Available online at www.sciencedirect.com
SCIENCE DIRECT®
Matrix Biology 23 (2004) 87–99
MATRIX BIOLOGY
www.elsevier.com/locate/matbio

Identification of a functional CBF-binding CCAAT-like motif in the core promoter of the mouse pro-α1(V) collagen gene (*Col5a1*)

Keiko Sakata-Takatani[a,b], Noritaka Matsuo[a], Hideaki Sumiyoshi[a], Tomiyasu Tsuda[b], Hidekatsu Yoshioka[a,*]

[a]*Department of Anatomy, Biology and Medicine, Faculty of Medicine, Oita University, 1-1 Hasama-machi, Oita 879-5593, Japan*
[b]*Department of Immunology and Allergy, Faculty of Medicine, Oita University, 1-1 Hasama-machi, Oita 879-5593, Japan*

Received 11 August 2003; received in revised form 26 March 2004; accepted 26 March 2004

Abstract

We used structure-function analysis of the core promoter region to elucidate the transcriptional features of the mouse α1(V) collagen gene (*Col5a1*). The core promoter, which lacks a typical TATA motif and has a high GC content, was defined within the 231 bp immediately upstream from the major transcription start site by transient transfection experiments. In this region, we identified three nuclear-factor binding sites by electrophoretic mobility shift assay: BS1 (−195 to −167), BS2 (−134 to −106), and BS3 (−110 to −80). Oligonucleotide competition and supershift assays revealed that Sp1, CBF, and Sp1-related protein specifically bind to BS1, BS2, and BS3, respectively. The CCAAT-like motif, CAAAT, and flanking sequences are conserved between the mouse and human gene. CBF, which recognizes this motif, activated the *Col5a1* promoter, as previously reported for *Col1a1* and *Col1a2*. Furthermore, overexpression of a wild-type and mutant forms of CBF-B subunit altered this activity. These results suggest that CBF is a key factor in the coordinated expression of type I and V collagen genes.

Keywords: α1(V) collagen promoter; CCAAT-binding factor CBF; Sp1; Gene regulation

1. Introduction

Fibril-forming collagens represent a structurally related group of molecules within the large family of collagens. This class of collagens includes nine distinct polypeptides that associate into five types of collagen: types I, II, III, V and XI (Vuorio and de Crombrugghe, 1990; van der Rest and Garrone, 1991; Brown and Timpl, 1995). They are essential for the integrity of the extracellular matrix scaffold. Type V collagen, which is a minor component, co-polymerizes with the major collagen, type I, and regulates the diameter of collagen fibers (Adachi and Hayashi, 1986; Linsenmayer et al., 1993; Andrikopoulos et al., 1995). Collagens are heterotrimer and/or homotrimer molecules made of three α chain. The predominant molecular form of type V collagen is the heterotrimer $[\alpha1(V)]_2\alpha2(V)$, although the α1(V)α2(V)α3(V) molecule has been extracted from human placenta and the homotrimer $[\alpha1(V)]_3$ also occurs in cultures of hamster lung cells (Burgeson et al., 1976; Sage and Bornstein, 1979; Haralson et al., 1980). Type V collagen production is elevated in many pathological conditions, such as inflammation (Narayanan et al., 1983), some forms of cancer (Barsky et al., 1982), and atherosclerosis (Ooshima, 1981).

Numerous studies have identified the *cis*-acting regulatory elements and the *trans*-acting factors that bind to the regulatory regions in mammalian collagen genes (Ramirez and Di Liberto, 1990). The transcriptional regulation of the α1(I) and α2(I) collagen genes has been most extensively studied (Slack et al., 1993). A common feature in the proximal regions of these promoters is the involvement of the ubiquitously expressed transcriptional factors Sp1, Sp3 (Inagaki et al., 1994; Tamaki et al., 1995; Ihn et al., 1996), and CBF (Maity et al., 1988; Coustry et al., 1995; Hatamochi et al., 1988; Maity and de Crombrugghe, 1998; Mantovani, 1998). Sp1 is a well-characterized zinc-finger transcrip-

Abbreviations: Sp1, specificity protein 1; CBF, CCAAT-binding factor; RACE, rapid amplification of cDNA ends; EMSA, electrophoretic mobility shift assay.

