Identification of an Upstream Regulatory Region Essential for Cell Type-specific Transcription of the pro-α2(V) Collagen Gene (COL5A2)*

Sharada Truter, Maurizio Di Liberto, Yutaka Inagaki, and Francesco Ramirez
From the Brookdale Center for Molecular Biology, Mt. Sinai School of Medicine, New York, New York 10029

(Received for publication, April 1, 1992)

The transcriptional features of the human α2(V) collagen gene (COL5A2) were examined by transfection experiments coupled to various DNA binding assays. This approach identified an upstream region essential for the cell type-specific expression of the COL5A2 promoter. Within this region are two nuclear factor-binding sites, FP-A and FP-B, responsible for the formation of distinct DNA-protein complexes. Mutations introduced across each of the two binding sites eliminated the formation of the cognate complex and decreased promoter activity by about 3-fold (FP-A) and 40-fold (FP-B) in transfection experiments. Competition experiments using recognition sequences for known transcription factors exhibiting some similarity to the FP-A- and FP-B-binding sites failed to inhibit COL5A2/protein interactions. Thus, COL5A2 expression appears to be under the positive control of a short regulatory sequence likely to harbor two novel nuclear factor-binding sites.

Fibril-forming collagens represent a structurally related group of molecules within the larger family of collagen proteins. The fibrillar class of collagens includes nine distinct polypeptides that associate into five different trimers (types I, II, III, V, and XI) (1, 2). Based on their relative levels of expression, these collagens are broadly divided into “major” and “minor” types (1, 2).

While the pathophysiology of the major fibrillar collagens (types I, II, and III) is somewhat understood, very little is known about the function and pathogenesis of the minor fibrillar collagens (types V and XI) (1, 2). Some evidence suggests that types V and XI may play a role in fibrillogenesis by regulating the growth and orientation of type I and II collagen fibrils in noncartilaginous and cartilaginous tissues, respectively (3, 4). Accordingly, modulation of the relative proportions of major and minor collagen types may be responsible for different tissue architectures during development and in the adult organism. Conversely, altered production of minor fibrillar collagens may play some role in connective tissue pathology. For example, type V collagen production is elevated in inflammatory and fibrotic conditions, some forms of cancer, atherosclerotic tissues, and in cultured cells treated with transforming growth factor-β (5–10).

Characterization of the cis-acting DNA sequences and trans-acting nuclear factors that modulate correct spatial/temporal patterns of collagen gene expression is an important step toward understanding connective tissue pathophysiology. Several investigations have recently indicated that proper expression of collagen genes is mediated by distinct arrangements of regulatory sequences often, but not always, residing both upstream and downstream of the transcription initiation site (for recent reviews, see Refs. 11 and 12). In our own work, we have previously analyzed some of the regulatory cis-acting elements present in the human type I collagen genes (13, 14). Here we extend these studies to COL5A2, the gene coding for the α2 chain of human type V collagen.

The transcriptional features of the COL5A2 promoter were analyzed by transfecting various chimeric plasmids into normal and transformed cell lines and by performing a variety of DNA binding assays. As a result, we identified a 52-bp long regulatory region that is essential for the cell type-specific expression of COL5A2.

**MATERIALS AND METHODS**

**Cell Culture and Transfection Experiments**—Human cells used in these studies were the fibrosarcoma line HT-1080 (15), the rhabdomyosarcoma line A-204 (16), Jurkat cells (17), and primary fetal fibroblasts (CF-37) prepared as previously described (13). Cells were grown and maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum.

Plasmid DNA was transfected into cells using the calcium phosphate procedure (18) or the electroporation technique (19). Conditions for the DNA preparation, cell transfection, and analysis of plasmid transcriptional activity have been previously detailed (13). Irrespective of the method employed and the cell line used, transfection efficiencies, calculated according to the activity of a standard amount of pSV2CAT (13), were consistently comparable.

**Construction of Chimeric Plasmids**—Isolation and characterization of the genomic clones used for the generation of the plasmids utilized in this study have been described elsewhere (43). Location of the major start site of transcription and numbering of the upstream sequence are as for Greenspan et al. (20).

Various portions of the COL5A2 promoter, sharing the same 3' end point, were inserted into the polylinker sequence of plasmid pBlCAT3 (21) upstream of the chloramphenicol acetyltransferase (CAT) gene. To generate the −2.3COL5A2/CAT construct, a 3.1-kb EcoRI/SacI genomic fragment containing exon 1 was subcloned into pUC18 (20). The DNA of this subclone was then amplified by the polymerase chain reaction (PCR) technique (22) using an oligonucleotide primer corresponding to the 5' sequence of the genomic subclone.

