Cloning and Characterization of a Transcription Factor That Binds to the Proximal Promoters of the Two Mouse Type I Collagen Genes*

Tadao Hasegawa‡‡, Akihide Takeuchi‡, Osamu Miyaishi‡, Ken-ichi Isobe‡, and Benoit de Crombrugghe‡‡

From the ‡Department of Molecular Genetics, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030 and the ‡Department of Basic Gerontology, National Institute for Longevity Sciences, 36-3, Gengo Morioka-cho, Obu, Aichi 474, Japan

We have used the yeast one-hybrid system to clone transcription factors that bind to specific sequences in the proximal promoters of the type I collagen genes. We utilized as bait the sequence between −180 and −136 in the pro-α2(I) collagen promoter because it acts as a functional promoter element and binds several DNA-binding proteins. Three cDNA clones were isolated that encoded portions of the mouse SPR2 transcription factor, whereas a fourth cDNA contained a potential open reading frame for a polypeptide of 775 amino acids and was designated BFCOL1. Recombinant BFCOL1 was shown to bind to the −180 to −152 segment of the mouse pro-α2(I) collagen proximal promoter and to two discrete sites in the proximal promoter of the mouse pro-α1(I) gene. The N-terminal portion of BFCOL1 contains its DNA-binding domain. DNA transfection experiments using fusion polypeptides with the yeast GAL4 DNA-binding segment indicated that the C-terminal part of BFCOL1 contained a potential transcriptional activation domain. We speculate that BFCOL1 participates in the transcriptional control of the two type I collagen genes.

Type I collagen is a protein that is abundantly synthesized by a discrete number of cell types including osteoblasts, odontoblasts, fibroblasts, smooth muscle cells, and mesenchymal cells. It is composed of two α1 chains and one α2 chain forming a characteristic triple helix. Expression of the genes for these polypeptides is coordinately regulated in a variety of physiological and pathological situations (1). Changes in the synthesis of type I collagen occur not only during embryonic development in specific tissues but changes also take place in disease states, for example during wound healing as well as in fibrotic diseases such as lung fibrosis, cirrhosis, and scleroderma. In many of these instances it is likely that the control of expression of the two type I collagen genes is mainly exerted at the level of transcription, but the precise mechanisms that control transcription of these genes are still poorly understood. Our long term goal is to identify the critical cis-acting elements in these two genes and both the cell-specific and ubiquitous transcription factors that presumably control their expression.

Recently, transgenic mouse studies have identified strong tissue-specific enhancer elements in the 5′-flanking regions of both type I collagen genes (2–5). These elements are located further upstream than the proximal promoter elements. For instance, in the mouse pro-α1(I) gene, a potent enhancer element for osteoblast and odontoblast expression was localized about 1.6 kilobases (kb) upstream of the start of transcription, whereas another strong element for expression in tendon and fascia fibroblasts was found between −2.3 and −3.2 kb (2). Similar experiments from other laboratories have produced analogous results (3, 4). These experiments strongly suggested that separate elements control the expression of this gene in different type I collagen-producing cells. In the pro-α2(I) gene, an element that strongly enhanced expression in fibroblasts and mesenchymal cells was located between 13.5 and 17.5 kb upstream of the transcription start (5). One can speculate that proteins binding to the upstream enhancers in both type I collagen genes cooperate with transcription factors binding to the proximal promoters to activate transcription in specific cell types.

Previous studies have identified several functional cis-acting elements in the 350-bp (base pair) proximal promoter of the mouse pro-α2(I) collagen gene (6). These included a binding site for the ubiquitous heterotrimeric CCAAT-binding factor (CBF), between −75 and −98 (7–9), redundant GC-rich binding sites for several proteins between −65 and −105, between −114 and −131, and between −152 and −176 (10). Several classes of proteins that are mainly ubiquitous proteins bind to these redundant sites. Transient expression and in vitro transcription experiments with wild-type and mutant templates indicated that the segment between −40 and −170 containing the three redundant elements was essential for promoter activation. Other studies identified three short cis-acting GC-rich elements in the human pro-α2(I) collagen gene between −323 and −264 (11) that were capable of binding SP1. Additional studies presented evidence that a protein complex which includes SP1 binds to this segment of the human promoter and participates in the transforming growth factor-β activation of this promoter (12). In the mouse promoter there is also a binding site for CTF/NF1 between −305 and −290 (13).

In the mouse pro-α1(I) collagen gene, the sequence between −220 and the TATA box presents strong homologies with the sequence of the pro-α2(I) gene in the same region. This DNA

* The abbreviations used are: kb, kilobase; bp, base pair; CBF, CCAAT-binding factor; GST, glutathione S-transferase; CAT, chloramphenicol acetyltransferase; IF-1, inhibitory factor 1; RT-PCR, reverse transcriptase-polymerase chain reaction.

