Structure, Chromosomal Localization, and Promoter Analysis of the Human Elastin Microfibril Interface Located protein (EMILIN) Gene *

(Received for publication, August 4, 1999)

Roberto Doliana†, Anna Canton‡, Francesco Bucciotti‡, Maurizio Mongiat‡, Paolo Bonaldo§, and Alfonso Colombatti†‡¶

From the †Divisione di Oncologia Sperimentale 2, Centro di Riferimento Oncologico, 33081 Aviano, Italy, the ‡Istituto di Istologia ed Embriologia, Università di Padova, 35100 Padova, Italy, and the §Dipartimento di Scienze e Tecnologie Biomediche, Università di Udine, 33100 Udine, Italy

Elastin microfibril interface-located protein (EMILIN) is an extracellular matrix glycoprotein abundantly expressed in elastin-rich tissues such as the blood vessels, skin, heart, and lung. It occurs with elastic fibers at the interface between amorphous elastin and microfibrils. In vitro experiments suggested a role for EMILIN in the process of elastin deposition. This multimodular protein consists of 995 amino acids; the domain organization includes a C1q-like globular domain at the C terminal, a short collagenous stalk, a region containing two leucine zippers, and at least four heptad repeats with a high potential for forming coiled-coil α-helices and, at the N terminus, a cysteine-rich sequence characterized by a partial epidermal growth factor-like motif and homologous to a region of multimerin. Here we report the complete characterization of the human and murine EMILIN gene, their chromosomal assignment, and preliminary functional data of the human promoter. A cDNA probe corresponding to the C terminal of EMILIN was used to isolate two genomic clones from a human BAC library. Sequencing of several derived subclones allowed the characterization of the whole gene that was found to be about 8 kilobases in size and to contain 8 exons and 7 introns. The internal exons range in size from 17 base pairs to 1929 base pairs. All internal intron/exon junctions are defined by canonical splice donor and acceptor sites, and the different domains potentially involved in the formation of a coiled-coil structure are clustered in the largest exon. The 5′-end of the EMILIN gene overlaps with the 5′-end of the promoter region of the ketohexokinase gene, whose chromosomal position is between markers D2S305 and D2S165 on chromosome 2. A 1600-base-pair-long sequence upstream of the translation starting point was evaluated for its promoter activity; five deletion constructs were assayed after transfection in primary chicken fibroblasts and in a human rhabdomyosarcoma cell line. This analysis indicates the existence of two contiguous regions able to modulate luciferase expression in both cell types used, one with a strong activatory function, ranging from positions −204 to −503, and the other, ranging from positions −504 to −683, with a strong inhibitory function.

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 extracellular matrix. These structures are composed of two distinct morphological elements: a more abundant amorphous core of which elastin, responsible for the elastic properties, is the major constituent; and microfibrillar structures of about a 10–12-nm diameter, which are located around the periphery of the amorphous component and consist primarily of fibrillin-1 and/or −2 (1, 2). To date, at least 10 distinct additional components that contribute to the elastic fiber organization have been identified and cloned, including microfibril-associated protein 1 to 4 (3–6), latent-transforming growth factor β-binding protein 1 to 4 (7–10), fibulins 1 and 2 (11–12), and microfibril-associated glycoprotein-2 (13). Whereas the amorphous elastic core is apparently poorly organized, fibrillin-containing microfibrils are highly organized structures. We had originally isolated from chicken aorta a protein, EMILIN, 1 that is broadly expressed in connective tissues, and it is particularly abundant in blood vessels, skin, heart, lung, kidney, and cornea (14–16). The protein is synthesized in vitro by chicken aorta smooth muscle cells and by tendon fibroblasts, it is deposited extracellularly as a fine network (17, 18), and once secreted, it is detected as S–S-bonded high molecular weight aggregates (18). EMILIN is found mainly at the interface between amorphous elastin and microfibrils (19), and its function might be to regulate the formation of the elastic fiber because the addition of EMILIN antibodies in the culture medium of smooth muscle cells greatly affects the process of elastin deposition at least in vitro (19).