*Corresponding author. Tel.: +81-97-586-5670; fax: +81-97-549-6302.
E-mail address: hidey@med.oita-u.ac.jp (H. Yoshioka).

0945-053X/04/$30.00
doi:10.1016/j.matbio.2004.03.003

tion factor that binds to the GC box or a similar motif. Sp1 and the related protein, Sp3, work as activators of the α1(I) and α2(I) collagen genes under basal or TGF-β-inducible conditions. CBF binds to the CCAAT motif, which is one of the common elements present in the proximal promoter. CBF, also called NF-Y or CP1, consists of three different subunits, CBF-A, CBF-B, and CBF-C, all of which are necessary to form the CBF-DNA complex (Maity et al., 1992). CBF acts as a transcriptional activator for the α1(I) and α2(I) collagen genes. The CCAAT element is located between nucleotides −84 and −80 in the mouse α2(I) gene, and between nucleotides −100 and −96 in the mouse α1(I) gene in a reverse orientation. Mutation of the CCAAT motif decreases basal promoter activity in both genes (Karsenty et al., 1988; Karsenty and de Crombrugghe 1990).

Lee and Greenspan (1995) characterized the promoter of the human α1(V) collagen gene. A minimal promoter region within 212 bp immediately upstream of the major transcription start site contains the region where Sp1 binds. Furthermore, they suggested that GAGA boxes in the promoter and the first exon may be implicated in affecting levels of transcription. In the present study, we characterized the core promoter of the mouse α1(V) collagen gene. Cell transfection experiments and DNA-binding assays have demonstrated that the factors which bind to the core promoter region are the same as those responsible for the expression of type I collagen genes. In particular, CBF binds to the CAAAT sequence, but not to the CCAAT sequence, and activates the gene.

2. Results

2.1. Structural analysis of the 5′ portion of Col5a1

To isolate the 5′ part of *Col5a1*, we screened a mouse genomic library using mHY217, a most 5′ clone of mouse α1(V) cDNA, which was isolated previously (Wu et al., 1998). Comparison of the genomic sequences of isolated clones with that of mHY217 revealed that the genomic clones did not contain the 5′ sequences of mHY217, but sequences downstream from the ATG codon, presumably due to a short portion of 5′ UTR of cDNA. Therefore, we decided to isolate a cDNA clone that included the 5′ region further upstream by using the 5′ RACE procedure. We performed 5′ RACE with two specific primers, RACE1 and RACE2. A 322-bp RACE product overlapped with mHY217, and contained 230 bp in the 5′ region. We re-screened the genomic library using this cDNA as a probe, and isolated a 15-kb genomic clone, mgKS1 (Fig. 1a). Southern blot and sequencing analyses revealed that mgKS1 contained only 682 bp upstream from the ATG codon. To isolate a clone that contained more of the 5′ region, we re-screened with a 0.55-kb fragment containing most of the 5′ region of mgKS1. We thus isolated another clone, mgKS3, encoding a further 13 kb upstream from the 5′ end of mgKS1. (The nucleotide sequence has been submitted to the DDBJ/EMBL/GenBank under accession number, AB098608.)

The location of the transcription start site was identified by 5′ RACE using different combinations of primer sets and a nuclease protection assay (Yoshioka et al., 1995). We performed 5′ RACE with two primer sets, RACE1 and RACE3 primers, and RACE2 and RACE4 primers. After subcloning, 14 positive RACE clones were selected and sequenced. As shown in Fig. 1b, the ten 5′ RACE products extended to the same upstream site were isolated from three independent PCRs. This information strongly suggested that this upstream site is the major start site of *Col5a1* transcription. We prepared a 360-nucleotide (nt) riboprobe containing this putative start site for use in a nuclease protection experiment. After RNase digestion, we observed one major resistant product of approximately 170 nt, and several minor products (Fig. 1c). The major protected band has a size consistent with protection starting at the major transcriptional start site identified by 5′ RACE. We concluded that the *Col5a1* gene contains multiple transcription start sites, with a major one located 361 bp upstream from the ATG codon. The nucleotide sequence of around mouse promoter is similar to that of human counterpart except BS3 region, which is missing in human gene (Fig. 1b and Fig. 3a). The similarity between both species from −231 to +70 is 85%. The *Col5a1* promoter lacks typical TATA and CCAAT boxes, and has a high GC content as shown in human gene (Lee and Greenspan, 1995).