‡ This work was supported by National Institutes of Health Grants HL41264 (to B. d. C.), CA16672 (for support of DNA sequences), and the Japan Foundation for Aging and Health (to T. H.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: Dept. of Molecular Genetics, Box 11, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030. Tel.: 713-792-2590; Fax: 713-794-4295.

\[\text{Received for publication, June 11, 1996, and in revised form, November 21, 1996}\]
Cloning of Protein Binding to the Pro-α2(I) Collagen Promoter

**Fig. 1.** Approximate location of known binding sites for DNA-binding proteins in the proximal promoters of the mouse pro-α1(I) and pro-α2(I) collagen genes. The cKrox binding sites in these promoters are from Refs. 31 and 32. Binding sites for DNA-binding proteins in the two corresponding human proximal promoters are discussed in the text.

| α1(I) | CBF | SP1 | c-Krox |
|-------|-----|-----|--------|
|       | +   | +   |        |

| α2(I) | CBF | SP1 | Proteins Binding to Consensus SP1 Binding Sites | Proteins Binding to Consensus Krox Binding Sites |
|-------|-----|-----|-----------------------------------------------|-----------------------------------------------|
|       |     |     | +                                               | +                                               |

Segment contains binding sites for DNA-binding factors that also bind to the proximal pro-α2(I) promoter. The DNA elements in the pro-α1(I) promoter include a binding site for CBF between −90 and −115, two apparently redundant sites between −190 and −170 and between −160 and −130 for a DNA-binding protein previously designated inhibitory factor 1 (IF-1), and two sites between −130 and −80 that flank the CBF-binding site and are binding sites for SP1 and probably other GC-rich binding proteins (14). DNA transfection experiments with the pro-α1(I) promoter showed that point mutations in the CBF-binding site decreased promoter activity, whereas small substitution mutations in some of the other sites resulted in an increase in transcription (14, 15). It was also shown that the sequence of the pro-α2(I) promoter between −173 and −143 was able to compete for the binding of a protein that was forming a major DNA-protein complex with two redundant elements in the pro-α1(I) promoter between −190 and −170 and between −160 and −130, suggesting that both type I promoters contained binding sites for the same proteins (15).

The purpose of the present study was to identify one or more trans-acting factors that bind to these proximal cis-acting elements in the type I collagen promoters. We have used the sequence between −180 to −136 of the pro-α2(I) collagen promoter to clone cDNAs for proteins binding to this segment using the yeast one-hybrid system (16, 17). This segment was chosen mainly because previous experiments with the pro-α2(I) proximal promoter showed that the DNA segment between −180 and −136 was capable of binding an array of DNA-binding proteins, many of which also bound to the same promoter between −133 and −105 and between −105 and −65; the sequence between −180 and −136 was also binding these proteins with greater efficiency than the more proximal sequences (10). One of the cDNAs that was cloned encodes a polypeptide of 775 amino acids, which also bound to two discrete sites in the pro-α1(I) collagen promoter.

**Materials and Methods**

**Cloning of DNA-binding Proteins Using the Yeast One-hybrid System**—The yeast strain BY 164 (MAT a his3 8 200 leu2-3, 112 ura3-52 lys2-801a trp1a) was provided by Dr. Stevan Marcus. The yeast reporter plasmid was constructed as follows. Six tandem copies of a double-stranded oligonucleotide corresponding to the sequence from −180 to −136 bp of the mouse pro-α2(I) collagen promoter were inserted into the BamHI site of the vector pRS315HIS containing the LEU2 gene as selectable marker (16, 18, 19) to generate pRS315HIS-6x160 (160 denotes the sequence between −180 and −136). The XhoI-SalI fragment of pRS315HIS-6x160 was then subcloned into the XhoI-SalI site of the vector pRS305 (16); this plasmid was designated pRS305HIS-6x160. After digestion with ClaI, this vector was used for transformation. Yeast transformation was performed by the polyethylene glycol/lithium acetate method (20). Plasmid integration in the genome of yeast strains was confirmed by Southern blot analysis using a 32P-labeled oligonucleotide from −180 to −136 bp. Cells were then plated on a minimal synthetic dextrose plate without histidine to verify background HIS3 gene activity. One of the yeast strains that had minimal HIS3 gene activity was also selected as the strain for the transformation after the initial selection. Plasmid pJL638−6x160 contained six tandem copies of the sequence from −180 to −136 of the mouse pro-α2(I) promoter in the PBlg-iacZ plasmid harboring the URA3 gene as a selectable marker (17). The yeast strain in which both pRS305HIS-6x160 and pJL638−6x160 plasmids were integrated was used for cDNA library transformation.

