THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 280, No. 16, Issue of April 22, pp. 15749–15760, 2005
Printed in U.S.A.

Analysis of Regulatory Regions of *Emilin1* Gene and Their Combinatorial Contribution to Tissue-specific Transcription*

Received for publication, November 5, 2004, and in revised form, January 28, 2005
Published, JBC Papers in Press, February 10, 2005, DOI 10.1074/jbc.M412548200

Carla Fabbro‡, Paola de Gemmis‡, Paola Braghetta‡, Alfonso Colombatti§‖, Dino Volpin‡, Paolo Bonaldo‡, and Giorgio M. Bressan‡¶

From the ‡Department of Histology, Microbiology, and Medical Biotechnologies, University of Padova, 35131 Padova and the §Division of Experimental Oncology, CRO-IRCCS, 33081 Aviano, and the ‖Department of Biomedical Sciences and Technologies, University of Udine, 33100 Udine, Italy

The location of regions that regulate transcription of the murine *Emilin1* gene was investigated in a DNA fragment of 16.8 kb, including the entire gene and about 8.7 and 0.6 kb of 5′- and 3′-flanking sequences, respectively. The 8.7-kb segment contains the 5′-end of the putative *2310015E02Rik* gene and the sequence that separates it from *Emilin1*, whereas the 0.6-kb fragment covers the region between *Emilin1* and *Ketohexokinase* genes. Sequence comparison between species identified several conserved regions in the 5′-flanking sequence. Most of them contained chromatin DNase I-hypersensitive sites, which were located at about −950 (HS1), −3100 (HS2), −4750 (HS3), and −5150 (HS4) in cells expressing *Emilin1* mRNA. *Emilin1* transcription initiates at multiple sites, the major of which correspond to two Initiator sequences. Promoter assays suggest that core promoter activity was mainly dependent on Initiator1 and on Sp1-binding sites close to the Initiators. Moreover, one important regulatory region was contained between −1 and −169 bp and a second one between −630 bp and −1.1 kb. The latter harbors a putative binding site for transcription factor AP1 matching the location of HS1. The function of different regions was studied by expressing *lacZ* constructs in transgenic mice. The results show that the 16.8-kb segment contains regulatory sequences driving high level transcription in all the tissues where *Emilin1* is expressed. Moreover, the data suggest that transcription in different tissues is achieved through combinatorial cooperation between various regions, rather than being dependent on a single *cis*-activating region specific for each tissue.

An important feature of the ECM[1] is the extreme variability of its composition and architecture, endowing tissues with specific mechanical and biological properties. A major factor contributing to this complexity is the tissue-specific transcription of different genes. In the past, several studies have investigated the regulatory sequences responsible for tissue-specific expression of ECM genes (1–6). The overall picture coming from these studies is a modular arrangement of regulatory regions (mostly enhancers), meaning that each region is a distinct and independent regulator of transcription in a specific tissue or in a set of tissues. Although some reports have indicated the importance of the interaction of distinct regulatory regions for high level tissue-specific transcription (7) and have shown that different basal promoters may influence expression driven by tissue-specific enhancers (8), the concept of modularity of regulatory regions is the key motif in most studies.

For several years the authors have been studying Emilin1, a protein of elastic fibers that is preferentially detected by electron microscopy between the amorphous core and the coat of microfibrils (the name is an acronym from Elastin microfibrils interface located protein) (9). The primary sequence shows that Emilin1 comprises five domains (10). A C1q-like domain similar to those of type VIII and type X collagens is located at the carboxyl-terminal end. A short collagenous domain separates the C1q domain from a long region with a high probability of coiled-coil conformation. Finally, two domains are found at the amino terminus: a signal peptide and a new type of domain called the EMI domain (11). The latter, a cysteine-rich sequence of about 75 amino acids, is shared by a number of genes (seven in mammals) that have been grouped into the *Emilin* gene family (www.gene.ucl.ac.uk/nomenclature/genefamily/emilin.html#HGNC_table2). Gene targeting experiments suggest that the protein has a role in the assembly of elastic fibers, particularly in blood vessels; in its absence, elastic lamellae of aorta are interrupted and their outline is irregular (12). A function of Emilin1 in elastogenesis is also indicated by the finding that the protein binds to other components of elastic fibers such as Elastin and Fibulin-5 (12). However, analysis of the distribution of *Emilin1* mRNA during mouse development induces us to suggest additional functions of Emilin1 not related to elastic fibers; in fact, early after implantation, *Emilin1* mRNA is detected not only in the cardiovascular system but also in extra-embryonic tissues (extra-embryonic visceral endoderm, ectoplacental cone, and spongiotrophoblast), epithelial tissues that do not produce elastic fibers (13). During organogenesis high level expression is found in interstitial connective tissue, perichondrium, and mesenchymal condensations, sites containing elastic fibers. This brief description of *Emilin1* expression during development testifies to a complex gene regulatory setup. Nevertheless, the *Emilin1* gene and flanking sequences are a compact unit; the gene, consisting of 8 exons, is about 8 kb long and is separated by the nearby *2310015E02Rik* and *Khk* genes by about 6.3 kb at the 5′-end and 0.7 kb at the 3′-end, respectively (14) (see Fig. 3*C*). This feature makes the *Emilin1* gene a good candidate for analysis of the mechanisms of tissue-specific transcriptional regulation.

In this report we describe the identification of *cis*-acting

* This work was supported by Progetto di Ricerca di Interesse Nazionale and Fondo per gli Investimenti della Ricerca di Base grants from the Italian Ministero dell'Istruzione Università e Ricerca. 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: Dipartimento di Istologia Microbiologia e Biotecnologie Mediche, Università di Padova, Viale G. Colombo 3, 35131 Padova, Italy. Tel.: 39-49-8276086; Fax: 39-49-8276079; E-mail: bressan@civ.bio.unipd.it.

[1] The abbreviations used are: ECM, extracellular matrix; ACH, active chromatin hub; CAT, chloramphenicol acetyltransferase; E, embryonic age as days post-coitum; *Khk*, *Ketohexokinase* gene; X-gal, 5-bromo-4-chloro-3-indolyl-β-D-galactoside; RACE, rapid amplification of cDNA ends.

regulatory regions of the *Emilin1* gene, including the core promoter and sequences involved in tissue-specific expression. The main conclusion coming from our results is that high level transcription in diverse tissues is the result of the differential contribution of several regulatory regions that act mostly in a combinatorial way.

MATERIALS AND METHODS

Plasmid Constructs—A 16.8-kb HindIII fragment, containing the entire *Emilin1* gene and 5′- and 3′-flanking sequences, was subcloned from a 135-kb BAC clone (12) into pBluescript vector (Stratagene, La Jolla, CA).

A fragment spanning positions −402 (corresponding to a BamHI restriction site) to −1 from the first codon was generated from the 16.8-kb fragment by PCR with the reverse primer comprising mutagenized bases that change the sequence corresponding to the translation start site ATGGCC to a PstI restriction site. After digestion with BamHI and PstI, the amplified 407-bp fragment and an 8249-bp HindIII/BamHI fragment extending from −403 to −8651 bp were ligated into the promoterless vector pBLCAT6 (15) to give plasmid p8.7-CAT. 5′-End deletions extending to about −6.0, −5.1, −4.0, −3.0, and −1.1 kb and to −630 bp were generated by exonuclease III treatment ("Erase a Base" kit, Promega) of the 8.7-kb fragment using the procedures recommended by the manufacturer.

The 630-bp deletion was inserted into pBLCAT6 to give p630-CAT. Deletions of p630-CAT from 5′- and 3′-ends were developed by Bal31 digestion following established protocols (16). p(630−470)-CAT and p(630−441)-CAT were generated from the 16.8-kb fragment by PCR with appropriate primer oligonucleotides.

For site-directed mutagenesis of Initiator1 and Initiator2 sequences (Fig. 1), the mutated p630-CAT constructs were obtained by ligation of a HindIII/NcoI fragment carrying sequences from −630 to −459 bp, an NcoI/XbaI fragment extending from −458 to −1, and the promoterless vector pBLCAT6. For p630Inr2-CAT, the mutant fragment was generated by PCR with a forward primer comprising mutagenized bases that change the native sequence of Initiator2 (−453/−447) TCACTCT to Tgtggtt (where lowercase letters indicate mutated bases) and a reverse primer comprising mutagenized bases that change the sequence corresponding to the translation start site ATGGCC to an XbaI restriction site TCTAGA. To derive the mutant fragment of p630Inr1-CAT, the Initiator1 sequence CCAGACA (−517/−511) was replaced with Cgtggtt (17) following a two-step PCR-based site-directed mutagenesis procedure (16) using the following primers: F1 (−630/−612), 5′-GAAGCTTCACATTTCTCTGTTTTTCC-3′, forward (HindIII site underlined); M1 (−522/−502), 5′-CCTTCCCCCaaccacGCTGGC-3′, reverse; M2 (−524/−504), CCGCCAGCgtggttGGGGGAA-3′, forward; R2 (−458/−441), 5′-GGGATCCGGCACAAGAGTGACCATG-3′, reverse (BamHI site underlined).

The transgenes p8.7-*lacZ*, p3.0-*lacZ*, p1.1-*lacZ*, and p0.6-*lacZ* were synthesized by insertion of the corresponding exonuclease III deletions (8.7, 3.0, and 1.1 kb and 630 bp) into the promoterless plasmid pNS*lacZ* in which the *lacZ* sequence of *Escherichia coli* is fused with the nuclear localization signal of SV40 (18). In the transgene p8.7Δ-*lacZ*, the 407-bp BamHI/PstI fragment was replaced with a 238-bp fragment extending from −402 to −170 generated by PCR with appropriate primer oligonucleotides.

To derive the transgene construct p8.7-*lacZ*-intron, the third intron of *Emilin1* gene was synthesized by PCR from the 16.8-kb fragment as DNA template and inserted downstream from the *lacZ* gene into the p8.7-*lacZ* plasmid.

