Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 5590–5594, August 1987
Biochemistry

Identification of a cell-specific transcriptional enhancer in the first intron of the mouse α_2 (type I) collagen gene

PELLEGRINO ROSSI* AND BENOIT DE CROMBRUGGHE†

Laboratory of Molecular Biology, National Cancer Institute, National Institutes of Health, Building 37, Room 2D20, Bethesda, MD 20892

Communicated by Jerome Gross, April 21, 1987

ABSTRACT **A transcriptional enhancer has been identified in the first intron of the mouse α_2 (type I) collagen gene in a region between +418 and +1524 base pairs from the transcriptional start site. The enhancer functions both when it is located 5′ and 3′ to the promoter that it activates and is independent of the orientation of the element. The enhancer stimulates both the homologous α_2 type I [α_2(I)] collagen promoter and the heterologous early simian virus 40 promoter. In transient expression experiments, enhancer-dependent transcription from the α_2(I) collagen promoter utilizes the same transcriptional start site as the one used in the endogenous α_2(I) collagen gene. The enhancer activates transcription at a distance of at least 3 kilobase pairs from the transcriptional start site. The α_2(I) collagen enhancer displays cell specificity, since it is functional in NIH 3T3 fibroblasts but completely inactive in a lymphoid cell line, in contrast to two immunoglobulin gene enhancers that show the opposite behavior. We find several areas of sequence homology with viral enhancers, particularly the enhancer of simian virus 40.**

Enhancer elements are cis-acting DNA sequences able to activate transcription from both homologous and heterologous promoters (for reviews, see refs. 1–3). This stimulation is independent of the orientation of the element and occurs whether the enhancer is located 5′ or 3′ to the promoter. Enhancers are active at distances of several kilobases (kb) from the transcriptional start site. First described in the simian virus 40 (SV40) genome (4–8), several enhancers or enhancer-like sequences were later reported in other viral genomes (1–3). Subsequently, enhancer elements were also found in several cellular genes. The first cellular enhancer was identified in the heavy-chain immunoglobulin gene (9, 10). Its peculiarity was its location in the first intron of the gene and its strict tissue specificity. Tissue specificity has also been observed for other cellular enhancers (1–3). With the exception of the immunoglobulin and β-globin gene enhancers (11–13), viral and cellular enhancers are in the vast majority of the cases located in a position 5′ to their promoter.

Type I collagen represents a major component of the extracellular matrix and constitutes one of the most abundantly synthesized proteins in vertebrates. It is produced in a variety of tissues, but mainly in tendons, skin, and bones. Its synthesis is normally restricted to cells of the mesenchymal lineage, mainly fibroblasts and osteoblasts, and to a lesser degree chondroblasts and cells of smooth and striated muscle. Under normal conditions, cells of ectodermal or endodermal origin, as well as cells of the hematopoietic lineage, do not produce type I collagen. The synthesis of type I collagen is also developmentally regulated, since it appears in the mouse embryo around day 8–9 (14, 15). Several collagen genes have been isolated. These genes are large and contain >50 exons (16–18). We have recently isolated and characterized the promoter region of the mouse α_2 type I [α_2(I)] collagen gene and presented evidence for its tissue-specific and stage-specific expression (19–21). Here we describe the identification of a strong cell-specific enhancer element in the first intron of this gene.

MATERIALS AND METHODS

DNA Constructions. To reconstruct a genomic DNA segment from +54 in the mouse α_2(I) collagen gene to ≈5 kb downstream of this site, a 694-base-pair (bp) *Sph* I/*Eco*RI fragment from the genomic subclone pAZ1001 (19), containing sequences between +54 and +748, was joined to a contiguous 4.3-kbp *Eco*RI/*Sac* I fragment derived from the genomic insert of phage λmcol 212 (19). λmcol 212 contains sequences of the mouse α_2(I) collagen gene that overlap with those of pAZ1001 over a distance of 200 bp. The 4.3-kbp *Eco*RI/*Sac* I fragment contains sequences from +748 to about +5000. After addition of *Bam*HI linkers, the 5-kbp fragment was subcloned in both orientations either in the *Bam*HI site of pAZ1003 (20), generating pR23 and pR24, or in the *Bgl* II site of pA10CAT$_2$ (22), generating pR21 and pR22. Table 1 lists all plasmids generated by inserting different α_2(I) collagen genomic subfragments either in the *Bgl* II site of pA10CAT$_2$ or in the *Bam*HI site of pAZ1003. The end points of these fragments are indicated. Fig. 1*B* shows the restriction sites that were used to generate the various fragments. The 2.2-kbp *Hin*cII fragment inserted in pA10CAT$_2$ to generate pR31 and pR32 was obtained by *Hin*cII digestion of pR25 and contained 100 bp of pBR322 DNA sequences on the 5′ side of the collagen genomic sequence. To generate pR42, pR36 was digested by *Taq* I and its ends were made blunt; subsequently, the plasmid was digested with *Bam*HI. The 2.6-kbp *Taq* I/*Bam*HI fragment was ligated to a 3.8-kbp *Bgl* II/*Bam*HI fragment from pA10CAT$_2$, isolated after the *Bgl* II site had been made blunt. The resulting insert contained sequences between +418 and +1188. The correctness of all constructions and the orientation of the genomic inserts were controlled by restriction analysis. All plasmids were grown in *Escherichia coli* N38 and purified by banding twice in cesium chloride. Plasmid preparations used in transfections contained <5% RNA and consisted essentially of only supercoiled DNA as tested by agarose gel electrophoresis.

