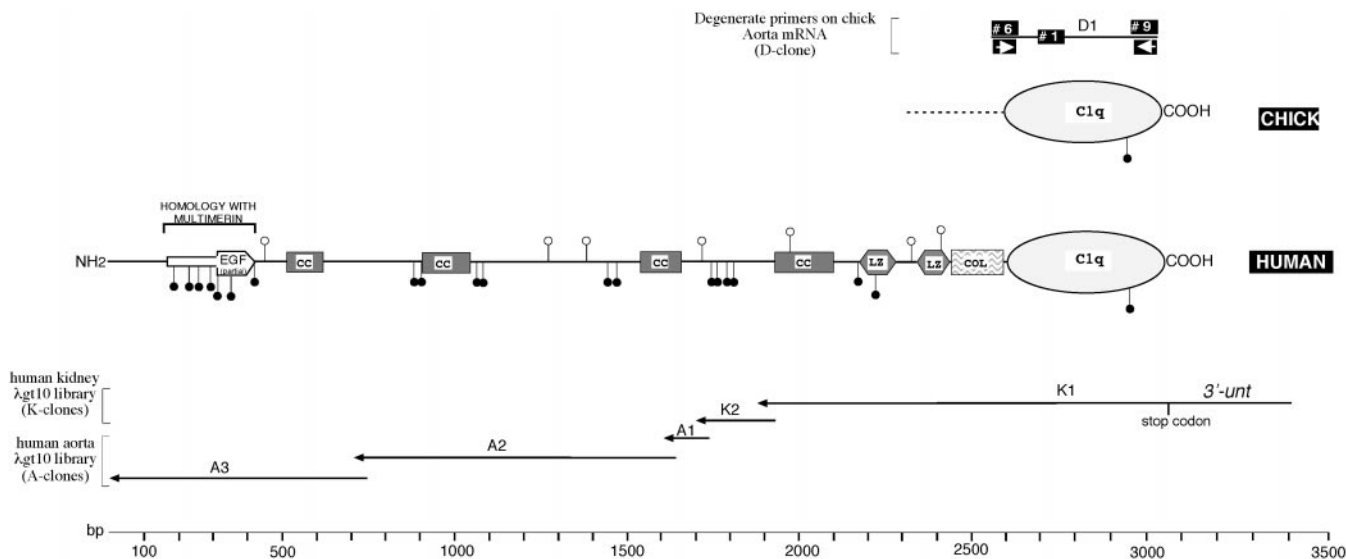


FIG. 2. A schematic diagram with the cloning strategy and the domain structure of chick and human EMILIN. The various types of clones (solid lines) used in determining the sequence are shown above (chick) and below (human) the respective diagrams. Each clone is identified by a capital letter followed by a number; letters refer to the tissue of origin (A, aorta; K, kidney) or the technique used (D, clones obtained by amplification with degenerate primers). 3'-unt indicates the 3'-untranslated region. The different domains are designated according to Bork and Koonin (54). EGF, epidermal growth factor; CC, coiled-coil; LZ, leucine zipper; COL, collagenous domain; C1q, gC1q-like domain. Cysteines and potential glycosylation sites are indicated by closed and open circles, respectively. The positions of peptides derived from the chick molecule and the corresponding degenerate primers are indicated by numbers and arrows, respectively, in reverse type.

peptide is at position 3, alignment between the N-terminal peptide sequence of the chick mature EMILIN and the deduced N terminus of the human EMILIN (see below) indicates the position 1 (Ala) as the most likely candidate for the initiation of the mature protein. Therefore, residues -21/-1 correspond most likely to a signal peptide, because the sequence agrees very well with the classical consensus sequence (31) and ends with a consensus signal cleavage site (32). Thus, the calculated molecular mass for the mature protein is 104.5 kDa. The human EMILIN contains 7 potential N-glycosylation sites and 20 cysteines with a number of them clustered as doublets, separated by none, one, or two residues that could be involved in intramolecular disulfide bonding. By applying the 3'-RACE, 383 additional bp corresponding to the 3'-untranslated sequence could be obtained. The observation that the most 3' end EST entry terminates 37 bp before the end of the present RACE product, together with a potential polyadenylation sequence 16 bp before the end of our sequence, strongly suggests that our RACE product includes the entire 3'-untranslated sequence.

Comparison between Chick and Human Sequences—The alignment of the 128-residue-long stretch of chick EMILIN and the C-terminal region of human EMILIN showed a high degree of homology (Fig. 4, panel C), the overall degree of amino acid sequence identity and identities plus similarities in this domain was calculated to be about 76 and 88%, respectively; furthermore, a partial sequence of the N terminus of chick EMILIN compares very well (9 residues are identical, and 2 are similar) with the deduced N terminus of the human EMILIN (Fig. 4, panel A). The high level of identities and similarities between chick and human EMILIN at both C and N termini indicates that the two sequences identify the same protein in these species.