To obtain the transgene construct p0.6-*lacZ* gene, the 8.0-kb fragment that contains the coding region of *Emilin1* from +78 in the first exon to the intergenic region between *Emilin1* and *Khk* genes (Fig. 3*C*) was generated by digestion of the 16.8-kb fragment with XhoI and HindIII and cloned into the SmaI site of pBluescript vector. The transgene 0.6-*lacZ* was excised with HindIII and NotI and inserted upstream of the XhoI/HindIII fragment of the above plasmid.

To derive the construct p8.7-*lacZ*-gene, the BamHI fragment of the p0.6-*lacZ* plasmid, including 402 bp of proximal promoter region and the *lacZ* gene, was cloned into the same site of pBluescript after removal of the HindIII site. The XhoI/HindIII fragment spanning positions +78 to +8111 was then inserted downstream the *lacZ* sequence. Finally, the BamHI fragment from −402 to +1299 of the 16.8-kb fragment was replaced with the BamHI fragment of the above plasmid containing 402 bp of the proximal promoter region, the *lacZ* gene, and the coding region from +78 to +1299. All constructs were sequenced in both directions to verify correct cloning.

Cells—The following cell lines were used: NIH3T3 fibroblasts, C2C12 myoblasts, EL4 lymphocytes (8), MC615 chondrocytes (19), BC3H1 smooth muscle cells (ATCC, Manassas, VA), and H.end (20) and cEnd (21) murine endothelial cell lines. The cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum, with the exception of EL4 cells, which were maintained in RPMI 1640, 10% fetal calf serum, 4 mM glutamine, and 20 mM 2-mercaptoethanol.

5′-RACE Analysis—To determine the transcription start sites of the *Emilin1* gene, 5′-RACE was performed using the "5′-RACE system for rapid amplification of cDNA ends" (Invitrogen) and total RNA from E14.5 mouse embryos. First strand cDNA was synthesized from 1 μg of total RNA using Superscript™ II reverse transcriptase (Invitrogen) and an antisense primer (RACE1, 5′-GCCGTGGTGAGTTCTGGG-3′, which corresponds to nucleotides 137–154) at 37 °C for 1 h. After (dC)-tailing with terminal dideoxynucleotide transferase at 37 °C for 10 min, the cDNA was PCR-amplified with internal antisense primers (RACE2, 5′-GCTTCCGTGGCTATGGTCAG-3′, which corresponds to nucleotides 43–62; RACE3, 5′-GCAGACAGCAGAGATAGCA-3′, which corresponds to nucleotides 25–43) and a 5′-anchor oligo(dG) primer provided with the kit. PCR products were resolved in a 2% agarose gel, isolated, and subcloned into pGEMT-Easy vector (Promega), and 40 clones were sequenced.

RNase Protection Assay—Four different partially overlapping DNA sequences were used to generate the riboprobes. Probe 1 was derived by subcloning a blunted 891-bp StuI/XhoI fragment extending from −814 to +77 into the SmaI site of pGEM3 vector (Promega). Probe 2 was similarly obtained from a 681-bp StuI/PstI fragment extending from −814 to −133. To generate probe 3 and probe 4, a 561-bp fragment from −815 to −254 and a 470-bp fragment from −815 to −345, respectively, were PCR-amplified and subcloned into pGEMT-Easy vector (Promega). Linear DNA templates were prepared by XbaI (probe 1), HindIII (probe 2), and SalI (probe 3 and probe 4) digestion. ^{32}P-Labeled RNA probes were transcribed using 40 units of T7 polymerase (Ambion, Inc.) in the presence of 1 μg of linearized plasmid, 10 μl of [α-^{32}P]UTP (400 Ci/mmol; Amersham Biosciences), ATP, GTP, and CTP (400 μM each), 10 mM dithiothreitol, and 40 units of RNasin (Promega) in a total volume of 20 μl. The hybridization of the probe with total RNA (10 μg from E14.5 mouse embryos) was carried out at 42 °C overnight according to the standard procedure of RPAII RNase protection assay kit (Ambion Inc.). RNase digestion was performed using an RNase A/RNase T_1 mixture in RNase digestion buffer (Ambion Inc.), and protected fragments were separated on a 6% nondenaturing polyacrylamide gel.

Transfections and Promoter Assays—NIH3T3 and C2C12 cells were grown as described above; 3 × 10^5 cells were plated into 10-cm Petri dishes and transfected the following day with the CAT constructs and the control plasmid pRSV-luciferase (15 and 1.2 μg, respectively) using the calcium phosphate method (22). All subsequent manipulations and assays were performed as described (23).

Generation and Analysis of Transgenic Mice—Fertilized B6D2F1 × B6D2F1 mouse oocytes were microinjected with *lacZ* constructs and implanted in the uterus of CD1 pseudopregnant mothers using standard procedures (24). Transgenic mice were identified by PCR analysis on genomic DNA from yolk sac or tail biopsies using primers derived from the *lacZ* sequence (forward, 5′-CGGTGATGGTGCTGCGTTGGA-3′; reverse, 5′-ACCACCGCACGATAGAGAGATTC-3′) and reaction conditions as described (25). Whole mount and histological examinations of β-galactosidase expression were carried out exactly as described (18).

Other Assays—Northern blotting and analysis of chromatin DNase I-hypersensitive sites were performed as described previously (8, 23).

RESULTS

Structural Features of the 5′- and 3′-Flanking Regions of Emilin1 Gene—The murine *Emilin1* gene is localized on chromosome 5 band B1 between markers D5Mit389 and Slc30a3, syntenic to human chromosome region 2p23.3 (14). The nearest gene in the 5′ position is *2310016E02Rik*, homologous to the human hypothetical protein FLJ21839 (Fig. 3*C*). The two genes are in a head-to-head orientation, with the two translation start sites 6380 bp apart. At the 3′-end, about 650 bp separate the polyadenylation signal of *Emilin1* from the first codon of the *Khk* gene (Fig. 3*C*).

In order to analyze the elements that control the transcriptional regulation of *Emilin1*, a 16,762-bp HindIII fragment (for brevity identified in this study as the 16.8-kb fragment) encom-