DNA Sequence. The sequence of the DNA between +54 and +1529 was determined by the chemical cleavage method (23) on both DNA strands, using the restriction sites indicated in Fig. 5*A*. Labeling of 5′ ends was done with [γ^{32}P]ATP and polynucleotide kinase. The *Taq* I/*Xmn* I fragment (indicated

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: SV40, simian virus 40; α_2(I) collagen, α_2 type I collagen; CAT, chloramphenicol acetyltransferase.

*Permanent address: Department of Public Health and Cell Biology, Second University of Rome, Tor Vergata Via Orazio Raimondo, Localita La Romanina, 00173, Rome, Italy.

†Permanent address: Department of Genetics, Box 11, The University of Texas System Cancer Center, M. D. Anderson Hospital and Tumor Institute, 1515 Holcombe Boulevard, Houston, TX 77030.

Table 1. Plasmids generated by inserting different α_2(I) collagen genomic subfragments in either pA10CAT$_2$ or pAZ1003 vector

Plasmid	Insert	Orientation	Vector	Relative CAT activity
pA10CAT$_2$	None	—	—	1
pSV$_2$CAT	SV40 enhancer	—	—	49
pR21	+54 to +5000	Direct	pA10CAT$_2$	1
pR22	+54 to +5000	Opposite	pA10CAT$_2$	67
pR25	+54 to +2500	Direct	pA10CAT$_2$	30
pR26	+54 to +2500	Opposite	pA10CAT$_2$	93
pR27	+2500 to +5000	Direct	pA10CAT$_2$	3
pR29	+54 to +948	Direct	pA10CAT$_2$	41
pR31	+54 to +2100	Direct	pA10CAT$_2$	96
pR32	+54 to +2100	Opposite	pA10CAT$_2$	65
pR34	+802 to +1800	Opposite	pA10CAT$_2$	32
pR35	+418 to +1524	Direct	pA10CAT$_2$	30
pR36	+418 to +1524	Opposite	pA10CAT$_2$	93
pR37	+418 to +1040	Opposite	pA10CAT$_2$	93
pR42	+418 to +1188	Opposite	pA10CAT$_2$	103
pAZ1003	None	—	—	1
pAZ1009	SV40 enhancer	—	—	10
pR23	+54 to +5000	Direct	pAZ1003	10
pR24	+54 to +5000	Opposite	pAZ1003	1
pR40	+418 to +1524	Direct	pAZ1003	11
pR41	+418 to +1524	Opposite	pAZ1003	11
pR43	+134 to +748	Direct	pAZ1003	1
pR44	+134 to +748	Opposite	pAZ1003	1

CAT activity is expressed as the ratio between the activity of the test plasmid and the basal level of activity of the plasmid containing the same promoter without enhancer element. Data represent averages of at least two independent transfection experiments.

in Fig. 5) was labeled at its 3′ end using the Klenow fragment of DNA polymerase and [α^{32}P]dCTP. The DNA sequence was analyzed using programs available in the NIH mainframe DEC-10 computer.

