

# Transcriptional Control of the Mouse $\alpha_2$ (I) Collagen Gene: Functional Deletion Analysis of the Promoter and Evidence for Cell-Specific Expression

AZRIEL SCHMIDT, PELLEGRINO ROSSI,<sup>†</sup> AND BENOIT DE CROMBRUGGHE\*  
*Laboratory of Molecular Biology, National Cancer Institute, Bethesda, Maryland 20892*

Received 15 July 1985/Accepted 4 November 1985

A chimeric gene was constructed in which sequences between 2,000 base pairs upstream of the start of transcription of the mouse  $\alpha_2$ (I) collagen gene and 54 base pairs downstream of this site were fused to the chloramphenicol acetyltransferase (CAT) gene. We present evidence suggesting that this collagen gene segment is sufficient for cell-specific expression of the chimeric gene. Indeed, the levels of CAT activity in transient expression experiments were at least 10 times higher after transfection of NIH 3T3 cells than after transfection of a mouse myeloma cell line, whereas much less difference was found after transfection of these two cell types with pSV2-CAT, a plasmid in which the early simian virus 40 promoter is fused to the CAT gene. Several deletions were introduced in the same 5'-flanking segment of the  $\alpha_2$ (I) collagen gene, and the effects of these deletions were examined after DNA transfection of the chimeric collagen-CAT gene into NIH 3T3 cells. At least two segments broadly located between -979 and -502 and between -346 and -104 are needed for optimal expression of the chimeric gene. These results were obtained both in transient expression experiments and by analysis of pools of NIH 3T3 cells that were stably transfected with the different mutants. In general, the effects of the deletions on the activity of the  $\alpha_2$ (I) collagen promoter were analogous, whether the plasmids harbored the simian virus 40 enhancer sequence or not, although the overall levels of expression of the chimeric gene were increased when the recombinant plasmids contained this sequence.

Eucaryotic promoters show a wide spectrum in the arrangement of their regulatory elements. In some genes, these elements are located within the gene (4, 24, 25); in others they have a compact distribution upstream of the start site of transcription (3, 10, 22), whereas in still others, some of these elements display a wider dispersion upstream of the start of transcription (30). At least three distinct transcriptional control elements have been described. One is the Goldberg-Hogness box, an AT-rich segment usually located about 30 base pairs (bp) upstream of the start of transcription (3, 6, 33). This segment appears to have a role in positioning the RNA polymerase enzyme so that the enzyme can start transcription from a given site in the promoter. Another segment is a GC-rich region found in several promoters. A protein, SP1, has been isolated which specifically binds to this segment in certain promoters and stimulates transcription of these promoters in a cell-free system (8). A third class of elements has been called enhancers because sequences in such elements are capable of stimulating transcription from both homologous and heterologous promoters. Enhancers also stimulate transcription even when they are removed by as much as several kilobases (kb) from the promoter segment, the activity of which they enhance. Furthermore, enhancers are generally active, regardless of their orientation (2, 13, 17).

We have been interested in the expression of several collagen genes (11, 15). They belong to a family of at least 20 genes which code for the polypeptide chains of the different collagen types (23). Many of these collagens are preferentially synthesized in certain types of cells. Type I collagen is

the most abundant of these collagens and constitutes an important part of the extracellular scaffold of bones, tendons, skin, and smooth muscles. Its synthesis is strongly induced around day 8 to 9 in the mouse embryo (1, 14). In fibroblasts in culture, type I collagen is one of the major biosynthetic products, but its synthesis is also greatly inhibited after transformation of these cells by *v-src*, *v-mos*, and *v-ras* (11, 15, 27). Since this inhibition is mediated by a transcriptional mechanism (26, 27), it provides an experimental system in which some aspects of the transcriptional control of this gene can be analyzed.

We initiated studies on the transcriptional and translational regulation of the gene for the  $\alpha_2$  chain of type I collagen as a paradigm for other genes which specify extracellular matrix proteins. Because the  $\alpha_2$ (I) collagen gene is very large (40 kb) (19) and therefore difficult to manipulate *in vitro*, we constructed a chimeric gene in which a segment of the  $\alpha_2$ (I) collagen gene between 2,000 bp upstream of the start site of transcription to 54 bp downstream of this site was fused to the bacterial gene for chloramphenicol acetyltransferase (CAT). Such a chimeric gene may not contain all of the regulatory elements which control transcription of the  $\alpha_2$ (I) collagen gene, but it allows us to ask, by performing DNA transfection experiments, whether the sequences between 2,000 bp upstream of the start of transcription and 54 bp downstream from this site can provide tissue specificity and whether defined deletions in this segment have an effect on expression of the chimeric gene. We have shown that in cell lines that stably express this collagen-CAT chimeric gene, transcription from the cloned  $\alpha_2$ (I) collagen promoter is strongly inhibited by *v-mos* transformation, in parallel to the inhibition of the expression of the endogenous type I collagen genes (27). Here we present evidence suggesting that the segment between 2,000 bp upstream of the start of transcription and 54 bp downstream from this site is suffi-

\* Corresponding author.

<sup>†</sup> Permanent address: Department of Public Health and Cell Biology, II University of Rome, Tor Vergata, Via Orazio Raimondo, Rome, 00173, Italy.

cient for cell-specific expression of the promoter. A deletion analysis of this segment shows that different subsegments are needed for optimal expression of the chimeric gene.

### MATERIALS AND METHODS

**Cells.** Mouse NIH 3T3 cells were maintained in Dulbecco modified Eagle medium (DMEM) supplemented with 10% calf serum, penicillin, and streptomycin. Mouse myeloma (MPC-11) cells were obtained from C. Queen (National Institutes of Health, Bethesda, Md.). The cells were maintained in DMEM supplemented with 20% horse serum.