The structural components of elastic fibers are defective in some heritable human syndromes: supra valvular aortic stenosis is associated with elastin gene defects (20), the Marfan’s syndrome, one of the most frequent genetic disorders, is associated with defects in the fibrillin-1 gene (21), and congenital contractural arachnodactyly has been linked to the fibrillin-2 gene (22). However, of the numerous heritable syndromes with clear alteration of elastic fibers listed in the Online Mendelian Inherited in the Man data base, most have not been linked to any gene yet.

Recently also the cDNA of human EMILIN has been cloned (23), and its structure revealed a unique modular organization. The presumed mature protein is composed of 991 amino acids and consists of a C terminal C1q-like globular domain (gC1q-like) endowed with cell adhesion-promoting functions, a short uninterrupted collagenous stalk, and a long segment of about 650 residues containing at least four heptad repeats with a  

† The abbreviation used is: EMILIN, elastin microfibril interface-located protein; g, globular; KHK, ketohexokinase; PCR, polymerase chain reaction; kb, kilobase(s); bp, base pair(s); RACE, rapid amplification of cDNA ends.

* This work was supported by grants (to A. C.) from the Associazione Italiana per la Ricerca sul Cancro, Ministero dell’Università e della Ricerca Scientifica e Tecnologica-Cofin 1998, Telethon Grant E.704, 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.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF162780. 

To whom correspondence should be addressed. Tel.: 0039-0434-659-365; Fax: 0039-0434-659-428; E-mail: acolombatt@ets.it.
Characterization of the Human EMILIN Gene

high potential for forming coiled-coil α-helices and two leucine zippers. At the N terminus there is a cysteine-rich sequence stretch characterized by the presence of a partial epidermal growth factor-like motif and highly homologous to a region of multimerin, a platelet and endothelial cell component (24). The presence of a gC1q-like domain and the recent identification that gC1q is structurally homologous to the tumor necrosis factor family of growth factors (25) allowed the inclusion of EMILIN in the C1q/tumor necrosis factor superfamily of proteins. Given the distribution of the protein and the characteristics of its domains, a proposed function for EMILIN might be that it plays a fundamental role in the process of elastogenesis also in vivo and in mediating cell-extracellular matrix interactions. Thus, the knowledge of the regulation of the EMILIN gene could provide us with tools to understand the molecular mechanisms of its expression, whereas the characterization of the EMILIN gene and of eventual molecular markers would help investigate the potential association of EMILIN with heritable elastic fiber disorders. In the current study, we have elucidated the primary structure of the human EMILIN gene, established the chromosomal position, and generated preliminary data on the function of the human proximal promoter as well as defined the general organization of the murine EMILIN gene.

Experimental Procedures

Isolation and Characterization of Human and Murine Genomic DNA Clones—A human genomic BAC library was screened for EMILIN genespecific clones at Genome System using a cDNA fragment prepared by PCR and corresponding to the 3’-end of EMILIN (HEM-3, from position 2580 to position 3048 of the cDNA as numbered in the GenBank™/EBI DNA data base, accession number AF088916. Two positive clones, hBAC-A and hBAC-B, were identified and further characterized. Both isolated clones were authenticated by successful PCR reamplification of insert fragments with a range of primer pairs derived from EMILIN cDNA sequences and partially characterized by restriction enzyme mapping and Southern blot. For each restriction endonuclease, 10 μg of DNA was digested to completion and the resulting fragments were resolved by electrophoresis in 0.8% agarose gels and transferred to nylon membranes (Hybond N+, Amersham Pharmacia Biotech). The filters containing the transferred DNA were hybridized in Rapid-hyb buffer (Amersham Pharmacia Biotech) under stringent conditions to 32P-labeled human EMILIN cDNA probes HEM-3 and HEM-5, the latter extending from position 240 to position 900 of the cDNA. Appropriate restriction fragments (B1, B2, and B3) were gel purified and subcloned in the pGEM 7z-e vector for further analysis (Fig. 1). A similar approach but with homologous cDNA probes was followed to identify the murine gene and to characterize the single BAC clone containing the entire EMILIN gene.