2.2. Functional analysis of the Col5a1 promoter region

To determine the smallest region required or sufficient to encompass the basal promoter of the *Col5a1* gene, a series of genomic fragments fused to a firefly luciferase reporter gene, pGL3-basic vector, were generated (Fig. 2a). These were transiently transfected into α1(V)-producing cells, NIH3T3 and A204 (Lee and Greenspan, 1995). Seven constructs containing progressive 5′ deletions of a 1.7-kb fragment were used to identify more closely the probable active promoter region (Fig. 2a). As shown in Fig. 2b, in the constructs from −1700/+39-Luc to −231/+39-Luc, luciferase activity was significantly higher than that of the pGL3-basic vector. Removal of the region between −231 and −84 reduced transcriptional activity to near basal levels. Confirming these results, the −1700/−232-Luc and −468/−232-Luc plasmids, which lack the −231 to +39 region, showed no transcriptional activity, suggesting that this region contains positive *cis*-acting elements (Fig. 2c). The plasmids containing a GAGA box in the 5′ untranslated region, −1700/+156-Luc and −231/+156-

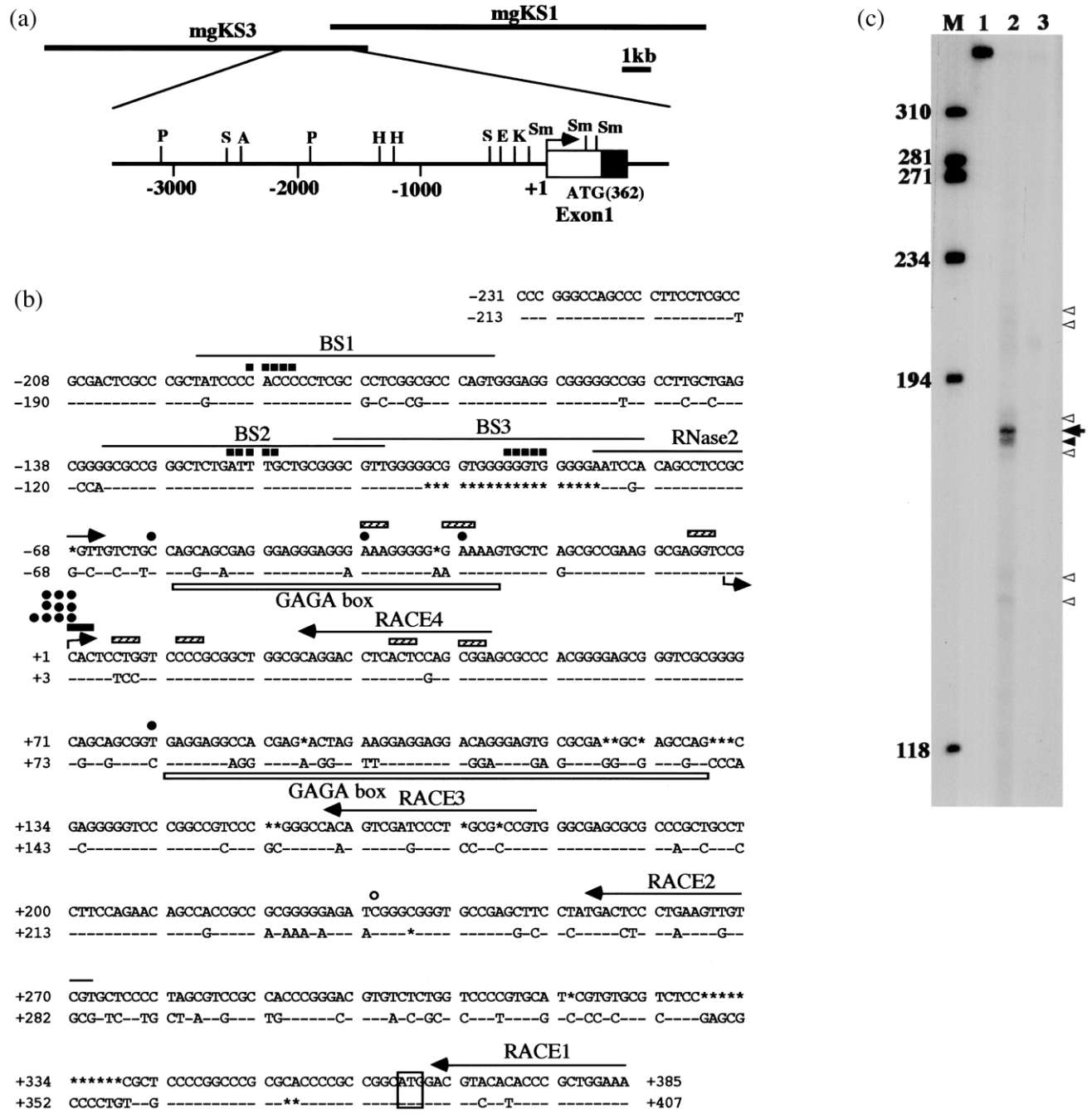

Fig. 1. Structural analysis of the 5′ region of *Col5a1*. (a) Schematic representation of the structure of the *Col5a1* gene. Top: Location of the genomic clones, mgKS1 and mgKS3. Bottom: Partial restriction map of a fragment from the mgKS3 clone. P, Pst I; S, Sac I; A, Apa I; H, Hind III; E, EcoR I; K, Kpn I; Sm, Sma I. Open and closed boxes represent untranslated and coding