**DNA constructions.** The construction of plasmid pAZ1009 has been described previously (27). The plasmid contains the mouse  $\alpha_2(I)$  collagen promoter region (-2000 to 54) fused to the gene for CAT. Plasmid pAZ1013 was constructed by digestion of pAZ1009 with *Bgl*II and *Sma*I enzymes, which each cleave once in pAZ1009. The end of the DNA was made blunt with the Klenow enzyme, and the DNA was circularized by ligation. Plasmid pAZ1018 was constructed by digestion of plasmid pAZ1009 with *Taq*I enzyme. After *Sma*I linkers were added, the DNA fragments were digested with *Bgl*II enzyme. The *Bgl*II-*Taq*I (-346 to -233) fragment was isolated and inserted into the large fragment of plasmid pAZ1009 which was produced by digestion with *Bgl*II and *Sma*I enzymes. Plasmid pAZ1019 was produced by digestion of plasmid pAZ1009 with *Xba*I and *Bgl*II enzymes, which both cleave pAZ1009 at unique sites. The ends of the large DNA fragment were made blunt by the Klenow enzyme, and the fragment was circularized by ligation. Plasmid pAZ1015 was constructed by digestion of plasmid pAZ1009 with *Xba*I enzyme, followed by partial digestion with *Hind*III. A 4.4-kb fragment was isolated and circularized after the addition of *Xba*I linkers. Plasmid pAZ1040 was produced by digestion of plasmid pAZ1009 with *Acc*I and *Xba*I enzymes, which both cleave at unique sites in pAZ1009. The ends were made blunt by reaction with the Klenow enzyme, and the large fragment was circularized. The series of deletions between -346 and -185 were made by using double-stranded exonuclease *Bal* 31. Plasmid pAZ1009 was linearized by digestion with *Bgl*II and treated for various periods of time with *Bal* 31. *Bgl*II linkers were added, and the plasmid was digested with the enzyme *Apa*I. The collection of short *Bgl*II-*Apa*I fragments was purified and ligated to the 6.35-kilobase pair *Bgl*II-*Apa*I fragment of pAZ1009. (For the approximate [ $\pm$ 5 bp] endpoints of these deletions [plasmid pAZ1057 to pAZ1065], see Fig. 2.) The size of the deletions was verified by the presence or absence of *Taq*I and *Hin*FI sites located between the *Bgl*II and *Apa*I sites and by the size of the fragments produced by these enzymes.

**DNA transfections.** DNA transfections and selection in G-418 medium were carried out as described previously (27). The transfection of the MPC-11 cells was done in a similar manner to that for NIH 3T3 cells with the following changes. A total of  $2 \times 10^6$  cells were plated in 10-cm-diameter dishes. One milliliter of DNA-calcium phosphate was added to the cells. Three to four hours later, the cells were transferred into a 50-ml centrifuge tube and diluted to 50 ml with DMEM. The cells were harvested by centrifugation, washed with DMEM, and suspended in 15% (wt/vol) glycerol in Hanks balanced salt solution. After 1 min, the cells were diluted with DMEM, washed, centrifuged, and cultured either for 2 days for transient expression experiments or for longer time periods to obtain stably transfected lines.

**Determination of the activity of the cloned collagen promoter.** The activity of the collagen-CAT transcription unit

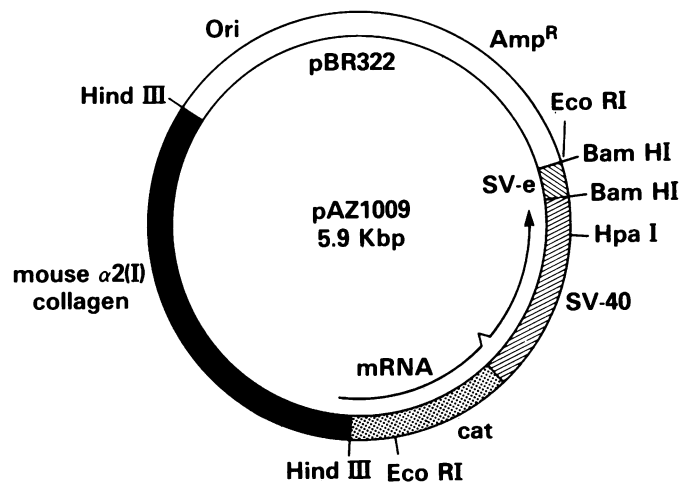


FIG. 1. Prototype recombinant plasmid used for the various constructions. The construction of this plasmid has been described previously (27). A DNA fragment between 2,000 bp upstream of the start of transcription of the mouse  $\alpha_2(I)$  collagen gene and 54 bp downstream from this site was fused to the bacterial gene for CAT (*cat*). The SV40 segment includes the SV40 small T splice sites, small T intron, and the SV40 early region polyadenylation site. SV-e is the segment between 113 and 270 in the SV40 map and includes one complete copy of the SV40 enhancer element and a portion of a second copy. Kbp, Kilobase pairs.

was determined by measuring both CAT enzyme activity and levels of CAT RNA by primer extension experiments. CAT enzyme levels were measured as described by Gorman et al. (9) with equal amounts of proteins in each assay and with a threefold higher concentration of acetyl coenzyme A. RNA was isolated by the guanidine thiocyanate-cesium chloride method (5). The primer extension experiment was carried out as described previously (27) with a synthetic 24-mer CAT primer and a 28-mer collagen primer. The hybridization conditions were similar to those described previously (27). RNA and a 50-fold excess amount of primer, which has been labeled at its 5' end with T4 polynucleotide kinase and [ $\gamma$ - $^{32}$ P]ATP, were mixed in 80% formamide-0.4 M NaCl-0.04 M PIPES (piperazine-*N,N'*-bis(2-ethanesulfonic acid), (pH 6.4)-0.001 M EDTA. The hybridization temperature was 45°C for the CAT primer. The reverse transcriptase reaction was carried out as described previously (27), but the concentration of deoxynucleotides was 250  $\mu$ M.

### RESULTS

**Evidence for cell-specific expression of the  $\alpha_2(I)$  collagen promoter.** Plasmid pAZ1009 is a recombinant plasmid in which a segment of the mouse  $\alpha_2(I)$  collagen gene, from 2,000 bp upstream of the start of transcription to 54 bp downstream of this site, is linked to the bacterial gene for CAT. This plasmid also contains a small segment of simian virus 40 (SV40) DNA, including the viral enhancer sequence (Fig. 1). We measured the levels of CAT activity 48 h after transfection of this plasmid into NIH 3T3 cells and into the MPC-11 cell line. NIH 3T3 cells synthesize the two type I collagen chains as two of their major biosynthetic products, whereas myeloma cells and other cells of the hematopoietic lineage synthesize none or very little of the type I collagen. In parallel, pSV2-CAT, a plasmid in which the early region promoter of SV40 is fused to CAT, was also transfected into both types of cells. Forty-eight hours after transfection with the plasmid containing the  $\alpha_2(I)$  collagen promoter, the

TABLE 1. Cell-specific expression of the  $\alpha_2(I)$  collagen promoter: CAT chimeric gene

Expt	Plasmid	CAT activity (cpm) in <sup>a</sup> :	
		NIH 3T3	MPC-11
1	pAZ1009	29,139	3,037
	pSV2-CAT	7,725	6,440
2	pAZ1009	45,858	100
	pSV2-CAT	11,732	2,535

<sup>a</sup> At 48 h after transfection of either NIH 3T3 cells or MPC-11 cells, extracts were made and assayed for CAT activity. CAT activity is expressed as counts per minute of [<sup>14</sup>C]chloramphenicol acetylated in a 1-h reaction.

levels of CAT activity were much higher in extracts of NIH 3T3 cells than in those of MPC-11 cells (Table 1). In one experiment, the difference was 10-fold; in the other experiments, this difference was much higher. In contrast, a much smaller difference in CAT activity was observed when these two cell types were transfected with pSV2-CAT. It should be noted that the levels of CAT enzyme after transfection of NIH 3T3 cells with pAZ1009 were consistently higher than those obtained with pSV2-CAT. These experiments suggest that the  $\alpha_2(I)$  collagen DNA sequences present in plasmid pAZ1009 are sufficient to confer cell specific expression to the  $\alpha_2(I)$  collagen gene.