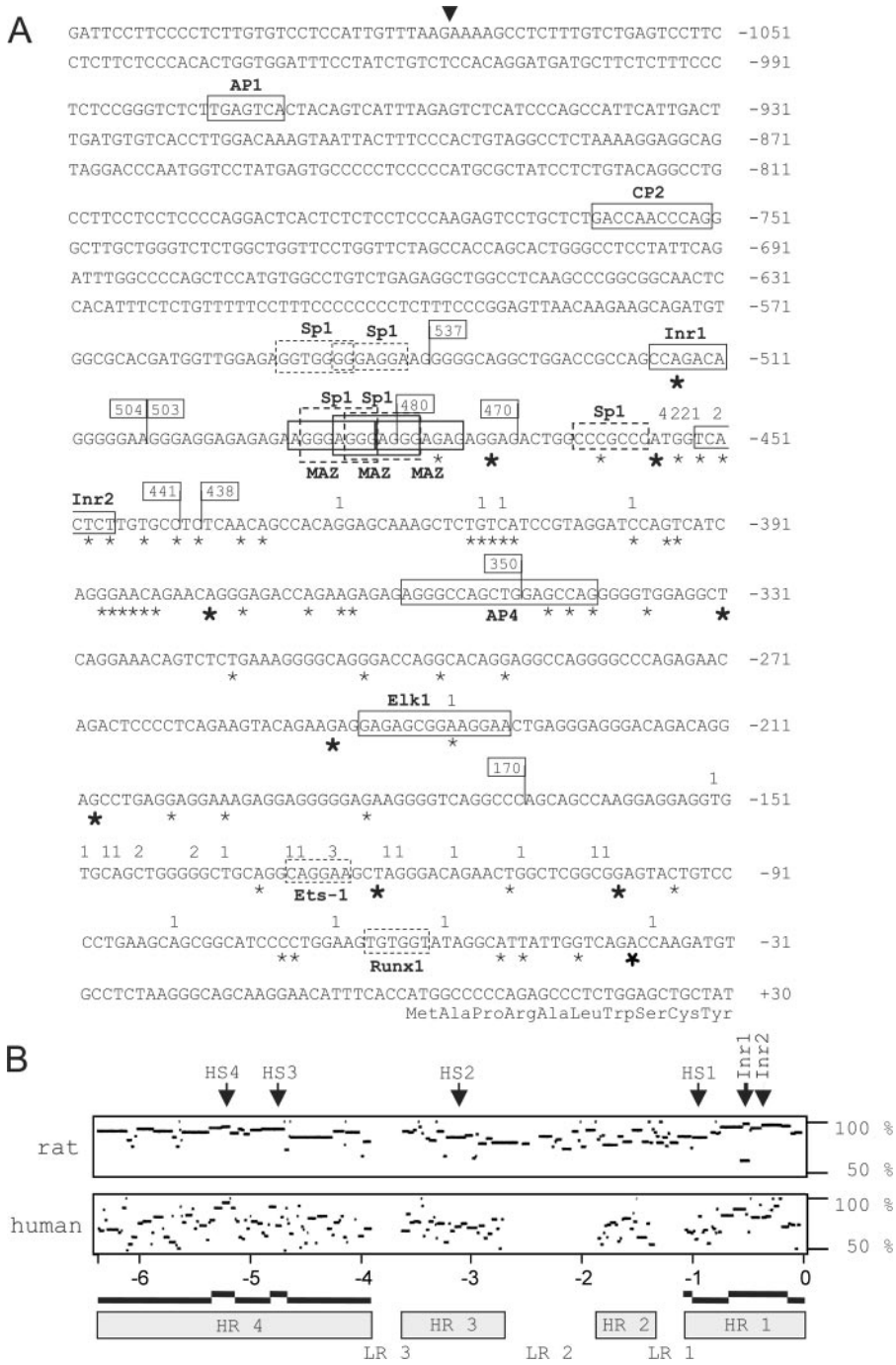

FIG. 1. **Analysis of the 5′-flanking region.** *A*, determination of transcription initiation sites of *Emilin1*. The figure shows the proximal 1110 nucleotides. Numbers above the sequence mark the 5′-end of RACE clones and indicate the number of clones with the same end. *Asterisks* below the sequence denote the position of the 5′-end of bands produced by RNase protection; *small* and *large asterisks* denote faint and strong bands, respectively. The sequences of putative transcription factor binding sites are *boxed*: *boxes* with *continuous lines* refer to binding sites conserved in the three species examined (mouse, rat, and human); *boxes* with *dashed line* delimit binding elements detected only in the mouse sequence. *Numbered flags* define the position of the 5′-end (*right flag*) or 3′-end (*left flag*) of deletion constructs defined in Fig. 2. *Arrowhead* marks the 5′-end of the exonuclease deletion identified in this paper as the 1.1-kb fragment. *B*, percent identity plot of the 5′-flanking region of murine *Emilin1* gene with the corresponding rat and human *Emilin1* genomic sequences, generated by pairwise alignment with the MultiPipMaker program. The percentage sequence identities (50–100%) are shown on the *y* axis. The scale (in kb) on the *x* axis refers to the position in the murine sequence. Stretches of sequence identity are indicated by *horizontal lines*, corresponding to length and percentage of nucleotide sequence identity. Position of the Initiator sequences (*Inr1* and *Inr2*) and of the DNase I-hypersensitive sites (*HS1* to *HS4*) is indicated by *vertical arrows*. The existence of three regions of low sequence (<50%) similarity (*LR1* to *LR3*) defines four regions of high similarity (*HR1* to *HR4*). The presence of peaks of very high sequence similarity can further divide high similarity regions into subregions, indicated in the figure by *alternating thick lines*.

passing the entire gene was subcloned from a BAC construct (12). In addition to the *Emilin1* gene, which spans 7479 bp and includes eight exons, the fragment contains 8651 bp of 5′-flanking sequence and 632 bp of the intergenic region between *Emilin1* and *Khk* (Fig. 3*C*). The characterization of the regulatory sequences contained in the 16.8-kb fragment started with the identification of transcription start sites and promoter assays in cell cultures using appropriate deletions.

Determination of the Transcription Start Sites—To map the transcription start site(s) of the *Emilin1* gene, 5′-RACE and RNase protection assays were employed. 40 clones obtained with 5′-RACE were sequenced. Although variable, the 5′-end of the clones was frequently found within two regions as follows: one is a broad region between −90 and −160 bp from the start codon, and the other corresponds to the short sequence of putative Initiator2 (Fig. 1). RNase protection analysis showed more than 60 different protection products falling between position −515 and the first codon. Three major transcription start sites were identified at −40, −101, and −123 bp; they were embedded in a region between −38 and −152 bp containing more than half of the transcription start sites determined by RACE (Fig. 1). A group of protected bands coincided with a major clustering of RACE products at putative Initiator2 (Fig. 1). A single protected band was located at position −515, corresponding to putative Initiator1. These results strongly suggest that multiple transcription start sites are utilized for mouse *Emilin1* transcription. They also indicate that transcription begins preferentially at putative Initiator2 but also at the other upstream putative Initiator1.

Functional Analysis of the 5′-Flanking Region—To characterize the functional role of *Emilin1* regulatory regions, different cell lines were transfected with CAT constructs, including

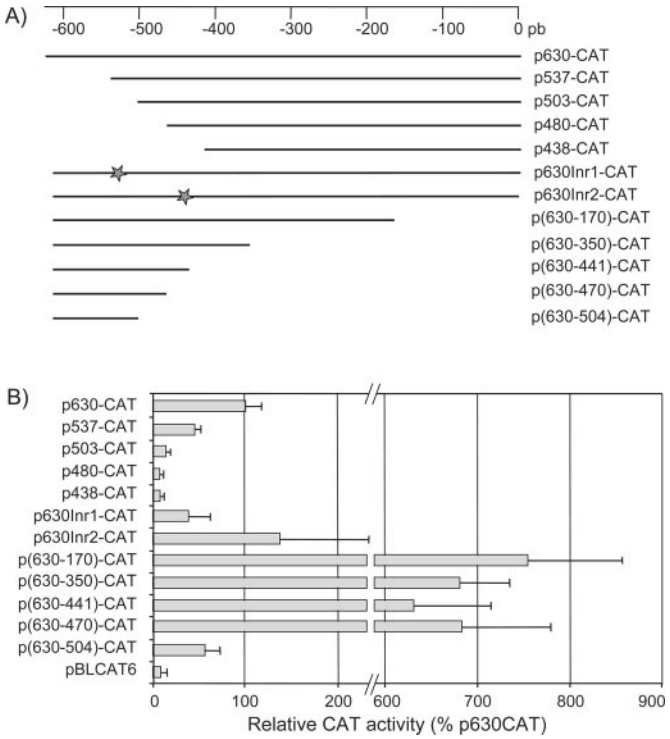

FIG. 2. **Functional promoter analysis of the 5′-flanking region of *Emilin1*.** *A*, representation of the sequence in different CAT constructs. The *numbers* indicate the limits of sequence present in the construct as nucleotides upstream of the first codon (see Fig. 1). *Stars* indicate the position of mutated Initiator sequences. The indication of base −1 in the 5′ deletions has been omitted. *B*, CAT promoter assays obtained with the constructs depicted in *A*.

various portions of the 5′-flanking sequence. Our initial analysis was focused on the proximal 630-bp fragment, from which several CAT chimeric constructs were synthesized carrying 5′ and 3′ deletions (Fig. 2*A*). The constructs were used for transient promoter assays in NIH3T3 cells, and the results are reported in Fig. 2*B*. Expression of the reporter gene was halved when the sequence upstream of −537 was cut out (compare constructs p630-CAT and p537-CAT). Further deletion of the region from −537 to −503 determined a considerable decrease of CAT activity (compare constructs p537-CAT and p503-CAT). Additional removal of 5′-sequences from −503 to −480 lowered gene reporter expression to background levels. This analysis indicates the presence of a core promoter region between −480 and −630. This sequence includes putative Initiator1 and several potential binding sites for transcription factor Sp1 (Fig. 1). The importance of the Initiator1 sequence was confirmed by the observation that mutation of the site significantly decreased CAT activity (Fig. 2, construct *p630Inr1-CAT*). On the contrary, the Initiator2 sequence did not contribute appreciably to core promoter function, as shown by the analysis of 5′ deletions (compare constructs *p480-CAT* and *p438-CAT* in Fig. 2) and the lack of reduction of promoter activity when the site was mutagenized (Fig. 2, *p630Inr2-CAT*).

To dissect more accurately the functional promoter elements, 3′ deletions of the 630-bp fragment were also analyzed. Removal of 169 bp from the 3′-end resulted in a 7-fold increase of reporter gene expression (Fig. 2, compare construct *p630-CAT* and construct *p(630–170)-CAT*). This high activity was not changed upon removal of an additional 300 bp (Fig. 2, constructs *p(630–350)-CAT*, *p(630–441)-CAT*, and *p(630–470)-CAT)*. On the contrary, activity was strongly reduced (more than 10-fold) with further deletion of 34 bp (Fig. 2, construct *p(630–504)-CAT*, which contains Initiator2 and some Sp1-binding sites). Finally, removal of sequences from −504 to −537, which include Initiator1, lowered CAT activity to background levels (data not shown). These results confirm the role of sequences comprising the two Initiators and Sp1-binding sites; in addition, they imply the presence of important regulatory elements within 169 bp upstream of the first codon.

We then investigated CAT constructs containing larger portions of 5′-flanking sequences (1.1, 3.0, 4.0, 6.0, and 8.7 kb). The plasmid carrying the 1.1-kb fragment exhibited 2-fold CAT expression compared with p630-CAT, indicating the presence of activating elements between −0.63 and −1.1 kb. On the other hand, no difference was detected between 1.1 kb and the longer constructs (data not shown). Because of this limitation, alternative methods were used to highlight regions potentially relevant for transcription, including the determination of chromatin DNase I-hypersensitive sites, sequence analysis using computer programs, and expression of promoter constructs in transgenic mice. These data are presented below.

Analysis of Chromatin DNase I-hypersensitive Sites—DNase I-hypersensitive sites in chromatin are structural landmarks indicative of control regions involved in constitutive and tissue- and/or stage-specific transcription (26, 27). For several ECM genes, DNase I-hypersensitive sites have been found associated with regions that control transcriptional regulation, particularly in a tissue-specific manner (7, 8, 28, 29).

DNase I-hypersensitive sites were analyzed within the 16.8-kb fragment using cell lines of different origin, including NIH3T3 fibroblasts, C2C12 myoblasts, MC615 chondrocytes, BC3H1 smooth muscle cells, EL4 lymphocytes, and endothelial cell lines cEnd and H.end. Northern blotting analysis revealed that all cell lines expressed *Emilin1* mRNA with the exception of H.end (Fig. 3*A*). In addition to the uncleaved 16.8-kb fragment, the probe hybridized to six species (11.6, 7.7, 5.5, 3.9, 3.5, and 2.5 kb) in DNA samples of NIH3T3 nuclei treated with DNase I (Fig. 3*B*). These species corresponded to hypersensitive sites at approximately +3000 (HSi), −950 (HS1), −3100 (HS2), −4750 (HS3), −5150 (HS4), and −6150 (HS5) from the first codon (Fig. 3*C*). Sites HSi and HS5 were present in all the cell lines analyzed with similar intensity, including the nonexpressing cell H.end, and are to be considered constitutive (*i.e.* present in chromatin independently from *Emilin1* gene expression) hypersensitive sites. Remarkable differences depending on the cell type could be noted for the other four sites. HS1 was present in all cell lines expressing *Emilin1* mRNA and appeared very strong in NIH3T3 and MC615 cells, less evident in C2C12 myoblasts, and very faint in BC3H1, cEnd, and EL4 cells. It was lacking in nonexpressing H.end cells. The band corresponding to hypersensitive site HS2 was very weak in NIH3T3, C2C12, and BC3H1, stronger in MC615, and absent in the other lines. HS3 was faint in all cell lines. HS4 was present only in NIH3T3, C2C12, and MC615 cells, being strongest in the last one. The presence of hypersensitive sites exclusively in *Emilin1*-expressing cells and the variable intensity of each band depending on the cell type suggest that the sites may correspond to regulatory regions involved in tissue-specific transcriptional regulation of the *Emilin1* gene.