**The abbreviations used are:** COL1A2, the gene coding for pro-α2(I) collagen; COL2A5, the gene coding for pro-α2(V) collagen; AP-1, activator protein-1; CAT, chloramphenicol acetyltransferase; CTF, CCAAT-binding transcription factor; NF-1, nuclear factor-1; PCR, polymerase chain reaction; kb, kilobase(s); bp, base pair(s).

*This work was supported by Grant AR-38648 from the National Institutes of Health. This is article number 98 from the Brookdale Center for Molecular Biology at the Mt. Sinai School of Medicine. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) M98761.
and a primer located immediately upstream to the ATG codon (nucleotide position +160) (20). To maintain the correct orientation of the PCR product in the expression vector, the two primers contained HindIII and BamHI recognition sequences, respectively. The 5' deletions constructs -1.7COL5A2/CAT, -1.1COL5A2/CAT, and -0.7COL5A2/CAT were derived from circularization of the -2.3COL5A2/CAT DNA cleaved at restriction sites with varied spacing and constructs with 5' deletions beginning at nucleotides -296 (-0.3COL5A2/CAT) and -99 (-0.1COL5A2/CAT) were generated by PCR amplifications using the 3.1-kb EcoRI/SacI genomic subclone using oligonucleotides specific for these upstream sequences and the afterfootprinted exon 1 primer. Internal deletions were obtained by linking 5' deletion constructs to appropriate genomic fragments derived by exonuclease III/mung bean nuclease treatment (23) of the -2.3COL5A2/CAT and -0.3 COL5A2/CAT plasmids DNA. Generation of the -0.15COL5A2/CAT construct was achieved by blunt-end subcloning of a PCR product (-150 to -100) into the -0.1COL5A2/CAT plasmid. Relative orientation of the PCR insert was determined by DNA sequencing. Likewise, relative orientation of the -150 to -100 PCR product, subcloned into the unique SacI site of the -777COL2A1 promoter (13), was ascertained by sequencing the relevant area of the resultant plasmid. Mutant forms of the sequence were generated by linker insertion/deletion mutagenesis. The subject of this study were generated by PCR amplification of cloned DNA using as primers mutated and wild-type oligonucleotides (22). The resultant molecules were purified and used for DNA binding assays, as well as for generating mutant plasmids for transfection experiments.

All PCR-derived constructs were fully sequenced and compared to the parental genomic subclones. Both DNA strands were sequenced by the dideoxynucleotide chain termination technique using universal primers or synthetic oligonucleotides (24). In some of the sequencing experiments, progressively overlapping deletions were generated with the exonuclease III/tung bean nuclease method (23). Conditions of PCR amplification were described previously (25).

Nuclear Extracts—Crude nuclear extracts were prepared at 4 °C according to the protocol of Morris et al. (26) with minor modifications. Cells were harvested, washed in phosphate-buffered saline, resuspended in 5 volumes of buffer I (0.32 M sucrose, 0.01 M Tris, pH 8.0, 5 mM MgCl₂, 1 mM EDTA, 1 mM spermidine, 1 mM dithiothreitol, 0.1 mM phenylmethylsulfonyl fluoride), and homogenized with 10-20 strokes in a Dounce homogenizer. Nuclei were centrifuged twice at 1,000 × g for 5 min in the same buffer, resuspended 10 volumes of buffer II (0.1 mM KCl, 20 mM HEPES, pH 7.5, 1 mM dithiothreitol, 0.1 mM EDTA, 0.1 mM phenylmethylsulfonyl fluoride), extracted with 0.4 M (NH₄)₂SO₄ for 30 min, and centrifuged at 35,000 rpm for 60 min in a 70 Ti rotor to pellet the chromatin. The supernatant was collected and the proteins were precipitated by the addition of 0.4 M (NH₄)₂SO₄ to a final concentration of 0.25 g/ml supernatant. Proteins were pelleted by centrifugation at 15,000 rpm for 1 hour in a Ti rotor. Pellets were resuspended in 0.5 volumes of buffer II containing 20% glycerol and dialyzed against the same buffer until the conductivity of the extract reached 0.1-0.15 M KCl. Nuclear extract, cleared by 10 min of centrifugation in a microcentrifuge, was stored in aliquots at -80 °C. The protein concentration of the extract was determined by the method of Bradford (27) using a commercial protein assay kit (Bio-Rad).