Domain Structure—The predicted domain structure of human EMILIN is composed, starting from the C-terminus, of a globular domain (gC1q-like), an uninterrupted stretch of 17 Gly-Xaa-Yaa triplets, indicating that EMILIN possesses a straight collagen stalk, and a 641-long amino acid sequence in which there are several heptad repeats separated by unrelated sequences with the potential for forming coiled-coil α -helices (see below). A 91-bp-long residue sequence that includes two

sequences corresponding to structures referred to as the "leucine zippers," which are typical of several gene regulatory proteins (33, 34), is located between the coiled-coil region and the collagenous stretch. This finding is rather unusual especially for an ECM protein, as there are only a few precedents in the literature for the presence of leucine zippers in the extra nuclear compartments: the *Drosophila* ECM protein pollux (35) and the cytoplasmic protein dystrophin, whose leucine zipper has been reported to interact with troponin (36). Although the leucine zipper pattern is far from being specific, it is not clear at the moment what might be the significance of its presence in the context of EMILIN. Finally, a search of the NCBI data bank indicated the existence of a region of homology between the N-terminal end of EMILIN and the platelet and endothelial specific protein, multimerin (24), spanning amino acids 33 and 108 (Fig. 4, panel B). This region contains several identical residues, including three conserved cysteines and a partial epidermal growth factor-like consensus sequence.

Coiled-coil Prediction—The 3-4-3-4 spacing of hydrophobic residues predicts that the region of human EMILIN spanning residues 152 and 792 (Figs. 2 and 3) will form an α -helical coiled-coil. When analyzed by different algorithms (37, 38), the probability for coiled-coil formation differed among the various heptad sequences. Using the PairCoil program (38), the highest probability score (around 1.0) was located at the level of the first heptad repeat (Fig. 5), and similar results were obtained when applying the Multicoil program (data not shown); furthermore, at least two other heptad repeats (positions 497-539 and 615-684) displayed a high probability of coiled-coil structures, and a fourth one had intermediate values.

Alignment of Sequences at the C-terminal Domain—The C terminus of EMILIN exhibits a striking homology to a gC1q-like domain of a number of proteins including the A, B, and C chains of human and mouse complement C1q protein (39, 40), the α 1 and α 2 chains of type VIII (41, 42), and the α 1 chain of type X (43) collagens, precerebellin (44), multimerin (24), ACRP-30/AdipoQ (45, 46), the HP-27 protein from Siberian chipmunks (47), and a sunfish saccular collagen (48). The length of this domain is included between 131 (gC1q-C) and 151 (EMILIN) residues, and the protein sequence comparison of the

prototype adhesive molecule, was assessed using the smooth muscle cell lines SK-UT-1 and SK-LMS-1 and the fibroblast cell line HT1080. The EMILIN gC1q-like domain was cloned in the pQE30 expression vector, and the protein was purified under native conditions and coated at different concentrations. The native state of the EMILIN gC1q-like domain was examined by CD spectra analysis. The peptide (Fig. 7) exhibited a classic β -sheet CD spectrum with a minimum at 217 nm. The evaluation of the content in β -sheet conformation for the peptide using the Menendez-Arias program (27) gives a value higher than 70%, therefore confirming that also the EMILIN gC1q-like domain fits very well with the structure determined for ACRP-30/AdipoQ (50). Cell adhesion to fibronectin was high and comparable for all cell lines with a cell binding of about 90%, the EMILIN gC1q-like domain promoted a high level of adhesion of SK-UT-1 cells, whereas HT1080 were negative, and SK-LMS-1 bound with a low percentage (Fig. 8A). However, although adhesion to fibronectin induced a strong cell flattening, cell adhesion to the EMILIN gC1q-like domain was not