DNA Transfections, Chloramphenicol Acetyltransferase (CAT) Assays, and RNA Analysis. NIH 3T3 mouse fibroblasts were transfected essentially as described (24) using 10-cm-diameter dishes and 1 ml of $Ca_3(PO_4)_2$·DNA precipitate, containing between 7 and 15 μg of plasmid DNA. A lymphoid cell line, S194 (a kind gift of C. Queen, National Institutes of Health), was transfected with a modified DEAE-dextran protocol as described (25). Equimolar amounts of each plasmid were used in all the experiments. Forty-eight hours after DNA transfection, the cells were harvested and analyzed either for CAT activity (26), using equal amounts of protein from each extract, or for RNA analysis. CAT activity was determined by counting strips cut from the thin-layer chromatography plates that correspond to the acetylated forms of [^{14}C]chloramphenicol identified by autoradiography. The data in the table represent the averages of at least two independent transfections. For the primer-extension experiment shown in Fig. 4, total RNA from NIH 3T3 fibroblasts was extracted by the guanidine thiocyanate/cesium chloride method (27) 60 hr after DNA transfection. Equal amounts of total RNAs were hybridized with an excess of a 5′-end-labeled CAT antisense synthetic oligonucleotide. The conditions for hybridization, reverse transcription, and electrophoretic analysis of the cDNA products were as described (20).

RESULTS

An Enhancer Within the α_2(I) Collagen Gene. The enhancing activity of several mouse α_2(I) collagen gene fragments was tested by inserting these fragments in the *Bgl* II site of pA10CAT$_2$ (22). This vector contains the early SV40 promoter, without the SV40 enhancer, linked to the CAT gene. The *Bgl* II site in pA10CAT$_2$ is immediately 5′ to the SV40 early promoter–CAT fusion (Fig. 1). In transient expression experiments, the pA10CAT$_2$ vector exhibits a low level of CAT expression. Some fragments were also inserted in the *Bam*HI site of pAZ1003, a vector in which a DNA segment of the mouse α_2(I) collagen gene (−2000 to +54) containing the promoter of this gene is fused to the CAT gene. The *Bam*HI site in pAZ1003 is ≈1.5 kbp 3′ to the collagen promoter. Table 1 indicates the endpoints in the α_2(I) collagen gene of each fragment that was tested for enhancer activity and summarizes the results of the transient expression assays performed with the various constructions. Fig. 2 and 3 show examples of representative CAT assays.

A 5-kbp fragment, spanning sequences from position +54 to approximately +5000, showed a clear enhancer activity with both the α_2(I) collagen promoter (Fig. 2, pR23) and the SV40 early promoter (Fig. 2, pR22). This fragment contains ≈200 bp of the first exon followed by the first intron and the continuation of the gene in the 3′ direction. The enhancing activity was seen when the fragment was cloned in the direct orientation in the pAZ1003 vector and in the opposite orientation in the pA10CAT$_2$ vector. However, the enhancing activity was lost when the orientation was inverted (Fig. 2, pR24 and pR21). A possible simple explanation for this observation is that an enhancer element is present in the 5-kb insert but that it is closer to the 5′ end of this fragment and that its activity is strongly reduced by increasing the distance between the enhancer and each of the two promoters that were used in the assays. Indeed, when fragments between +54 and +2500, (Table 1, pR25 and pR26) and between +2500 and +5000 (Table 1, pR27) were subcloned in pA10CAT$_2$, only the +54 to +2500 fragment showed enhancing activity. This time, the enhancing activity was observed when the fragment was cloned in one or the other orientation, even though the effect was somewhat higher in the opposite than in the direct orientation.

Enhancing activity was also observed when a smaller fragment was assayed removing 400 bp from the 3′ end of the 2.5-kbp fragment (Table 1, pR31 and pR32). Again the activity is seen in the two orientations. A fragment from +802 to +1800 (Table 1, pR34) and a fragment from +54 to +948 (Table 1, pR29), each cloned in pA10CAT$_2$, both showed enhancer activity. This could indicate that the enhancer is

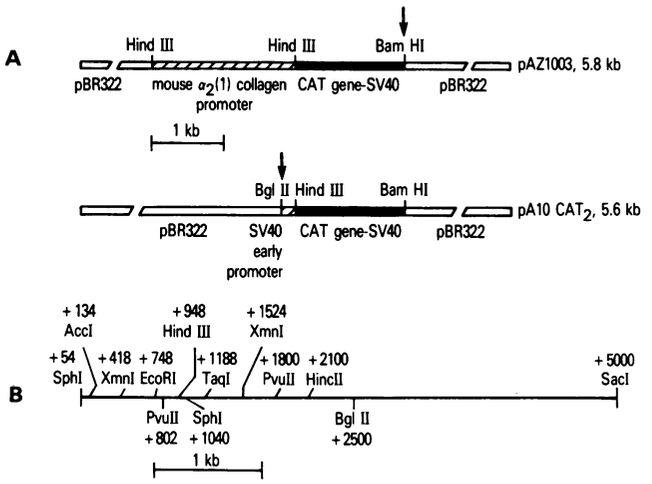


FIG. 1. (*A*) Schematic representation of the DNA vectors, pAZ1003 and pA10CAT$_2$, used to construct the various plasmids described in this paper. Arrows indicate the site in which the α_2(I) collagen gene fragments were inserted. (*B*) Restriction map of mouse α_2(I) collagen gene between nucleotide +54 and +5000 from the start site of transcription. The relevant restriction sites used to map the enhancer activity are indicated.

