Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 8869–8873, December 1987
Biochemistry

Regulatory elements in the first intron contribute to transcriptional control of the human α1(I) collagen gene

(DNA sequence/transcription factors/promoter)

PAUL BORNSTEIN*†, JOHN MCKAY*, JUDITH K. MORISHIMA*, SREELEKHA DEVARAYALU*, AND RICHARD E. GELINAS‡

*Department of Biochemistry, University of Washington, and ‡Fred Hutchinson Cancer Research Center, Seattle, WA 98195

Communicated by Seymour J. Klebanoff, August 26, 1987 (received for review June 10, 1987)

ABSTRACT **Several lines of evidence have suggested that the regulation of type I collagen gene transcription is complex and that important regulatory elements reside 5′ to, and within, the first intron of the α1(I) gene. We therefore sequenced a 2.3-kilobase *Hin*dIII fragment that encompasses 804 base pairs of 5′ flanking sequence, the first exon, and most of the first intron of the α1(I) human collagen gene. A 274-base-pair intronic sequence, flanked by *Ava* I sites (A274), contained a sequence identical to a high-affinity decanucleotide binding site for transcription factor Sp1 and a viral core enhancer sequence. DNase I protection experiments indicated zones of protection that corresponded to these motifs. When A274 was cloned 5′ to the chloramphenicol acetyltransferase (CAT) gene, driven by an α1(I) collagen promoter sequence, and expression was assessed by transfection, significant orientation-specific inhibition of CAT activity was observed. This effect was most apparent in chicken tendon fibroblasts, which modulate their level of collagen synthesis in culture. We propose that normal regulation of α1(I) collagen gene transcription results from an interplay of positive and negative elements present in the promoter region and within the first intron.**

Type I procollagen is a major secreted product of a variety of cells, and the synthesis of this protein is subject to a wide range of control, as part of developmental programs and in cellular differentiation. However, relatively little is known concerning the cis-acting regulatory elements in either of the α1(I) and α2(I) genes that code for the two homologous polypeptides in the protein.

Early studies of the reduction in collagen synthesis observed in cells transformed with retroviruses provided evidence that regulation occurred at the transcriptional level (1, 2). Schmidt *et al.* (3) subsequently showed that transformation of NIH 3T3 cells with a mixture of Moloney sarcoma and murine leukemia viruses significantly reduced the transcription of transfected plasmids containing the chloramphenicol acetyltransferase (CAT) gene driven by either the chicken or mouse α2(I) promoter and upstream elements. Schmidt *et al.* (3) therefore concluded that retroviral transformation inhibited the expression of endogenous and transfected α2(I) promoters in a similar fashion. These investigators have proceeded to examine the expression of plasmids in which the CAT gene was driven by the mouse α2(I) collagen promoter, from −2000 base pairs (bp) to +54 bp from the start site of transcription of the α2(I) gene. Deletions in this promoter suggest that several regions in this sequence may be important for expression of the α2(I) gene (4).

An interesting clue to the regulation of the α1(I) gene in mice has come from the studies of Jaenisch and coworkers. Schniecke *et al.* (5) and Harbers *et al.* (6) found that the Mov-13 strain, which is an embryonic lethal mutation in the homozygous form, results from insertion of Moloney murine leukemia virus into the first intron of the α1(I) collagen gene. The mouse α1(I) gene (7) as well as the human α1(I) gene (8) and the chicken α2(I) gene (9) contain DNase I-hypersensitive sites in the first intron. Virus insertion causes changes in the methylation pattern (10) and in the chromatin conformation of the mutated gene. Specifically, a DNase I-hypersensitive site 5′ to the start of transcription, which is associated with active transcription of the normal collagen gene, is absent in the mutant allele (7). Hartung *et al.* (11) have compared the activity of the α1(I) collagen promoter in cells derived from wild-type and Mov-13 embryos by nuclear run-on transcription and S1 nuclease mapping. They showed that initiation of transcription of the mutant gene was reduced by a factor of 20–100. These findings, taken together, suggest that the structure of the first intron of the α1(I) gene could participate in the regulation of its transcription.

In an initial analysis of the function of the first intron of the human α1(I) gene,§ we have observed that a 274-bp *Ava* I–*Ava* I fragment is capable of inhibiting the transcription of the CAT gene, under control of a collagen promoter sequence, in transient transfection experiments. Such inhibitory activity could contribute to the array of regulatory influences to which the type I collagen genes are subject.