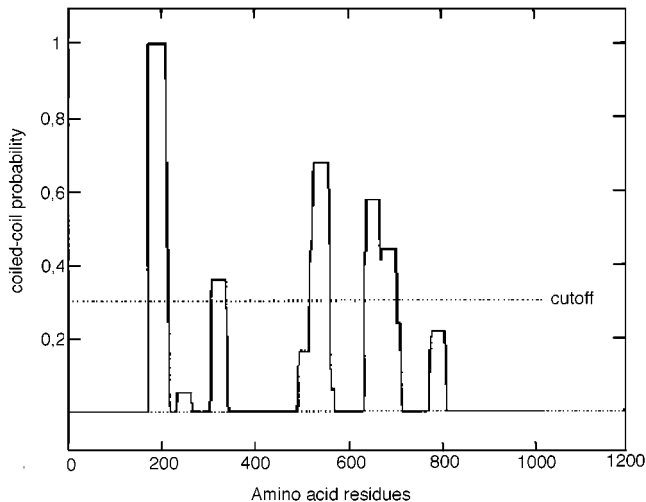
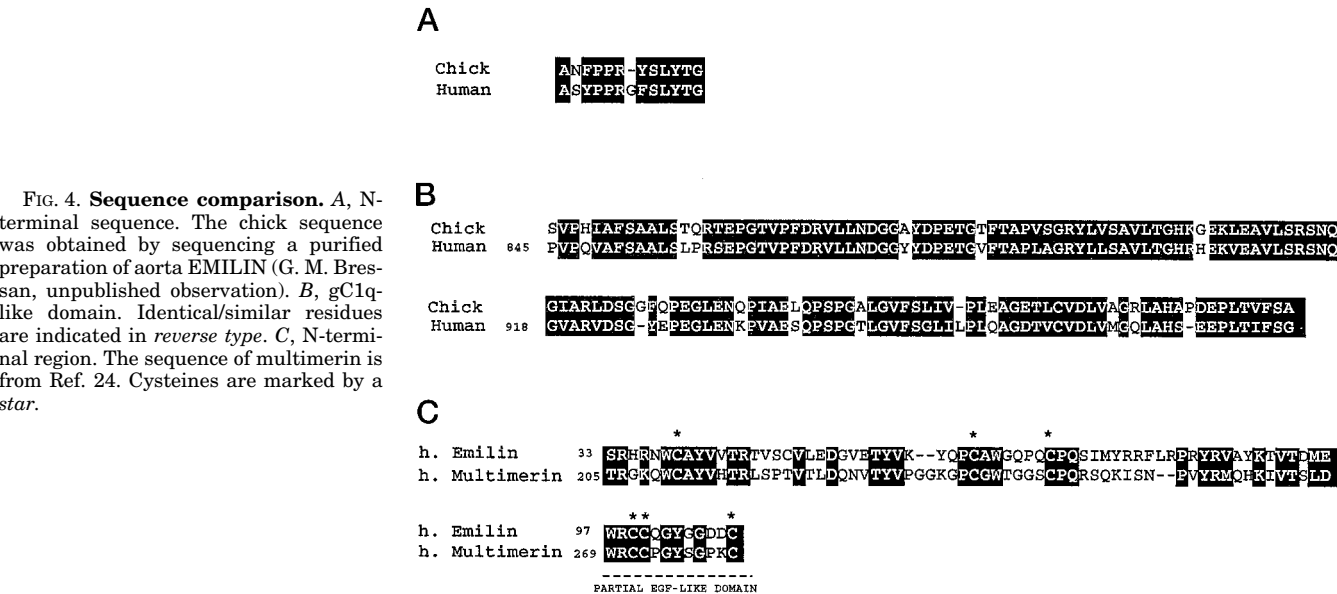


FIG. 5. The probability of coiled-coil α -helical structures. Several highly probable coiled-coil regions are identified using the PairCoil program (37) between residues 152 and 684.

followed by a significant level of cell spreading (Fig. 8B). The SK-UT-1 cells were for the most part round with small blebs or short projections with only a very low percentage of cells displaying a flat morphology. These results suggest either that cell adhesion to the EMILIN gC1q-like domain uses different mechanisms/receptors than adhesion to fibronectin and is not followed by cytoskeletal rearrangements or, less likely, that the standard cell adhesion conditions of time, temperature, and medium composition used to assay adhesion to fibronectin are not appropriate to measure cell adhesion to the C1q-like domain of EMILIN.

DISCUSSION

The results provided in this report concern a new human cDNA whose major structural elements were a gC1q-like C-terminal domain, a short uninterrupted collagenous domain, and an extended domain containing sequences with the potential of forming amphipathic coiled-coil α -helices. The determination of the primary structure through cDNA cloning and the assignment of this novel sequence to human EMILIN was made possible by the very close identity at the C terminus between the chick EMILIN sequence and that of the corresponding human cDNA. In fact, the finding that the deduced

sequence of clone D1 of chick EMILIN contained peptide 1 of the tissue-purified EMILIN confirmed that the sequence amplified from chick aorta mRNA corresponded to that of the genuine EMILIN described by us (18–23) as an elastin-associated protein with undefined function. Thus, the sequence similarity at the gC1q-like domain, the near identity between the N-terminal residues of the mature chick protein, and the deduced amino acids of the presumed mature human protein support the conclusion that we are dealing with the same ECM constituent in the two species.

Evidence that chick EMILIN stained strongly with PAS (18) and that in biosynthetic studies a treatment with tunicamycin reduced the apparent molecular mass of about 20–25 kDa (23) indicated that chick EMILIN was highly glycosylated. The present identification of seven potential N-glycosylation sites in the human EMILIN sequence is in accord with the previous experimental data using the chick system (18, 23). Similarly, the presence of 20 cysteine residues with a high potential for intermolecular S-S bonding is also in accord with the finding that newly synthesized and secreted chick EMILIN migrated as a monomer in SDS gels under reduced conditions but was present as a large aggregate that did not enter the gel in the absence of reducing agents (23).