Probes for DNA/Protein Binding Assays—Fragments containing COL5A2 promoter sequences used as probes or competitors in DNA/protein binding assays were derived from the -0.3COL5A2/CAT plasmid using the 5'-mer probe (nt 0 to +160), parental DNA was cleaved with HindIII, end-labeled with either Klenow DNA polymerase or T4 polynucleotide kinase, digested with BamHI, and purified by polyacrylamide gel electrophoresis (28). The same treatment was applied for preparing the 73-bp probe (-152 to -79) which was generated by digesting the -0.3COL5A2/CAT construct with MluI and PstI. Oligonucleotides corresponding to the footprinted regions of the COL5A2 promoter extend from nucleotides -91 to -115 (FP-A) and from nucleotides -114 to -150 (FP-B) (17). Their DNA sequences along with those of other oligonucleotides used in the DNA binding experiments are listed in Table I. Oligonucleotides were synthesized by an Applied Biosystems model 380 synthesizer and purified in a denaturing polyacrylamide gel electrophoresis (28). One strand was annealed to the complementary strand and purified from single-stranded polyacrylamide gel electrophoresis in 1× TBE (1× TBE, 0.089 M Tris borate, 2 mM EDTA) (28). Computer search for sequence homology to known nuclear factor–binding sites was carried out using the program MacVector (International Biotechnologies Inc., New Haven, CT).

DNase I Footprinting—DNase I footprinting was performed using a modification (29) of the technique described by Gaiss and Schmitz (30). About 10,000 cpm of end-labeled DNA were incubated at 4 °C for 1 hour with poly( dl- dC) in a 25-µl reaction volume containing 0.1 M Tris, pH 7.5, 0.08 M NaCl, 4% glycerol, 0.001 M β-mercaptoethanol, and 1 mM EDTA. After incubation, samples were brought to 5 mM MgCl₂ and 2.5 mM CaCl₂ and digested with 0.1-1 unit of DNase I (Boehringer Mannheim) for 60 s at room temperature. Reactions were terminated by the addition of 3 µl of 500 mM EDTA, 1 µl of 5 µg/µl tRNA, and 0.1 volume of 3 M sodium acetate, extracted with an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) and ethanol-precipitated. The samples were resuspended in denaturing buffer (10 mM NaOH, 80% formamide, 0.1% xylene cyanol, 0.1% bromphenol blue), heated at 90 °C for 2 min, and electrophoresed through 5 or 8% polyacrylamide-7 M urea sequencing gels (19:1 acrylamide:bisacrylamide ratio) in 1× TBE buffer. Gels were exposed to x-ray film overnight at -80 °C with an intensifying screen.

Cell Type-specific Expression of the COL5A2 Promoter—In order to elucidate the transcriptional features of the COL5A2 promoter, a series of chimeric constructs containing different lengths of human sequence linked to the CAT gene were analyzed after transfection of cultured cells. In these initial functional assays, two different tumor lines were used as recipient cells. The first is the human rhabdomyosarcoma line A-204 which produces significant amounts of type V collagen (16). The second is the human fibrosarcoma line HT-1080 which does not synthesize any of the known fibrillar collagen types (15).

The transcriptional activity of a COL5A2/CAT plasmid containing 2.3 kb of upstream sequence was first compared in the two tumor lines. High levels of CAT expression were obtained with A-204 cells, whereas HT-1080 extracts exhibited only basal CAT activity (Fig. 1A). This suggested that the -2.3COL5A2 promoter contains cis-acting DNA elements capable of sustaining high and cell type-specific gene expression.

Second, the effects of progressive 5' deletions on the transcriptional activity of the COL5A2/CAT plasmid were tested in A-204 and HT-1080 cells (Fig. 1A). Removal of nearly half of the COL5A2 promoter sequence resulted in a 2-fold increase in CAT expression. This enhancement may conceivably result from the elimination of negative cis-acting element(s) located between -2.3 and -1.0 kb. Likewise, the presence of both negative and positive cis-acting sequences may be inferred by comparing the transcriptional activity of plasmids -0.66COL5A2/CAT and -0.3COL5A2/CAT (Fig. 1A). The most dramatic effect on COL5A2-driven CAT expression was,
Regulation of pro-α2(V) Collagen Transcription

A-204 nuclear extracts. This resulted in the protection of at least four distinct regions that extend from about nucleotides −290 to −90 (Fig. 3). In the present study, we focused on the characterization of the two most proximal footprints (A and B in Fig. 3).

To confirm the footprints and delineate more clearly their boundaries, a DNase I protection assay was performed using a shorter probe that spans from nucleotides −152 to −79 (Fig. 2A). Binding of A-204 nuclear proteins to this 73-bp long probe was challenged by molar excess of the same sequence and an unrelated segment of the COL5A2 promoter (Fig. 4). This documented the binding specificity of the footprints which were more accurately mapped from nucleotides −97 to −100, observed with the removal of nucleotides −296 to −100, in that the resultant plasmid −0.1COL5A2/CAT directs only background levels of CAT activity in A-204 cells (Fig. 1A).