Nucleotide Sequencing and Sequence Analysis—Sequencing of exons, introns, and gene flanking regions was carried out on purified subclones or, in some cases, directly on the BAC clones, using the D8Rhodamine Kit (Perkin-Elmer). All reported sequences have been confirmed by sequencing of both strands.

Northern Blot Analysis—Northern blot was performed using a human multiple tissue blot from CLONTECH. A 32P-labeled probe was synthesized using as template the HEM-3’-clone and the multiprime labeling kit (Amersham Pharmacia Biotech). Hybridization was performed at 65 °C in Rapid-hyb buffer. All the other procedures were performed using standard techniques.

Rapid Amplification of cDNA Ends (RACE)—The 5’-RACE Kit for rapid amplification of cDNA ends (Roche Molecular Biochemicals) and oligo(dT)-tailed using terminal transferase (Promega Corp.) were used following the manufacturer’s instructions. Briefly, first-strand cDNA was synthesized from 10 μg of human uterus total RNA using avian myeloblastosis virus reverse transcriptase (40 units) and 10 pt of the reverse primer RACE1 (positions -248 to +270 of the EMILIN gene) (see Fig. 6). The first-strand cDNA was purified using a glass fiber spin cartridge (Roche Molecular Biochemicals) and oligo(dT)-tailed using terminal transferase (Promega Corp.). The cDNA was amplified by PCR through two rounds of 40 cycles each at 95, 50, and 72 °C using the oligo(dT)-anchor primer provided with the kit as the sense primer and with RACE2 (first round, positions +149 to +168) and RACE3 (second round, positions +118 to +138) as antisense primers. The resulting amplification products were subcloned into pGEM-T (Promega Corp.) and sequenced. To confirm the results obtained by the RACE technique, additional PCR experiments were performed with two sense primers (TSPE, positions -18 to -1 and TSP1, positions +1 to +20) located immediately upstream and downstream from the transcription initiation site, in combination with an antisense primer located in exon 2 (TSP2, positions +1466 to +1660), using the same template as in the RACE experiments.

Deletion Constructs for Promoter Analysis—To study the regulation of the human EMILIN gene proximal promoter, transient transfection experiments were performed with various promoter 5’-deletion fragments cloned into the promoterless vector (pGL3-Basic) (Promega Corp.). Regions of the EMILIN proximal promoter were amplified by PCR using sense primers with a KpnI-containing tail (P1 to P8 in Fig. 6) and an antisense primer with a SacI-containing tail (RP1 in Fig. 6). PCR products were double-digested to create cohesive ends and all fragments were purified by a Qiagen midi prep kit (Qiagen GmbH, Germany) before ligation into identical sites in the pGL3-Basic. Large scale plasmid isolation and purification was carried out using the Qia-gen maxi prep kit (Qiagen GmbH, Germany). Plasmids were quantified by both spectrophotometry and ethidium bromide-agarose gel.

Cell Culture and Cell Transfections Assays—Transient cell transfection of the human RD rhabdomyosarcoma cell line and of primary cultures of chicken embryo tendon fibroblasts were performed with LipofectAMINE (Life Technologies, Inc.). Briefly, cultured cells at about 80% confluence were co-transfected in a 6-well plate with 1.5 μg of plasmid DNA and with 0.2 μg of the pEGFP (CLONTech) plasmid DNA to monitor for transfection efficiency. In each experiment pGL3-Basic and pGL3-luc, in which the luciferase gene was driven by the SV40 promoter and enhancer, were used as negative and positive controls, respectively. The -1200/-160 reference constructs were transfected in parallel with the other EMILIN promoter constructs. Each promoter construct was transfected in triplicates in at least three different experiments. After 48 h of incubation, the cells were rinsed once with phosphate-buffered saline and the fluorescence activity of the co-transfected pEGFP vector was measured in situ with a fluorometer (Tecan AG, Switzerland). The cells were then harvested by scraping, lysed in 200 μl of reporter lysis buffer (Promega Corp.), and tested for luciferase activity.