**DNA Sequencing**—DNA sequencing was carried out using a primer present in the DNA for the GAL4 transactivation domain (5′-GGAT- GTTTAATACCACT-3′) or T3, T7, and SP6 primers.

**Expression of Cloned cDNAs by In Vitro Transcription and Translation**—Three different recombinant polypeptides corresponding to the full-length, the N-terminal part, and the C-terminal part of BFCOL1 were generated using the Tn7-coupled reticulocyte lysate system (Promega Corp.). For the full-length polypeptide, the SalI-NotI fragment of pC86-BFCOL1 was inserted into the SalI-NotI site of the pBluescript KS vector (pBS-BFCOL1-full). For the N-terminal and C-terminal polypeptides, the SalI-NotI fragment and the SalI-NotI fragment of pC86-BFCOL1 DNA were inserted into the SalI-PstI site of the pGem Zf (+) and the EcoRV-NotI site of the pBluescript KS to generate pZf-BFCOL1-N and pBS-BFCOL1-C, respectively. 35S-Labeled polypeptides were analyzed using 10% sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis. These were exposed to Fuji RX film.

**Generation of Fusion Polypeptides with Glutathione S-Transferase (GST)**—Three different fusion polypeptides were generated. For the full-length and N-terminal fusion polypeptides, the SalI-NotI fragment of pC86-BFCOL1 and the SalI-NotI fragment of pZf-BFCOL1-N were
Cloning of Protein Binding to the Pro-α(I) Collagen Promoter

RESULTS

Cloning of cDNAs for Polypeptides That Bind to the −180 to −136-bp Segment of the Mouse Pro-α(I) Collagen Promoter—Previous studies indicated that the GC-rich DNA segment between −180 and −136 of the mouse pro-α(I) collagen gene was able to bind several different DNA-binding proteins in vitro and that this segment was also able to compete for the binding of proteins to two other redundant but discrete sites closer to the transcription start site of this promoter (10). A deletion of this same segment resulted in substantial decrease in promoter activity. To begin to identify some of the proteins that bound to this segment, we used this DNA as bait in the yeast one-hybrid system and screened two mouse embryo fibroblast cDNA libraries, one primed with oligo(dT) (library 1) and the other primed with a random hexamer oligonucleotide (library 2). In the yeast strain that was used for selection, plasmid pRS305-HIS-αx160 was integrated into the genome. In this plasmid, six tandem copies of the sequence of the pro-α(I) gene between −180 and −136 were cloned upstream of a minimal yeast GAL1 promoter itself linked to the HIS3 gene. After screening six million independent colonies from library 1 and three million from library 2, an initial 81 histidine-positive colonies from library 1 and 44 histidine-positive colonies from library 2 were picked; 17 cDNA plasmids from library 1 and 12 cDNA plasmids from library 2 gave positive colonies upon retransformation of the parental strain. However, most of these also gave histidine-positive colonies with the yeast strain in which the control plasmid pRS305-HIS was integrated, a plasmid that contains the minimal GAL1 promoter but not six tandem copies of the −180 to −136 sequence. Only four cDNAs, all from library 1, could specifically activate the HIS3 gene of the yeast strain containing the pRS305-HIS-αx160 plasmid without activating the HIS3 gene of the strain with the pRS305-HIS control plasmid. This suggested the possibility that the recombinant fusion polypeptides encoded by each of these four cDNAs might bind specifically to the −180 to −136-bp segment of the pro-α(I) collagen gene and not to the DNA sequence of the GAL1 minimal promoter. One cDNA clone was designated as BFCOL1 (see below). Two other cDNAs contained an almost full-length coding sequence for mouse SP-2 (28), and the fourth cDNA was a shorter partial cDNA for SP-2.

Fig. 4A presents the sequence of BFCOL1 cDNA. The open reading frame starting from the first methionine codon in the cDNA is 2328 nucleotides long and encodes a putative polypeptide of 775 amino acids. The amino acid sequence corresponding to the N-terminal 400 amino acids of BFCOL1 presents a 95% identity with human hTβ, a protein previously identified as binding to the promoter of the gene for the Vβ8.1 chain of the...
The amino acid sequence corresponding to the four zinc finger motifs are underlined. The sites of the two primers used for RT-PCR (12n and 12c) are shown in bold and are underlined. Numbers correspond to the DNA sequence. B, partial restriction map of BFCOL1. The location of the four zinc finger domain, the primers used for RT-PCR, and the probe used for Northern hybridization (EB) are also indicated.