Since 5′ deletion of promoter sequence could result in read-through transcription originating from the upstream vector region, CAT activity was also determined using −2.3COL5A2/CAT constructs harboring three different internal deletions between nucleotides −660 and −99 (Fig. 1B). These experiments confirmed the contribution of the −296 to −100 region to the transcriptional activity of the COL5A2 promoter (Fig. 1B). Finally, none of the deletion mutants displayed substantial changes in background CAT activity when assayed in the type V collagen-nonproducing cell line HT-1080 (Fig. 1). From these functional assays we concluded that cis-acting elements capable of directing high and cell type-specific transcription from the COL5A2 promoter are located between nucleotides −296 and −100.

Identification of Nuclear Factor-binding Sites—To identify DNA/protein binding sites within the transcriptionally active −0.3COL5A2 promoter, the 456-bp region extending from nucleotide −296 to immediately 5′ of the ATG codon (+160) (Fig. 2A) was subjected to DNase I footprinting analysis using A-204 nuclear extracts. The noncoding (lanes 1 and 2) and coding strands (lanes 3 and 4) of the 456-bp fragment incubated either without (lanes 1 and 3) or with (lanes 2 and 4) A-204 nuclear extracts. After DNase I digestion, products were analyzed on a 5% polyacrylamide sequencing gel. Solid vertical lines (A through D) represent the regions protected from DNase I digestion. M, radiolabeled φX174 HaeIII-digested DNA with fragment sizes indicated on the left of each autoradiogram. On the right side are the approximate locations of the end points of the footprint-containing region. G, Maxam and Gilbert G + A sequencing reaction (31).

Identification of Nuclear Factor-binding Sites—To identify DNA/protein binding sites within the transcriptionally active −0.3COL5A2 promoter, the 456-bp region extending from nucleotide −296 to immediately 5′ of the ATG codon (+160) (Fig. 2A) was subjected to DNase I footprinting analysis using A-204 nuclear extracts. The noncoding (lanes 1 and 2) and coding strands (lanes 3 and 4) of the 456-bp fragment incubated either without (lanes 1 and 3) or with (lanes 2 and 4) A-204 nuclear extracts. After DNase I digestion, products were analyzed on a 5% polyacrylamide sequencing gel. Solid vertical lines (A through D) represent the regions protected from DNase I digestion. M, radiolabeled φX174 HaeIII-digested DNA with fragment sizes indicated on the left of each autoradiogram. On the right side are the approximate locations of the end points of the footprint-containing region. G, Maxam and Gilbert G + A sequencing reaction (31).
observed occasionally (see lanes containing region with A-204 nuclear extracts). These footprinted regions of the COL5A2 promoter were also protected by HT-1080 nuclear extracts, as well as by nuclear extracts from other cell lines that do not produce type V collagen (data not shown).

The footprinting data were then verified by the gel mobility shift assay using A-204 nuclear extracts incubated with the same 73-bp probe. This gave rise to two clear sets of retarded DNA/protein complexes, termed I and II (Fig. 5A). Moreover, competition with FP-A or FP-B sequences correlated the slower and faster moving complexes with nuclear factors binding to FP-A and FP-B, respectively (Fig. 5A). This conclusion was independently corroborated by an experiment in which the labeled FP-A and FP-B oligonucleotides were used separately as probes in the gel mobility shift assay (Fig. 5, C and D). In each of these two assays, specificity of binding was documented by competition with unlabeled FP-A and FP-B oligonucleotides and an unrelated sequence as well (Fig. 5, C and D).

Finally, a methylation interference experiment was performed for each of the two retarded complexes to determine which bases of the COL5A2 elements participate in binding to nuclear proteins (Fig. 6). Points of contact were determined with clarity only for G residues (Fig. 6). In the FP-A element, four contacts were mapped between nucleotides −105 and −113, whereas two Gs complementary to the C residues of a duplicated ATCA motif were protected by the factor binding to FP-B (Fig. 2B).

**Structural Analysis of the FP-A and FP-B Elements**—

Inspection of the sequences around the protein contact points of the COL5A2 elements revealed some level of homology to the consensus recognition sequences of two distinct families of transcription factors. Between nucleotides −105 and −109 of the FP-A noncoding strand is a DNA motif (GCCAAG) closely resembling a CTF/NF-1 binding site (GCCAAT) (32), while a potential AP-1 binding site (TGAC/GTCA) (32) is located between nucleotides −124 and −118 of the FP-B coding strand (TCAATCA) (Fig. 4).