**cis-Acting mutations in the  $\alpha_2(I)$  collagen promoter.** Several deletion mutations were introduced in plasmid pAZ1009; this removed various segments between -2,000 and -104 bp. All deletion mutations leave the DNA sequences around the CAT box (-87 to -82) and TATA box (-33 to -26) intact.

TABLE 2. Levels of CAT activity assayed in transient expression experiments

Plasmid	CAT activity		No. of expt
	(% of wild type $\pm$ SEM) <sup>a</sup>		
1009	100		6
1015	18 $\pm$ 9		5
1019	84 $\pm$ 23		2
1018	100 $\pm$ 1		2
1013	19 $\pm$ 14		5
1040	11 $\pm$ 1		2
1064	23 $\pm$ 9		4
1062	23 $\pm$ 5		4
1057	136 $\pm$ 33		3
1059	88 $\pm$ 24		4
1060	71 $\pm$ 21		6
1067	50 $\pm$ 23		6
1065	27 $\pm$ 9		6

<sup>a</sup> The values of CAT activity in extracts of cells transfected with the various deletions are expressed as the percentage of the value obtained in extracts of cells transfected with pAZ1009, the plasmid which contained the intact promoter and which was used as the control in each experiment.

The first series of deletions was produced by using restriction sites located at various distances upstream of the CAT box. Figure 2 schematically represents the various mutants produced by using a combination of restriction enzymes. The strategy of their construction is also briefly described above. We assayed the activity of the various deletion mutants in transient expression assays in NIH 3T3 cells and also in NIH 3T3 cells in which pAZ1009 or its derivatives were stably integrated.