regions, respectively. The major start site of transcription is marked with an arrow. (b) Alignment of the sequences of mouse *Col5a1* (upper) and human *COL5A1* (lower) (Lee and Greenspan, 1995) promoter regions and parts of exon 1. Only the nucleotides of the human that differ from the mouse are shown. The small bars and asterisks indicate identical and missing nucleotides, respectively, in the corresponding regions of both species. The start site determined by 5′ RACE is denoted by black dots. The approximate 5′ ends of the major and minor RNase-protected bands are indicated by a black and striped bars, respectively. Nucleotide numbering of both species start with +1, which correspond to the major transcription start sites indicating with small arrows. The primers used for 5′ RACE and the RNase protection assay, the ATG codon of the translation start site, and the 5′ end of mgHY217 (Wu et al., 1998) are indicated by arrows, a box, and an open dot, respectively. The binding sites for nuclear factors, BS1–BS3, and the GAGA box identified in the human promoter are indicated with horizontal lines and open bars, respectively. The critical nucleotides in BS1-BS3 are emphasized with black squares. (c) RNase protection assay of the 360-nt riboprobe using RNA from mouse lung (lane 2) and yeast (lane 3). The undigested riboprobe is shown in lane 1, and size makers (in bp) are indicated in lane M. The major protected products is indicated by an arrow and, the minor products by open and closed arrowheads, respectively.