Partial nucleotide and deduced amino acid sequences of mouse BFCOL1, the human homologue of BFCOL1, its human homologue, and htbβ, are also shown. Dashes represent nucleotides that are identical in BFCOL1, its human homologue, and htbβ. Sal, Sali; E, EcoRI; Sma, SmaI; Sca, ScaI; Ns, NsiI; No, NotI.

human T cell receptor (29). Subsequent to the codon for amino acid 400 in BFCOL1, the reported nucleotide sequence of htbβ DNA displays two translational frame changes compared with that of BFCOL1 immediately followed by a termination codon (29). The sequence of the 340 amino acids at the C terminus of BFCOL1 has no significant amino acid sequence homology with
other polypeptides present in Genebank data bases. The nucleotide sequence preceding the initial methionine codon of BFCOL1 is also similar to that of htβ except that the first 25 nucleotide residues at the 5’ end of the cDNA of BFCOL1 are different from comparable residues in htβ. As reported for htβ, the deduced amino acid sequence of BFCOL1 contains four potential zinc finger motifs.

Since the reported nucleotide sequence of the cDNA of htβ following the termination codon is also about 90% identical to that of BFCOL1 DNA, we performed RT-PCR experiments in order to verify the nucleotide sequence of the human homologue of BFCOL1 RNA. We used two primers that bracketed the sequence containing the reported frame-shifts and stop codon in htβ and RNAs from two different human cell lines, the colon carcinoma cell line HT29 and the stomach cell line HSC34. The location of the primers that were used (12n, 12c) are indicated in Fig. 3A. The sequence of the PCR product from the RNAs of these human cells is presented in Fig. 3B and shows a continuous open reading frame without the frame-shifts that were reported earlier.

We then asked whether the entire open reading frame shown in Fig. 3A was translated into a polypeptide of the expected size, and we constructed three plasmids for in vitro transcription-translation. One plasmid encoded the full-length BFCOL1 (pBS-BFCOL1-full, from 1 to 2426), whereas the others encoded the N-terminal part (p5Zf-BFCOL1-N, from 1 to 1156) and the C-terminal segment (pBS-BFCOL1-C, from 1395 to 2426) of BFCOL1. The major product of pBS-BFCOL1-full (Fig. 4, lane 1) was a single polypeptide, whereas the DNA of p5Zf-BFCOL1-N (lane 2) gave rise to two major polypeptides and several fainter species. The major polypeptide species appeared to run more slowly by SDS-polyacrylamide gel electrophoresis than expected from their estimated molecular sizes (expected sizes are 89 kDa for pBS-BFCOL1-full and 43 kDa for p5Zf-BFCOL1-N) maybe due to increased SDS binding. The major product of pBS-BFCOL1-C (Fig. 4, lane 3) was a single polypeptide that had about the expected size (estimated size is 37 kDa).

**DNA-binding Experiments**—Gel shift experiments were performed with the −180 to −136 DNA segment of the mouse pro-α2(I) collagen promoter that was used in the one-hybrid screen to verify whether BFCOL1 was able to form a DNA-protein complex under the conditions of this assay and to determine which segment of the BFCOL1 polypeptide contained a DNA-binding domain. In these experiments, the products of in vitro transcription-translation shown in Fig. 4 were tested. With the full-length BFCOL1, one major protein-DNA complex was detected (Fig. 5, lane 4), whereas two protein-DNA complexes were seen with the N-terminal BFCOL1 (Fig. 5, lane 3), the upper complex being less intense than the lower complex, although both complexes were migrating faster than the complex with full-length BFCOL1. The presence of these two complexes could possibly be a result of the heterogeneity of protein products observed with the cDNA for N-terminal BF-COL1 (see Fig. 4, lane 2). With the polypeptide corresponding to the C-terminal part of BFCOL1, no protein-DNA complexes were detected other than nonspecific bands (Fig. 5, lane 2). These results suggested that the N-terminal part of BFCOL1 contained a DNA-binding domain and that DNA binding might be mediated by the four tandem zinc fingers, consistent with previous results obtained with htβ (29).

The proximal promoter of the mouse pro-α2(I) collagen gene contains several redundant GC-rich elements. To examine whether other segments of the 350-bp proximal promoter of this gene contained binding sites for BFCOL1, in vitro DNase I footprints were performed using a recombinant GST-full-length BFCOL1 fusion polypeptide and the promoter segment from −350 to +7. As shown in Fig. 6, the recombinant GST-full-length BFCOL1 fusion polypeptide protected the region between −180 and −152 (lane 2) and only this segment. When a similar DNase I footprint was performed using the promoter labeled on the other strand (lane 4), again no other protected regions were observed. Hence, recombinant BFCOL1 binds only to one specific sequence in this proximal promoter and not to other GC-rich sequences.