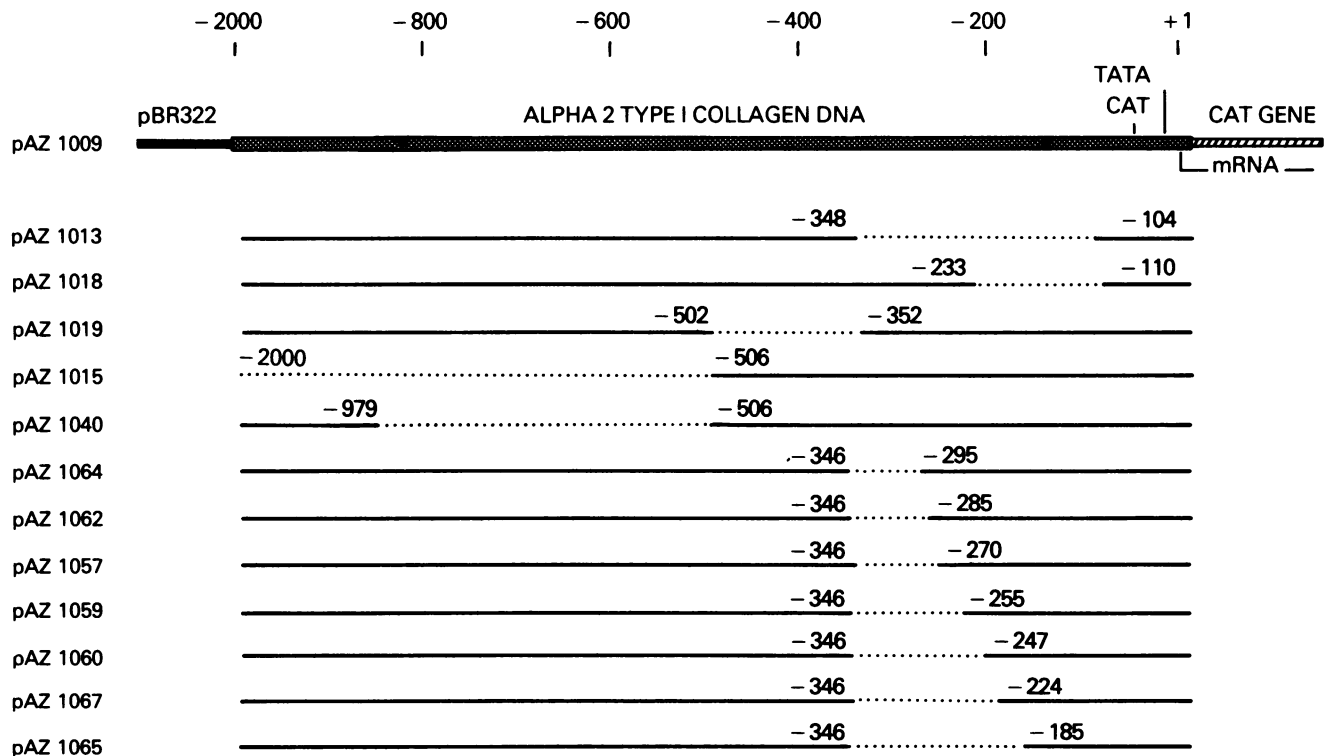


FIG. 2. Recombinant plasmids containing different deletions in the promoter region of  $\alpha_2(I)$  collagen gene fused to the CAT gene. The details of each construction are described in the text. The following restriction sites were used in these constructions: *Hind*III at -2000; *Acc*I at -983; *Xba*I at -507; *Bgl*II at -352; *Taq*I at -236; *Xma*I at -110. Positions -295 to -185 correspond to 3' endpoints of deletions of increasing size generated with *Bal*I nuclease. The correctness of the deletions was verified by DNA restriction analysis and, for some deletions, by DNA sequencing.

```

                                -1300                                -1250
CCTACTTGAA TCTCCTCACC GGAGTCGAAG GGGAGGTTTC CAGGAGTTC GCCAGTTCCT CTCATGAACA TAAAATGGC CTAAGATGC TGTGTCTGCT
                                -1200                                -1150
GTGACTTTGG GGACATAATG AGAGTCTGAA TAGGTCTCTC CAGCCTCATG ATGGATGCTC GGCAGGCCAA GGCACTGCGG CCCCTGGAAC TTGGAAGACT
                                -1100                                -1050
CCAAGAAACT TTGGGAAGGA GAATTTACTA TGTCTCTGGC TCGTCTGAC AGCCAAAGTC TCCTACTCCA ATTCTTCAGA GTCAAAGTCT CTTCCAAGA
                                -1000                                -950
TGGGGAGAGA TTGCATCTGT TCTGGAGGGG ACAGCTTGGG ATGTTAAGGA AGAGGTTGTA TACGCCCTAG GGATGATGGA AGGTGTGAAG ATGCAGGCAG
                                -900                                -850
GGGTGCTGGC AGTGGCTGTG GAAGGGGTCA TAAATAGGCA GCAAGGCCAA GAAGGCTGAG TGTTCACTGA TCTGTTATCT GATCTAGGCT CGGCATCTGA
                                -800                                -750
GTCTGGCTGG TAATCCTTCA TGGGGTCTGT ACATGGGGCT GGAAATGGGG CTTAACCCAA GTAAATGTAA TAATGCACAG TTTATCTCTT GATGCCTGCT
                                -700                                -650
CCCCTCATCA ACGAGAGAGA GGGGGGTACA GACAGAGAGA GTCAGAGAGA GGATGGAGGA GGGAGGGAGG AAGGGAGGGA GGAAGGGAGG GAGGGAGGGA
                                -600                                -550
GGGAGGAGGG AGGGAGGGAG GGAGGGAGAG ACAGAGAGAT AGAGAGACAG AGGAGAGAGG AGATTATGGC ACACAGGCCC CCCCAACTG GGGAAATTAG
                                -500                                -450
GGGGCAGGGG TGGAGCCAAC CCATGAGGGC AGCTCTAGAG TCGCCACGC TGTCACCCTC CTCCCCTCAC ACCCCAGCCC CCTCCAGCTT GCAGATGGCC
                                -400                                -350
CGCAGGGCAC AGAGTGAAGC GGGACTGGAC AGCTCCTGCC TTTTCATCTT CGGCCTCCAG GTCCTCAAC ACCCGCCCCG CCGCCCGGAG ATCTGTAAGG
                                -300                                -250
AGCCCACGTA GGTGTCCTAA AGTGCTTCCA AACTTGGCAA GGGCGAGAGA GGGCGGGTGG CTGGGGAGGG CGGAGGTATG CAGACAGGGA GTCAGAGTTC
                                -200                                -150
CCCCTCGAAA GCCTTCAAAA GTGTCCACGT CCTCAAAAAG AATGGAACCA ATTTAAGAAG CCCCAGTACC ACGTCCCTCC CCCCTCGGCT CCCTCCCCTG
                                -100                                -50
CTCCCCGCA GTCTCTCCC AGCACTGAGT CCCGGGCCCC TAGCCCTAGC CCTCCCATG GTGGAGACGT TTTTGGAGGC ACCCTCCGGC TGGGGAAACT
                                1                                50
TTTCCCATAT AAATAAGGCA GGTCTGGGCT TTATTATTTT AGCACCCAGG CAGCAGGAGG TTTCAGCTAA GTTGGAGGGA ACGGTCCACG ATTGCATGCC

```

FIG. 3. DNA sequence of the 5'-flanking region and part of the first exon of the mouse  $\alpha_2(I)$  collagen gene. The DNA sequence between -1340 and -507 was determined by the method of Maxam and Gilbert (16) by 5'-end-labeling the following sites: *Xba*I (-507), *Acc*I (-983), and *Pvu*II (-137). The DNA sequence between -507 and 60 was previously reported (27). Underlined sequences are the CAT box (-87 to -82) and the TATA box (-34 to -26). 1 refers to the start of transcription.

Table 2 shows the levels of CAT enzyme activity found in extracts of NIH 3T3 cells 48 h after transfection by the various deletion mutants. The results are average values of between two to seven experiments in which at least two different DNA preparations were used. The results are expressed as a percentage of the value of CAT activity obtained with pAZ1009, the plasmid containing the wild-type promoter which was assayed in each experiment. The absolute values of CAT activity show a certain degree of variation from one experiment to another, especially over a long period of time. This is attributable to the state of the cells, particularly to the number of passages in culture. Deletion of the DNA sequences between -348 and -104 (pAZ1013) or between -2000 and -506 (pAZ1015) results in a strong decrease in CAT enzyme levels. The residual activity obtained after transfection of these plasmids was in the range of 4 to 35% for pAZ1013 and 8 to 27% for pAZ1015 compared with that for the control plasmid. The effect of deletions between -233 and -110 (pAZ1018) and between -502 and -352 (pAZ1019) was much less marked. The length of the deletion between -2000 and -506 was narrowed by assaying a deletion of 473 bp between -979 and -506 (pAZ1040). Transfection of NIH 3T3 cells with this plasmid, followed by a transient expression assay, resulted in a low level of expression of the chimeric gene, which was similar to that found with the larger deletion from -2000 to -506. Results of all of these experiments suggest that under our experimental conditions, at least two different segments in the  $\alpha_2(I)$  collagen promoter are required for optimal expression of this promoter. The DNA sequence of the segment between -1340 and -506 was determined and is shown in Fig. 3, together with the previously determined sequence between -506 and 1 (28). We noted that the tetranucleotide AGGG is tandemly repeated 15 times in this sequence, with additional repeats in which a G residue is replaced by an A residue. In

fact, the segment between -710 and -610 of the coding strand is very rich in purines and contains only four pyrimidine residues.

To determine more specifically which DNA sequences between -346 and -104 are important for promoter activity, several deletions were made between -346 and -189 using the double-stranded exonuclease *Bal* 31. A series of deletion mutants was generated in which the 5' end of the deletion was fixed, but in which the 3' end of the deletion varied (Fig. 2). The 3' endpoints of these deletions (plasmids pAZ1057 to pAZ1065; Fig. 2) are approximations, with a possible error of  $\pm 5$  bp. When these plasmids were tested for promoter activity by the transient expression assay, the following results were obtained. Transfections with plasmids pAZ1064 and pAZ1062, which contain short deletions extending from -346 to -295 and to -285, respectively, gave rise to low CAT enzyme activity. With plasmids with longer deletions extending to -270 (pAZ1057), the level of expression of the mutated collagen promoter was similar to or greater than that of the intact collagen promoter. However, if the length of the deletions was further increased, the level of expression of the chimeric gene showed a gradual reduction. With the larger deletions extending to -185 (pAZ1065), the activity of the collagen promoter was severely reduced. The segment between -346 and -104 appears, therefore, to have complex regulatory effects, the possible significance of which will be discussed below.