Several oligonucleotides containing high affinity CTF/NF-1 binding sites (33) and the AP-1 recognition sequence of the human collagenase gene (34) were tested for the ability to compete for the formation of the two DNA/protein complexes (Table I). Since CCAAT proteins have been categorized on the basis of their affinity to different DNA binding sites (33), four distinct CTF/NF-1 oligomers were used in these competition experiments (Table I). Aside from the high affinity CP-1, CP-2, and NF-1 binding sites, a modified version of the adenovirus NF-1 sequence was also used. This sequence, previously shown to compete effectively NF-1 binding to the COL1A2 promoter (10, 35), displays a DNA motif (GCCAAG) identical with that of FP-A (Table I). None of these potential competitor molecules affected binding of the 73-bp probe to A-204 nuclear proteins (Fig. 5B). It should be noted that in this figure only the result of the modified adenovirus NF-1 oligonucleotide is shown as a representative example of all CTF/NF-1 competitors. In conclusion, these data strongly suggest that FP-A and FP-B may represent novel nuclear factor-binding sites.

On the basis of the methylation interference data, five nucleotide substitutions were introduced across the contact points of FP-A and FP-B (Fig. 2B). These mutant sequences were then assayed for their ability to inhibit the interaction between A-204 nuclear extracts and wild-type 73-bp probe. Neither of the two mutant sequences, added alone or in combination, was capable of affecting binding of the factor to the cognate wild-type sequence (Fig. 7). There was also no discernible difference in complex formation when a combi-
Sequences are aligned according to the best fit to the consensus-element motif (underlined). For each element, only the sequence of the relevant DNA strand is shown.

| Name, origin | Sequence |
|--------------|----------|
| FP-A, COL5A2 promoter | 5’-CATCCAGTGTCGGCAAGCAACGG-3’ |
| NF-1, adeno-origin (40) | 5’-TTTTTGGCTGAACCCAATATGAG-3’ |
| NF-1, modified adeno-origin (10) | 5’-TTGGATGAAGCCAGAAG-3’ |
| CP-1, adeno-major late promoter (41) | 5’-CTACACCTTATAADDACATCAACTT-3’ |
| CP-2, fibrinogen promoter (42) | 5’-TGAACCATTCCAGCCACCTTTTA-3’ |
| FP-B, COL5A2 promoter | 5’-CGTGTCTGATTGGTGTGTCTTACATCATGCAAGC-3’ |
| AP-1, collagenase promoter (34) | 5’-TGAACATAGTCAGACACCTTC-3’ |

**Fig. 7. Competition of complex formation with wild-type and mutant sequences.** DNA binding was analyzed by the gel retardation assay using the labeled 73-bp probe. Lane 1, no competitor (0); lane 2, competition with 100-fold molar excess of wild-type FP-A oligonucleotide and FP-B oligonucleotide (Au, Bu); lane 3, wild-type FP-A oligonucleotide (Au); lane 4, wild-type FP-B oligonucleotide (Bu); lane 5, mutant FP-A oligonucleotide (Am); lane 6, mutant FP-B oligonucleotide (Bm). Lanes 7 to 9, competition with 100-fold molar excess of a 54-bp segment (see Fig. 4) containing only wild-type sequence (AuBu, lane 7), only mutant sequences (AmBm, lane 10), and combinations of wild-type FP-A and mutant FP-B sequences (AuBm, lane 8) and vice versa (AuBu, lane 9). Lane I and II designate the FP-A and FP-B complexes, respectively.

**Fig. 8. Transfection experiments with DNA binding probes.** The minimal COL5A2 promoter directs substantial CAT activity when transfected into A-204 but not into HT-1080 cells (Fig. 8). Thus, although nuclear proteins of both A-204 and HT-1080 nuclei bind in vitro to the FP-A and FP-B sequences, only those from the former are capable of forming transcriptionally active complexes. Lastly, the effects of the FP-A and FP-B mutations on the activity of the minimal promoter were assayed functionally by DNA transfection of A-204 cells. This revealed that -0.15COL5A2/CAT plasmids containing the mutated FP-A and FP-B elements display a 3- and 40-fold drop in CAT activity, respectively, compared to the wild-type minimal promoter construct (Fig. 9). Collectively, these results demonstrate that optimal activity of the minimal COL5A2 promoter transiently expressed in A-204 cells depends on the structural integrity of each of these two positively acting upstream elements.