To further confirm the specific binding of BFCOL1 to a discrete site in the proximal promoter of the pro-α2(I) collagen gene, gel shift experiments were performed using a 32P-labeled −180 to −136 oligonucleotide as probe and several competitor DNA oligonucleotides corresponding to specific sequences present in the mouse pro-α2(I) collagen proximal promoter (Fig. 7). In this experiment, the product of full-length BFCOL1 DNA was used, generated by in vitro transcription-translation. As expected, the −180 to −136 oligonucleotide (lane 2) competed for binding, as did, with somewhat less efficiency, the shorter −176 to −152 oligonucleotide (lane 3), which is included in the
Cloning of Protein Binding to the Pro-α2(I) Collagen Promoter

Fig. 6. DNase I footprint of the mouse pro-α2(I) collagen proximal promoter. The BamHI-NarI fragment of plasmid pLAG23, which contains the −350 to +7-bp fragment of the mouse pro-α2(I) collagen promoter, was labeled at the NarI 5′-end (A) and BamHI 5′-end (B), incubated without or with recombinant GST-full-length BFCOL1 fusion protein, and treated with DNase I. Lanes 1 and 3, DNase I digestion pattern of the DNA incubated with GST protein. Lanes 2 and 4, DNase I digestion pattern of DNA incubated with the GST-BFCOL1 full-fusion polypeptide protein. Protected regions are marked by the white vertical boxes. The numbers on the right correspond to base pairs upstream of the start of transcription based on a G + A Maxam and Gilbert reaction (not shown).

Fig. 7. DNA-binding assays. Double-stranded 32P-labeled oligonucleotides containing the sequences between −180 and −136 of the mouse pro-α2(I) collagen promoter were incubated with the products of in vitro transcription and translation without or with various unlabeled double-stranded oligonucleotide competitors. Competitor oligonucleotides are indicated at the top of each lane and were used in 100-fold excess.

FIG. 7. DNA-binding assays. Double-stranded 32P-labeled oligonucleotides containing the sequences between −180 and −136 of the mouse pro-α2(I) collagen promoter were incubated with the products of in vitro transcription and translation without or with various unlabeled double-stranded oligonucleotide competitors. Competitor oligonucleotides are indicated at the top of each lane and were used in 100-fold excess.

The results of previous gel shift experiments using crude nuclear extracts of NIH/3T3 fibroblasts were consistent with the hypothesis that a DNA-binding protein was binding to discrete sites in both the proximal pro-α1(I) and pro-α2(I) collagen promoters (14, 15). This protein was tentatively designated inhibitory factor 1 (IF-1) based on the result that substitution mutations in its binding sites in each promoter, which inhibited its binding, resulted in an increase in promoter activity. The binding site of IF-1 in the pro-α2(I) promoter corresponded to the binding site for BFCOL1, whereas the binding sites in the pro-α1(I) promoter were located between −190 and −170 and between −160 and −130. To test whether these sites in the pro-α1(I) promoter could also bind BFCOL1, gel shift assays were performed using two labeled oligonucleotides from −194 to −168 and from −168 to −129 in the mouse pro-α1(I) collagen promoter in conjunction with the GST-full-length BFCOL1 fusion polypeptide. When this GST-fusion polypeptide was used with an oligonucleotide corresponding to the sequence between −180 to −136 in the pro-α2(I) promoter, two complexes were observed, a slower migrating complex and a more intense faster migrating complex (Fig. 8, lane 1). The difference in the pattern of DNA-protein complexes between those observed with GST-BFCOL1 fusion polypeptides synthesized in E. coli and those seen with BFCOL1 synthesized in the reticulocyte lysate (see Fig. 5) could possibly be due to the heterogeneity of the GST-BFCOL1 fusion polypeptides as examined by SDS-polyacrylamide gel electrophoresis (data not shown). Fig. 8 shows that the recombinant GST-BFCOL1 full-length fusion polypeptide was also able to bind to the −168 to −129 pro-α1(I) oligonucleotide and with much less efficiency to the −194 to −168 pro-α1(I) oligonucleotide DNAs (Fig. 8, lanes 9, and 17).