**Effect of deletions in the  $\alpha_2(I)$  collagen promoter on the expression of the collagen-CAT chimeric gene in stably transfected cells.** We also wanted to test the effect of the various deletions after they had been stably introduced into the genome of NIH 3T3 cells to form permanent cell lines. Presumably, the state of the transfected DNA is more similar to that of the endogenous gene when it is stably incorporated into the genome than it is in transient expression experi-

TABLE 3. Levels of CAT activity in stably transfected cells

Plasmid	CAT activity (% of wild type $\pm$ SEM) <sup>a</sup>	No. of pools assayed
1009	100	6
1015	18 $\pm$ 10	5
1019	62 $\pm$ 17	5
1013	4 $\pm$ 2	4
1018	52 $\pm$ 17	5
1064	49 <sup>b</sup>	
1062	68 <sup>b</sup>	
1057	106 <sup>b</sup>	
1059	80 $\pm$ 32	6
1060	62 $\pm$ 19	4
1067	30 $\pm$ 8	4
1065	18 $\pm$ 5	3

<sup>a</sup> The values of CAT activity are expressed as the percentage of the value obtained with pools of cells that had been stably transfected with pAZ1009, which was used as the control in each experiment.

<sup>b</sup> Five or more isolated colonies were assayed, and the CAT levels were averaged.

ments. The plasmids were introduced into NIH 3T3 cells by cotransfection with plasmid pSV2-neo followed by selection with G-418 (18). In plasmid pSV2-neo the gene for aminoglycoside phosphotransferase was fused to the early promoter of SV40 (31). G-418-resistant cells were assayed for CAT enzyme activity as pools of approximately 300 colonies. In a few instances several (five or more) individual colonies were also tested. Table 3 summarizes the results of these experiments.

The effect of the various deletions in these cells was very similar to those that were observed in the transient expression assays. Cells which contained plasmids pAZ1013 (-348 to -104) and pAZ1015 (-2000 to -506) had very low levels of CAT enzyme activity. Cells which contained plasmids

pAZ1018 (-233 to -110) or pAZ1019 (-502 to -352) had 52 and 62% as much CAT enzyme activity as the cells with the intact collagen promoter. In permanent expression assays, the mutations which deleted various DNA segments between -346 and -185 also had effects that were similar to those observed in the transient expression assay. Deletions of DNA sequences from -346 to -295 and to -285, respectively, in pAZ1064 and pAZ1062 caused a decrease in promoter activity, although the decrease caused by these two deletions was less pronounced than in transient expression experiments. When the deletions were extended to -270 (pAZ1057), the promoter activity was restored to the level of the control plasmid. With still larger deletions, a gradual decrease in promoter activity was observed. We conclude that the various deletions have a similar effect on expression of the  $\alpha_2(I)$  collagen promoter-CAT chimeric gene whether they are assayed in permanently transfected NIH 3T3 cells or after transient expression in the same cells.

**Effects of deletions in the  $\alpha_2(I)$  collagen promoter on RNA transcription.** We wanted to examine whether the lower levels of CAT enzyme activity caused by various deletions in the  $\alpha_2(I)$  collagen promoter were caused by lower levels of CAT RNA and also whether the deletions had any effect on the location of the start site of transcription. Therefore, we performed primer extension experiments, using as primer a synthetic 24-mer oligonucleotide which hybridizes with a segment of CAT RNA located close to the 5' end of this RNA (Fig. 4). RNAs were extracted from cells 48 h after transfection of NIH 3T3 cells with plasmid DNAs containing either the full-length promoter or derivatives with deletions in this promoter. Figure 4 shows the results of these experiments. The extended primer ran as a doublet of 145 to 146 nucleotides. Its size corresponds precisely with that of a collagen-CAT mRNA [54 nucleotides of  $\alpha_2(I)$  collagen mRNA plus 75 nucleotides of CAT RNA] that starts at the

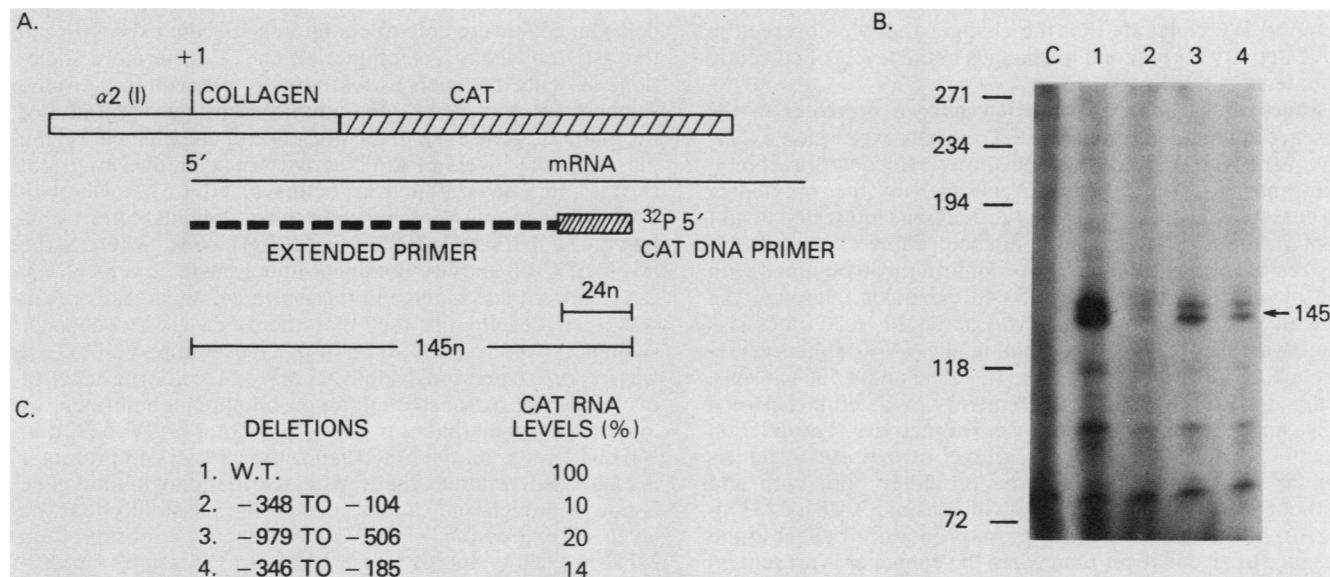


FIG. 4. Effect of deletions in the mouse  $\alpha_2(I)$  collagen promoter on the levels of correctly initiated CAT RNA. NIH 3T3 cells were transfected with the following plasmids: pAZ1009 (lanes 1 in parts B and C); pAZ1013 (lanes 2 in parts B and C); pAZ1040 (lanes 3 in parts B and C); pAZ1065 (lanes 4 in parts B and C). Forty-eight hours later the RNAs were purified and analyzed by primer extension as described in the text. (A) The primer extension experiment. (B) Electrophoretic fractionation of the labeled cDNA products on a 7 M urea-5% polyacrylamide gel. (C) RNA from untransfected cells. Numbers on left of autoradiograph indicate size of markers in nucleotides. (C) After autoradiography the counts per minute of  $^{32}P$  in the 145n extended primers was measured by liquid scintillation spectrometry. The results are expressed as the percentage of the value obtained with the wild-type (W.T.) plasmid pAZ1009.