The domain organization of EMILIN is unique; it bears features shared with several other members of the C1q/TNF superfamily (Fig. 9), i.e. C1q (A, B, C), collagens VIII, X, saccular collagen, ACRP-30/AdipoQ, and HP-27, such as the gC1q-like domain and a collagenous domain, but also EMILIN displays an extended discontinuous and potentially coiled-coil region that is absent in all the other members of the C1q/TNF superfamily, except multimerin, a large soluble glycoprotein found in platelets α -granules and endothelial Weibel-Palade bodies. Multimerin forms disulfide-linked homomultimers of variable sizes (55) and interacts with factor V, which is stored complexed with multimerin in the α -granules (56). There is experimental evidence that several members of the C1q/TNF superfamily trimerize to form either heterotrimeric collagen triple helices that are expressed as soluble plasma proteins or type II membrane-bound molecules such as in C1q (A, B, C) (57, 58) or to form homotrimers as in collagen X (59) and ACRP-30/AdipoQ (46). EMILIN is also likely to form similar trimers; it possesses a gC1q-like domain, which is highly homologous to those of the other members of the superfamily and an uninterrupted collagenous domain, which can form a collagen-like stalk region.

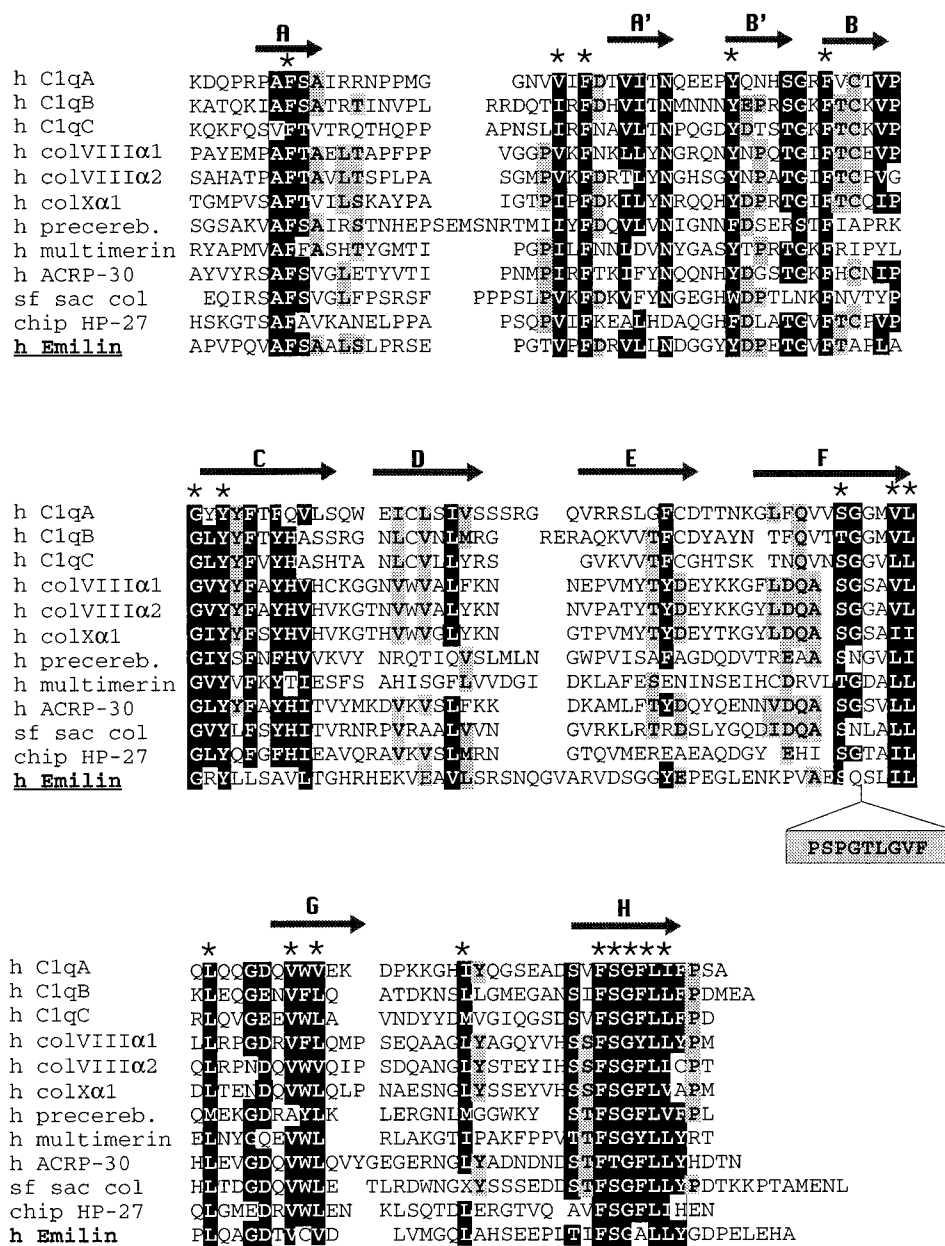


FIG. 6. Sequence alignment of the gC1q-like domains. Groups of similar residues are defined as follows: tiny, S, G, A; aliphatic, I, L, V, M; aromatic, H, F, Y, W; positive, R, H, K; negative, D, E; hydroxyl, S, T; amide, N, Q. Identical or similar residues shared by 50% or more of the sequences are shaded in gray, and those shared by more than 75% of the sequences are in reverse type. An asterisk above the sequences marks residue conservation in all the sequences. The locations of the β strands according to the crystal of the gC1q-like domain of ACRP-30/AdipoQ (52) are indicated above the sequence by arrows and capital letters. An insertion of 10 residues unique to EMILIN is boxed in gray. The sequences of C1qA, C1qB, C1qC, $\alpha 1$ (VIII), $\alpha 2$ (VIII), $\alpha 1$ (X), precerebellin (cerebellin), multimerin, ACRP-30/AdipoQ, sunfish saccular collagen (*sf sac coll*), and HP-27 (chipmunk (*chip*) HP-27) are from accession numbers P02745, K03430, P02747, X57527, M60832, X65120, P23435 and P02682, U27109, D45371, P38085, A41752, respectively. *h*, human; A-H indicate the locations of the β strands.

Curiously, among the members of the family, only EMILIN, ACRP-30/AdipoQ, and Hib27 possess an uninterrupted collagenous domain. The EMILIN gC1q-like domain, when compared with the other gC1q-like domains, has a much longer F β -strand because of a 10-residue insertion. However, the residues conserved throughout both the C1q and TNF families of proteins and important in the packing of the hydrophobic core of the individual monomer (50) are present in EMILIN gC1q-like domain in the same relative positions, and this appears sufficient to predict a similar trimeric and spatial organization also for the domain of EMILIN. The potential structural homology between the TNF family of growth factors and the gC1q-like domains of the C1q family of proteins suggested that these diverse members might derive from ancestral elements