MATERIALS AND METHODS

Plasmid Construction. The 2.3-kilobase (kb) *Hin*dIII–*Hin*dIII fragment (E) of the α1(I) collagen gene, derived from the cosmid CG103 (8), was subcloned into the *Hin*dIII site of pUC19 to produce pUC19-E. An *Ava* I digest of pUC19-E was fractionated by acrylamide gel electrophoresis, and a 274-bp fragment containing a unique *Bgl* II site (A274; bases 820–1093, see Fig. 1) was isolated by electroelution. The fragment was blunt-ended with the Klenow fragment of DNA polymerase I, ligated to synthetic *Bam*HI linkers, and cloned into pUC19 to produce A274-pUC19.

pSV1-CAT was a gift of F. Yoshimura and T. Hollon (University of Washington, Seattle). The plasmid was derived from pUC18 and pSV2-CAT (12) and contains a multiple cloning polylinker and the CAT gene driven by the simian virus 40 (SV40) early promoter but lacks the 72-bp SV40 enhancer contained in pSV2-CAT. A274 was cloned into the *Bam*HI site of pSV1-CAT in both orientations to produce A274(+)pSV1-CAT and A274(−)pSV1-CAT.

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: CAT, chloramphenicol acetyltransferase; SV40, simian virus 40; BHK, baby hamster kidney; CTF, chicken tendon fibroblast(s).

†To whom reprint requests should be addressed.

§This sequence is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03559).

The SV1 promoter was removed from pSV1-CAT and a 430-bp fragment from −332 to 98 in the collagen gene (Fig. 1) was substituted to produce pCOL-CAT (Fig. 2). A274 was cloned into the *Bam*HI site of pCOL-CAT in both orientations to produce A274(+)pCOL-CAT and A274(−)pCOL-CAT (see Fig. 2).

DNA Sequence Analysis. The 2.3-kb *Hin*dIII fragment was subcloned into the filamentous phage vector M13mp19 (17) and subclones were recovered that carried the 2.3-kb *Hin*dIII fragment in both orientations, as shown by hybridization analysis of single-stranded DNA (18). Double-stranded DNA was subjected to unidirectional digestion with exonuclease III; this procedure facilitated the generation of an overlapping set of deletions by a concerted enzymatic method (19). This set of subclones was sequenced by the chain-termination method of Sanger *et al.* (20). The sequence of each strand was determined independently. The position of the cap site and the junction for the first exon were taken from Chu *et al.* (21).

DNase I Protection ("Footprinting"). HeLa and human lymphoid cell (BJAB) nuclear extracts were prepared as described by Dignam *et al.* (22) and Wildeman *et al.* (16) and dialyzed against 20 mM Hepes, pH 7.9/20 mM KCl/1 mM $MgCl_2$/0.5 mM dithiothreitol/17% glycerol. Nuclear extracts of mouse L cells, prepared by minor modifications of the method described by Piette *et al.* (23), were dialyzed against the same buffer.

The *Bam*HI-linkered A274 collagen intronic fragment, cloned into the *Bam*HI site of pUC19, was 5′ end-labeled with ^{32}P on the coding or noncoding strand with T4 polynucleotide kinase. DNase I footprinting was performed by modifications of the methods described by Wildeman *et al.* (16) and Augereau and Chambon (24). Nuclear extract (4–6 μl) was incubated on ice for 15 min in 10 μl containing 10% glycerol, 12 mM Hepes (pH 7.9), 50 mM KCl, 2.5 mM $MgCl_2$, 0.5 mM dithiothreitol, and 80–100 ng of sonicated salmon sperm DNA. One microliter of ^{32}P-labeled template (6–9 × 10^4 cpm, ≈50 fmol) was then added and the samples were incubated for