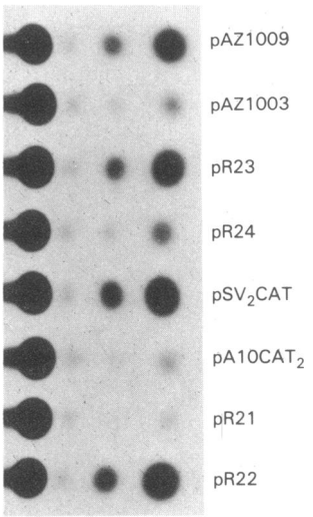


FIG. 2. Identification of an enhancer sequence within α_2(I) collagen gene. The boundaries of the inserts of the various plasmids are shown in Table 1. In this and other experiments, CAT activities were measured after transfection of three types of plasmids: (*i*) the basal levels of activities of pA10CAT$_2$ or pAZ1003; (*ii*) the levels of activities of the plasmids containing the α_2(I) collagen intron segments; (*iii*) the levels of activities of constructions containing the SV40 enhancer—i.e., pSV$_2$CAT or pAZ1009—as positive controls. Plasmid pSV$_2$CAT contains the full SV40 enhancer element immediately 5′ to the early SV40 promoter (26). Plasmid pAZ1009 is a pAZ1003 derivative in which a 135-bp fragment containing the SV40 enhancer was inserted 3′ to the CAT gene (24).

located between +802 and +948. Alternatively, the enhancer might be divided in two distinct domains, one located between +54 and +948, the other between +802 and +1800. Strong enhancement was also achieved with fragments spanning from +418 to +1524 (Table 1, pR36 and pR35) and from +418 to +1188 (Table 1, pR42). Since the segment flanking both sides of a *Sph* I site at +1040 presents clear homologies with part of the SV40 enhancer element, we tested the fragment between +418 and +1040, removing at least part of the sequences that are homologous with the SV40 enhancer. However, high enhancing activity was still found with this construction (Table 1, pR37).

The 1.1-kbp fragment from +418 to +1524 also stimulates the α_2(I) collagen promoter when cloned in either orientation 3′ to the CAT gene (Fig. 3, pR40, pR41). The slight difference between the two orientations observed in the experiment shown in Fig. 3 was not observed in other experiments.

Other experiments showed that the activity of the mouse α_2(I) collagen promoter was not increased when a fragment of 15 kbp preceding the start of transcription was placed in the vector instead of the 2-kbp mouse collagen fragment in pAZ1003. Also, no increase of basal expression was detectable when a fragment from +134 to +748 was inserted in the *Bam*HI site of pAZ1003 (see Table 1, plasmids pR43 and pR44).

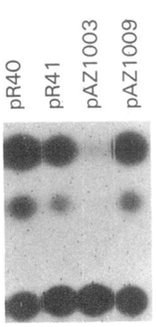


FIG. 3. Enhancing activity of sequences between +418 and +1524 tested by transient expression in NIH 3T3 expression in NIH 3T3 fibroblasts of constructions containing the mouse α_2(I) collagen promoter. Representative CAT assays are shown.

The 1.1-kbp fragment from +418 to +1524 stimulates the formation of correctly initiated CAT transcripts from the α_2(I) collagen promoter. This was shown by a primer-extension experiment (Fig. 4). Lane 2 shows the presence of a cDNA doublet ≈145 nucleotides long that corresponds to transcripts starting from the same place as previously identified as the transcriptional start sites in the endogenous mouse α_2(I) collagen gene. This doublet is not detectable with RNA extracted from mock-transfected cells (lane 3).

DNA Sequence of the α_2(I) Collagen Gene Enhancer Element. The sequence of the mouse α_2(I) collagen gene, between +1 to +1529 is shown in Fig. 5*B*. This figure also shows the sequencing strategy used for determining the sequence of both DNA strands (Fig. 5*A*).