Mutations were then introduced into these DNA segments, i.e. 5′ CGCGCCCCCCC 3′ → 5′ CGCGCTTTTCCC 3′ in the −194 to −168 sequence of the pro-α1(I) (sequence represents lower strand) and 5′ CCTCCCGCCCTC 3′ → 5′ GGTCCGCCCTC 3′ in both the −168 to −129 sequence of pro-α1(I) and the −180 to −136 sequence of the pro-α2(I) promoter, and the mutant oligonucleotides tested in DNA-binding assays with recombinant BFCOL1. Lanes 8, 16, and 24 of Fig. 8 show that each of these mutations abolished the binding of the recombinant GST-BFCOL1 full-length polypeptide. We also performed competition experiments using the wild-type oligonucleotides as probes. With each of the three wild-type-labeled oligonucleotide probes, the wild-type −180 to −136 sequence of pro-α2(I) (Fig. 8, lanes 2, 10, and 18) and the wild-type −168 to −129 sequence of pro-α1(I) (Fig. 8, lanes 4, 12, and 20) acted as strong competitors. In contrast, the mutant −180 to −136 sequence of pro-α2(I) (Fig. 8, lanes 3, 11, and 19) and the mutant −168 to −129 sequence of pro-α1(I) (Fig. 8, lanes 5, 13, and 21) were unable to compete. The wild-type −194 to −168 oligonucleotide of the pro-α1(I) collagen promoter had little effect as competitor
when the other two oligonucleotides were used as probes (Fig. 8, lanes 6 and 14), confirming the notion that this sequence binds BFCOL1 much less efficiently than the other two sites. A mutant oligonucleotide corresponding to the −194 to −168 sequence of the pro-α1(I) promoter had no effect as competitor with all oligonucleotide probes (Fig. 8, lanes 7, 15, and 23). Hence, BFCOL1 binds to two sites in the pro-α1(I) proximal promoter with different efficiencies and to one site in the pro-α2(I) collagen proximal promoter. The locations of these binding sites are the same as those previously identified as binding to IF-1. The same substitution mutations that inhibited IF-1 binding in crude nuclear extracts (14, 15) also inhibited BFCOL1 binding.

Functional Analysis—To test whether BFCOL1 could either activate or inhibit transcription, DNAs for the “full-length” and N-terminal segment of BFCOL1 were cloned in a mammalian expression vector, and these DNAs were cotransfected with a pro-α2(I) collagen promoter (−350 to +54) linked to the luciferase reporter gene in BALB 3T3 fibroblasts. A plasmid containing DNA for the C-terminal part of BFCOL1, lacking the DNA-binding domain and driven by the same mammalian expression promoter, served as control. No activation occurred with any of the three BFCOL1 constructions (data not shown). At higher concentrations the full-length BFCOL1 and the C-terminal BFCOL1 segment caused inhibition of the 350-bp pro-α2(I) collagen promoter, presumably as a result of squelching, i.e. titration of another transcription factor that is important for expression of this promoter (30). Very similar results were observed when the promoter contained six tandem repeats of the −180 to −136 pro-α2(I) sequence cloned upstream of a minimal pro-α2(I) promoter (−40 to +54). Again at higher concentrations of BFCOL1, inhibition occurred, but this took place even when the reporter plasmid contained mutations that abolished binding of BFCOL1, strongly suggesting that the inhibition was not dependent on binding of BFCOL1 to DNA and hence presumably due to squelching. Similar results were also obtained after cotransfection of the BFCOL1 plasmids and the reporter plasmids in S194 B cells (data not shown).

To determine whether segments of BFCOL1 contained a potential transactivation domain, three mammalian expression plasmids were constructed coding for fusion polypeptides with the yeast GAL4 DNA-binding domain. The DNAs for full-length, N-terminal, and C-terminal fusion polypeptides were cotransfected along with a plasmid containing a GAL4-binding site upstream of an SV40 promoter itself linked to the CAT gene. Activation of the reporter gene occurred only with the plasmid coding for the GAL4-BFCOL1 C-terminal fusion polypeptide. No transcriptional activation was detected with either the GAL4-BFCOL1 full-length or the GAL4-BFCOL1 N-terminal fusion polypeptides (Fig. 9). This experiment suggested that the C-terminal segment of BFCOL1 contained a potential transcription activation domain. This does not exclude the possibility that the other segments of BFCOL1 might contain additional activation domains. Indeed, the presence of the BFCOL1 DNA-binding domain in the two other fusion polypeptides might eventually interfere with binding to the GAL4-binding site in the promoter of the reporter plasmid.

Northern Blot Analysis—To determine the size of the BFCOL1 RNA transcripts, a Northern hybridization experiment was performed (Fig. 10). Three RNA transcripts were identified as follows: one transcript had a size of about 9 kb, another of about 5.5 kb, and a third RNA, which migrated more slowly than the 9-kb species, was seen in S194 B cells. These RNAs are all larger than the size of our cDNA. This pattern of RNAs is analogous to that previously shown to hybridize to the hTβ DNA probe although the shorter RNA was shown to have a mobility of 4 to 4.2 kb in humans. It is possible that in the human RNA either the 3′-untranslated segment or the 5′-untranslated segment or both are shorter than in the mouse RNA (29).