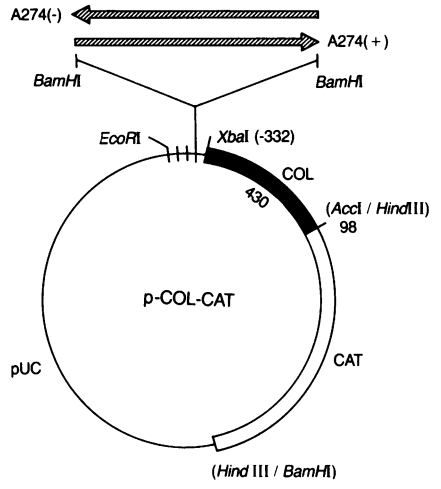

FIG. 2. Structure of pCOL-CAT. The CAT gene, driven by a 430-bp collagen 5′ flanking sequence from −332 to +98, was inserted in pUC18. The CAT sequence contains, in addition to the CAT structural gene, the small t intron and a poly(A) signal (12). *Bam*HI-linkered A274 was inserted in either orientation into the *Bam*HI site to produce A274(+)pCOL-CAT or A274(−)pCOL-CAT.

```
                10        20        30        40        50        60        70        80        90
                .         .         .         .         .         .         .         .         .
                                                                                          AAGC
-800  TTTCTTCTGCAGAGGCCACCTAGTCATGTTTCTCACCCTGCACCTCAGCCTCCCCACTCCATCTCTCAATCATGCCTAGGGTTTGGAGGAAGGCATTTGA
-700  TTCTGTTCTGGAGCACAGCAGAAGAATTGACATCCTCAAAATTAAAACTCCCTTGCCTGCACCCCTCCCTCAGATATCTGATTCTTAATGTCTAGAAAGG
-600  AATCTGTAAATTGTTCCCCAAATATTCCTAAGCTCCATCCCCTAGCCACACCAGAAGACACCCCCAAACAGGCACATCTTTTTAATTCCCAGCTTCCTCT
-500  GTTTTGGAGAGGTCCTCAGCATGCCTCTTTATGCCCCTCCCTTAGCTCTTGCCAGGATATCAGAGGGTGACTGGGGCACAGCCAGGAGGACCCCCTCCCC
-400  AACACCCCCAACCCTTCCACCTTTGGAAGTCTCCCCACCCAGCTCCCCAGTTCCCCAGTTCCACTTCTTCTAGATTGGAGGTCCCAGGAAGAGAGCAGAG
-300  GGGCACCCCTACCCACTGGTTAGCCCACGCCATTCTGAGGACCCAGCTGCACCCCTACCACAGCACCTCTGGCCCAGGCTGGGCTGGGGGGCTGGGGAGG
-200  CAGAGCTGCGAAGAGGGGAGATGTGGGGTGGACTCCCTTCCCTCCTCCTCCCCCTCTCCATTCCAACTCCCAAATTGGGGGCCGGGCCAGGCAGCTCTGA
-100  TTGGCTGGGGCACGGGCGGCCGGCTCCCCCTCTCCGAGGGGCAGGGTTCCTCCCTGCTCTCCATCAGGACAGTATAAAAGGGGCCCGGGCCAGTCGTCGG
CAP SITE +1
     AGCAGACGGGAGTTTCTCCTCGGGGTCGGAGCAGGAGGCACGCGGAGTGTGAGGCCACGCATGAGCGGACGCTAACCCCCTCCCCAGCCACAAAGAGTCT
                                MetPheSerPheValAspLeuArgLeuLeuLeuLeuLeuAlaAlaThrAlaLeuLeuThrHisGlyGlnGluGluGlyGln
 101 ACATGTCTAGGGTCTAGACATGTTCAGCTTTGTGGACCTCCGGCTCCTGCTCCTCTTAGCGGCCACCGCCCTCCTGACGCACGGCCAAGAGGAAGGCCAA
                ValGluGlyGlnAspGluAsp
 201 GTCGAGGGCCAAGACGAAGACAGTAAGTCCCAAACTTTTGGGAGTGCAAGGATACTCTATATCGCGCCTTGCGCTTGGTCCCGGGGGCCGCGGCTTAAAA
 301 CGAGACGTGGATGATCCGGAGACTCGGGAATGGAAGGGAGATGATGAGGGCTCTTCCTCGGCGCCCTGAGACAGGAGGGAGCTCACCCTGGGGCGAGGTT
 401 GGGGTTGAACGCGCCCCGGGAGCGGGAGGTGAGGGTGGAGCGCGGCGTGAGTTGGTGCAAGAGAGAATCCCGAGCGCGCAACCGGGGAAGTGGGGATCTT
 501 GGGTGCAGAGTGAGGAAAGCACGTCGAAGATGGGATGGGGGCGCCGAGCGGGGCATTTGAAGCCCAAGATGTAGAAGCAATCAGGAAGGCCGTGGGATGA
 601 TTCATAAGGAAAGATTGCCCTCTCTGCCGGCTAGAGTGTTGCTGGGGCCGTGGGGGTGCTGGGCAGCCGCGGAGGGGGTGCGGAGCGTGGGCGGGTGGAG
 701 GATGAGAAACTTTGGCGCGGACTCGGCGGGGCGGGGTCCTTGCGCCCCCTGCTGACCGATGCTGAGCACTGCGTCTCCCGGTCCAACGCTTACTGGGGCA
 801 GGAGCCGGAGCGGGAAGACCCGGGTTATTGCTGGGTGCGGACCCCCACCTGCTAGATCTGGAAAGTAAAGCCAGGGATGGGGCAGCCCAAGCCTCTTAAA
 901 GAGGTAGTCGGGCCGGTGAGGTCGGCCCCGCCCCGGCCCCATTGCTTAGCGTTGCCCGACACCTAGTGGCCGTCTGGGGAGCCGCTAGCGCGGTGGGAGT
1001 GGTTAGCTAACTTCTGGACTATTTGCGGACTTTTTGGTTCTTTGGCTAAAAGTGACCTGGAGGCATTGGCTGGCTTTGGGGACTGGGGATGGCCCGAGAG
1101 CGGCTTTTAAGTTGTCTAGGTGCTGGAGGTTAGGGTGTCTCCTAATTTAGGTACATTTCAAGTCTTGGGGGGGCCTCCCTTCCAATCAGCCGCTCCCATT
1201 CTCCTAGCCCCGCCCCCGCCACCCCACCTGCCCAGGGAATGGGGGCGGGATGAGGGCTGGACCTCCCTTCTCTCCTCCCTCGCCCTCCTCCTGTCTCTAC
1301 CACGCAAGCCACTCCCCACGAGCCTGCCCTCCCGATGGGGCCCCTCCTATTCTCCCCCCGCCCTCCCCCTCTCACCCTGTGGTTTTTATTTCACTTGGCT
1401 TCAGCGCCAATGGGCTGAGGTTGGAGTTGGAAGCCACCGCGGACTAAAGCTT
```