Sequence analysis of the segment between +54 and +1524 reveals several interesting features: (*i*) The sequence contains two enhancer "core" elements (GTGG$\overset{AAA}{TTT}$G) as found in several viral enhancers (1, 2), one from +295 to +302, the other from +1301 to +1308. Both core sequences are outside a segment that has maximal enhancing activity (+418 to +1040). (*ii*) The pentanucleotide (CCCTG) is present 12 times on one or the other DNA strand. (*iii*) A 12-bp sequence between +952 and +963 is found as a direct repeat from +1218 to +1224 with 1 bp mismatch. Another direct repeat is found on each side of the *Sph* I site at +1040. (*iv*) Two inverted repeats are present, one between +781 and +806, the other between +820 and +845. (*v*) Several areas show striking homologies with the viral SV40 enhancer region, particularly around the *Sph* I site (at +1040) and 3′ to the *Taq* I site (at +1188) in a region where one of the two enhancer core elements is found. The areas of homology between the α_2(I) collagen enhancer and the SV40 enhancer element fall outside a segment of the viral enhancer that is common with a sequence in the IgG κ-chain enhancer. This sequence binds a nuclear factor found in cells of the B-lymphoid lineage (28). Homologies were also found with other enhancer sequences, such as the enhancers of polyomavirus and of the long terminal repeat of Rous sarcoma virus.

No clear homologies were found between the sequence of the first intron of the mouse α_2(I) collagen gene and the sequence of the first intron of the corresponding chicken gene (17). However, several areas of homology were found between the sequence reported in Fig. 5 and the sequence spanning the second and third intron of the avian gene. The possibility that a similar enhancer element is present in the intronic sequences of the chicken gene is under study.

Cell Specificity of the α_2(I) Collagen Gene Enhancer. To test for cell specificity of the α_2(I) collagen gene enhancer element, the following plasmids were compared: (*i*) Two derivatives of pA10CAT$_2$ (pR22 and pR36) containing the α_2(I) collagen enhancer with the SV40 promoter. (*ii*) Two derivatives of pAZ1003 (pR23 and pR40) containing the α_2(I) collagen gene enhancer with the α_2(I) collagen promoter. (*iii*) Two constructions harboring either the κ-chain immunoglob-

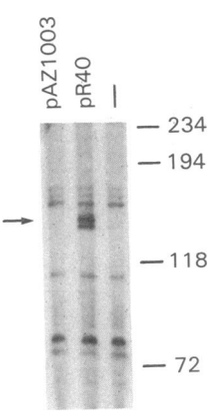
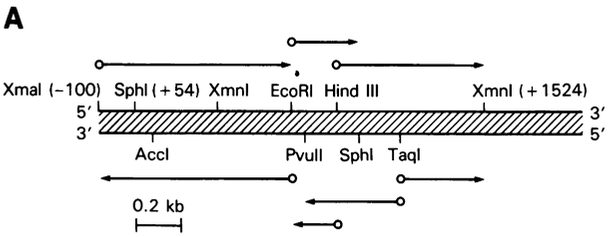


FIG. 4. Primer extensions of RNAs extracted from NIH 3T3 fibroblasts transfected with the indicated plasmids. A 5′-end-labeled CAT antisense synthetic oligonucleotide was hybridized to equal amounts of total RNA and extended by reverse transcription. Arrow indicates position of cDNA doublet found in cells transfected with pR40. The size of this cDNA is 145 nucleotides. The cDNA products were fractionated on a 6% acrylamide sequencing gel. The lane on the right represents mock-transfected NIH 3T3 fibroblasts. Numbers on right represent nucleotides.