Sequence Comparisons—The 5′-flanking region was compared with the corresponding rat and human sequences using the MultiPipMaker (bio.cse.psu.edu/) (30) (Fig. 1*B*) and Blast softwares (data not shown). Three regions of low similarity (<50%) divided the sequence into four regions of high similarity (Fig. 1*B*). Moreover, two of the high similarity regions could be further subdivided into subregions by the presence of peaks of very high similarity; as a consequence, the sequence comparison distinguished 14 sequences in the 5′-flanking region as summarized in Fig. 1*B* and in Table I. One important result was apparent from

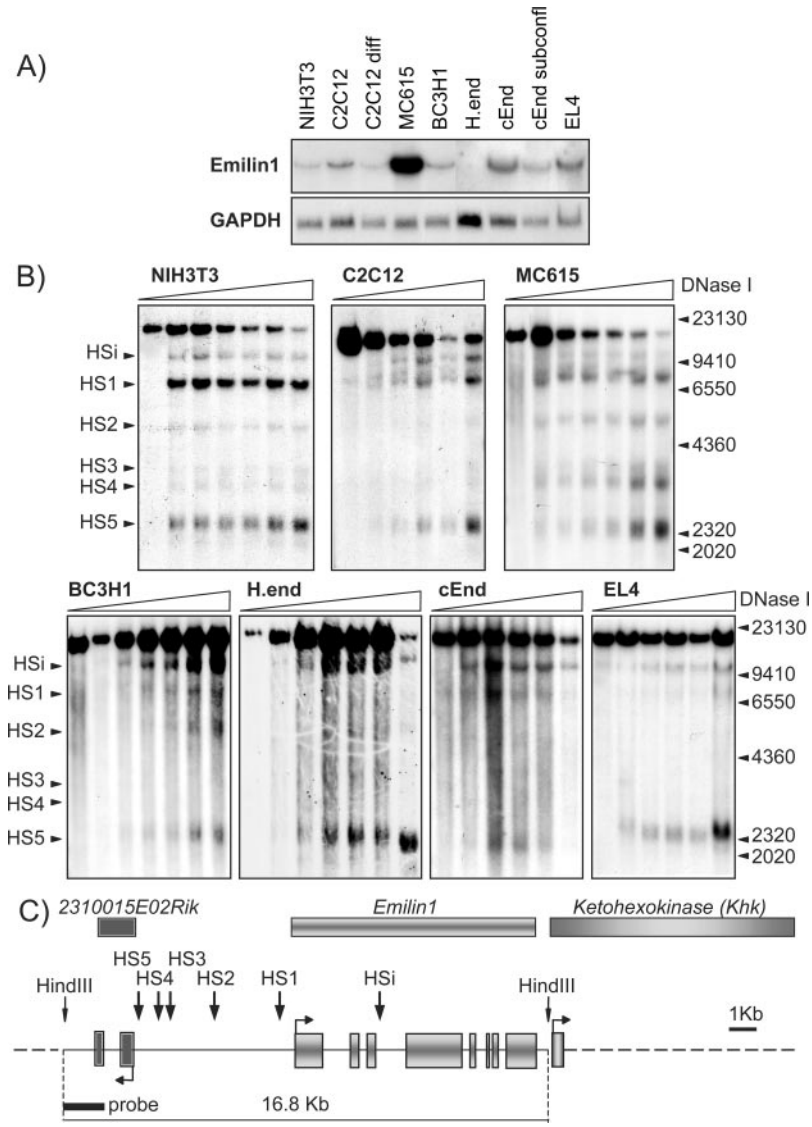

FIG. 3. **Identification of chromatin DNase I-hypersensitive sites.** *A*, expression of *Emilin1* mRNA in the different cell lines. *B*, Southern blotting analysis of DNA from nuclei digested with different doses of DNase I using the probe indicated in *C*. *C*, genomic organization and location of chromatin DNase I-hypersensitive sites within the 16.8-kb fragment investigated in this study. *Subconf*, subconfluent culture; *diff*, cultures grown under differentiation conditions; *GAPDH*, glyceraldehyde-3-phosphate dehydrogenase.

this analysis: three out of four chromatin DNase I-hypersensitive sites mapping in the 5′-flanking region (HS1, HS3, and HS4) coincided with peaks of higher similarity (Fig. 1*B* and Table I).

We next analyzed the sequences of the three species for the presence of conserved transcription factor binding sites using MatInspector (www.genomatix.de/cgi-bin//matinspector_prot-.mat_fam.pl) and TRANSFAC (www.biobase.de/cgi-bin/biobase/transfac/7.2/match/bin/match.cgi) search programs. Several conserved sites could be detected with both programs (Table I and Fig. 1*A*), and the data can be summarized as follows. 1) No conserved putative sites could be detected in the low similarity regions. On the contrary, each high similarity region contained at least two sites. 2) The sequence −156 to −679, representing the longest stretch of very high similarity, contained several sites, including the two Initiator sequences. As *Emilin1* lacks canonical TATA and CAAT boxes, it is possible that the two Initiators contribute to the localization of the transcription initiation sites in different species. 3) The peaks of very high similarity coincident with chromatin-hypersensitive sites HS1, HS3, and HS4 enclosed at least one putative binding element (Table I). In particular, an AP1 recognition sequence (Fig. 1*A*) was found in the 144-bp-long peak where HS1 was mapped; a cluster of five binding elements were located in the 132-bp fragment that overlapped HS3; one SRF binding sequence was found in the 240-bp peak where HS4 was located. 4) Outside the very high similarity peaks, conserved sites could be detected in some, but not all, subregions.

Analysis of Regulatory Regions Using Transgenic Mice—To locate transcriptionally significant regions, we also exploited expression of promoter-reporter gene constructs *in vivo*. The series of constructs tested is depicted in Fig. 4. One group of constructs included 5′-flanking sequences extending for different lengths upstream of the first codon (constructs 0.6-*lacZ*, 1.1-*lacZ*, 3.0-*lacZ*, and 8.7-*lacZ*). Deletion of the proximal 169 bp from 8.7-*lacZ* produced construct 8.7Δ-*lacZ* that was derived in order to investigate the role of the small sequence that showed a dramatic influence on promoter-CAT construct activities *in vitro* (see Fig. 2). Finally, a group of constructs included sequences downstream of the first codon. The third intron was included in one construct (8.7-*lacZ*-intron) because of a relatively higher density of putative transcription factor binding sites detected in the mouse in this intron compared with the other introns (data not shown). Two additional constructs (0.6-*lacZ*-gene and 8.7-*lacZ*-gene) contained the entire region of the *Emilin1* gene (excepting for the first 77 coding bases included in the first exon) and most of the intergenic sequence between *Emilin1* and *Khk* genes. Expression of promoter-*lacZ* constructs was studied mainly on founder embryos at E14.5. Mouse lines were derived with some of the constructs for analysis of transgene expression at different developmental stages,

TABLE I
Organization of the 5′-flanking sequence of Emilin1 revealed by sequence comparison among species

Regions[a]	Sub-regions (length in bp)[a]	Transcription factor binding site (position)[b]	Localization of chromatin DNaseI hypersensitive sites[c]	Similarity[d]
HR 1: from -1 to -1055	-1 to -155 (155)	No conserved sites	None	55
	-156 to -679 (524)	Elk1 (-243) AP4 (-359) Initiator (-453) MAZ (-481) MAZ (-485) MAZ (-489) Initiator (-517)	None	83
	-680 to -911 (232)	CP2 (-761)	None	55
	-912 to -1055 (144)	AP1 (-978)	Contains HS1	64
LR 1	-1056 to - 1339 (284)	No conserved sites	None	<50
HR 2	-1340 to -1847 (508)	GATA3 (-1699)	None	60
LR 2	-1848 to -2703 (856)	No conserved sites	None	<50
HR 3: from -2704 to -3619	-2704 to -3619 (916)	STAT (-2792) Ets-1 (-2794)	Contains HS2	50
LR 3	-3620 to -3893 (274)	No conserved sites	None	<50
HR 4: from -3894 to -6379	-3894 to -4684 (791)	ER (-4370) CP2 (-4414)	None	56
	-4685 to -4816 (132)	Nkx3.2 (-4758) Myogenin/NF1 (-4779) FoxJ (-4808)	Contains HS3	78
	−4817 to -5115 (299)	C/EBP (-4945) STAT (-4960)	None	61
	−5116 to -5355 (240)	SRF (-5340)	Contains HS4	84
	-5356 to -637 (1024)	No conserved sites	None	64

[a] Interspecies (mouse, rat, and human) sequence comparison was performed with the MultiPipMaker program (Fig. 1*B*). The analysis has identified three low similarity regions (LR) and four high similarity regions (HR) and several subregions within some HR regions (see Fig. 1*B*).
[b] The list of putative transcription factor-binding sites is limited to those identified with both programs used (MatInspector and TRANSFAC). The position of the site is indicated by the 5′-nucleotide of the consensus sequence.
[c] Chromatin DNase I-hypersensitive sites are indicated as defined in Fig. 1.
[d] Average of percentage of similarity of mouse *versus* human sequences within the indicated region is shown.

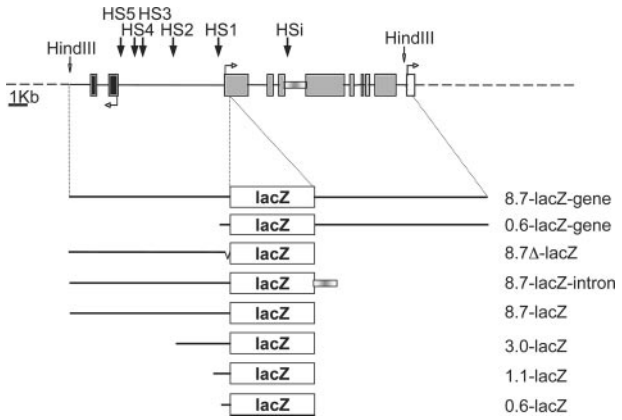

FIG. 4. **Constructs tested for promoter activity in transgenic mice.** The reporter gene *lacZ,* coding for the β-galactosidase gene of *E. coli*, is fused to the indicated sequences derived from the 16.8-kb fragment. The intron sequence of construct 8.7-*lacZ*-intron corresponds to intron 3.

considering that expression of *Emilin1* mRNA begins soon after implantation (13).