with close functional activity (50). Likely targets for the proteins containing C1q/TNF domains are cell surface receptors; these are well studied in TNF, but initial data are also available for the C1q complement component (51–53). In fact, several cell types are endowed with the capability to attach to the C1q complement component via cell surface binding sites; two types of structures have been described, a binding protein that recognizes the collagenous domain (60) and another component that binds to the gC1q domain (51). However, more recently the effective nature of this second type of binding molecule has been disputed (61), and further studies are necessary. The finding that EMILIN gC1q-like domain displayed cell pro-adhesive capacity for some smooth muscle cells but seemed to be much less reactive for fibroblastic cells is consistent with the

above evidence and suggests that cell recognition of this domain might be exerted via specific cell surface receptors. The adhesion was high for SK-UT-1 cells but, within the time frame of the cell adhesion assay, was not followed by a consistent spreading. Thus, it is possible that "receptors" distinct from classical integrins such as those recognized by fibronectin are involved here. Neither the physiological significance of the observed adhesion is clear yet nor whether this adhesion plays a primary or an auxiliary role. Close contacts between amorphous elastin and smooth muscle cells in the aorta of 16-day embryos have been reported (62), and the ultrastructural localization of EMILIN (19) does not exclude that the interaction between the elastin amorphous core and the cells could also take place via an EMILIN intermediate. As EMILIN was detected in early stages of aorta development, in association with

a network of thin fibrils likely representing maturing microfibrils (19), EMILIN deposition can be considered an early event in elastogenesis, and this conclusion is reinforced by the observation that the process of elastic fiber formation *in vitro* was greatly affected by the addition of anti EMILIN antibodies (19). Whether the process of elastin deposition and elastic fiber formation is regulated through cell adhesion via the EMILIN gC1q-like domain remains to be seen.

EMILIN, like multimerin (55), is heavily disulfide-linked and thus can be found as large aggregates in the culture medium of aorta smooth muscle cells (23). The possibility to form coiled-coil α -helices could further amplify its potential to asso-

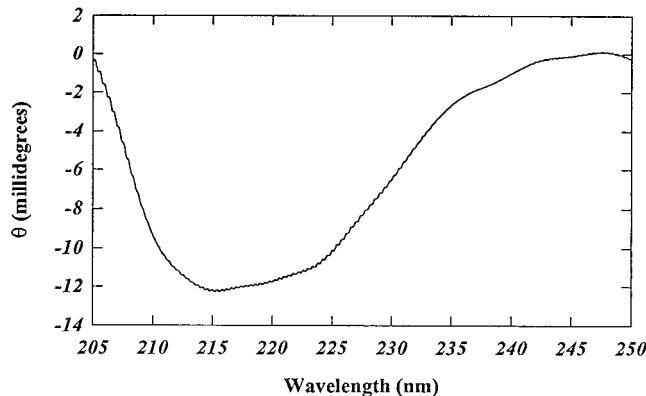


FIG. 7. Circular dichroism spectrum of the C1q-like domain. The spectra were recorded at a protein concentration of 100 μ M in 25 mM Na_3PO_4 , 150 mM NaCl, pH 6.0, at 10 $^\circ\text{C}$ and 0.2-cm path-length. The β -sheet content was estimated from the mean residue molecular ellipticity (θ) ($\text{deg} \times \text{cm}^2 \times \text{dmol}^{-1}$), based on a mean residue weight of 104.3 and analyzed using the Menendez-Arias program (27).