B

```
-1                                                                          +100
AGCACCACGG CAGCAGGAGG TTTCGACTAA GTTGGAGGGA ACGGTCCACG ATTGCATGCC TGCGCCCGCC AGGTGATACC TCCGCTGGTG ACCCAGGGGC

                                                                            +200
TCTGCAACAC AAGGAGTCTG CATGTCTAAG TGGTAGACAT GCTCAGCTTT GTGGATACGC GGACTCTGTT GCTGCTTGCA GCAACTTCGT GCCTAGCAAC

                                                                            -300
ATGCCAATGT AAGTGTCTTC TGCTGTTTTT GGGGGAGAGT TGAAGGAGAG AGGCTGGATA GGTGGGTATC CAACAGAGGG AAACTTCACA CATTCATTCC

                                                                            -400
ACATACAGCA AGTGAGGTCG CCCTCCAATC TGTCGGAGAA AACACTAGGG TTCACGAGAT ACATGGAAAA AGAATGGAAA TATGAAAGAC TTGACTGTTG

                                                                            -500
CCAAGAGTTG GGGGAAGTTT TTCCCTTAAA ACTAGACTGT TAGCGGATGC CCCGGTTGGC GAAGCTCACA AGAGGCAAGA AGGTTGCCTC AAAGATATAA

                                                                            -600
ACCTGCTGAC AGCATGAACG ACTGAAGAGG GCTCCACTCT TCAATTCCCA GAATTATGCC CGTGGTCACG TTGTGAAACT TGAGGCCTAG TTTAGAAAGC

                                                                            +700
CCCGTAATAG AATAATACAA AACTCACAGG ACTCCATTTT CTTGTGGGCG GCCGAGTTGC TCTGGATGGC GTTGATGAGA TAATGATAAA AATACGCCCC

                                                                            -800
CTGGAGGGGG TAAAGTGCAG GAACAGGGAC CAGGCCTTAA GCGCTAGGAA TTCCCTTTGG GTCCAGGGAA TCCAATGTTG CCAGCTTCAC CTCTTGGTCC

                                                                            -900
AGCTGGGCCG ATTGGGCCAG GTCTGCCATC AATGCCAGGG AGACCCTGAA ATCAGAATTT CACCTTTATT CCCCACCCCC CTGGTTTCCA GCTATCAGGA

                                                                            -1000
TGAGGATTTT GTCAAATGAT CAAACAAGGT AGTTTTCTGT ATGCGTAAAG CTTCCTGTAT AAAGTGTCAA TAAATTATGA CTCCATTCCC CTGTTTTAAA

                                                                            -1100
TAGGATGTCA GTACACAGAC AGACAGTGCA ATCTTGCATG CACGCAATCT GTCCCTCTCC CTTTTCTGTC CCCTTGTCAC CTGCTTCAAT TTTCTCAGTC

                                                                            -1200
ATGGTTTCCT TACCCCTTCC CTGCTTTCAG CCCTGCATCC ATTTAGGCTC CGTAATTACA ATCTATTACC TTCTGTTCCC AGATGATGTC GAGGTGGATG

                                                                            -1300
ATATTTATTT TCAATGTATT CATGTATAAA TAAGCCTCAA AAATAGACTT GAGGAGGGTG CAATAAGTGT TCTTGAAAAC GTAATGGAAA ATGAGCAATA

                                                                            -1400
CTTACCACAG GGCCTTCTTT ACCAGATGGG CCAACATTTC CAGGAGACCC AGGAAGACCC TGTAAAATAA ATGAGGGTCC TTTCATATGG GTGTGTTAGC

                                                                            -1500
TCCTACACTC AAGGCTCACA TTAATTGAAT CTCACCCACG TCTCTCTTCT CACTTTTGTT TGCTGCTGAG GGGATTCATT AGAGCTAGCA GCCTTGCTGG

AAAGTATAAA AAACTCTTAG AATCNNTTC
```

FIG. 5. Restriction map (*A*) and nucleotide sequence (*B*) of the mouse α_2(I) collagen gene between position +1 and +1529 (+1 corresponds to the start of transcription). The sequencing strategy is indicated with arrows in the restriction map. Two consensus enhancer core elements are boxed. Inverted repeat elements are underlined with arrows. Direct repeats are marked above the sequence. Dots above the sequence indicate the position of a highly repeated pentanucleotide. Vertical arrow indicates the splice site after the first exon of the gene. The sequence from +1 to +290 was reported previously (19).

ulin gene enhancer (pKCAT K) or the heavy-chain immunoglobulin gene enhancer (pKCAT H) driving the expression of the κ-chain promoter (29). (*iv*) One plasmid containing the SV40 enhancer with its own promoter (pSV$_2$CAT) and one plasmid containing the SV40 enhancer with the α_2(I) collagen promoter (pAZ1009). All 10 plasmids were transfected both in NIH 3T3 fibroblasts and in cells of a lymphoid line (S194). Fig. 6 shows that every construction that contains the α_2(I) collagen gene enhancer shows strong activity in fibroblasts but no activity in the lymphoid cell line. Constructions harboring the immunoglobulin κ-chain and heavy-chain gene enhancer are essentially only active in lymphoid cells with the possible exception of pKCATK, which has a very reduced level of activity in fibroblasts. Finally, constructions containing the SV40 enhancer are active in both types of cells.

DISCUSSION

We have identified a strong tissue-specific enhancer element in the first intron of the mouse α_2(I) collagen gene. The enhancer activates transcription from both the homologous α_2(I) collagen promoter and from the SV40 early promoter. In our experiments, we have observed the enhancing effect both 5′ to the SV40 early promoter and 3′ to the mouse α_2(I) collagen gene promoter. The element is generally active in either orientation. The enhancing effect is relatively independent from the distance between the element and the promoter, since it is observed both when placed immediately 5′ to the SV40 early promoter or ≈1.6 kbp 3′ to the α_2(I) collagen promoter.