The results obtained with different constructs at E14.5 are summarized in Table II, which reports the frequency of expression in a specific tissue over the total number of expressing mouse lines for each construct. The tissues considered in Table II are those that exhibited positive staining by *in situ* hybridization (13) and immunohistochemistry.[2] The following features of transgene expression are deduced from Table II.

1) The frequency of expression of the largest construct (8.7-*lacZ*-gene) was 100% in every tissue. Moreover, expression at positive sites was strong, as most cells were labeled (see some examples in Fig. 5), and staining became evident very quickly during incubation in the X-gal solution (usually 10–20 min). This suggests that the 16.8-kb fragment analyzed contains all or at least the major regulatory sequences responsible for tissue-specific transcription of *Emilin1*.

2) The absence of ectopic expression with the 8.7-*lacZ* and 8.7-*lacZ*-intron indicates that the 5′-flanking region contains sequences limiting transcription in tissues where *Emilin1* is not produced, like embryonic epithelia and central nervous system (13). One of these sequences is likely located within the proximal 169 bp, as deletion of this stretch from the 8.7-kb fragment (construct 8.7Δ-*lacZ*) gives rise to ectopic expression of the transgene at high frequency. However, this is not the only sequence involved, because the shorter constructs (0.6-*lacZ*, 1.1-*lacZ*, and 3.0-*lacZ*) also exhibit frequent ectopic expression. Therefore, it can be hypothesized that sequences limiting ectopic expression are located in the proximal 169 bp and between −3.0 and −8.7 kb. Paradoxically, ectopic expression was present in 50% of lines generated with the largest construct, 8.7-*lacZ*-gene (see "Discussion").

3) Only in tendons and ligaments was a single region (the proximal 630 bp) sufficient to produce maximal expression frequency, although the staining was weak. The same region was frequently active in connective tissue associated with muscle (Fig. 5*D*). Additional sequences, however, were necessary to increase staining intensity in these tissues.

4) Expression of the *lacZ* transgene in all other tissues requires more than one region. Applying the same reasoning used above for ectopic expression, the regions necessary to achieve maximal frequency of expression in different tissues can be

[2] A. D'Urso, P. Braghetta, and G. M. Bressan, unpublished data.

TABLE II

Frequency of expression of Emilin1 promoter-lacZ constructs in different tissues

Constructs are defined in Fig. 4. Mouse embryos were analyzed at E14.5. After whole mount X-gal staining, the embryos were observed *in toto* and then processed for histological analysis. Staining was examined in tissues where *Emilin1* mRNA was observed to be expressed at the same age (13).

	0.6-*lacZ*	1.1-*lacZ*	3.0-*lacZ*	8.7-*lacZ*	8.7-*lacZ*-intron	8.7Δ-*lacZ*	0.6-*lacZ*-gene	8.7-*lacZ*-gene
Skin and annexes								
Subepidermal mesenchyme	2/6	3/7	3/7	8/8	13/13	5/6	1/6	4/4
Vibrissae mesenchyme	1/6	5/7	4/7	8/8	12/13	6/6	1/6	4/4
Cornea and sclera mesenchyme	1/6	3/7	3/7	8/8	13/13	6/6	1/6	4/4
Skeletal system								
Perichondrium	0/6	5/7	4/7	8/8	11/13	6/6	0/6	4/4
Cartilage	2/6	6/7	4/7	8/8	12/13	6/6	2/6	4/4
Tendons/ligaments	6/6	6/7	6/7	8/8	13/13	6/6	5/6	4/4
Intervertebral disks	0/6	4/7	4/7	8/8	11/13	6/6	0/6	4/4
Limb bud mesenchyme	1/6	6/7	6/7	8/8	13/13	6/6	0/6	4/4
Muscle (fasciae and interstitial tissue)	3/6	6/7	4/7	6/8	8/13	5/6	4/6	4/4
Circulatory system:								
Endocardium	0/6	0/7	1/7	3/8	8/13	0/6	0/6	4/4
Myocardium	0/6	0/7	0/7	2/8	3/13	0/6	1/6	4/4
Pericardium (visceral)	0/6	0/7	1/7	3/8	5/13	0/6	1/6	4/4
Endocardial cushions/valves	0/6	0/7	2/7	3/8	1/13	2/6	1/6	4/4
EC[a] of embryonic blood vessels	2/6	4/7	3/7	8/8	13/13	5/6	1/6	4/4
Media of medium/large vessels	0/6	1/7	2/7	7/8	11/13	5/6	1/6	4/4
Digestive system								
Mesenchyme and SMC of mucosae[b]	0/6	5/7	2/7	8/8	13/13	6/6	1/6	4/4
Large cells in liver[c]	0/6	0/7	0/7	0/8	2/13	0/6	1/6	4/4
Lung mesenchyme	0/6	1/7	0/7	3/8	8/13	0/6	2/6	4/4
Kidney mesenchyme	0/6	0/7	0/7	2/8	4/13	0/6	1/6	4/4
Submaxillary gland and pancreas mesenchyme	0/6	0/7	0/7	0/8	0/13	0/6	2/6	4/4
Mesenchyme at other locations								
Perineural mesenchyme	0/6	3/6	3/7	8/8	13/13	5/6	1/6	4/4
Mesenchymal condensations[d]	2/6	6/7	6/7	8/8	13/13	6/6	4/6	4/4
Organ capsules	1/6	3/7	3/7	8/8	9/13	6/6	0/6	4/4
Extra-embryonic tissues								
Fetal blood vessels of placenta	0/6	0/7	0/7	8/8	12/13	0/6	1/6	4/4
Umbilical cord mesenchyme	0/6	3/7	2/7	8/8	13/13	6/6	1/6	4/4
Spongiotrophoblast	0/6	0/7	0/7	0/8	0/13	0/6	2/6	4/4
Ectopic expression[e]	6/6	5/7	5/7	0/8	0/13	5/6	5/6	2/4

[a] The abbreviations used are as follows: EC, endothelial cells; CNS, central nervous system; SMC, smooth muscle cells.
[b] More frequently the intestine.
[c] These cells likely represent fetal megakaryocytes (38).
[d] Positive locations include mesenchyme of branchial arches, frontal region, pinnae of ear, and nasal folds.
[e] *Emilin1* mRNA is not expressed in all embryonic epithelia and in cells of the nervous tissue (13,38).

identified. This allows tentative assignment of various tissues to groups differing from the regions involved in tissue-specific regulation, as shown in Table III.

5) The region included between −1.1 and −3.0 kb does contribute significantly to tissue-specific transcription only in endocardial cushions, where expression frequency raises from 0% for 1.1-*lacZ* to 28% for 3.0-*lacZ*. In all other tissues, the expression frequency was essentially the same for 1.1-*lacZ* and 3.0-*lacZ* constructs, indicating a marginal effect for the region −1.1 to −3.0 kb.

6) For some tissues, expression is critically dependent on the presence of the proximal 169 bp of the 5′-flanking region. Notable examples are the heart (Fig. 5, *A–C*), kidney, and lung mesenchyme (data not shown). Deletion of the 169 bp from the 8.7-kb promoter region abolished completely transgene activity (compare expression frequency of 8.7-*lacZ* with that of 8.7Δ-*lacZ* in Table II). Of interest, the sharp margins of staining obtained with the nuclear *lacZ* marker gene revealed a previously unrecognized feature of *Emilin1* expression in the heart, whereas endocardial cells of both left and right ventricle are equally active in transcription, only the myocardium of the right ventricle expresses the gene at high levels (Fig. 5, *B* and *C*).

7) One striking feature has been observed for some cell types that require different regulatory regions for maximal expression, depending on the particular anatomical district. One example is interstitial mesenchyme associated with organs such as skin, glands, and lung. In the skin, the presence of the entire 5′-flanking region was sufficient for high frequency and high level expression (Table II and data not shown). On the contrary, expression in lung required the simultaneous involvement of the 5′-flanking and the gene region, as apparent in Fig. 5; individually, the two regions activate only scattered cells (Fig. 5, *E* and *F*) in a limited number of transgenic lines; in the presence of both regions, all mesenchymal cells of lung are intensely stained (Fig. 5*G*) in each mouse line. A second example is provided by endothelial cells, which fall into two groups (Table II). Endothelial cells of blood vessels within the embryo apparently require regions enclosed between −0.17 and −1.1 and between −3.0 and −8.7 kb, with the region −0.17 and −1.1 making a significant contribution (Fig. 5*H* and Table II). On the other hand, endothelial cells of fetal blood vessels of labyrinthine placenta depend on the simultaneous contribution of the proximal 170-bp sequence and the −3.0 to −8.7 region; the 0.17 to −1.1 region apparently was ineffective (Fig. 5, *J–L* and Table II).