**COL5A2 Regulation in Nontransformed Cells—**Having established the transcriptional features of the minimal COL5A2 promoter in a tumor cell line, we re-examined them in primary fetal fibroblasts (CF-37). To this end, CF-37 and Jurkat cells were transfected with plasmids -0.3COL5A2/CAT, -0.15COL5A2/CAT, and -0.1COL5A2/CAT, as well as with -0.3 COL5A2 promoter construct that harbors an internal deletion of the (FP-A, FP-B)-containing region (Fig. 10). Consistent with the endogenous levels of COL5A2 expression, the transcriptional activity of the chimeric constructs was significantly less in fibroblasts than rhabdomyosarcoma cells (Figs. 1, 8, and 10). There were also some differences in the relative levels of expression of individual constructs transfected into the two cell lines. To be precise, -0.3COL5A2/CAT and -0.1COL5A2/CAT exhibited lower and higher transcriptional activities in fibroblasts compared to A-204 cells, respectively (Figs. 1 and 10). Conceivably, these differences may reflect the intrinsic diversity of the regulatory programs that modulate COL5A2 expression in the two cell lines. Regardless of this point and consistent with the A-204 functional data, elimination of the 52-bp region substantially reduced transcription from the COL5A2 promoter in nontransformed fibroblasts as well (Fig. 10).

Based on these results, the (FP-A, FP-B)-containing probe was subjected to the gel mobility shift assays after incubation with nuclear extracts from CF-37 and Jurkat cells. This resulted in the formation of two complexes seemingly identical to those previously seen with A-204 nuclear proteins and, like them, specifically competed by FP-A and FP-B unlabeled oligonucleotides (Fig. 11A). A difference was, however, noted in the nature of the complexes formed between the FP-B element (complex II) and factors present in nuclear extracts of Jurkat and CF-37 cells (compare lanes 1 and 2 with 4 and 5 in Fig. 11A). The difference was confirmed by additional gel mobility shift assays in which Jurkat and CF-37 nuclear extracts were incubated with labeled FP-A and FP-B oligo-
The major objective of this study was to identify and characterize the shortest DNA sequence capable of directing high and cell type-specific transcription from the human COL5A2 promoter. This goal was pursued by correlating and integrating data obtained from cell transfection experiments with the histograms on the right indicating the percentage of CAT conversions in transfected CF-37 (hatched) and Jurkat (black) cells, relative to pSV2CAT. The autoradiograms at the bottom are representative duplicate CAT assays of the constructs schematized above, with number 5 showing the activity of pSV2CAT. It should be noted that the enzymatic activity of this plasmid was derived from an amount of protein extract 1/10 less than those of the other constructs. The data represent an average of three independent tests ± S.E.

**Fig. 10.** COL5A2 expression in normal fibroblasts. On the top is a schematic representation of the COL5A2/CAT chimeric constructs used in the functional assays, with the histograms on the right indicating the percentage of CAT conversions in transfected CF-37 (hatched) and Jurkat (black) cells, relative to pSV2CAT. The autoradiograms at the bottom are representative duplicate CAT assays of the constructs schematized above, with number 5 showing the activity of pSV2CAT. It should be noted that the enzymatic activity of this plasmid was derived from an amount of protein extract 1/10 less than those of the other constructs. The data represent an average of three independent tests ± S.E.

**Fig. 11.** Gel mobility shift assays of the (FP-A, FP-B)-containing region with CF-37 and Jurkat nuclear extracts. The probes used in each experiment are indicated above each set of autoradiograms. Panel A, competitor molecules are shown above each lane; lanes 1 and 4, incubation without competitor DNA (O); lanes 2 and 5, competition with 100-fold molar excess of the FP-A oligonucleotide; lanes 3 and 6, competition with 100-molar excess of the FP-B oligonucleotide. In both panels, the letters F and J signify nuclear extracts from CF-37 fibroblasts and Jurkat cells, respectively. 1 and II identify the two DNA-protein complexes.

**Discussion**

The major objective of this study was to identify and characterize the shortest DNA sequence capable of directing high and cell type-specific transcription from the human COL5A2 promoter. This goal was pursued by correlating and integrating data obtained from cell transfection experiments and in vitro DNA binding assays. As a result, the 5' boundary of the minimal COL5A2 promoter was located 152 bp upstream of the start site of transcription. Previously described features of this promoter (20, 43) include the absence of a CCAAT motif and the presence of a TATA-like element (TATTATA) located between residues −32 and −27 (Fig. 2A).

Cell type-specific transcription of the minimal promoter is under the positive control of two distinct, nearly overlapping regulatory sequences, termed FP-A (−115 to −99) and FP-B (−149 to −118). The specificity of DNA/protein interactions at these two sites was extensively documented by competition experiments which utilized homologous, unrelated, and mutant sequences. Collectively, these experiments suggest that the two COL5A2 elements are likely to be novel nuclear protein binding sites. Based on the available evidence, we cannot, however, exclude that either or both of the trans-acting factors binding to FP-A and FP-B may represent an alternative form of an already characterized nuclear protein or a structurally related but genetically distinct product.