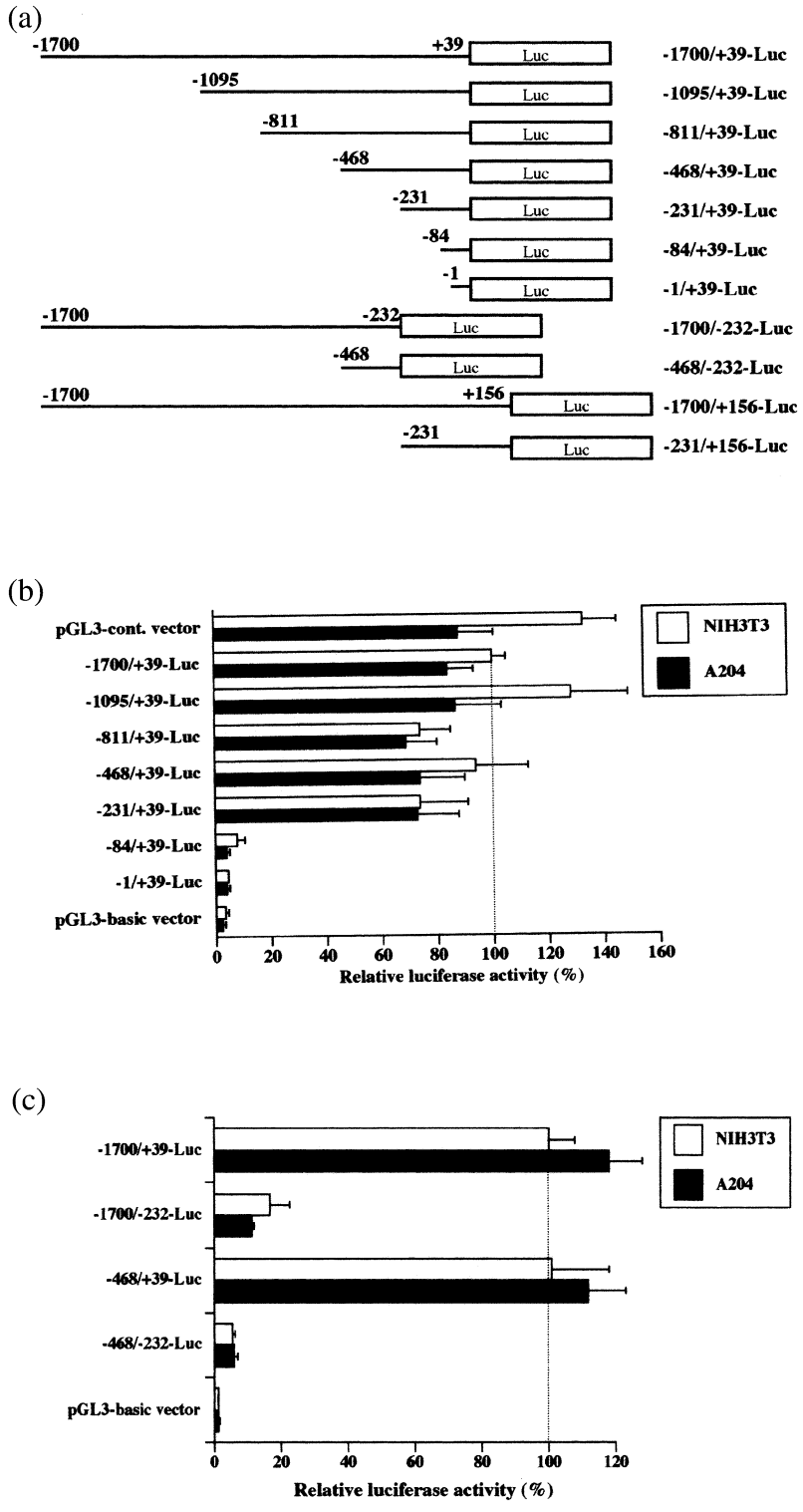

Fig. 2. Luciferase assay of the *Col5a1* promoter. (a) Schema of luciferase reporter constructs. The assay was performed using 5′ (b) and 3′ (c) stepwise-deleted constructs. A firefly luciferase reporter construct was cotransfected with the *Renilla* vector, pRL-TK, used as an internal control for transfection efficiency. The histograms indicate percentage activity normalized to the internal control and shown relative to that of the −1700/+39-Luc construct transfected into NIH3T3 cells. The data represent the mean of three independent tests ±S.D. Two different cell lines, NIH3T3 (white histogram) and A204 (black histogram), were used for transfection studies.

Luc, did not elevate promoter activity, as was reported for the human gene (data not shown) (Lee and Greenspan, 1995).