8) In three tissues, expression was strongly dependent on

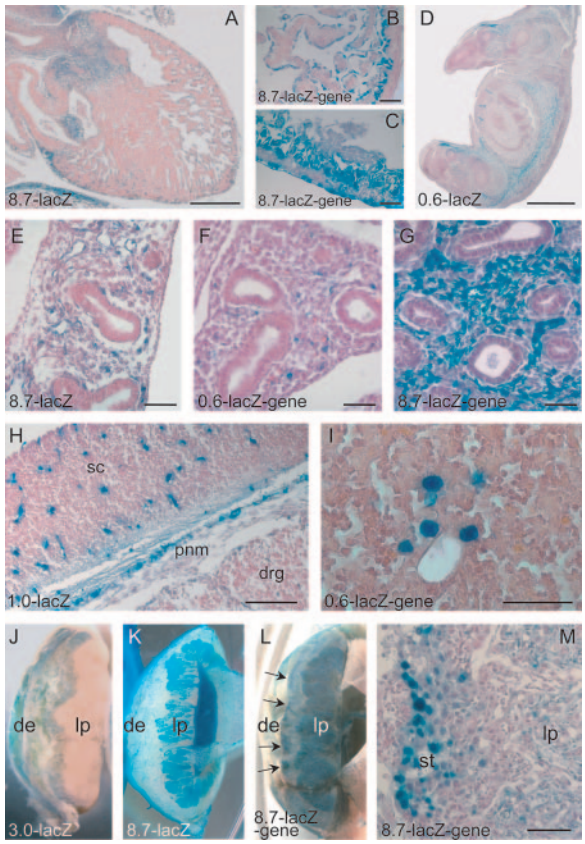

FIG. 5. **Expression of *lacZ* fusion constructs in E14.5 mouse embryos.** Embryos were incubated with X-gal to reveal *lacZ* expression (*blue color*), sectioned, and stained with hematoxylin and eosin. The promoter construct used to generate the transgenic mice is indicated in each panel. *A,* section of heart showing staining of pulmonary and aortic valves, ventricular endocardium, and pericardium. *Bar* = 250 μm. *B* and *C,* expression of 8.7-*lacZ*-gene construct in left (*B*) and right (*C*) ventricle. Staining is found in endocardium of both ventricles, whereas only myocardium of right ventricle is positive. *Bars* = 50 μm. *D,* expression in cells of tendons, ligaments, and fasciae of the shortest construct 0.6-*lacZ*. *Bar* = 1 mm. *E–G,* X-gal staining of lung from transgenic mice carrying the indicated transgene; high expression in mesenchymal cells is achieved only with the construct that includes the complete 5′-flanking and the gene region. *Bars* = 50 μm. *H*, intense staining of endothelial cells of blood vessels of spinal cord (*sc*) and perineural mesenchyme (*pnm*). *drg*, dorsal root ganglion. *Bar* = 100 μm. *I,* group of cells with X-gal-positive large nuclei in liver. *Bar* = 50 μm. *J–L,* whole mount preparations of placenta comparing expression of different transgenes. The organ has been cut in two parts, and the section surface is shown in each panel. The 5′-flanking region upstream of 3.0 kb is necessary for transgene expression in blood vessels of labyrinthine region of placenta (*lp*). Staining in spongiotrophoblast cell clusters (*arrows*) is seen only when both the −3.0 to −8.7 and the gene region are present (*L*). *de*, decidua. *M,* histological section of placenta of *L* showing high level expression in spongiotrophoblast (*st*) and endothelial cells of blood vessels in the labyrinthine region of placenta (*lp*). *Bar* = 100 μm.

regulatory regions contained within the *Emilin1* gene; no transgene activity was found in the absence of sequences of the gene region, even when the entire 5′-flanking sequence was present. These tissues include large cells in liver, likely fetal megakaryocytes (Fig. 5*I*), mesenchyme associated with some glands (data not shown), and spongiotrophoblast (Fig. 5, *L* and *M*). The regulatory regions enclosed in the gene sequence are likely not the same for the three tissues. Indeed, some activating elements for megakaryocytes are contained in the third intron, as suggested by X-gal staining of the cells in some mouse lines derived with construct 8.7-*lacZ*-intron. On the contrary, spongiotrophoblast and gland mesenchyme express the transgene only when the entire *Emilin1* gene sequence is present (construct 8.7-*lacZ*-gene).

Role of Regulatory Regions during Development—Emilin1 expression during mouse development exhibits a complex stage-specific and tissue-specific regulation (13). Therefore, experiments were carried out to define the role of different regulatory regions in the transcriptional activation of *Emilin1* at different stages of development, with the following results.

1) *Emilin1* mRNA expression begins soon after implantation at E6.5–7.5 in the ectoplacental cone and extra-embryonic visceral endoderm. These tissues exhibited high frequency X-gal staining with the 8.7-*lacZ*-gene construct (3 out of 3 lines) (Fig. 6*A*), no staining in all five 8.7-*lacZ* (Fig. 6*B*) and two 0.6-*lacZ*-gene mouse lines, and low frequency (1/7) labeling in 8.7-*lacZ*-intron mouse lines (data not shown). During further development, the two tissues give rise to spongiotrophoblast and yolk sac epithelium, respectively. The simultaneous contribution of the two regions (5′-flanking and gene) to high level expression in these tissues was also evident at E8.5; the yolk sac epithelium was positive in all 8.7-*lacZ*-genes (Fig. 6*C*) but not in 8.7-*lacZ* (Fig. 6*D*) or 0.6-*lacZ*-gene transgenics and at low frequency (1/7) in the 0.6-*lacZ*-gene (data not shown). A similar condition was detected at E12.5 (data not shown). These results suggest that the simultaneous presence of the 5′-flanking region and the gene region is necessary for high frequency early transcription in ectoplacental cone and extra-embryonic visceral endoderm and derived tissues. At E7.5, transgene expression was found also in cells of embryonic and extra-embryonic mesoderms (Fig. 6*A*). However, activation in these tissues, which was particularly strong in allantoid, depended solely on the presence of the entire 5′-flanking sequence (Fig. 6*B*).

2) Soon after beginning of gastrulation, *Emilin1* mRNA is intensely produced in the cardiovascular system, where expression persists at high levels during further development. At that stage, 8.7-*lacZ* and 8.7-*lacZ*-gene transgenes were strongly activated in endothelial cells of blood islands (Fig. 6, *D* and *C*) and of blood vessels, and in all cells of the endocardium (Fig. 6*E*) at high frequency (5/5 and 3/3 lines respectively). Myocardial cells were stained with local differences; the most intense reaction was detected in the bulbous arteriosus, whereas cells of the common ventricle and atrium were weakly or not labeled (Fig. 6*E*). Moreover, staining of the truncus arteriosus wall was very strong (Fig. 6*E*). No significant differences could be noted on the expression pattern of the 8.7-*lacZ* and 8.7-*lacZ*-gene constructs at E8.5 in heart tissues. At later stages (E11.5, E12.5, and E14.5), however, the frequency of expression of the latter construct remained maximal, although it decreased significantly for the former, becoming 50–60% in endo-myocardium and cardiac cushions at E11.5 and E12.5 and about 37% at E14.5 (see Table II). The behavior of the shorter constructs was also examined (0.6-*lacZ*, 1.1-*lacZ*, and 3.0-*lacZ*). The transgene 0.6-*lacZ*-gene was not expressed in early cardiovascular systems (data not shown). At later stages expression was never detected in the heart and appeared at low frequency at E14.5 in blood vessels. The results were different for expression of transgenes 1.1-*lacZ* and 3.0-*lacZ* in the heart. At variance with what was observed at E14.5, when frequency of expression in the endocardium was 0% (see Table II), expression was seen in scattered cells of the endocardium and cardiac cushions at E12.5 (2/2 and 3/3 lines for the two constructs respectively), although no transgene activity was detected in the myocardium of these mouse lines. Similarly, whereas the sequence −1 to −169 bp from the first codon was indispensable for expression in heart tissues at E14.5 (Table II, compare construct 8.7-*lacZ* and 8.7Δ-*lacZ*), removal of this region decreased intensity (less cells and fainter staining) but did

TABLE III
Contributions of regulatory regions to tissue-specific transcription of the Emilin1 gene[a]

	from ATG to -169 bp	from -169 to -0.6 kb	from -0.6 to -1.1 kb	from -1.1 to -3.0	from -3.0 to -8.7 kb	3rd intron and/or other sequences in gene
Tendons/ligaments		*				
Perichondrium, intervertebral disks, perineural M[b], SMC of media of blood vessels, digestive system M, umbilical cord M			*		*	
Cartilage, limb bud M, subepidermal M, mesenchymal condensations, capsules, cornea, embryonic EC		*	*		*	
Fetal blood vessels of labirinthine placenta, suppression of ectopic expression	*				*	
Muscle M		*	*			*
Endocardium, myocardium, pericardium, lung M, kidney M	*				*	*
Endocardial cushions				*		*
Large cells in liver, glands M, spongiotrophoblast[c]			*			*

[a] The scheme has been deduced from data of Table II considering the frequency of expression of different constructs in the indicated tissues and the minimal set of regulatory regions necessary to reach the maximal (~100%) frequency. An asterisk marks the suggested contribution of the regulatory region to transcription in the indicated tissues.

[b] The abbreviations used are as follows: EC, endothelial cells; SMC, smooth muscle cells; M, mesenchyme.

[c] The role of 5′-flanking sequences in these tissues is inferred from our data but is not directly proven. When examined singularly, these sequences have always given 0% expression frequency (see Table II). On the other hand, the gene region linked to the proximal 5′-flanking sequence (construct 0.6-*lacZ*-gene) has produced only limited expression (frequency ≤ 33%). As expression frequency from the construct including both the entire 5′-flanking and the gene region is maximal, it must be assumed that the 5′-flanking region increases the transcriptional activity of the gene region.

not abolish expression of the transgene in endocardium and cardiac cushions at E12.5 (2/3 embryos scored positive); however, no staining was present in the myocardium (data not shown). These results suggest that timely transcriptional activation of *Emilin1* in the heart during development does not require the proximal 169 bp of the 5′-flanking sequence and the gene region. Nevertheless, both regions are necessary for sustained transcription at later stages.

3) A second tissue of early *Emilin1* expression is mesenchyme at almost any localization. Also for this type of cells, initial gene transcription was strongly dependent on the presence of the full set of 5′-flanking sequences (Fig. 6*E*), although it became more and more differentially regulated by various sets of regions during organogenesis (see Table II).

DISCUSSION

Transcription of genes such as *Emilin1*, whose expression is restricted to a given set of tissues and modulated during development, depends on several types of *cis*-regulatory elements. These include the following: the core promoter, which recruits the transcription machinery and directs accurate initiation of transcription; enhancer or silencer sequences, which activate or inhibit transcription in different tissues; locus control regions, originally functionally defined as dominant activating sequences that confer position-independent and copy number-dependent expression on a linked transgene in transgenic mice (31); and insulators/boundary regions, proposed to be required at the borders of a regulatory domain of a gene to counteract inappropriate effects of nearby heterochromatin and/or distal enhancers (32, 33).