In our experiments, the enhancing activity is ≈100-fold when tested with the SV40 promoter and ≈10-fold when tested with the α_2(I) collagen promoter. This difference could be due to the different distances between the enhancer and the promoter in the two types of constructions. In addition, the basal level of activity of the mouse α_2(I) collagen promoter is generally 5–10 times higher than the basal level of activity of the SV40 early promoter in NIH 3T3 fibroblasts. Since the SV40 early promoter–CAT construction has a higher sensitivity, it was used in our attempts to map the limits of the enhancer region.

A 1106-bp fragment from +418 to +1524 appears to contain full enhancer activity. No enhancer activity was detected between +134 and +748 and no substantial decrease in enhancing activity was observed by narrowing the 3′ end of the insert to +1040. It is, therefore, possible that at least one fully functional enhancer element is present between +748 and +1040. It is also possible that more than one element is present in this enhancer, as is the case in several viral enhancers (1, 2). Additional experiments are needed to delineate the boundaries of the enhancer sequences more precisely. The enhancing region is presumably confined within the first intron of this gene, since the splice site after

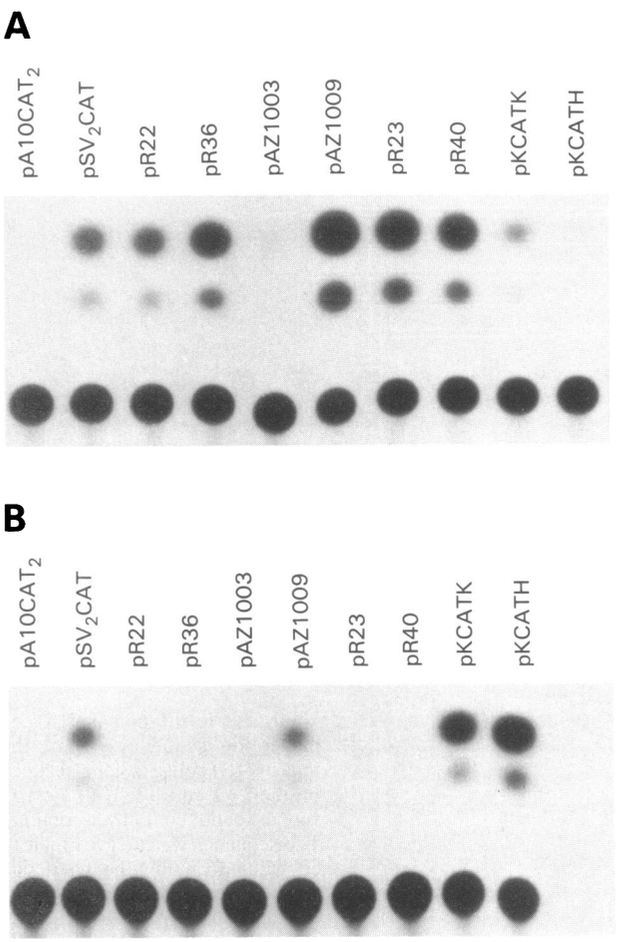


FIG. 6. Cell-specific expression of the α_2(I) collagen enhancer. Plasmids were transfected either in NIH 3T3 fibroblasts (*A*) or in the mouse lymphoid cell line S194 (*B*). In both plasmids pKCATK and pKCATH, the mouse κ-chain immunoglobulin promoter drives the CAT gene; pKCATK contains a mouse κ-chain immunoglobulin gene enhancer and pKCATH contains the mouse heavy-chain immunoglobulin gene enhancer. These two plasmids are described in detail the ref. 29.

the first exon maps at position +208, and the sequence between +54 and +1524 does not contain the 11-nucleotide sequence of the second exon, previously identified around position +2100 in the chicken α_2(I) collagen genomic sequence (17) and also present in the human α_2(I) collagen gene (F. Ramirez, personal communication).

The activity of the mouse α_2(I) collagen gene enhancer was completely inhibited when a 5-kbp insert containing sequences between +54 and +5000 was cloned in the direct orientation 5′ to the SV40 early promoter or in the opposite orientation 3′ to the cognate promoter. In these constructions, the sequences needed for full enhancement (between nucleotides +418 and +1524) are 3.5-kbp away from the SV40 early promoter or 5 kbp from the α_2(I) collagen promoter. This distance per se may render the enhancer inactive or, alternatively, sequences within this segment may interfere with the activity of the enhancer.