2.3. Identification of nuclear factor binding sites in the basal Col5a1 promoter

On the basis of the above data, we concentrated on the −231 to −84 region to characterize the minimum promoter for *Col5a1*. EMSA were performed to examine the transcription factors that bind in this region, using nuclear extracts from NIH3T3 cells. First, we used two oligonucleotides, corresponding to the sequences from −250 to −167 and from −176 to −80, as probes (Fig. 3a). The −250/−167 probe produced one major protein–DNA complex and the −176/−80 probe produced two major protein–DNA complexes (Fig. 3b). Each band was competitively inhibited by corresponding unlabeled excess oligonucleotide. Furthermore, the band in the −250 to −167 region and the upper band in the −176 to −80 were competitively inhibited by the −176/−80 and the −250/−167 probes, respectively. Because both bands migrated at the same position, the results suggest that the DNA-binding proteins might be identical or similar. To identify the DNA-binding sites of these proteins more closely, we prepared shorter oligonucleotides (Fig. 3a). The protein–DNA complex in the −250 to −167 region was detected by the −216/−167 and −195/−167 probes, but not by the −250/−207 probe, indicating that the binding site is located in the −195 to −167 region (binding site 1: BS1) (Fig. 3a,c lanes 1–3). The band detected by the −195/−167 probe migrated more slowly than that detected by the −216/−167 probe, and became a doublet (Fig. 3c lane 3Fig. 5c lane 2). This may reflect a slight difference in the conformation of the DNA–protein complex induced by each probe. In the −166 to −80 region, the lower band was detected by the −166/−93 probe, the −166/−106 probe, and the −134/−106 probe, indicating that the protein–DNA complex corresponding to the lower band is located in the −134 to −106 region (BS 2) (Fig. 3a,c lanes 5, 6 and 8). However, the band was detected by neither the −166/−121 nor the −120/−80 probe (Fig. 3c lanes 7 and 9). This result suggests that the region around −120 is critical for the formation of the protein–DNA complex. However, the upper band was detected by the −120/−80 and −110/−80 probes, indicating that the binding site is located in the −110 to −80 region (BS 3) (Fig. 3a,c lanes 9 and 10). No upper band was detected by the −166/−93 probe (Fig. 3c lane 5), implying that the −93 to −80 region is critical in the formation of the protein–DNA complex corresponding to the upper band.

The two sequences, 5′-CCCCACCCCC-3′ in the −195 to −167 region (BS 1), and 5′-GGGGGTGGGG-