Preliminary analysis of the ionic requirements for DNA binding indicates that formation of complexes I and II is not inhibited by high concentrations of EDTA nor is it improved by the addition of zinc ions. In contrast, binding of the two complexes differs somewhat from each other with respect to temperature and optimal Mg²⁺ concentration. Complex I dissociates at lower Mg²⁺ concentration and lower temperature than complex II. The distinct behavior of the two complexes under these experimental conditions adds support to the
notion that the FP-A and FP-B sequences are recognized by different transcription factors.

It is of interest to note that the corresponding 52-bp region of the mouse promoter differs from the human sequence for only three nucleotides, none of which is contacted by nuclear factors (43). This phylogenetic conservation of sequence indirectly substantiates the importance of this upstream regulatory segment for COL5A2 expression. Finally, our DNA binding experiments left unresolved whether or not the occupancy of the two COL5A2 sites by transcription factors is mutually exclusive. This is because of the lack of reproducibility and specificity of the third, slower migrating DNA-protein complex seen occasionally in the DNA binding assays.

Additional information about the transcriptional features of the minimal promoter was obtained in functional assays that utilized various manipulations of the 52-bp sequence and different recipient cells. From these experiments, several conclusions can be drawn. First, the formation of transcriptionally active complexes depends greatly on the steric conformation of the 52-bp region, in that this DNA element is substantially more active when inserted into the CAT-plasmid in its natural orientation. Second, transcriptionally active complexes can be formed irrespective of the surrounding sequences, as shown by the behavior of the COLIA2/COL5A2 mosaic promoter. Third, the contribution to transcriptional activity of the FP-B element is significantly greater than FP-A, since a combination of sequences in which only the integrality of the former sequence is preserved still confers substantial activity to the COL5A2 promoter. Lastly and in contrast to type V-nonproducing cells, nuclear proteins from type V-producing cells that bind to the 52-bp region appear to be in transcriptionally active form. The conclusion is obviously based on the assumption that in vitro binding of FP-A and FP-B is to the same proteins in different nuclear extracts. This seems not to be the case at least for the DNA-protein complex that is formed in vitro between the FP-B sequence of COL5A2 and proteins present in Jurkat nuclei (Complex II). Indeed, from the available evidence, one might even argue that, in addition to being the major contributor to COL5A2 transcription in type V-producing cells, FP-B is also capable of binding to negative trans-acting factors that repress COL5A2 expression in type V-nonproducing cells.

Such a redundancy of binding and duality of effects are both well documented phenomena (36). Similarly, there are in the literature numerous examples of putative gene-specific transcription factors that are expressed more broadly than the target genes (36). Relevant to the collagens, a case in point is the recently characterized CCAAT-binding factor, CBF (37). Despite its postulated specificity for coordinating transcription of the two type I collagen genes, the pattern of expression of this heteromeric activator is not confined to tissues of mesenchymal origin (38, 39).

Additional characterization of the trans-acting factors binding to the two cis-acting elements of the minimal COL5A2 promoter will eventually clarify the nature of these regulatory interactions in type V-producing and -nonproducing cells. This will also enable us to detail the interplays between these two regulatory elements as well as between them and additional control elements, such as those mapped further upstream of the 52-bp region.

Acknowledgments—We are indebted to Drs. S. Chen-Kiang, B. de Crombrugghe, G. Karsenty, and R. Raghow for providing us with reagents used in this study. We also thank Drs. L. Pick and G. Sonnenhein for many helpful suggestions and R. Lingeza for typing the manuscript.