FIG. 1. Nucleotide sequence of the 2.3-kb *Hin*dIII fragment of the α1(I) collagen gene. The amino acid sequence of the translated part of the first exon, corresponding to the signal peptide and 12⅓ amino acids of the NH_2-terminal propeptide, is also given. The start of transcription at base +1 (cap site) is indicated. TATAAA and CCAAAT sequences are boxed. Also boxed are four decanucleotide GC motifs shown to have high or medium affinity for the transcription factor Sp1 (13). The junction between exon 1 and intron I (bases 222–223) is indicated by a vertical line. A sequence that resembles the adenovirus E1A enhancer element (14) is overlined and two viral core enhancer motifs (15) are underlined. The 274-bp *Ava* I intronic fragment, A274, is flanked by brackets.

10 min at 20°C. One microliter of DNase I (50–500 ng in nuclear extraction buffer, freshly diluted from a 1 mg/ml solution in 50% glycerol/50 mM Tris·HCl, pH 7.5/50 mM KCl/5 mM $MgCl_2$/0.5 mM dithiothreitol) was added and digestion was allowed to proceed for 90 sec at 20°C. Control digestion of template in the absence of nuclear extract was performed with 5–10 ng of DNase. The reaction was terminated by the addition of 250 μl of 0.3% $NaDodSO_4$/0.15 M NaCl and phenol/$CHCl_3$ extraction. Nucleic acids were precipitated from ethanol in the presence of 15 μg of glycogen and analyzed on 6% polyacrylamide gels containing 50% urea.

Gene Transfer into Mammalian Cells by Transient Transfection. Baby hamster kidney (BHK) cells, HeLa cells, L cells, and chicken tendon fibroblasts (CTF) were grown in monolayer culture in 60-mm dishes. CTF were isolated from 17-day chicken embryos as described (25) with the modification that partially purified bacterial collagenase (CLSP-A, Worthington) was substituted for trypsin. Cells were grown in Dulbecco's modification of Eagle's medium with 10% fetal calf serum. CTF were transfected after four or five passages in medium without added ascorbate, conditions under which the initially high level of collagen synthesis is markedly reduced (26). Plasmid DNA was isolated by twice-banding in CsCl gradients, which was followed by proteinase K and RNase A digestion and precipitation from ethanol. DNA (3 or 5 μg per dish) was added as a calcium phosphate precipitate (27). The culture medium was changed 24 hr later and cells were harvested 48 hr after addition of DNA. Cell lysates were obtained by three cycles of freeze-thawing and supernatants were assayed for CAT activity by acetylation of [^{14}C]chloramphenicol (27). In experiments in which a plasmid containing the β-galactosidase gene was cotransfected as a control for the efficiency of transfection, cell lysates were assayed for β-galactosidase activity by fluorescence with 4-methylumbelliferyl β-D-galactoside as a substrate.

RESULTS

Nucleotide Sequence of the 2.3-kb *Hin*dIII Fragment of the α1(I) Collagen Gene. The nucleotide sequence of the 2256 bases comprising the *Hin*dIII–*Hin*dIII fragment at the 5′ end of the human α1(I) collagen gene is shown in Fig. 1. The predicted amino acid sequence for the translated portion of the first exon is also shown. The sequence consists of a 5′ flanking sequence of 804 bases, the first exon of 222 bases, and 1230 bases of the first intron, which is ≈1.4 kb long (28). The start site of transcription, previously determined by Chu *et al.* (21) by S1 nuclease mapping, is located at base +1 and is preceded by TATAAA and CCAAAT sequences, which are boxed in Fig. 1. Two canonical CCAAT sequences are present on the opposing strand at −97 to −101 and −122 to −126. The sequence is largely in agreement with the sequence −275 to 281 published by Chu *et al.* (21), although a number of differences exist, particularly 5′ to the CCAAT motif.

Four GC motifs, shown to have high or medium affinity for the transcription factor Sp1 (13), are also enclosed within boxes in Fig. 1. One of these is located within the intronic fragment A274, shown below to have regulatory properties for collagen gene transcription. There are also two sequences, underlined in Fig. 1, that closely resemble the viral core enhancer consensus sequence GTGG$\overset{AAA}{TTT}$G (15). The 5′ flanking sequence also contains a sequence, overlined in Fig. 1, that is reminiscent of the adenovirus E1A enhancer element (14). Two very similar pyrimidine-rich sequences are found in the 5′ flanking sequence at −167 to −136 and in the intron at 1262–1298.