In our experiments, the two major species of 9 and 5.5 kb were seen in two fibroblast cell lines, a B cell line and a T cell line.
DISCUSSION

We have used the yeast one-hybrid system to clone a cDNA for a protein, designated BFCOL1, that binds to the segment between −180 and −136 in the promoter of the mouse pro-α2(I) collagen gene. This DNA-binding protein, which appears to be a ubiquitous protein, does not bind to other sites within the proximal 350 bp of this promoter but binds to two discrete sites in the proximal promoter of the mouse pro-α1(I) collagen gene. The cDNA of BFCOL1 contains an open reading frame for a polypeptide of 775 amino acids. The sequence of the N-terminal 400 amino acids of BFCOL1 shows a 95% identity with that of hTβ, a human polypeptide that binds to the promoter of a β-subunit of the human T cell receptor gene (29). The reported open reading frame of hTβ codes for only 454 amino acids due to two translational frameshifts and a termination codon in the sequence compared with that of BFCOL1. In contrast, our RT-PCR experiments indicated a continuous open reading frame in this segment of the human RNA encoding an amino acid sequence essentially identical to that specified by the mouse BFCOL1 RNA. It is, therefore, almost certain that BFCOL1 and hTβ correspond to one and the same gene. The C-terminal amino acid segment of BFCOL1 corresponds to a new unique sequence containing at least one serine-threonine-rich segment and is overall much more hydrophilic than the N-terminal half. Our experiments using fusion polypeptides with the DNA-binding domain of GAL4 indicate that the C-terminal part of BFCOL1 has the potential to serve as a transcriptional activator domain. Preliminary experiments with the yeast two-hybrid system also indicated that the C-terminal part of BFCOL1 could interact with the TATA-binding protein-associated factor TAF110 (not shown). The N-terminal portion of BFCOL1 includes its DNA-binding domain, and its sequence contains four potential zinc fingers of the class Cys2-His2.

The sequence in the two type I collagen genes to which BFCOL1 binds do not contain the CACCC box sequence to which hTβ was proposed to bind (29). The sequence between −176 and −152 in the pro-α2(I) collagen promoter and the sequence between −168 and −129 in the pro-α1(I) collagen promoter contain an 11-bp identical sequence (5′ CACCCTCCCTCTC 3′) that must be part of the binding site as mutations within this sequence abolish binding of BFCOL1. The other binding site in the pro-α1(I) promoter between −194 and −168 is a much weaker binding site, and although it is pyrimidine-rich, it does not contain the same 11-bp sequence. The same mutations in each of the two identical 11-bp conserved sequences abolish binding of BFCOL1 to larger pro-α2(I) and pro-α1(I) oligonucleotides. In earlier experiments, these same mutations were shown to increase the activity of each of the two promoters 3- to 4-fold in transient DNA-transfection experiments (14, 15). Mutations in the −194 to −168-binding site of the pro-α1(I) promoter also caused an increase in promoter activity, but this increase was clearly much smaller than with mutations in the 11-bp identical sequence of the pro-α1(I) and pro-α2(I) promoters. Hence, there is a correlation between the efficiency of binding of BFCOL1 to its binding sites in the two type I collagen promoters and the previously reported effects of mutations in these sites on increase in promoter activity.

Another recently identified DNA-binding protein, cKrox (31), was also shown to bind to the same sequences in the pro-α1(I) promoter as those to which BFCOL1 is binding. However, cKrox appeared to bind more efficiently to the −194 to −168 sequence than to the −168 to −129 sequence.

Recent experiments from our laboratory have shown that several classes of proteins could bind to the segment of the pro-α2(I) collagen promoter between −180 and −136 (10). These proteins included SP1, proteins different from SP1 that bind to an SP1 consensus binding site, proteins that bind to a Krox consensus binding site, and probably others. Many of these proteins could also bind to two discrete sites that are closer to the start of transcription in this promoter. A deletion of the −180 to −136 segment decreased promoter activity about 2-3-fold. If the three redundant sites are deleted together, no promoter activation occurred above that of a minimal promoter containing a TATA box and a transcription start site (−40 to +54). Hence, the overall activity of the −180 to −136 element in the pro-α2(I) promoter is clearly that of an
enhancer, implying that at least some of the several proteins that bind to this element should act as transcription activators. We speculate that as with nuclear extracts in vitro (10), several proteins in vivo also compete for binding to this element. In experiments using fusion polypeptides with the yeast GAL4 DNA-binding domain, we showed that the C-terminal part of BFCOL1 had the potential of being a transcriptional activator domain; the degree of activation was, however, weaker than with the activation domain of two other DNA-binding proteins that were similarly tested as GAL4 fusion polypeptides.2 One possibility for the increase in promoter activity which took place with mutations that abolished the binding of BFCOL1 (14, 15) is that the binding of other transcription factors that bind to the same area of the promoter to binding sites which would overlap partly with that of BFCOL1 would now occur more efficiently. If these transcription factors were more potent activators than BFCOL1, then the net result would be an increase in promoter activity.

In brief, BFCOL1 appears to be one of several ubiquitous proteins that bind to discrete sites in the proximal promoters of the two type I collagen genes and probably control the activity of these promoters in conjunction with other ubiquitous DNA-binding proteins, as well as tissue-specific transcription factors.