REFERENCES

1. Burgesson R. (1988) Annu. Rev. Cell Biol. 4, 551-557
2. Vuorio, E., and de Crombrugghe, B. (1989) Annu. Rev. Biochem. 58, 857-872
3. Birch, E. D., Fisch, J. M., Babiarski, J. P., and Linenmayer, T. (1986) J. Cell Biol. 106, 999-1008
4. Mendler, M., Eich-Bender, S. G., Vaughn, L., Winterhalter, K., and Brockner, P. (1988) J. Cell Biol. 106, 709-719
5. Neumann, R., Engel, S., and PACE, R. C. (1983) Colagen Res. Relat. 3, 323-334
6. Sage, H., Priftal, P., and Bornstein, P. (1981) Biochemistry 20, 3778-3784
7. Barski, S. H., Rao, C. N., Grotendorst, G. R., and Liotta, L. A. (1982) Am. J. Pathol. 108, 276-283
8. Inagaki, Y., Tomokawa, Y., Sakamoto, H., Hirohashi, S., Kobayashi, K., Hottori, N., Ramires, F., Torada, M., and Sugimura, T. Biochem. Biophys. Res. Commun. 148, 829-835
9. McCollough, K. A., and Balint, G. (1975) Nature 258, 73-75
10. Rossi, P., Karsenty, G., Roberts, A. B., Roche, N. S., Sporn, M. B., and de Crombrugghe, B. (1987) Nature 325, 405-414
11. Bornstein, P., and Sage, H. (1989) Prog. Nucleic Acid Res. Mol. Biol. 37, 67-106
12. Rossier, F., and Di Liberto, M. (1990) FASEB J. 4, 1616-1623
13. Bosar, S., Su, M. W., Ramires, F., Sanchez, M., and Avedimento, E. V. (1988) J. Biol. Chem. 263, 13321-13325
14. Kanari, V. M., Chen, Y. Q., Su, M. W., Ramires, F., and Uitto, J. (1990) J. Clin. Invest. 86, 1492-1495
15. Altaiar, K., Vahteri, A., Krieg, T., and Timpl, R. (1984) Eur. J. Biochem. 140, 2126-2122
16. Altaiar, K., Myhlyik, R., Sage, H., Priftal, P., Vahteri, A., and Bornstein, P. (1982) J. Biol. Chem. 257, 9015-9024
17. Gillis, S., and Watson, J. (1980) J. Exp. Med. 152, 1709-1719
18. Ramirez, F., and Van der Eb, A. (1973) Viruslog 62, 456-457
19. Chu, G., Hayakawa, H., and Berg, P. (1987) Nucleic Acids Res. 15, 1311-1327
20. Calespan, D., Lee, S. T., Lee, B. S., and Hoffman, G. G. (1991) Gene Exp. I, 29-39
21. Laskow, B., and Schutz, G. (1987) Nucleic Acids Res. 15, 5430
22. Ehrlich, H. A. (1989) PCR Technology: Principles and Applications for DNA Amplification, Stockton Press, New York
23. Henikoff, S. (1984) Gene (Amst.) 30, 351-359
24. Zagursky, R. J., Herman, M. L., Baumsteiner, K., and Lomax, N. (1986) Gene Anal. Tech. 2, 89-94
25. Lee, B., Godfrey, M., Vitale, E., Horii, H., Mattei, M. G., Sarfaz, F., Tsoupons, R., Ramires, F., and Hollister, D. W. (1991) Nature 352, 330-334
26. Morris, G. F., Price, D. H., and Marzluff, W. F. (1986) Proc. Natl. Acad. Sci. U.S.A. 83, 1-5
27. Bradford, M. M. (1976) Anal. Biochem. 72, 248-254
28. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
29. Di Liberto, M., Le, Z. C., and Childs, G. (1989) Genes Dev. 3, 973-985
30. Galas, D., and Schmitz, A. (1978) Nucleic Acids Res. 5, 3175-3179
31. Mazan, A., and Gilbert, W. (1986) Methods Enzymol. 105, 499-506
32. Mitchell, P. J., and Tjian, R. (1988) Science 245, 371-378
33. Chirico, L. A., Baldwin, A. S., Caruth, R. W., and Sharp, P. A. (1988) Cell 53, 11-24
34. Angel, P., Imagawa, M., Chiu, R., Stein, B., Imam, R. J., Rahmsdorf, H. J., Jun, C., Herrlich, P., and Karrin, M. (1987) Cell 49, 729-738
35. Oikarinen, J., Hatamoto, A., and de Crombrugghe, B. (1987) J. Biol. Chem. 262, 1064-1070
36. Johnson, P. F., and McKnight, S. L. (1989) Annu. Rev. Biochem. 58, 799-834
37. Hatamoto, A., Goliumbek, P. T., Van Schaftingen, E., and de Crombrugghe, B. (1986) J. Biol. Chem. 261, 5940-5947
38. Vuorio, E., Marty, S. N., and de Crombrugghe, B. (1989) J. Biol. Chem. 264, 25489-25496
39. Matyi, S. N., Vuorio, E., and de Crombrugghe, B. (1990) Proc. Natl. Acad. Sci. U.S.A. 87, 5376-5380
40. Rosenfeld, P. J., O'Neill, P. A., Wides, R. J., and Kelly, T. J. (1987) J. Cell Biol. 7, 875-886
41. Carthew, R. W., Chodosh, L. A., and Sharp, P. A. (1985) Cell 43, 439-448
42. Morgan, J. G., Courtis, G., Foul, G., Chodosh, L. A., Campbell, L., Evans, E., and Crabtree, G. R. (1986) Mol. Cell. Biol. 6, 1455-1460
43. Truter, S., Andrikopoulos, K. D., Di Liberto, M., Wonsack, L., and Ramirez, F. (1992) Connect. Tissue Res. in press