Protein Binding to a Collagen Intronic Sequence. The 274-bp intronic sequence, flanked by *Ava* I restriction sites (A274, bases 820–1093, Fig. 1), was chosen for initial study because this sequence contained a complete high-affinity GC motif as well as a viral core enhancer sequence. This region also encompasses a DNase I-hypersensitive site in fibroblast and placental chromatin; this site was mapped 0.5 kb from the 3′ end of the 2.3-kb *Hin*dIII fragment—i.e., at about base 950 (8). The collagen intronic fragment A274 was 5′ end-labeled with ^{32}P on the noncoding strand and subjected to partial digestion with DNase I in the absence or presence of nuclear extracts (Fig. 3). The fragmentation pattern clearly shows three regions of protection (footprints) flanked by DNase I-hypersensitive sites. Based on coelectrophoresis of the products of an A + G sequence reaction and a 5′ ^{32}P-end-labeled *Msp* I digest of pBR322, these regions correspond to bases 919–944 (I), 951–978 (II), and 986–1009 (III). Region I contains a high-affinity (but inverted) decanucleotide GC motif, GCCCCGCCCC (29) (bases 925–934), and regions II and III are (G + C)-rich; in addition, region III contains a viral core enhancer element (15).

We used nuclear extracts from HeLa and lymphoid cells as sources of ubiquitously distributed transcription factors and extracts of L cells, which are fibroblastic and may preferentially express factors involved in regulation of collagen gene transcription. All three nuclear extracts induced similar zones of protection; however, it would appear that the L-cell extract was most effective since the footprint in region II was more complete with this extract, even though a smaller amount of protein was used.

A274 was also 5′ end-labeled with ^{32}P on the coding strand and subjected to DNase I footprinting (Fig. 4). The fragmentation pattern again shows regions of protection that parallel those observed with the noncoding strand. In the case of the coding strand, region I is represented by a footprint extending

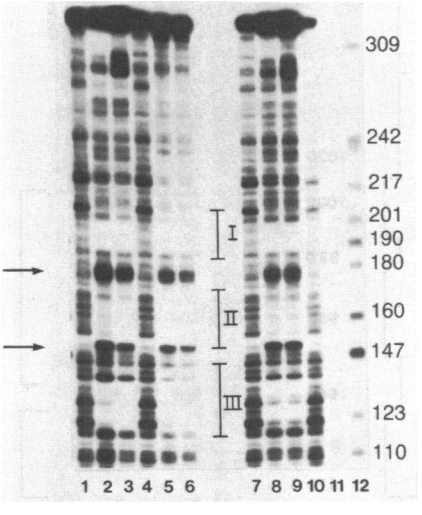

FIG. 3. DNase I cleavage pattern of the noncoding strand of the collagen intronic sequence A274 in the presence of nuclear extracts. The vertical lines mark three regions of protection (footprints): I, II, and III. Unlabeled arrows indicate DNase I-hypersensitive sites. Sizes are given in bp. Lane 1, cleavage with 5 ng of DNase I in the absence of nuclear extract (naked DNA); lane 2, 40 μg of HeLa extract, 200 ng of DNase I; lane 3, 40 μg of HeLa extract, 150 ng of DNase I; lane 4, naked DNA; lane 5, 24 μg of L-cell extract, 200 ng of DNase I; lane 6, 24 μg of L-cell extract, 150 ng of DNase I; lane 7, naked DNA; lane 8, 40 μg of lymphoid cell extract, 500 ng of DNase I; lane 9, 40 μg of lymphoid cell extract, 300 ng of DNase I; lane 10, naked DNA; lane 11, A + G sequence ladder of A274 (a longer exposure of the autoradiogram permitted visualization of the sequence to beyond the position of region I); lane 12, 5′ end-labeled (with ^{32}P) *Msp* I digest of pBR322.

from base 920 to 940 and regions II and III are represented by a zone of partial protection extending from base 950 to 1000 (Fig. 4). DNase I-hypersensitive sites at about base 950 can be seen in some of the lanes in Fig. 4.

Orientation-Specific Inhibition of CAT Gene Expression by a Collagen Intronic Sequence. When pCOL-CAT was introduced into BHK cells, HeLa cells, or CTF by calcium phosphate-mediated transfection, substantial CAT activity was detected in all three cell types. Although extensive comparisons were not performed, the collagen promoter appeared more effective in directing the transcription of the CAT structural gene than was the SV1 promoter, and this effectiveness was more pronounced in CTF, cells known to be capable of synthesizing high levels of collagen (under some conditions) *in vitro* (26). Indeed, in CTF the collagen promoter was more effective than was the SV40 promoter in combination with the 72-bp SV40 enhancer (data not shown).

Consistently, the presence of A274 in the (−) orientation 5′ either to the SV40 basal promoter or to the collagen promoter inhibited the transcription of the CAT gene when plasmids were transfected into CTF (Table 1). The fold inhibition with the collagen promoter varied from 4 to 20 and averaged about 5 in 12 experiments. Occasionally, when CAT activity was high in cells transfected with pCOL-CAT, less inhibition was achieved by addition of the collagen intronic sequence. In no experiment was stimulation observed. In contrast, less inhibition (about 2- to 3-fold in 7 experiments) was observed when A274 was introduced in the (+) orientation. Similar results were obtained with the SV40 basal promoter (Table 1). The inhibition observed with A274 is not likely to be due to titration of limiting amounts of an essential transcription factor necessary for binding to the promoter (e.g., SP1) since cotransfection of a 12-fold excess of A274-pUC19 did not change CAT activity (data not shown).