Results and Discussion

Exon-Intron Organization of the Human and Marine EMILIN Gene—Screening of the human genomic BAC library with a human EMILIN cDNA probe led to the identification and purification of two clones, hBAC-A (58 kb) and hBAC-B (33 kb), which yielded PCR products of the expected size when amplified using several EMILIN cDNA-specific oligonucleotide primers. The authenticity of clones hBAC-A and hBAC-B was definitively proved by successful direct sequencing using primers in the coding region of EMILIN. For better sequencing some subclones were prepared by digesting the clones hBAC-A and hBAC-B with a panel of restriction endonucleases followed by Southern blot and filter hybridization with probes derived from different regions of the EMILIN cDNA. One SauI fragment (A1) derived from clone hBAC-A and BamHI and HindIII fragments derived from clone hBAC-B (B1 and B3, respectively) (Fig. 1) were subcloned in the pGEM 7 vector and entirely sequenced. To overcome some sequencing problems, an additional clone (B2) was obtained by PCR using as template clone hBAC-B and primers 35 and 37 (Fig. 1). The subclones generated from the hBAC clones overlapped and covered the entire EMILIN gene. This was assessed by a series of successful PCRs performed with a set of primers that amplified the expected regions at the extremities of the EMILIN cDNA. Furthermore, clone B3 contained additional 1371 bp upstream of the translation start site (see Fig. 6), that included the 5’-untranslated region and at least part of the EMILIN promoter region (see below).

The exon/intron boundaries of the protein-coding region of the EMILIN gene were identified by comparison with the cDNA sequence (23). Considering the size of the coding sequence (about 3 kb) the EMILIN gene is remarkably compact, spanning only about 8.0 kb in length. It is organized in 8 exons
interrupted by 7 introns (Fig. 2). Introns vary in length from 151 to 1126 bp and exons vary from 17 to 1929 bp. The average size of the exons is 500 bp and that of introns 550 bp, resulting in an overall coding capacity for this gene close to 50%, a percentage significantly higher than the average of 10% found in most vertebrate genes. The genomic organization of few constituents of the elastic fiber are known, and it is quite distinct from that of EMILIN. For instance, the microfibril-associated glycoprotein-2 has a cDNA size of about 1 kb, and its genome is around 10 kb (26–29); the elastin gene (30, 31) as well as the fibrillin-1 gene (32) have a similar 1:10 ratio between the coding and the noncoding regions. In the case of fibrillin-1 recent evidence indicates that for this gene the ratio between coding and noncoding sequences is even higher than 1:20 (33). The translation initiation codon of EMILIN is in exon 1, and the last exon contains 380 bp of noncoding sequences. Furthermore, all the sequences at the exon/intron boundaries are in full agreement with the consensus rules established for the canonical splice donor and acceptor sites of vertebrate genes (Fig. 3) and with the exception of the first two introns, that are in phase 2, all the other introns are in phase 1.

A comparison of the exon/intron organization with that of the protein deduced from the cDNA sequence reveals that there is no obvious correlation between the exon/intron structure and the modular organization of the protein. As schematized in Fig. 4A, the first exon contains the 5′-untranslated region and the signal peptide, and the cysteine-rich region showing homology to the N terminus of the blood protein multimerin is shared between exon 2 and part of exon 3. Instead, a remarkable correlation between gene structure and protein domains is evident in the very large exon 4, which includes a cluster of four coiled-coil structures and of two leucine zippers; the 5′- and 3′-exon borders correspond exactly to the beginning of the first coiled-coil stretch and to the end of the second leucine zipper, respectively. As both coiled-coil and leucine zipper structures might be involved in interchain association, this large exon might represent a sort of functional unit. The organization in this part of the gene resembles that found in many extracellular matrix genes in which often one or more protein domains are encoded within a single exon (for a review see Ref. 34).