TABLE 4. Levels of CAT activity assayed in transient expression experiments of NIH 3T3 cells

Plasmid <sup>a</sup>	CAT activity <sup>b</sup>	No. of expt
1009-A	100	3
1015-A	12 ± 1	3
1018-A	9 ± 7	3
1013-A	3 ± 1	3
1040-A	2	1
1062-A	17 ± 10	3
1057-A	102	1

<sup>a</sup> Plasmid numbers with the letter A (apposed) represent the same plasmids as represented in Fig. 1 and 2 but without the SV40 enhancer sequence. They were obtained by treatment of the respective parent plasmids with *Bam*HI enzyme, which removes the enhancer element, followed by circularization.

<sup>b</sup> Transfection experiments were performed and CAT activities are expressed as described in footnote *a* of Table 2. In each experiment, the values of CAT activity are expressed as the percentage of the value obtained with extracts of cells transfected with pAZ1009-A, the plasmid which contained the intact promoter. Extracts were assayed for 12 h.

correct place in the  $\alpha_2(I)$  collagen promoter. By performing a similar primer extension experiment with RNA extracted either from NIH 3T3 cells or from tails of mice and a DNA primer specific for endogenous  $\alpha_2(I)$  collagen mRNA, a similar cDNA doublet was found. This suggests that the RNA in the endogenous gene also starts at two adjacent nucleotides. This property is clearly preserved in the transfected promoter. The data shown in Fig. 4 indicate that the three deletions of the collagen promoter that were tested do not alter the start of CAT RNA transcription.

Figure 4 also shows that the levels of CAT RNA are strongly reduced in cells transfected with three different deletions. The radioactive bands corresponding to the extended primer were excised and counted. The results are tabulated in Fig. 4 and indicate that the levels of correctly initiated CAT RNA extracted from cells transfected with plasmids containing deletions in the  $\alpha_2(I)$  collagen promoter are reduced to the same extent as those of CAT enzyme activity. We conclude that the changes that we observed in CAT activity correspond to changes in the levels of correctly initiated CAT RNA.

**Effect of deletions in the  $\alpha_2(I)$  collagen promoter in the absence of an enhancer element.** Transient expression experiments of NIH 3T3 cells were also performed with recombinant plasmids which carry a number of the previously discussed  $\alpha_2(I)$  collagen promoter deletions but which do not contain the SV40 enhancer sequence (Table 4). In general, the results of these experiments confirm those obtained with the plasmids containing the SV40 enhancer, although the overall level of CAT activity was about 10 to 20 times less than with the plasmids containing the SV40 enhancer sequence. A large deletion between -2000 and -506 causes a sharp drop in CAT activity. Similarly, a deletion between -348 and -104 strongly decreases this activity. The deletion from -346 to -270 produces a level of expression that is similar to that of the wild-type promoter both with and without the enhancer. Deletion of plasmid 1018 (-233 to -110), however, has a much more pronounced effect in the absence of an enhancer than when the enhancer is present on the plasmid.

## DISCUSSION

Two types of plasmid constructions containing the mouse  $\alpha_2(I)$  collagen gene 5'-flanking sequences were used in our experiments. In most experiments, in addition to the  $\alpha_2(I)$  collagen promoter segment which was directly fused to the

CAT gene, the plasmid also contained a small segment of the SV40 genome identified as the SV40 enhancer. In a number of previously reported studies in which the effects of regulatory sequences on the expression of cloned genes were examined, similar enhancer sequences were present in the plasmid constructions that were used in DNA transfection experiments (18, 20, 21, 24, 29). The major reason for the presence of this viral enhancer in our experiments was that in the absence of enhancers the levels of correctly initiated CAT RNA, as measured by a primer extension assay, were difficult to detect in transient expression experiments, although the levels of CAT enzyme could be measured.

One possible reason for this low level of expression is that the segment of the  $\alpha_2(I)$  collagen promoter between -2000 and 54 does not contain all of the transcriptional regulatory elements of the  $\alpha_2(I)$  collagen gene. All *Eco*RI fragments covering the chick  $\alpha_2(I)$  collagen gene and ~5 kb of 5' and ~5 kb of 3'-flanking sequence therefore were subcloned to test whether in transient expression assays these fragments contain enhancer activity. In these experiments the chick  $\alpha_2(I)$  collagen promoter (-1300 to 110) was used as the test promoter in a construction similar to that shown in Fig. 1, in which this promoter was fused to the CAT gene. The various fragments were subcloned in an *Eco*RI site 3' to the CAT gene. Although some fragments slightly stimulated the activity of this promoter (two- to threefold), none were found which stimulated this activity fivefold or more (Chiaki Setoyama, personal communication). One cannot exclude the fact that a putative enhancer element contains an *Eco*RI site and therefore would be inactivated by the cloning procedure. We considered the possibility that the sum of several weak enhancers, which were distributed throughout this 38-kb gene, could stimulate transcription of this gene.

Results of our experiments show that sequences between -2000 and 54 in the  $\alpha_2(I)$  collagen gene have an effect that is dominant over the effect produced by the SV40 enhancer. Indeed, despite the presence of an SV40 enhancer in pAZ1009, sequences between -2000 and 54 in the  $\alpha_2(I)$  collagen gene were able to confer cell-specific expression to the  $\alpha_2(I)$  collagen-CAT chimeric gene. Furthermore, deletions of defined sequences that are far upstream in the same segment resulted in a strong inhibition in the expression of the chimeric gene even in the presence of an enhancer on the plasmid. The levels of CAT activity measured in transient expression assays after transfection of NIH 3T3 cells with pAZ1009 were much higher than those obtained after transfection of the same plasmid in MPC-11 cells, whereas the levels of CAT activity obtained after transfection of pSV2-CAT showed less change in these two cell lines. This result was seen despite the fact that the same SV40 enhancer sequences were present in both pAZ1009 and pSV2-CAT. It was reported previously that in MPC-11 cells the level of expression of a transfected kappa immunoglobulin gene is much higher than that of the early promoter of SV40 (32). In NIH 3T3 cells, on the other hand, the early SV40 promoter is expressed, whereas the transfected immunoglobulin gene is not. Although one cannot exclude the possibility that the level of expression of pSV2-CAT that we observed in MPC-11 cells is due to a high level of expression of pSV2-CAT coupled to a low efficiency of transfection, we consider this possibility to be unlikely. We postulate, therefore, that some factors that are present in NIH 3T3 cells but that are not present, or present in much lower concentrations, in MPC-11 cells interact with sequences in the  $\alpha_2(I)$  collagen promoter segment and that these interactions are necessary for the chimeric  $\alpha_2(I)$  collagen promoter-CAT gene to be