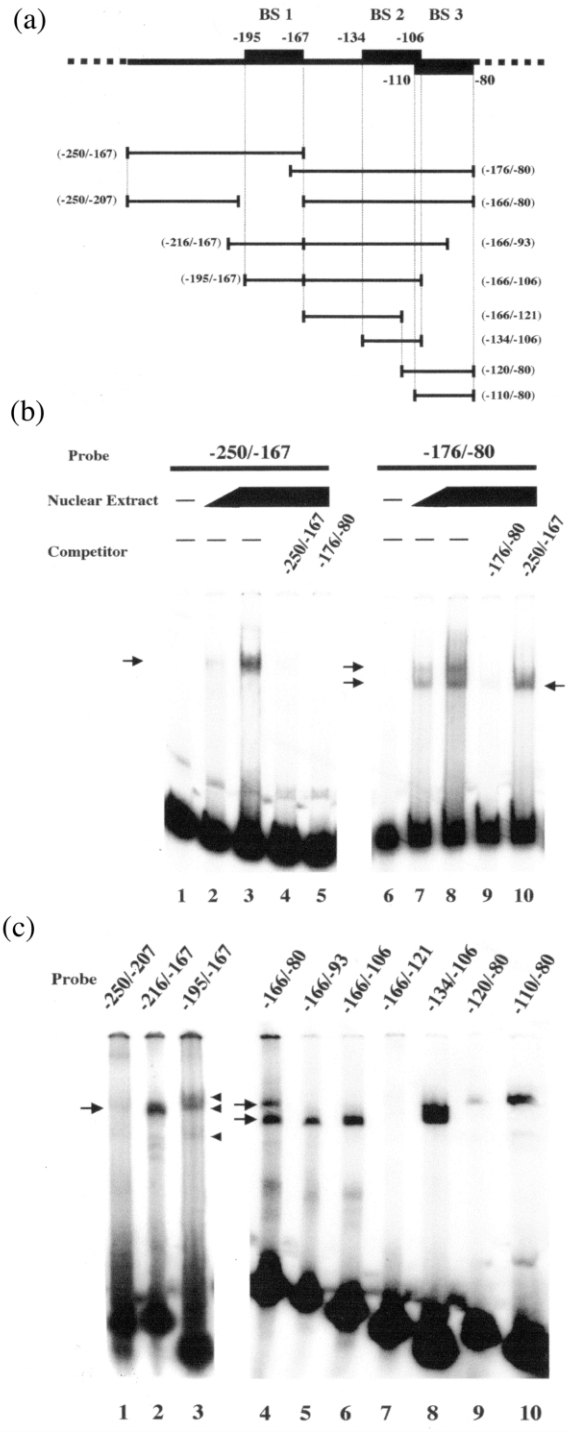

Fig. 3. Binding of nuclear factors to the *Col5a1* proximal promoter. (a) Positions of the designed oligonucleotides corresponding to *Col5a1* nucleotides −250 to −80 for EMSA. BS1, BS2, and BS3 indicate nuclear-factor binding sites. (b) Detection of protein–DNA complexes with two oligonucleotides covering the sequence from −250 to −80. (c) Determination of the binding sites of nuclear factors to the *Col5a1* proximal promoter. The arrows indicate the binding of specific complexes in (b) and (c). Note that three specific bands, indicated by arrowheads, can be seen in lane 3. Compare with lanes 2 and 7 in Fig. 5c.