Our report describes the first characterization of the regulatory setup of mouse *Emilin1*. The DNA fragment investigated comprised the gene and the 5′- and 3′-flanking sequences separating it from adjacent *2310016E02Rik* and *Khk* genes. The major transcription start sites were located in association with two Initiator consensus sequences identified as Initiator1 (−517 to −511) and Initiator2 (−453 to −447). Besides Initiator1, promoter assays in cell cultures have assigned important functional roles to sequences just upstream of it and between Initiator1 and Initiator2. These sequences contain putative binding sites for transcription factor Sp1 (nucleotides −551 to −540 and −489 to −479). Sp1-binding sites are often located close to the transcription initiation site in TATA-less promoters and contribute to transcriptional initiation (34). On the basis of these results, a reasonable suggestion is that the core promoter of the *Emilin1* gene is located between nucleotides 551 and 447 upstream of the first codon.

In addition to providing information on the core promoter region, assays in cultured cells using CAT deletion constructs have identified the proximal 169 bp as an important regulatory fragment. *In vitro* the activity of the fragment is repressive; *in vivo* its function is more complex, as it stimulates transgene expression in some tissues (mainly the heart) and, at the same time, it lowers the frequency of ectopic expression. It is likely that the fragment contains both stimulatory and inhibitory elements and that its overall function depends on the cellular context. This conclusion is also suggested by the presence of matches for transcription factors Ets-1 and Runx1, the latter of which is often involved in transcriptional repression (35). These sites are not conserved in rat and human, and the entire 167-bp sequence also exhibits species variations. Hence, the fragment may have a relevant function only in the mouse.

Promoter assays *in vitro* have also located an activating region between about −0.6 and −1.1 kb from the first codon. This region contains a conserved putative binding site for transcription factor AP1. The actual function of this element has not been investigated. However, the close matching of its sequence with the position of DNase I-hypersensitive site HS1 suggests that the AP1 recognition motif may be functional in

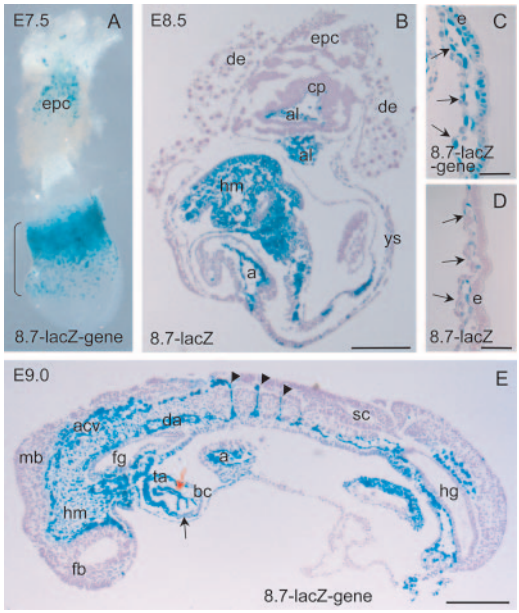

FIG. 6. **Expression of *lacZ* fusion constructs in mouse embryos at early post-implantation developmental stages.** The transgenic line analyzed is indicated in each panel. The embryonic age of some preparations is also reported. *A,* whole mount staining of E7.5 embryo of one 8.7-*lacZ*-gene line. X-gal-positive tissues include the ectoplacental cone (*epc*) and groups of cells at the embryonic-extra-embryonic border (*square bracket*). Histological analysis of this embryo revealed that the latter positive region includes cells of embryonic and extra-embryonic mesoderm and extra-embryonic visceral endoderm (data not shown). *B,* section of E8.5 embryo of one 8.7-*lacZ* line. Ectoplacental cone (*epc*) is not stained. The allantois (*al*) is strongly labeled, whereas the chorionic plate (*cp*) is not. At this age the two tissues contact and adhere to each other starting the formation of the chorio-allantoic placenta. *a*, atrium; *de*, decidua; *hm*, head mesenchyme; *ys*, yolk sac. *Bar* = 250 μm. *C* and *D,* section of yolk sac from E8.5 embryos. The yolk sac epithelium (*e*) is stained in 8.7-*lacZ*-gene (*C*) but not in the 8.7-*lacZ* (*D*) line, whereas endothelial cells of blood islands (*arrows*) are positive in both lines. *Bars* = 50 μm. *E*, sagittal section of E9.0 embryo. Strong *lacZ* staining is found in endothelial cells of blood vessels (*da*, dorsal aorta; *acv,* anterior cardinal vein; *arrowheads,* intersegmental arteries) and endocardium (*red arrow*). The myocardial layer of heart is also positive (*black arrow*). Intense expression is found in mesenchyme (*hm*, head mesenchyme). No expression is detectable in epithelial cells, like those of gut (*fg*, foregut; *hg*, hindgut) and in the central nervous system (*fb*, forebrain; *mb*, midbrain; *sc*, spinal cord). *a*, atrium; *bc*, bulbous cordis; *ta*, truncus arteriosus. *Bar* = 250 μm.

the regulation of *Emilin1*. The −0.6 to −1.1 region is also important for expression in mesenchymal and endothelial cells at several locations *in vivo* (Table II).

Unfortunately, *in vitro* promoter assays were not informative for the sequence upstream of −1.1 kb. Nevertheless, a combination of different methods, namely mapping of chromatin DNase I-hypersensitive sites, sequence comparisons, and expression of *lacZ* promoter constructs in transgenic mice, has succeeded in outlining the functional and structural features of this region. Overall, the three methods have given concurrent results that are also in agreement with the data obtained for the proximal 1.1-kb fragment using promoter assays. Sequences hypersensitive to DNase I are believed to reflect a local rearrangement of nucleosomes and possibly a local distortion in DNA topology due to binding of a large number of transcription factors (36). This suggestion was matched for hypersensitive sites of the *Emilin1* 5′-flanking region, which mostly map in small fragments with very high sequence conservation and are enriched for transcription factor DNA-binding motifs.

Sequence comparisons between species (mouse, rat, and human) have revealed four regions with higher similarity in the 5′-flanking sequence (Fig. 1*B* and Table I). These regions could be further subdivided into subregions because of the presence of short peaks of very high similarity (Table I). The region proximal to the first codon is 1055 nucleotides long and corresponds largely to the sequence that we have characterized more thoroughly using promoter assays *in vitro* (see above). The transcription start sites and two conserved Initiator elements, as well as putative binding sites for other transcription factors, are located within −156 to −679, a subregion of high interspecies similarity. The segment further upstream, which stimulates CAT activity by 2-fold in promoter assays, harbors the conserved AP1 recognition site mapping very close to DNase I-hypersensitive site HS1. The function of the second and third high similarity regions (nucleotides −1340 to −1847 and −2704 to −3619, respectively) were not well established in our analysis because of the fact that constructs were not derived in which the two regions were separated. In fact, the information from these regions comes from the comparison of the expression pattern of 3.0-*lacZ* (that contains the entire region two and part of regions three) with 1.1-*lacZ* and 8.7-*lacZ* (that lack and include both regions respectively. Nevertheless, the observation that construct 3.0-*lacZ* was expressed at slightly increased frequency in endocardial cushions compared with 1.1-*lacZ* suggests that either region may contribute to transcription levels in this tissue. The fourth region (from −3894 to −6379) is of key importance for expression in several tissues, particularly the circulatory system, and for achievement of high transcription levels in transgenic mice. Two hypersensitive sites (HS3 and HS4) are enclosed in this region. Most strikingly, both map within short sequences (132 and 241 bp, respectively), exhibiting very high similarity between species (about 80%) and containing conserved consensus sequences for a few transcription factors binding sites. The function of each individual hypersensitive site has not been investigated here. As a consequence, it is not known whether they contribute to regulation in specific tissues or act just to boost transcription levels.

The examination in transgenic mice has also shown that relevant activating regions lie within the gene sequence. However, no DNase I-hypersensitive sites related with expression of the gene have been identified within the gene sequence. The explanation may be due to either the complete absence of such sites within the gene or to the fact that hypersensitive sites are detectable only in tissues that are strongly dependent on the gene region for expression, and these tissues have not been examined here.

It is clear from our data that none of the different regulatory regions have the peculiar property necessary and sufficient for transcription at high frequency and intensity in a given tissue. The exception of the proximal 0.6-kb 5′-flanking region, which drives production of the transgene in tendons and ligaments with maximal frequency (see Table II), is only apparent, as staining in 0.6-*lacZ* lines was usually limited to a portion of potentially expressing cells and was of variable intensity. Instead, more than one region was required for high level expression in different tissues, and the same region could cooperate with other region(s) to activate transcription in a set of distinct tissues.

If regulatory regions act in such a combinatorial way, how is tissue specificity achieved? The best explanation comes from a model derived from studies on the hemoglobin genes (27). This hypothesis proposes that the key feature to establish an independent expression profile is the ability of the core promoter for a specific gene to communicate with strongly activating *cis*-regulatory elements in a particular tissue. In the model, intervening chromatin stretches are supposed to loop out even when they contain important regulatory sites for other nearby genes,

the reason being that these regions cannot positively interact with the promoter of the specific gene. The spatial unit of activating regulatory DNA regions for a specific gene is referred to as an active chromatin hub (ACH). Productive ACH formation underlies the correct gene expression, requiring the presence of protein factors with the appropriate affinities for each other bound to their cognate DNA sequences. The data reported here can be explained by suggesting that different regulatory sequences participate in the formation of an ACH, depending on the set of protein factors produced in each tissue where *Emilin1* is expressed.