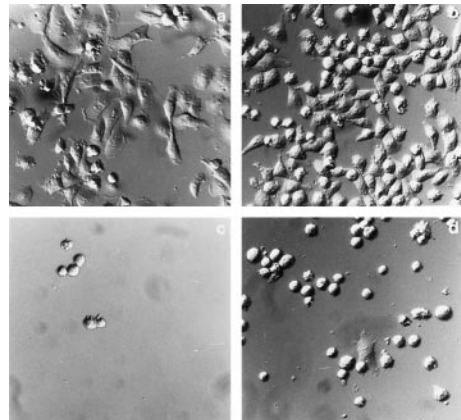


FIG. 8. Cell adhesion of SK-UT-1 and SK-LMS-1 smooth muscle and HT1080 fibroblastic cell lines. Cell adhesion to intact human fibronectin (a and b) and to the recombinant human EMILIN gC1q-like domain (c and d) purified under native conditions. The ligand proteins were plated at 10 μ g/ml, and the cell adhesion assay was carried out in phosphate-buffered saline containing 1 mM Ca^{2+} and 1 mM Mg^{2+} . Adhesion to wells coated with denatured bovine serum albumin alone was negligible. Nomarski optics microscopy images of HT-1080 (left) and SK-UT-1 (right) cells are shown.

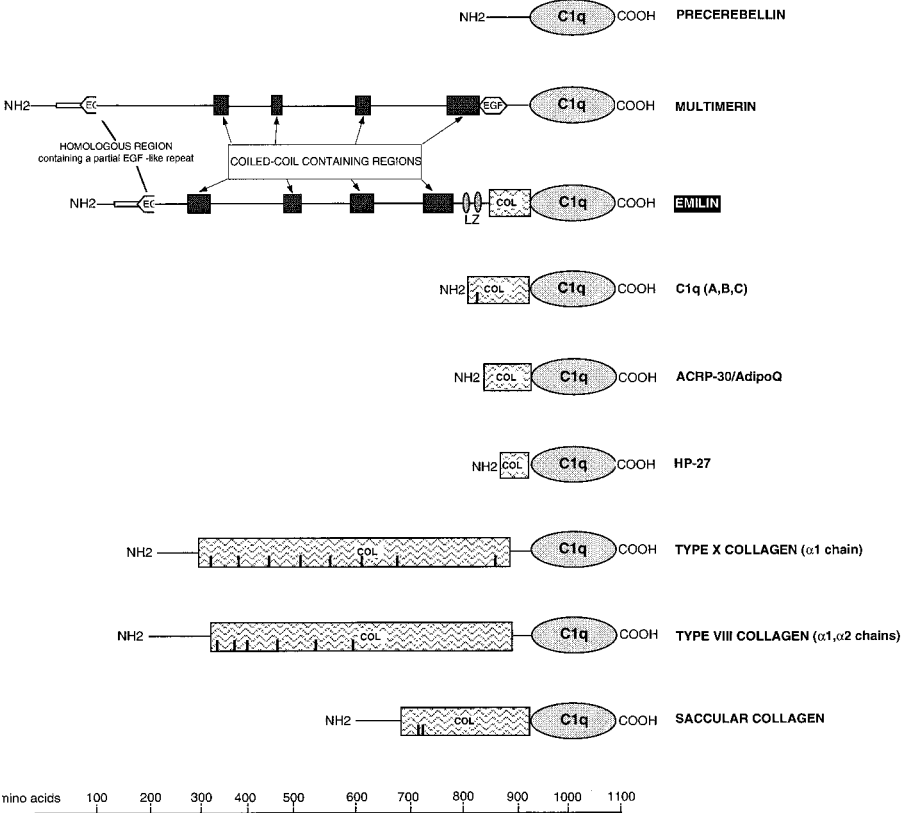


FIG. 9. The C1q/TNF superfamily of proteins. The different domains are designated as in Fig. 2. Bars within the collagenous domains (COL) indicate short interruptions or imperfections in the Gly-Xaa-Yaa sequence. The order in which the different members are depicted highlights that EMILIN, in addition to the C1q-like domain common to all the family members, shears a short collagenous domain with only a few of the members and the coiled-coil-containing region only with multimerin. EGF, epidermal growth factor.