As a preliminary effort to prepare constructs suitable for homologous recombination in ES cells, the murine EMILIN gene has also been isolated from a BAC library. A single clone of about 100 kb was identified in the screening, and the EMILIN gene was then characterized and completely sequenced2 and is schematically shown in Fig. 4. The murine gene is remarkably similar to the human gene; not only is the size of the exons (except for the first and the last that have not been completely characterized yet) almost identical between the two species, but also the size of the introns is nearly superimposable.

The 3′-Untranslated Region and Chromosomal Assignment—Searches of the GenBankTM/EBI DNA data base showed a virtual identity between the 3′-end of the EMILIN gene and the 5′-end of the human ketohexokinase (KHK) promoter (35). The homology extends for 659 bp, and there are very few mismatches, likely because of polymorphisms and/or sequence reading errors. Fusion artifacts in the BAC clones used can be excluded as the same type of homology has been found in the kidney aq10 clones used to characterize the EMILIN cDNA (23). Moreover, a CA dinucleotide repeat sequence, independently localized about 5 kb upstream of the KHK gene (36), corresponds to the one found here by us in intron 3 of the

---

2 P. Bonaldo and R. Doliana, unpublished data.

---

Fig. 1. Schematic representation of the BAC clones used to characterize the EMILIN gene. The two human BAC clones are indicated as hBAC-A and hBAC-B. Subclones generated by digestion with restriction enzymes are also shown underneath each hBAC clone and are marked A.1 and B.1 and B.3, respectively. B.2 was generated by PCR cloning using oligonucleotides 35 and 37. tsp, transcription start site; tep, transcription ending point.
EMILIN gene (see Fig. 4A). Because the human KHK and the closely linked glucokinase regulatory protein genes have been mapped on chromosome 2p23, between the markers D2S305 and D2S165 (36), we can conclude that EMILIN and KHK lie in close proximity in a head-to-tail orientation, the EMILIN gene being localized upstream of the KHK gene with its 3′-end overlapping with the 5′-end of the promoter region of the KHK gene (Fig. 4B). In addition, a similar KHK glucokinase regulatory protein gene organization has been found in the mouse and in the rat (35). We have evidence, based on sequencing of the murine BAC clone, that indeed the murine EMILIN gene is similarly localized in close proximity to the KHK gene. Mapping of human YAC contigs indicated that the KHK gene lies 400 kb downstream from the phosphatase gene PPP1CB (36, 37); however, the murine homolog of human PPP1CB maps, like several other human genes located on 2p, to mouse chromosome 12D (38). Thus, the co-localization of KHK and glucokinase regulatory protein genes on mouse chromosome 5 (39), despite their nonsyntenic relationship with PPP1CB, represents a strong indication that the murine EMILIN gene might be similarly localized to chromosome 5.

**Dot Blot and Northern Blot Analysis**—A blot containing poly(A+) RNA of human tissues was hybridized with the EMILIN cDNA probe HEM-3 and the results indicated that several tissues including heart, aorta, small intestine, colon, lung, kidney, uterus, and ovary gave a positive signal, whereas all parts of the brain analyzed were consistently negative (data not shown). Next, Northern blots of poly(A+) RNAs from selected human tissues were hybridized with the same probe as above and a single transcript of approximately 4 kb was detected (Fig. 5). The expression pattern of EMILIN mRNA in human tissues as demonstrated by dot-blot and Northern blot is in

---

**FIG. 2.** Complete sequence of the human EMILIN gene. The DNA sequence was determined as described under “Experimental Procedures.” The initiation codon, ATG, and the stop signals are in reverse mode. Sequences of exons are in uppercase letters; sequences of introns are in lowercase letters. A microsatellite sequence in intron III is in bold and underlined. Arabic and roman numbers to the left indicate exons and introns, respectively. The donor and acceptor sites are boxed.