It was interesting that in addition to BFCOL1 the yeast one-hybrid system identified SPR2, a DNA-binding protein related to SP1. Our earlier experiments had suggested that the −180- to −136-bp segment of the pro-α2(I) gene was capable of binding SP1 and other proteins different from SP1 that bind to a consensus SP1-binding site (10). SPR2 could be one of these proteins.

Acknowledgments—We are indebted to Xin Zhou for many useful suggestions with the yeast one-hybrid system. We thank Patricia McCauley for editorial assistance.

REFERENCES
1. Vuorio, E., and de Crombrugghe, B. (1990) Annu. Rev. Biochem. 59, 837–872
2. Rossert, J., Eberspaecher, H., and de Crombrugghe, B. (1995) J. Cell Biol. 129, 1421–1432
3. Pavlin, D., Lichter, A. C., Bedalov, A., Kream, B. E., Harrison, J. R., Thomas, H. F., Gronowicz, G. A., Clark, S. H., Woody, C. O., and Rowe, D. W. (1992) J. Cell Biol. 116, 226–236
4. Slack, J. L., Liakka, D. A. J., and Bornstein, P. (1991) Mol. Cell. Biol. 11, 2066–2074
5. Bou-Gharios, G., Garrett, L. A., Rossett, J., Niederreither, K., Eberspaecher, H., Smith, C., Black, C., and de Crombrugghe, B. (1996) J. Cell Biol. 134, 1333–1344
6. Schmidt, A., Rossi, P., and de Crombrugghe, B. (1986) Mol. Cell. Biol. 6, 347–354
7. Hatamochi, A., Paterson, B., and de Crombrugghe, B. (1986) J. Biol. Chem. 261, 11310–11314
8. Hatamochi, A., Golubbek, P. T., Van Schaftingen, E., and de Crombrugghe, B. (1988) J. Biol. Chem. 263, 5940–5947
9. Maity, S. N., Golubbek, P. T., Karsenty, G., and de Crombrugghe, B. (1988) Science 241, 582–585
10. Haasegawa, T., Zhou, X., Garrett, L. A., Ruteshouser, E. C., Maity, S. N., and de Crombrugghe, B. (1996) Nucleic Acids Res. 24, 3253–3260
11. Tamaki, T., Ohnishi, K., Hartl, C., LeRoy, E. C., and Trojanowska, M. (1995) J. Biol. Chem. 270, 4299–4304
12. Inagaki, Y., Truter, S., and Ramirez, F. (1994) J. Biol. Chem. 269, 14828–14834
13. Oikarinen, J., Hatamochi, A., and de Crombrugghe, B. (1987) J. Biol. Chem. 262, 11064–11070
14. Karsenty, G., and de Crombrugghe, B. (1990) J. Biol. Chem. 265, 9934–9942
15. Karsenty, G., and de Crombrugghe, B. (1991) Biochem. Biophys. Res. Commun. 177, 538–544
16. Wang, M. M., and Reed, R. R. (1993) Nucleic Acids Res. 21, 605–612
17. Li, J. J., and Herskowitz, I. (1993) Mol. Cell. Biol. 13, 4137–4144
18. Harper, J. W., Adami, G. R., Wei, N., Keyomarsi, K., and Elledge, S. J. (1993) Nature 363, 121–126
19. Wilson, T. E., Fahrner, T. J., Johnston, M., and Milbrandt, J. (1991) Science 252, 1296–1300
20. Gietz, D., St. Jean, A., Woods, R. A., and Schiestl, R. H. (1992) Nucleic Acids Res. 20, 1425
21. Deleted in proof
22. Deleted in proof
23. Nardelli, J., Gibson, T., and Charnay, P. (1992) Nucleic Acids Res. 20, 4137–4144
24. Sadowski, I., Ma, J., Triezenberg, S., and Ptashne, M. (1988) Nature 335, 563–564
25. Harper, J. W., Adami, G. R., Wei, N., Keyomarsi, K., and Elledge, S. J. (1993) Nature 363, 121–126
26. Chen, C., and Okayama, H. (1987) Mol. Cell. Biol. 7, 2745–2752
27. Neumann, J. R., Morency, C. A., and Russian, K. O. (1987) BioTechniques 5, 444–447
28. Hagen, G., Muller, S., Beato, M., and Suske, G. (1992) Nucleic Acids Res. 20, 5519–5525
29. Wang, Y., Kohori, J. A., and Hood, L. (1993) Mol. Cell. Biol. 13, 5691–5701
30. Gill, G., and Ptashne, M. (1988) Nature 334, 721–724
31. Galea, P., Mune, M., Duy, P., and Karsenty, G. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 9372–9376
32. Galea, P., Park, R.-W., Ducy, P., Mattei, M.-G., and Karsenty, G. (1996) J. Biol. Chem. 271, 21331–21339

S. Sinha and L. Garrett, personal communication.