expressed, even when an SV40 enhancer is present in the transfected plasmid. Alternatively, the MPC-11 cells could contain an inhibitor which would block expression of the chimeric gene. The explanation that would attribute the difference in expression in the two cell types to the differential stability of CAT mRNA or CAT protein is unlikely. Indeed, much less difference in CAT activity was seen when the two cell types were transfected with pSV2-CAT. One might also consider the possibility that the low level of expression of the  $\alpha_2(I)$  collagen-CAT chimeric gene is caused by the fact that MPC-11 cells are transformed cells by analogy with the very low levels of expression of the  $\alpha_2(I)$  collagen-CAT chimeric gene in *v-mos*-transformed NIH 3T3 cells compared with its expression in the parent NIH 3T3 cell line (27). However, expression of the chimeric SV40 CAT gene is also inhibited in *v-mos*-transformed NIH 3T3 cells approximately to the same extent as the  $\alpha_2(I)$  collagen-CAT chimeric gene (27). Hence, it is likely that the reason why the expression of the  $\alpha_2(I)$  collagen promoter is inhibited in MPC-11 cells is different from the reason why this chimeric gene is inhibited in *v-mos*-transformed fibroblasts.

Our hypothesis that the  $\alpha_2(I)$  collagen promoter-CAT chimeric gene exhibits cell-specific expression in tissue culture cells is strengthened by the tissue-specific expression of the same gene in transgenic mice, in which the same recombinant plasmid as pAZ1009 minus the SV40 enhancer sequence was stably introduced in the germ line. In six of eight transgenic strains, the  $\alpha_2(I)$  collagen promoter-CAT chimeric gene shows a pattern of expression that is parallel to that of the endogenous type I collagen genes (H. Westphal, P. A. Overbeek, J. S. Khillan, A. B. Chpelinsky, A. Schmidt, K. A. Mahon, K. E. Bernstein, J. Piatigorsky, and B. de Crombrughe, Cold Spring Harbor Symp. Quant. Biol., in press). We believe, therefore, that both in tissue culture cells and in intact mice, the sequences present between -2000 and 54 are able to confer cell- or tissue-specific expression to the  $\alpha_2(I)$  collagen gene.

The second set of experiments described here is a deletion analysis of the 5'-flanking sequences of the  $\alpha_2(I)$  collagen gene. The results obtained with the different deletions were very reproducible in different experiments. For several deletions these results were confirmed by measuring the levels of correctly initiated CAT mRNA in transient expression experiments. Finally, the deletions had very similar effects whether they were measured in transient expression assays or after the DNAs harboring these deletions were stably incorporated in the cellular genome. All deletions left the region around the CAT box and the TATA box, as well as the segment between these conserved elements, intact. We have therefore not yet examined what effect deletions in these segments would have. It is clear that at least two segments, located between -979 and -506 and between -346 and -104, are necessary for the efficient expression of the chimeric gene. Although we cannot exclude the possibility that the sequences that are deleted are those that are needed to allow the enhancer to stimulate the  $\alpha_2(I)$  collagen promoter, we consider this to be unlikely. Indeed, we noted that when plasmids missing the SV40 enhancer were used, the effects of the deletions, in general, were similar to those obtained when plasmids containing the enhancer were used. We postulate that factors which are needed for optimal activity of this promoter interact with these DNA segments.

When an attempt was made to localize the DNA sequences between -342 and -98, which are important for optimum promoter activity, a complex picture emerged. Small-sized deletions from -346 to -285 caused a decrease

in the expression of the chimeric gene. However, when the 3' endpoint of the deletions reached the DNA sequence at -270, this inhibition was overcome, but when the size of the deletion was further increased in the 3' direction, a gradual decrease in the expression of the chimeric gene was observed. The activity of the promoter was greatly reduced when the deletion reached as far as -185. Interestingly, although a deletion between -233 and -110 (pAZ1018) had no effect by itself, deletions extending from -346 displayed a strong inhibitory effect, only if part of the -233 to -110 sequence was also deleted. The importance of the DNA sequences between positions -233 and -110 was only revealed if the upstream sequences between -346 and -235 were removed.

There are several possible interpretations of these results. One is that between -346 and -104 there are at least three segments that play different roles in the control of  $\alpha_2(I)$  collagen gene transcription. The first segment, between -346 and -285, would act as an interaction site for a positive regulatory factor. Removal of this sequence would result in a decrease in the activity of the collagen promoter. A second segment around -270 might interact with a negative regulatory factor. When this segment is removed, this negative regulation would be abolished, resulting in an increase in the level of expression of the chimeric gene. By deleting more DNA sequences downstream from this DNA segment, other segments which could be sites for the same or other positive regulatory factors would be removed. The net result would be a strong inhibition in the expression of the chimeric gene. An alternative explanation would be that the exact sizes of the deletions may play a critical role. Indeed, these deletions may result in the placement of crucial interaction sites for positive factors on the same or opposite faces of the DNA helix. Several procaryotic regulatory factors are composed of two interacting symmetrical subunits which interact with two symmetrically arranged DNA sites separated from each other by a fixed distance (7). Also, if two different proteins would interact with two different sites on the promoter but also with each other, the distance between their interaction sites may be important. In *Escherichia coli* the *gal* repressor has two interaction sites which are separated by approximately 100 bp (12). A mutation in either one of these two sites results in constitutive expression of the *gal* genes, suggesting that there is an interaction between the two sites.

We have not yet formally evaluated whether the sequences in the  $\alpha_2(I)$  collagen promoter, which are important for optimal expression of the  $\alpha_2(I)$  collagen-CAT gene, are bona fide enhancer sequences. This hypothesis will be tested by inverting these sequences or by placing them at other locations in this or other expression plasmids. Furthermore, competition experiments in which subsegments of the  $\alpha_2(I)$  collagen promoter are cotransfected with pAZ1009 should provide evidence for the existence of *trans*-acting factors that interact with specific sequences in this promoter.

Between -300 and -249 the sequence AGGCG was repeated four times (with one exception in which the C is replaced by an A). Removal of three of these repeats resulted in increased expression of the chimeric gene (see 1057). One of the few homologies between the  $\alpha_1(I)$  and  $\alpha_2(I)$  collagen promoter sequences was the sequence GGCAAG-GGCG which overlapped with these repeats at the 5' end of the segment. One of the few homologies between the  $\alpha_1(III)$  and  $\alpha_2(I)$  collagen promoter sequences was found on each side of these repeats at the 3' end of the segment. Another repeated sequence, AAAG, which was found in segments,

the removal of which results in decreased promoter activity, was found twice between -344 and -320 and three times between -243 and -207.