**FIG. 3.** Sequences of the intron:exon junctions of the human EMILIN 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.
agreement with the previously reported distribution of the protein as detected in chicken tissues (14–16). Furthermore, given the overall high sequence homology between the human and the murine EMILIN cDNAs, it is likely that this protein will be distributed, as already described for chicken tissues, in a similar pattern in human and in murine tissues and that it might play an equivalent function in the different species.

Determination of the Transcription Initiation Site—On the basis of the length of the coding sequence (3048 bp), of the 3' end untranslated region (380 bp), of the poly(A) tail (100–200 bp), and of the size of the EMILIN mRNA as suggested by Northern blot analysis, the 5' end untranslated region should encompass about 500 bp. To precisely determine the transcription initiation site we tried repeatedly to perform RNase digestions of RNA isolated from human tissues and protected with different 5' end probes, but all attempts proved unsuccessful and no protected bands were detected. We resorted then to a combination of 5' end RACE-PCR and reverse transcriptase-PCR on human uterus, a tissue that has been shown to express a high level of EMILIN mRNA in the mouse. 5' end RACE-PCR was performed on human uterus total RNA using primers selected at about 150 bp downstream from the presumed transcription starting point as indicated by Northern blots and cDNA size. Cloning and sequencing of the 5' end RACE-PCR products allowed mapping of the transcription initiation site at 511 bp upstream of the AUG start codon, corresponding to the g labeled as +1 in Fig. 6. In fact, of the ten clones sequenced, nine gave the same result and started with the g at position 511, and only one clone started 20 bases downstream, strongly suggesting that the major transcription site corresponds to that indicated as +1 in Fig. 6 and that no transcription starting points are present upstream of position +1. The finding of only one RACE product ending 20 bases downstream from position +1 suggests that if there is a second downstream transcription start site, it must be rarely used. However, because in the RACE method the sense primer used in PCR is represented by a poly(dT) there is an unavoidable unbalance in base composition and consequently in the annealing temperatures of the two primers, resulting in a PCR reaction of low efficiency. This intrinsic weakness of the RACE method implicates that rare RNA messages might be missed resulting in under-representation of the actual transcription

starting point(s). To circumvent this problem, a second type of experiment was set up; in this case the same RNA template was subjected to a PCR reaction with an antisense primer located in exon 2 (TSPR), to distinguish between genomic- and RNA-initiated sequences, and two sense primers located upstream (TSPE) and downstream (TSPI) of the presumed transcription initiation site identified by the RACE method. In accord with the conclusions reached above, the PCR performed with a downstream primer produced a fragment of the expected size, whereas the upstream primer failed to do so, confirming that the transcription start site must be localized between TSPE and TSPI (data not shown).

Sequence Analysis of the 5'-End Flanking and Untranslated Region of the Human and Murine EMILIN Gene—To identify the DNA elements that control the transcription of the EMILIN gene, a 1.3-kb region immediately upstream of the putative transcription initiation site was sequenced. As shown in Fig. 6, this DNA sequence does not present characteristics typical of GpC islands (40), lacks TATA and CCAAT elements close to the transcription initiation site, and does not show any homology to other known transcriptional control motifs. These features are often found in genes that are not constitutively active but rather are specifically regulated during differentiation or development and initiate transcription at only one or a few closely clustered sites (41). TATA-less promoters are a common characteristics of genes involved in elastic fiber structure and function such as elastin (30), fibrillin-1 (32), microfibril-associated protein 1 (27, 28), and microfibril-associated glycoprotein-2 (26).

Further analysis of the 5'-flanking region using the SIGSCAN 4.05 program at the IMAS web site revealed a variety of potential binding sites for several transcription factors. The sites with a consensus sequence of at least 6 bases and with a match ratio higher than 0.9 are shown in Fig. 6. A number of potential transcription factors binding sites proximal to the transcription initiation site, and does not show any homology to other known transcriptional control motifs. These features are often found in genes that are not constitutively active but rather are specifically regulated during differentiation or development and initiate transcription at only one or a few closely clustered sites (41). TATA-less promoters are a common characteristics of genes involved in elastic fiber structure and function such as elastin (30), fibrillin-1 (32), microfibril-associated protein 1 (27, 28), and microfibril-associated glycoprotein-2 (26).