An interesting aspect of our study is that the DNA segment analyzed contains information for transcription in every tissue where *Emilin1* products have been detected. This may suggest that the fragment harbors the complete set of regulatory regions necessary for appropriate *Emilin1* transcription. To establish whether this is the case, additional studies must be carried out for investigating the quantitative aspects of transcription stimulated by the 16.8-kb fragment as part of a suitable transgene. Such transgenes should exhibit position-independent and copy number-dependent expression, as expected for a locus control region. In the ACH model the locus control region is viewed as a part of ACH itself, contributing combinations of *cis*-regulatory elements that can establish position-independent gene expression. These studies are underway in our laboratory.

One feature of the ACH model is that stable formation of ACH is the critical event buffering against position effects in transgenic experiments and that stable enhancer-promoter interactions, rather than the presence of insulating borders, determine appropriate gene expression. This proposition explains an unexpected finding in our study concerning ectopic expression. Constructs comprising short stretches (less than 3 kb) of the 5′-flanking region were incorrectly transcribed at high frequency (more than 70% of transgenic mouse lines). This can be attributed to interaction of activating sequences unable to form an ACH with nearby promoters. The frequency dropped to zero with longer constructs, including 8.7 kb of 5′-flanking sequence, likely due to the fact that regulatory regions of the transgene were involved in an ACH and were therefore not available for interactions with promoters at the insertion site. Deleting the 5′ proximal 169 bp causes destabilization of the ACH and reappearance of ectopic expression. Most unexpectedly, the addition of the gene region (construct 8.7-*lacZ*-gene) reactivated ectopic expression in some lines (50% frequency). This paradox can be accounted for by observing that the extremities of the investigated 16.8-kb fragment contain possible regulatory regions for the two genes contiguous to *Emilin1*, *2310015E02Rik,* at the 5′-end and *Khk* (a housekeeping gene) at the 3′-end. In the above model, these sequences are assumed not to interact with the ACH, as they belong to a different functional transcription unit and can consequently synergize with local promoters. Consistent with this suggestion, unpublished work[3] has shown that the short intergenic region between *Emilin1* and *Khk* contains an activating sequence for *Khk* that is located within the 16.8-kb fragment studied here.

During development, the set of regulatory regions that induces maximal levels of *Emilin1* expression in some tissues may change. This conclusion is suggested by the analysis of transgenic embryos at different stages. One example is heart, where transgene production at early organ formation is driven mainly by the 5′-flanking region, whereas it becomes dependent on the gene region at mid-advanced organogenesis. A second example is mesenchyme, where initial activation relies mainly on the 5′-flanking region, whereas different arrays of regulatory regions direct expression in the various organs at later stages. These observations are also consistent with the ACH model, where the key event is a productive assembly of regulatory regions rather than the presence of a *cis*-acting module responsible for tissue specificity.

The analysis of transgenic mice has also revealed details of *Emilin1* expression that were not fully appreciated by using *in situ* hybridization (13). A significant one concerned myocardium, where expression was particularly high in the bulbous cordis and in the right ventricle, which largely derives from it (37). The functional meaning of this finding remains unknown, considering that the heart is normal in *Emilin1*-deficient mice (12). Nevertheless, it is tempting to speculate that Emilin1 may be involved in the acquisition of local differences of the right ventricle.

In addition to suggesting that transcription of *Emilin1* in different tissues is achieved through combinatorial cooperation between various regions, rather than being dependent on a single *cis*-activating region specific for each (or a small group of) tissue(s), the results reported here are relevant as they define a strong and compact regulatory region that can be exploited for expression of genes in tissues overlapping with *Emilin1* expression domains. Examples of such tissues are the ectoplacental cone, extra-embryonic visceral endoderm, megakaryocytes, and endothelial and mesenchymal cells during development. Further characterization of the various regulatory regions identified in this study will allow the design of more detailed sets of regulatory sequences for better tissue- and stage-specific gene expression.

Acknowledgments—We thank Dr. F. Bussolino for supplying the H.end cell line and Dr. K. Frka for technical assistance.

REFERENCES

1. Bridgewater, L. C., Lefebvre, V., and de Crombrugghe, B. (1998) *J. Biol. Chem.* **273,** 14998–15006
2. Leung, K. K., Ng, L. J., Ho, K. K., Tam, P. P., and Cheah, K. S. (1998) *J. Cell Biol.* **141,** 1291–1300
3. Liska, D. J., Reed, M. J., Sage, E. H., and Bornstein, P. (1994) *J. Cell Biol.* **125,** 695–704
4. Niederreither, K., D'Souza, R. N., and de Crombrugghe, B. (1992) *J. Cell Biol.* **119,** 1361–1370
5. Rossert, J., Eberspaecher, H., and de Crombrugghe, B. (1995) *J. Cell Biol.* **129,** 1421–1432
6. Rossert, J. A., Chen, S. S., Eberspaecher, H., Smith, C. N., and de Crombrugghe, B. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 1027–1031
7. Bou-Gharios, G., Garrett, L. A., Rossert, J., Niederreither, K., Eberspaecher, H., Smith, C., Black, C., and Crombrugghe, B. (1996) *J. Cell Biol.* **134,** 1333–1344
8. Fabbro, C., Braghetta, P., Girotto, D., Piccolo, S., Volpin, D., and Bressan, G. M. (1999) *J. Biol. Chem.* **274,** 1759–1766
9. Bressan, G. M., Daga-Gordini, D., Colombatti, A., Castellani, I., Marigo, V., and Volpin, D. (1993) *J. Cell Biol.* **121,** 201–212
10. Colombatti, A., Doliana, R., Bot, S., Canton, A., Mongiat, M., Mungiguerra, G., Paron-Cilli, S., and Spessotto, P. (2000) *Matrix Biol.* **19,** 289–301
11. Doliana, R., Bot, S., Bonaldo, P., and Colombatti, A. (2000) *FEBS Lett.* **484,** 164–168
12. Zanetti, M., Braghetta, P., Sabatelli, P., Mura, I., Doliana, R., Colombatti, A., Volpin, D., Bonaldo, P., and Bressan, G. M. (2004) *Mol. Cell. Biol.* **24,** 638–650
13. Braghetta, P., Ferrari, A., de Gemmis, P., Zanetti, M., Volpin, D., Bonaldo, P., and Bressan, G. M. (2002) *Matrix Biol.* **21,** 603–609
14. Doliana, R., Canton, A., Bucciotti, F., Mongiat, M., Bonaldo, P., and Colombatti, A. (2000) *J. Biol. Chem.* **275,** 785–792
15. Boshart, M., Kluppel, M., Schmidt, A., Schutz, G., and Luckow, B. (1992) *Gene (Amst.)* **110,** 129–130
16. Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
17. Burke, T. W., and Kadonaga, J. T. (1996) *Genes Dev.* **10,** 711–724
18. Braghetta, P., Fabbro, C., Piccolo, S., Marvulli, D., Bonaldo, P., Volpin, D., and Bressan, G. M. (1996) *J. Cell Biol.* **135,** 1163–1177
19. Mallein-Gerin, F., and Olsen, B. R. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 3289–3293
20. Ghigo, D., Arese, M., Todde, R., Vecchi, A., Silvagno, F., Costamagna, C., Dong, Q. G., Alessio, M., Heller, R., Soldi, R., Trucco, F., Garbarino, G., Pescarmona, G., Mantovani, A., Bussolino, F., and Bosia, A. (1995) *J. Exp. Med.* **181,** 9–19

[3] P. de Gemmis, C. Fabbro, D. Volpin, P. Bonaldo, and G. M. Bressan, manuscript in preparation.

21. Williams, R. L., Courtneidge, S. A., and Wagner, E. F. (1988) *Cell* **52,** 121–131
22. Wigler, M., Sweet, R., Sim, G. K., Wold, B., Pellicer, A., Lacy, E., Maniatis, T., Silverstein, S., and Axel, R. (1979) *Cell* **16,** 777–785
23. Piccolo, S., Bonaldo, P., Vitale, P., Volpin, D., and Bressan, G. M. (1995) *J. Biol. Chem.* **270,** 19583–19590
24. Nagy, A., Gersenstein, M., Virtersten, K., and Behringer, R. (2003) *Manipulating the Mouse Embryo: A Laboratory Manual*, 3rd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
25. Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K., and Mattick, J. S. (1991) *Nucleic Acids Res.* **19,** 4008
26. Reinke, H., and Horz, W. (2004) *Biochim. Biophys. Acta* **1677,** 24–29
27. de Laat, W., and Grosveld, F. (2003) *Chromosome Res.* **11,** 447–459
28. Girotto, D., Fabbro, C., Braghetta, P., Vitale, P., Volpin, D., and Bressan, G. M. (2000) *J. Biol. Chem.* **275,** 17381–17390
29. Antoniv, T. T., De Val, S., Wells, D., Denton, C. P., Rabe, C., de Crombrugghe, B., Ramirez, F., and Bou-Gharios, G. (2001) *J. Biol. Chem.* **276,** 21754–21764
30. Schwartz, S., Zhang, Z., Frazer, K. A., Smit, A., Riemer, C., Bouck, J., Gibbs, R., Hardison, R., and Miller, W. (2000) *Genome Res.* **10,** 577–586
31. Grosveld, F., van Assendelft, G. B., Greaves, D. R., and Kollias, G. (1987) *Cell* **51,** 975–985
32. West, A. G., Gaszner, M., and Felsenfeld, G. (2002) *Genes Dev.* **16,** 271–288
33. Labrador, M., and Corces, V. G. (2002) *Cell* **111,** 151–154
34. Butler, J. E., and Kadonaga, J. T. (2002) *Genes Dev.* **16,** 2583–2592
35. Taniuchi, I., and Littman, D. R. (2004) *Oncogene* **23,** 4341–4345
36. Leach, K. M., Nightingale, K., Igarashi, K., Levings, P. P., Engel, J. D., Becker, P. B., and Bungert, J. (2001) *Mol. Cell. Biol.* **21,** 2629–2640
37. Kaufman, M. H., and Bard, J. B. L. (1999) *The Anatomical Basis of Mouse Development*, Academic Press, London, UK
38. Leimeister, C., Steidl, C., Schumacher, N., Erhard, S., and Gessler, M. (2002) *Dev. Biol.* **249,** 204–218