In brief, our data suggest that transcriptional regulatory signals in the 5'-flanking sequence of the  $\alpha_2(I)$  collagen promoter do not have a compact distribution but are located at relatively long distances upstream of the start of transcription.

#### ACKNOWLEDGMENT

P.R. is the recipient of a UICC Cancer Research Campaign fellowship.

#### LITERATURE CITED

- Adamson, E. D., and S. E. Ayers. 1979. The localization and synthesis of some collagen types in developing mouse embryos. *Cell* **16**:953-965.
- Banerji, J., S. Rusconi, and W. Schaffner. 1981. Expression of a  $\beta$ -globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**:299-308.
- Benoit, C., and P. Chambon. 1981. In vivo sequence requirements for the SV40 early promoter region. *Nature (London)* **290**:304-309.
- Bogenhagen, D. F., S. Sakonju, and D. D. Brown. 1980. A control region in the center of the 5S RNA gene directs specific initiation of transcription. II. The 3' border of the region. *Cell* **19**:27-35.
- Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**:5294-5299.
- Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, D. Kedinger, and P. Chambon. 1980. Promoter sequences of eukaryotic protein-coding genes. *Science* **209**:1406-1413.
- DeCrombrughe, B., S. Busby, and H. Buc. 1984. Cyclic AMP receptor protein: role in transcription activation. *Science* **224**:831-838.
- Gidoni, D., W. S. Dynan, and R. Tjian. 1984. Multiple specific contacts between a mammalian transcription factor and its cognate promoters. *Nature (London)* **312**:409-413.
- Gorman, C. M., L. F. Moffat, and B. Howard. 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* **2**:1044-1051.
- Gruss, P., R. Dhar, and G. Khoury. 1981. Simian virus 40 tandem repeated sequences as an element of the early promoter. *Proc. Natl. Acad. Sci. USA* **78**:943-947.
- Howard, B. H., S. L. Adams, M. E. Sobel, I. Pastan, and B. de Crombrughe. 1978. Decreased levels of collagen mRNA in Rous sarcoma virus-transformed chick embryo fibroblasts. *J. Biol. Chem.* **253**:5869-5874.
- Irani, M. H., L. Orosz, and S. Adhya. 1983. A control element within a structural gene: the gal operon of *Escherichia coli*. *Cell* **32**:783-788.
- Khoury, G., and P. Gruss. 1983. Enhancer elements. *Cell* **33**:313-314.
- Leivo, I., A. Vaheri, R. Timpl, and J. Wartiovaara. 1980. Appearance and distribution of collagens and laminin in the early mouse embryo. *Dev. Biol.* **76**:100-114.
- Liau, G., Y. Yamada, and B. de Crombrughe. 1985. Coordinate regulation of the levels of type III and type I collagen mRNA in most but not all mouse fibroblasts. *J. Biol. Chem.* **260**:531-536.
- Maxam, A., and W. Gilbert. 1980. Sequencing end-labeled DNA with base specific chemical cleavages. *Methods Enzymol.* **65**:499-560.
- Moreau, P., R. Hen, B. Wasylyk, R. Everett, M. P. Gaub, and P. Chambon. 1981. The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res.* **9**:6047-6068.
- Ohno, S., and T. Taniguchi. 1982. Inducer-responsive expression of the cloned human interferon  $\beta_1$  gene introduced into cultured mouse cells. *Nucleic Acids Res.* **10**:967-977.
- Okhubo, H., G. Vogeli, M. Mudryj, V. E. Avvedimento, M. Sullivan, I. Pastan, and B. de Crombrughe. 1980. Isolation and characterization of overlapping genomic clones covering the chicken  $\alpha_2$  (type I) collagen gene. *Proc. Natl. Acad. Sci. USA* **77**:7059-7063.
- Ostrowski, M. C., A. L. Huang, M. Kessel, R. G. Wolford, and G. L. Hager. 1984. Modulation of enhancer activity by the hormone responsive regulatory element from mouse mammary tumor virus. *EMBO J.* **3**:1891-1899.
- Page, M. J., and M. G. Parker. 1983. Androgen-regulated expression of a cloned rat prostatic C3 gene transfected into mouse mammary tumor cells. *Cell* **32**:495-502.
- Pelham, H. R. 1982. A regulatory upstream promoter element in the *Drosophila* hsp 70 heat-shock gene. *Cell* **30**:517-528.
- Prockop, D. J., and K. I. Kivirikko. 1984. Heritable diseases of collagen. *N. Engl. J. Med.* **311**:376-386.
- Queen, C., and D. Baltimore. 1983. Immunoglobulin gene transcription is activated by downstream sequence elements. *Cell* **33**:741-748.
- Sakonju, S., D. F. Bogenhagen, and D. D. Brown. 1980. A control region in the center of the 5S RNA gene directs specific initiation of transcription. I. The 5' border of the region. *Cell* **19**:13-25.
- Sandmeyer, S., B. Gallis, and P. Bornstein. 1981. Coordinate transcriptional regulation of type I procollagen genes by Rous sarcoma virus. *J. Biol. Chem.* **256**:5022-5028.
- Schmidt, A., C. Setoyama, and B. de Crombrughe. 1985. V-mos transformation regulates the expression of the  $\alpha_2(I)$  collagen promoter introduced in NIH 3T3 cells. *Nature (London)* **314**:286-289.
- Schmidt, A., Y. Yamada, and B. de Crombrughe. 1984. DNA sequence comparison of the regulatory signals at the 5' end of the mouse and chick  $\alpha_2$  type I collagen genes. *J. Biol. Chem.* **259**:7411-7415.
- Seiler-Tuyns, A., J. D. Eldridge, and B. M. Paterson. 1984. Expression and regulation of chicken actin genes introduced into mouse myogenic and nonmyogenic cells. *Proc. Natl. Acad. Sci. USA* **81**:2980-2984.
- Shore, D., M. Squire, and K. A. Nasmyth. 1984. Characterization of two genes required for the position-effect control of yeast mating-type genes. *EMBO J.* **3**:2817-2823.
- Southern, P. J., and P. Berg. 1982. Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. *J. Mol. Appl. Genet.* **1**:327-341.
- Stafford, J., and C. Queen. 1983. Cell-type specific expression of a transfected immunoglobulin gene. *Nature (London)* **306**:77-79.
- Wasylyk, B., R. Derbyshire, A. Guy, D. Molko, A. Roget, R. Teoule, and P. Chambon. 1980. Specific *in vitro* transcription of conalbumin gene is drastically decreased by single-point mutation in T-A-T-A-box homology sequence. *Proc. Natl. Acad. Sci. USA* **77**:7024-7028.