Functional Deletion Analysis of the Putative Proximal Promoter of the Human EMILIN Gene—To define the promoter region and the functional subregions that controls EMILIN expression, several 5'-deletion constructs of the 1.3-kb fragment that is located immediately upstream of the transcription initiation site were generated. The deletion constructs, fused 5' to the luciferase gene of a reporter plasmid, were transfected into two distinct cell types for transient expression analysis (Fig. 7). Chick embryo fibroblasts are primary cells that have previously shown to be active in endogenous expression of the EMILIN gene (17, 18), whereas RD is a cell line derived from a human rhabdomyosarcoma that expresses EMILIN (4). The transfection efficiencies were controlled by a cotransfection with a green fluorescent protein-containing vector and were then normalized according to these values. Results of the luciferase assays in both cell types indicate that the shortest construct (Promo 6) exerts only a limited positive effect (+12%) with respect to the pGL3-luc positive control. Promo 5, which spans from −503 to +166, induced a significant increase in promoter activity (40–75% of the positive control in chick embryo fibroblasts and RD, respectively) compared with Promo 6. This indicated that the basic EMILIN promoter elements necessary for expression of the luciferase gene in these cells are contained within the −204 to −503 region. The promo 4 construct from −504 to −683 had no activity suggesting that it contained a strong down-regulatory cis-element. The remain-
ing deletion constructs (Promo 3 to Promo 2) and the undeleted construct (Promo 1) were also negative indicating either that there are no additional positive control elements or that the inhibitory region is very strong and able to shut down any positive cis-element. The finding that two cell lines of so differ-
ent origin and species behaved nearly equally when transfected with the Promo 4 construct strengthens the concept that this region plays an important regulatory function. In the elastin gene, which is the only one among the elastic fiber-associated genes that has been characterized to some extent, similar strong cis-acting inhibitory elements have not been found (42, 43).

Conclusions—In this study we have elucidated the complete structure of the EMILIN gene and its chromosomal location and provided a partial characterization of the proximal promoter region. The EMILIN gene is unusually very compact with very short introns and, with the exception of the very large exon 4 where all the potential coiled-coil motifs are present, is not strictly organized in the one exon one module fashion.

Expression studies revealed that the main positive and negative regulatory elements responsible for transcription lie within a fragment extending from position –2683 to –2204 relative to the translation start site. This short sequence contains several potential binding sites for transcription factors and will be soon characterized in more detail.

The EMILIN gene showed a partial overlap with the promoter of the gene encoding for the enzyme KHK. These overlapping gene-within-a-gene systems are rare but not unique. One of the most unusual examples is that represented by the