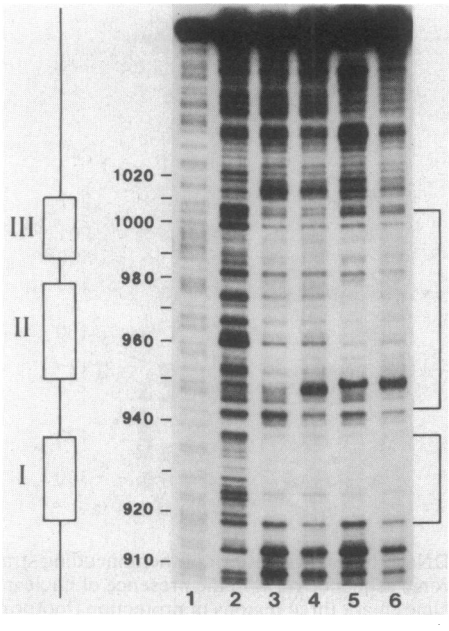

FIG. 4. DNase I cleavage pattern of the coding strand of the collagen intronic sequence A274 in the presence of nuclear extracts. Lane 1, A + G sequence ladder; lane 2, cleavage with 5 ng of DNase I in the absence of nuclear extract (naked DNA); lane 3, 36 μg of L-cell extract, 100 ng of DNase I; lane 4, 24 μg of L-cell extract, 50 ng of DNase I; lane 5, 60 μg of HeLa extract, 200 ng of DNase I; lane 6, 40 μg of HeLa extract, 100 ng of DNase I. The regions of protection from DNase I cleavage—I, II, and III (determined as indicated in Fig. 3)—are indicated by open rectangles. Base numbers (with base 1 representing the start of transcription) were derived from comparison of the A + G reaction with the DNA sequence in Fig. 1.

Table 1. Modulation of CAT activity by the collagen intronic sequence A274 in transient transfection experiments

		CAT activity*					
		CTF		BHK cells		HeLa cells	
Group	Plasmid	%	*n*	%	*n*	%	*n*
A	pSV1-CAT	100	3	100	3		
	A274(+)pSV1-CAT	58 ± 3.8	3	65 ± 11.9	3		
	A274(−)pSV1-CAT	21 ± 4.2	3	38 ± 6.5	3		
B	pCOL-CAT	100	12	100	3	100	2
	A274(+)pCOL-CAT	44 ± 5.6	7	100	3	84	2
	A274(−)pCOL-CAT	19 ± 2.6	12	63 ± 10.2	3	106	2

Cells received 4 μg of CAT plasmid and 1 μg of Rous sarcoma virus β-galactosidase plasmid per 60-mm dish in group A and 2 μg of CAT plasmid and 1 μg of Rous sarcoma virus β-galactosidase plasmid per 60-mm dish in group B.

*CAT activity (conversion of chloramphenicol to acetylchloramphenicol) was measured after a 90-min incubation and was corrected in each case to reflect an equal degree of β-galactosidase activity. The % conversion was set at 100 for pSV1-CAT and pCOL-CAT, and the effects of A274(+) and A274(−) were indicated as a percentage ± the standard error of the mean. *n*, Number of determinations.

When CAT gene constructs containing either the collagen or SV40 promoter were transfected into BHK or HeLa cells, two major differences were observed in comparison with transfection of CTF. The collagen promoter was more effective in directing transcription of the CAT gene in CTF than it was in BHK or HeLa cells. Thus, in one of the experiments summarized in Table 1, using the same level of specific transfected DNA (2 μg per 60-mm culture dish), 10 μg of CTF cell protein was sufficient to produce 41% conversion of chloramphenicol to acetylchloramphenicol, whereas 100 μg of HeLa cell protein was required to produce 5% conversion. These numbers were normalized with reference to cotransfected β-galactosidase activity and therefore cannot be explained by differences in the effectiveness of transfection in these cells. BHK cells were more responsive to the collagen promoter than were HeLa cells but far less so than CTF (data not shown).

A274 in the (+) or (−) orientation produced little inhibition of CAT gene transcription in either BHK or HeLa cells when used in conjunction with the collagen promoter (Table 1). However, BHK cells did respond negatively to the A274 element, particularly in the (−) orientation, when the CAT gene was driven by the SV40 basal promoter, although the effect was less than in CTF (Table 1). Thus, CTF, which are capable of producing collagen, recognize the collagen promoter more efficiently than do BHK or HeLa cells, which synthesize little or no collagenous protein. In addition, the collagen intronic sequence is more effective as a modulator of transcription in CTF than in BHK or HeLa cells.