The α_2(I) collagen enhancer is very active in NIH 3T3 fibroblasts but completely inactive in a lymphoid cell line, in contrast to the κ-chain or heavy-chain immunoglobulin enhancer, which are inactive in fibroblasts but active in lymphoid cells. The SV40 enhancer, on the other hand, is active in both cell types. Previously, we have shown that the α_2(I) collagen promoter displayed tissue-specific expression. Evidence for this view mainly came from experiments using transgenic mice (21) and also from DNA transfection of tissue culture cells (20). Thus, both the promoter and the enhancer of the α_2(I) collagen gene contain sequences that confer tissue specificity to this gene, presumably restricting its expression to cells derived from mesenchymal cells.

We wish to thank Paul Golumbek for help in the early stages of this work, Azriel Schmidt for sequencing the *Xma* I/*Eco*RI fragment, Cary Queen for the generous gift of plasmids pKCATK and pKCATH and the cell line S194, and Joyce Sharrar for editorial assistance. P.R. is a recipient of an Arthritis Foundation Fellowship.

1. Khoury, G. & Gruss, P. (1983) *Cell* **33,** 313–314.
2. Serfling, E., Jasin, M. & Schaffner, W. (1985) *Trends Genet.* **1,** 224–230.
3. Voss, S. D., Schlokat, V. & Gruss, P. (1986) *Trends Biochem. Sci.* **11,** 287–289.
4. Banerji, J., Rusconi, S. & Schaffner, W. (1981) *Cell* **27,** 299–308.
5. Benoist, C. & Chambon, P. (1981) *Nature (London)* **290,** 304–310.
6. Fromm, M. & Berg, P. (1982) *J. Mol. Appl. Genet.* **1,** 457–481.
7. Gruss, P., Dhar, R. & Khoury, G. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 943–947.
8. Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M. P. & Chambon, P. (1981) *Nucleic Acids Res.* **9,** 6047–6068.
9. Gillies, S. D., Morrison, S. S., Oi, V. T. & Tonegawa, S. (1983) *Cell* **33,** 717–728.
10. Banerji, J., Olson, L. & Schaffner, W. (1983) *Cell* **33,** 729–740.
11. Queen, C. & Baltimore, D. (1983) *Cell* **33,** 741–747.
12. Choi, O. R. & Engel, J. D. (1986) *Nature (London)* **323,** 731–734.
13. Hesse, J. E., Nickol, J. M., Lieber, M. R. & Felsenfeld, G. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 4312–4316.
14. Adamson, E. D. & Ayers, S. E. (1979) *Cell* **16,** 953–965.
15. Leivo, I., Vaheri, A. & Timpl, R. (1980) *Dev. Biol.* **76,** 100–114.
16. Ohkubo, H., Vogeli, G., Mudryj, M., Avvedimento, V. E., Sullivan, M., Pastan, I. & de Crombrugghe, B. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 7059–7063.
17. Boedtker, H., Finer, M. & Aho, S. (1985) *Ann. N.Y. Acad. Sci.* **460,** 85–116.
18. Ramirez, F., Bernard, M., Chu, M. L., Dickson, L., Sangiorgi, F., Weil, D., De Wet, W., Junien, C. & Sobel, M. (1985) *Ann. N.Y. Acad. Sci.* **460,** 117–129.
19. Schmidt, A., Yamada, Y. & de Crombrugghe, B. (1984) *J. Biol. Chem.* **259,** 7411–7415.
20. Schmidt, A., Rossi, P. & de Crombrugghe, B. (1986) *Mol. Cell. Biol.* **6,** 347–354.
21. Khillan, J. S., Schmidt, A., Overbeek, P. A., de Crombrugghe, B. & Westphal, H. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 725–729.
22. Laimins, L. A., Gruss, P., Pozzatti, R. & Khoury, G. (1984) *J. Virol.* **49,** 183–189.
23. Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
24. Schmidt, A., Setoyama, C. & de Crombrugghe, B. (1985) *Nature (London)* **314,** 286–289.
25. Stafford, J. & Queen, C. (1983) *Nature (London)* **306,** 77–79.
26. Gorman, C. M., Moffat, L. & Howard, B. H. (1982) *Mol. Cell. Biol.* **2,** 1044–1051.
27. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
28. Sen, R. & Baltimore, P. (1986) *Cell* **46,** 705–716.
29. Garcia, J. V., Bich-Thyy, L. T., Stafford, J. & Queen, C. (1986) *Nature (London)* **322,** 383–385.