### Table I

<table>
<thead>
<tr>
<th>Factor</th>
<th>Consensus (number)</th>
<th>Human sequence (position)</th>
<th>Mouse sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2A</td>
<td>GCAGGTG (22)</td>
<td>GCAGGTG (-1182)</td>
<td></td>
</tr>
<tr>
<td>CAC-BF</td>
<td>CCCACCCT (9)</td>
<td>CCCCACCCT (-1059)</td>
<td></td>
</tr>
<tr>
<td>NF-BAl</td>
<td>ACCTTT (4)</td>
<td>ACCTTT (-985)</td>
<td>ACCTTT (-944)</td>
</tr>
<tr>
<td>Elk-1</td>
<td>SMGAAAG (5)</td>
<td>CAGGAAG (-972)</td>
<td>CAGGAAG (+949)</td>
</tr>
<tr>
<td>NF-κE4</td>
<td>CCAGGTG (3)</td>
<td>CCAGGTG (-922)</td>
<td>CAGGAAG</td>
</tr>
<tr>
<td>Pu.1</td>
<td>GAGGAA (9)</td>
<td>GAGGAA (-909)</td>
<td>GAGGAA (+315)</td>
</tr>
<tr>
<td>LyF-1</td>
<td>TTTGGGAGR (10)</td>
<td>TTTGGGAGR (-643)</td>
<td></td>
</tr>
<tr>
<td>Pu-box</td>
<td>AAAGAGGA (4)</td>
<td>AAAGAGGA (-594)</td>
<td>AAAGAGGA</td>
</tr>
<tr>
<td>TFII-I</td>
<td>GTCTATC (3)</td>
<td>GTCTATC (-584)</td>
<td>GTCTATC (+116)</td>
</tr>
<tr>
<td>ILF</td>
<td>AGAAGGAG (3)</td>
<td>AGAAGGAG (-509)</td>
<td></td>
</tr>
<tr>
<td>COUP</td>
<td>CYBTGACCY (12)</td>
<td>CTCCTGACC (-413)</td>
<td></td>
</tr>
<tr>
<td>TFII-I</td>
<td>CTCCTT (5)</td>
<td>CTCCTT (-281)</td>
<td></td>
</tr>
<tr>
<td>E-box</td>
<td>CAGGTGGGC (4)</td>
<td>CAGGTGGC (-159)</td>
<td>CACCTGGC</td>
</tr>
<tr>
<td>ENKTF-1</td>
<td>TGCCGTA (4)</td>
<td>TGCCGCA (-57)</td>
<td></td>
</tr>
<tr>
<td>EH-1</td>
<td>GACAGGAA (3)</td>
<td>GGGAGGAA (-34)</td>
<td></td>
</tr>
<tr>
<td>NF-S</td>
<td>TGTCASC (2)</td>
<td>TGTCACC (+98)</td>
<td></td>
</tr>
<tr>
<td>PEB-1</td>
<td>TCTCTCTT (3)</td>
<td>TCTCTCT (+126)</td>
<td></td>
</tr>
<tr>
<td>IAF</td>
<td>GCCATCTG (10)</td>
<td>GCCAGGTG (+152)</td>
<td>GCCAGGTG</td>
</tr>
<tr>
<td>NRF-2</td>
<td>CTTCGGCT (7)</td>
<td>CTTCGGCT (+281)</td>
<td>CTTCGGCT</td>
</tr>
<tr>
<td>CORE-BF</td>
<td>TGCGGT (3)</td>
<td>TGCGGT (+1380)</td>
<td></td>
</tr>
</tbody>
</table>
locus containing the Tenascin X gene; in only 160 kb there are at least 13 transcriptional units, including several genes that overlap in the same or in the opposite orientation (44). Whereas in some cases, i.e. the α1 and α2 chains of type IV collagen, the close proximity of two genes allows a coordinate expression (45, 46), this is unlikely for EMILIN and KHK, given the diverse functions and the distinct and preferential tissue expression patterns of the two gene products. The chromosomal assignment of the EMILIN gene was deduced from that established for KHK, which mapped on chromosome 2p23.2–23.3 between the markers D2S305 and D2S165 (36). The complete characterization of the exon-intron boundaries and of the introns and the knowledge of the chromosomal position of the EMILIN gene will allow also analysis of the gene in diseases. The Marfan’s syndrome phenotype represents one end of a phenotypic continuum (47), and the knowledge that several Marfan-related phenotypes are not related to other known genes of the microfibril-elastic fiber system implies that mutations in other genes such as EMILIN, whose products are structural components of the microfibril-elastic fiber system, may contribute to this phenotypic continuum.

REFERENCES

Structure, Chromosomal Localization, and Promoter Analysis of the Human Elastin MicrofibrillInterfase Located proteIN (EMILIN) Gene
Roberto Doliana, Anna Canton, Francesco Bucciotti, Maurizio Mongiat, Paolo Bonaldo and Alfonso Colombatti

doi: 10.1074/jbc.275.2.785

Access the most updated version of this article at http://www.jbc.org/content/275/2/785

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 47 references, 19 of which can be accessed free at http://www.jbc.org/content/275/2/785.full.html#ref-list-1