DISCUSSION

It is likely that the mechanisms that have developed for regulation of type I collagen gene expression are complex and are based in part on transcriptional control. [See Bornstein *et al.* (30) for a brief review.] Since the available evidence suggests that the α1(I) and α2(I) genes coding for type I collagen, as well as the α1(III) gene coding for type III collagen, are usually coordinately expressed (31), clues to the regulation of α1(I) may be gained by a study of the promoter and 5′ flanking sequences of the α2(I) and α1(III) genes. Comparison of the 5′ regions of the mouse and chicken α2(I) genes shows several areas of substantial similarity (32). The most striking is a sequence that surrounds the translational initiation site. It has been suggested that this sequence could code for an inverted complement in the mRNA, which may

assume a relatively stable stem–loop structure (33). Similarly conserved sequences have been discovered in the α1(I) and α1(III) genes (21, 34). Other conserved features include TATAAA and CCAAT motifs, inverted G/C-rich complementary sequences bracketing the CCAAT motif, and a pyrimidine-rich stretch. The latter sequence, in the mouse and chicken genes, is sensitive to S1 nuclease when introduced into supercoiled bacterial plasmids (35). A DNase I-hypersensitive site has also been found in this region in chromatin (9). These data collectively implicate the 5′ flanking region of collagen genes in transcriptional control.

In this work we provide evidence that a 274-bp sequence within the first intron of the human α1(I) gene (bases 820–1093; Fig. 1) contributes to the inhibition of transcription of this gene. The intronic sequence contains a high-affinity Sp1 binding site (bases 925–934) and a viral core enhancer motif (bases 999–1007). As shown by DNase I protection footprinting of the coding and noncoding strands of A274 (Figs. 3 and 4), DNA binding proteins that recognize sequences within the intronic segment occur in nuclear extracts of HeLa cells, lymphoid cells, and L-cell fibroblasts. This protection matches the location of the GC box and viral core enhancer motif.

It would at first appear paradoxical that a collagen-synthesizing cell such as the CTF should recognize an inhibitory element within the collagen gene. However, it should be emphasized that under the conditions used for growth and transfection of cells, collagen synthesis was reduced substantially compared with the rate of synthesis in these cells immediately after their removal from the tissue (ref. 26; unpublished observations). The hypothesis that the ability of A274 to negatively regulate the activity of a collagen promoter correlates with the level of collagen synthesis in CTF awaits testing. Thus, it may be possible, with ascorbate supplementation, to reduce the effectiveness of A274 as an inhibitor of collagen gene transcription.

Since nuclear extracts from fibroblast-like cells (L cells) as well as HeLa and lymphoid cells provide protection of the intronic segment of A274 from DNase I digestion (Figs. 3 and 4), the relevant protein factors would seem to be broadly distributed. Similar considerations apply to the proteins that bind to the immunoglobulin heavy chain enhancer (36). There is, indeed, increasing evidence that many cis-acting elements may be shared in viral and cellular genes and that common trans-acting factors may function differently in different genes depending on a combinatorial arrangement of binding sites and protein factors (37).

We suggest that a limited number of cis-acting elements, and their corresponding proteins, combine to confer appropriate regulation of cellular genes and to provide tissue specificity. This assumption requires that transcriptional complexes will differ as a consequence of the number and arrangement of regulatory elements and that each gene will possess a unique combinatorial arrangement of these elements. In the case of the α1(I) collagen gene, regulatory elements in the first intron, although in themselves not unique in character, would interact through DNA binding proteins with promoter and upstream elements to provide transcriptional control appropriate for that gene. The ability of DNA binding proteins to act at a distance could be explained by a looping model, as discussed by Ptashne (38).

If this analysis is correct, an inhibitory or silencer function for the intronic elements might occur only if the levels of the relevant DNA binding proteins in the cell were sufficiently high. Conceivably, the same sequences could act in either a stimulatory or inhibitory fashion depending on the relative effective concentrations of these proteins. Complex cellular control, appropriate to the needs for type I collagen synthesis in different circumstances, could therefore be effected by the relative abundance of the several DNA binding proteins and competition for binding provided by other genes in an "open" conformation in chromatin. Posttranscriptional regulatory mechanisms would serve to augment this repertoire.

Note. While this manuscript was under review, Rossi and de Crombrugghe (39) reported related results. A related paper also appears in this issue (40).

We thank F. Yoshimura and T. Hollon for plasmid pSV1-CAT. We also thank D. Duksin for assistance with the early phases of the transfection experiments, R. Palmiter and H. Sage for useful discussions, and K. Doehring for secretarial assistance. This work was supported by National Institutes of Health Grants AM112248, AM31232, DE02600, HL18645, and HL37073.