ciate into even larger aggregates. In fact, one of the heptad repeat sequences has a probability to form α -helices near 1.0, and in two other regions the probability is above 0.6. Although formal proof that these heptad repeats can form trimers is still lacking, the chances are high given that the presumed trimerization process can initiate from the C1q-like domain at the C terminus, proceeding then through the collagenous domain next to it as in collagen X (59) and ACRP-30/AdipoQ (46). Further studies are required to define how EMILIN subunits are assembled into the large disulfide-linked multimers, *i.e.* whether the EMILIN gC1q-like domain is a likely site for initial interchain association and whether the heptad repeats associate only intramolecularly into trimers or can also associate intermolecularly, *i.e.* with other EMILIN trimers. To investigate these possibilities, the preparation of full-length, truncated, and point-mutated EMILIN recombinant molecules in eukaryotic cells is in progress.³

Acknowledgments—We thank Maria Teresa Mucignat and Gabriella Mungiguerra for technical assistance, Dr. Paola Spessotto for her help in performing the cell adhesion assays, and Dr. Gianluca Tell for performing the CD spectra analysis.

REFERENCES

- Partridge, S. M. (1962) *Adv. Protein Chem.* **17**, 227–302
- Greenlee, T. K., Jr., Ross, R., and Hartman, J. L. (1966) *J. Cell Biol.* **30**, 59–71
- Cleary, E. G. (1987) in *Connective Tissue Disease* (Uitto, J., and Parejda, A. J., eds) pp. 55–81, Marcel Dekker, Inc., New York
- Rosenbloom, J., Abrams, W. R., and Mecham, R. (1993) *FASEB J.* **7**, 1208–1218
- Sakay, L. Y., Keene, D. R., and Engvall, E. (1986) *J. Cell Biol.* **103**, 2499–2509
- 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
- Henderson, M., Polewski, R., Fanning, J. C., and Gibson, M. A. (1996) *J. Histochem. Cytochem.* **44**, 1389–1397
- 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
- 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
- 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
- Kanzaki, T., Olofsson, A., Moren, A., Werntedt, C., Hellman, U. K., Claesson-Welsh, L., and Heldin, C. H. (1990) *Cell* **6**, 1051–1061
- Gibson, M. A., Hatzinikolas, G., Davis, E. C., Baker, E., Sutherland, G. R., and Mecham, R. P. (1995) *Mol. Cell. Biol.* **15**, 6932–6942
- Yin, W., Smiley, E., Germiller, J., Mecham, R. P., Florer, J. R., Wenstrup, R. J., and Bonadio, J. (1995) *J. Biol. Chem.* **270**, 10147–10160
- Saharinen, J., Taipale, J., Monni, O., and Keski-Oja, J. (1998) *J. Biol. Chem.* **273**, 18459–18469
- 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
- Hatzinikolas, G., and Gibson, M. A. (1998) *J. Biol. Chem.* **273**, 29309–29314
- Cleary, E. G., and Gibson, M. A. (1996) in *The Structure and Function of Extracellular Matrix* (Comper, W. D., ed) Vol. 2, pp. 95–140, Gordon and Breach Science Publisher, New York
- Bressan, G. M., Castellani, I., Colombatti, A., and Volpin, D. (1983) *J. Biol. Chem.* **258**, 13262–13267
- Bressan, G. M., Daga-Gordini, D., Colombatti, A., Castellani, I., Marigo, V., and Volpin, D. (1993) *J. Cell Biol.* **121**, 201–212
- Colombatti, A., Bressan, G. M., Castellani, I., and Volpin, D. (1985) *J. Cell Biol.* **100**, 18–26
- Colombatti, A., Bressan, G. M., Volpin, D., and Castellani, I. (1985b) *Collagen Relat. Res.* **5**, 181–191
- Colombatti, A., Poletti, A., Bressan, G. M., Carbone, A., and Volpin, D. (1987) *Collagen Relat. Res.* **7**, 259–275
- Colombatti, A., Bonaldo, P., Volpin, D., and Bressan, G. M. (1988) *J. Biol. Chem.* **263**, 17534–17540
- Hayward, C. P. M., Hassell, J. A., Denomme, G. A., Rachubinski, R. A., Brown, C., and Kelton, J. G. (1995) *J. Biol. Chem.* **270**, 18246–18251
- Pearson, W. R., and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 2444–2448