1. Avvedimento, E., Yamada, Y., Lovelace, E., Vogeli, G., de Crombrugghe, B. & Pastan, I. (1981) *Nucleic Acids Res.* **9**, 1123–1131.
2. Sandmeyer, S., Gallis, B. & Bornstein, P. (1981) *J. Biol. Chem.* **256**, 5022–5028.
3. Schmidt, A., Setoyama, C. & de Crombrugghe, B. (1985) *Nature (London)* **314**, 286–289.
4. Schmidt, A., Rossi, P. & de Crombrugghe, B. (1986) *Mol. Cell. Biol.* **6**, 347–354.
5. Schniecke, A., Harbers, K. & Jaenisch, R. (1983) *Nature (London)* **304**, 315–320.
6. Harbers, K., Kuehn, M., Delius, H. & Jaenisch, R. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1504–1508.
7. Brindl, M., Harbers, K. & Jaenisch, R. (1984) *Cell* **38**, 9–16.
8. Barsh, G. S., Roush, C. L. & Gelinas, R. E. (1984) *J. Biol. Chem.* **259**, 14906–14913.
9. Merlino, G. T., McKeon, C., de Crombrugghe, B. & Pastan, I. (1983) *J. Biol. Chem.* **258**, 10041–10048.
10. Jähner, D. & Jaenisch, R. (1985) *Nature (London)* **315**, 594–597.
11. Hartung, S., Jaenisch, R. & Brindl, M. (1986) *Nature (London)* **320**, 365–367.
12. Gorman, C., Moffat, L. & Howard, B. (1982) *Mol. Cell. Biol.* **2**, 1044–1051.
13. Kadonaga, J. T., Jones, K. A. & Tjian, R. (1986) *Trends Biochem. Sci.* **11**, 20–23.
14. Hearing, P. & Shenk, T. (1983) *Cell* **33**, 695–703.
15. Weiher, H., Konig, M. & Gruss, P. (1983) *Science* **219**, 626–631.
16. Wildeman, A. G., Sassone-Corsi, P., Grundström, T., Zenke, M. & Chambon, P. (1984) *EMBO J.* **3**, 3129–3133.
17. Norrander, J., Kempe, T. & Messing, J. (1983) *Gene* **26**, 101–106.
18. Messing, J. (1983) *Methods Enzymol.* **101**, 20–78.
19. Henikoff, S. (1984) *Gene* **28**, 351–359.
20. Sanger, F., Nicklen, S. & Coulson, A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
21. Chu, M.-L., de Wet, W., Bernard, M. & Ramirez, F. (1985) *J. Biol. Chem.* **260**, 2315–2320.
22. Dignam, J. D., Lebowitz, R. M. & Roeder, R. G. (1983) *Nucleic Acids Res.* **14**, 1475–1489.
23. Piette, J., Kryszki, M.-H. & Yaniv, M. (1985) *EMBO J.* **4**, 2675–2685.
24. Augereau, P. & Chambon, P. (1986) *EMBO J.* **5**, 1791–1797.
25. Duksin, D. & Bornstein, P. (1977) *J. Biol. Chem.* **252**, 955–962.
26. Schwarz, R. I. & Bissell, M. J. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4453–4457.
27. Gorman, C. (1985) in *DNA Cloning, A Practical Approach*, ed. Glover, D. M. (IRL, Oxford), Vol. 2, pp. 143–190.
28. Chu, M.-L., de Wet, W., Bernard, M., Ding, J.-F., Morabito, M., Myers, J., Williams, C. & Ramirez, F. (1984) *Nature (London)* **310**, 337–340.
29. Briggs, M. R., Kadonaga, J. T., Bell, S. P. & Tjian, R. (1986) *Science* **234**, 47–52.
30. Bornstein, P., Hörlein, D. & McPherson, J. (1984) in *Myelofibrosis and the Biology of Connective Tissue*, eds. Berk, P., Castro-Malaspina, H. & Wasserman, L. R. (Liss, New York), pp. 61–80.
31. Liau, G., Yamada, Y. & de Crombrugghe, B. (1985) *J. Biol. Chem.* **260**, 531–536.
32. Schmidt, A., Yamada, Y. & de Crombrugghe, B. (1984) *J. Biol. Chem.* **259**, 7411–7415.
33. Yamada, Y., Mudryj, M. & de Crombrugghe, B. (1983) *J. Biol. Chem.* **258**, 14914–14919.
34. Liau, G., Mudryj, M. & de Crombrugghe, B. (1985) *J. Biol. Chem.* **260**, 3773–3777.
35. McKeon, C., Schmidt, A. & de Crombrugghe, B. (1984) *J. Biol. Chem.* **259**, 6636–6640.
36. Schlokat, U., Bohmann, D., Schöler, H. & Gruss, P. (1986) *EMBO J.* **5**, 3251–3258.
37. Schlokat, U. & Gruss, P. (1986) in *Oncogenes and Growth Control*, eds. Graf, T. & Kahn, P. (Springer, Berlin), pp. 226–234.
38. Ptashne, M. (1986) *Nature (London)* **322**, 697–701.
39. Rossi, P. & de Crombrugghe, B. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 5590–5594.
40. Horton, W., Miyashita, T., Kohno, K., Hassell, J. R. & Yamada, Y. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8864–8868.