- Altschul, S. F., Gish, W., Miller, M., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410
- Menendez-Arias, L., Gomez-Gutierrez, J., Garcia-Fernandez, M., Garcia-Tejedor, A., and Moran, F. (1989) *Comput. Appl. Biosci.* **4**, 479–482
- Giacomello, E., Neumayer, J., Colombatti, A., and Perris, R., (1999) *Biotechniques*, in press
- Spessotto, P., Giacomello, E., and Perris, R. (1999) *Methods in Molecular Biology: Extracellular Matrix*, Humana Press Inc., Totowa, NJ, in press
- Kozak, M. (1981) *Nucleic Acids Res.* **9**, 5233–5252
- Nielsen, H., Engelbrecht, J., Brunak, S., and van Heijne, G. (1997) *Protein Eng.* **10**, 1–6
- Perlman, D., and Halvorson, H. O. (1983) *J. Mol. Biol.* **167**, 391–409
- Landshultz, W. H., Johnson, P. F., and McKnight, S. L. (1988) *Science* **240**, 1759–1764
- Bush, S. J., and Sassone-Corsi, P. (1990) *Trends Genet.* **6**, 36–40
- Zhang, S.-D., Kassis, J., Olde, B., Mellerick, D. M., and Odenwald, W. F. (1996) *Genes Dev.* **10**, 1108–1119
- Pearlman, J. A., Powaser, P. A., Elledge, S. J., and Caskey, C. T. (1994) *FEBS Lett.* **354**, 183–186
- Lupas, A. (1996) *Methods Enzymol.* **266**, 513–525
- 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
- Sellar, G. C., Blake, D. J., and Reid, K. B. M. (1991) *Biochem. J.* **274**, 481–490
- Reid, K. B. M. (1985) *Biochem. J.* **231**, 729–735
- Muragaki, Y., Mattel, M.-G., Yamaguchi, N., Olsen, B. R., and Ninomiya, Y. (1991) *Eur. J. Biochem.* **197**, 615–622
- Muragaki, Y., Jacenko, O., Apte, S., Mattel, M.-G., Ninomiya, Y., and Olsen, B. R. (1991) *J. Biol. Chem.* **266**, 7721–7727
- Reichenberger, E., Beler, F., LuValle, P., Olsen, B. R., von der Mark, K., and Bertling, W. M. (1992) *FEBS Lett.* **311**, 305–310
- Urade, Y., Oberdick, J., Molinar-Rode, R., and Morgan, J. I. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 1069–1073
- Scherer, P. E., Williams, S., Fogliano, M., Baldini, G., and Lodish, H. F. (1995) *J. Biol. Chem.* **270**, 26746–26749
- Hu, E., Liang, P., and Spiegelman, B. M. (1996) *J. Biol. Chem.* **271**, 10697–10703
- Takamatsu, N., Ohba, K., Kondo, J., Kondo, N., and Shiba, T. (1993) *Mol. Cell. Biol.* **13**, 1516–1521
- Davis, J. G., Burns, F. R., Navaratnam, D., Lee, A. M., Ichimiya, S., Oberholtzer, J. C., and Greene, M. I. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 707–712
- Smith, K. F., Haris, P. I., Chapman, D., Reid, K. B. M., and Perkins, S. J. (1994) *Biochem. J.* **301**, 249–256
- Shapiro, L., and Scherer, P. E. (1998) *Curr. Biol.* **8**, 335–338
- Ghebrehwet, B., Boon-Leong, L., Peerschke, E. I., Willis, A. C., and Reid, K. B. M. (1994) *J. Exp. Med.* **179**, 1809–1821
- Peerschke, E. I., Reid, K. B. M., and Ghebrehwet, B. (1994) *J. Immunol.* **152**, 5896–5901
- Peerschke, E. I., Smyth, S. S., Teng, E. I., Dalzell, M., and Ghebrehwet, B. (1996) *J. Immunol.* **157**, 4154–4158
- Bork, P., and Koonin, E. V. (1996) *Curr. Opin. Struct. Biol.* **6**, 366–375
- Hayward, C. P. M., Warkentin, T. E., Horsewood, P., and Kelton, J. G. (1991) *Blood* **77**, 2556–2560
- Hayward, C. P. M., Furmaniak-Kazmierczak, E., Cieutat, A.-M., Moore, J. C., Bainton, D. F., Nesheim, M. E., Kelton, J. G., and Côté, G. (1995) *J. Biol. Chem.* **270**, 19217–19224
- Reid, K. B. (1989) *Behring Inst. Mitt.* **84**, 8–19
- Kaul, M., and Loos, M. (1995) *J. Immunol.* **155**, 5795–5802
- Frischoltz, 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
- Peerschke, E. I., Malhotra, R., Ghebrehwet, B., Reid, K. B. M., Willis, A. C., and Sim, R. B. (1993) *J. Leukocyte Biol.* **53**, 179–184
- Dedio, J., Jähnen-Dechent, W., Bachmann, M., and Müller-Esterl, W. (1998) *J. Immunol.* **160**, 3534–3542
- Daga-Gordini, D., Bressan, G. M., Castellani, I., and Volpin, D. (1987) *Histochem. J.* **19**, 623–632

³ M. Mongiat, R. Doliana, and A. Colombatti, unpublished observations.