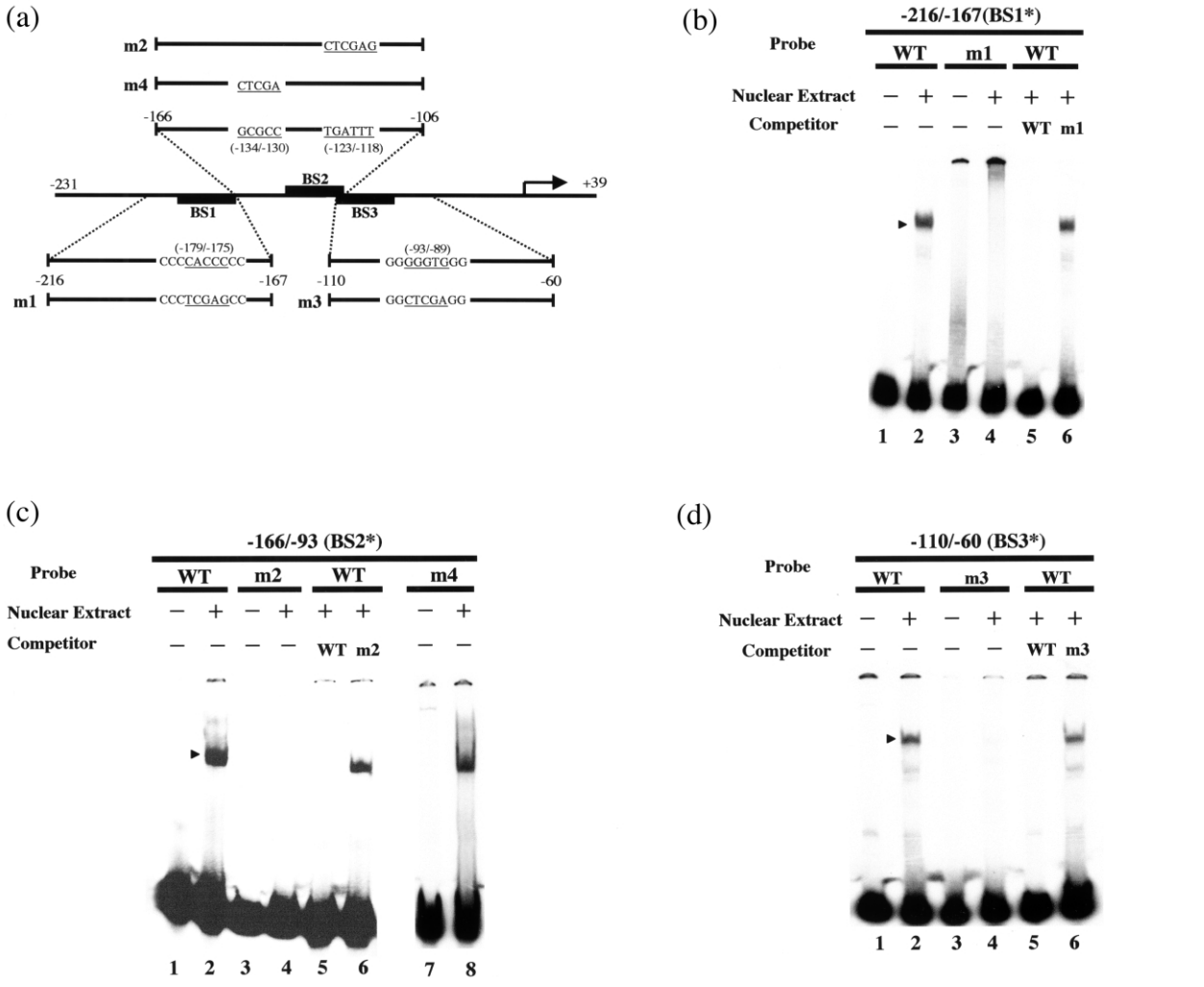

Fig. 4. Mutation analysis of the binding of nuclear factors to the *Col5a1* promoter. (a) Schematic representation of the locations and nucleotide substitutions of mutant oligonucleotides, m1–m4. The binding of nuclear factors was examined in the −216 to −167 (b), the −166 to −93 region (c), and −110 to −60 regions (d), with labeled wild-type (WT) or mutant (m1–m4) nucleotides in the presence or absence of excess unlabeled WT or mutant competitor. The specific binding of nuclear factors is indicated with arrowheads. BS1*, BS2* and BS3* indicate the regions containing BS1, BS2 and BS3, respectively.

3′ in the −110 to −80 region (BS 3), are inverted complementary motifs (Fig. 1b and Fig. 4a). This suggests that identical or closely related factors are involved in both regions. On the basis of these data, we generated mutant oligonucleotides substituted at critical positions, m1–m3, for a DNA-binding assay to assess the three BS regions (Fig. 4a). In the regions containing BS1 and BS3, the mutant oligonucleotides, m1 and m3, failed to form protein–DNA complexes (Fig. 4b, lane 4 and Fig. 4d, lane 4). Moreover, these mutant oligonucleotides did not compete with isotope-labeled wild-type oligonucleotides at the level at which the wild-type oligonucleotides inhibited their binding (Fig. 4b, lanes 5 and 6 and Fig. 4d, lanes 5 and 6). In the region containing BS2, the mutant oligonucleotide, m2, also failed to form a protein–DNA complex (Fig. 4c, lane 4), but the wild-type sequence and m4, which is mutated around −132, formed these complexes (Fig. 4c, lanes 2 and 8). A competition assay confirmed that m2 did not inhibit the binding of the normal protein–DNA complex (Fig. 4c lane 6).

2.4. Identification of nuclear factors in the basal Col5a1 promoter

To identify the nuclear factors associated with the *Col5a1* promoter, competition experiments were carried out by adding oligonucleotides containing the binding sites for several known transcription factors that occur in BS1, BS2, and BS3. Protein–DNA complexes in the −216 to −167 region (BS1) competed against excess unlabeled Sp1- and CBF-binding oligonucleotides (Fig. 5a, lanes 5 and 6). In the −166 to −80 region, the faster migrating complex in BS2 was competitively

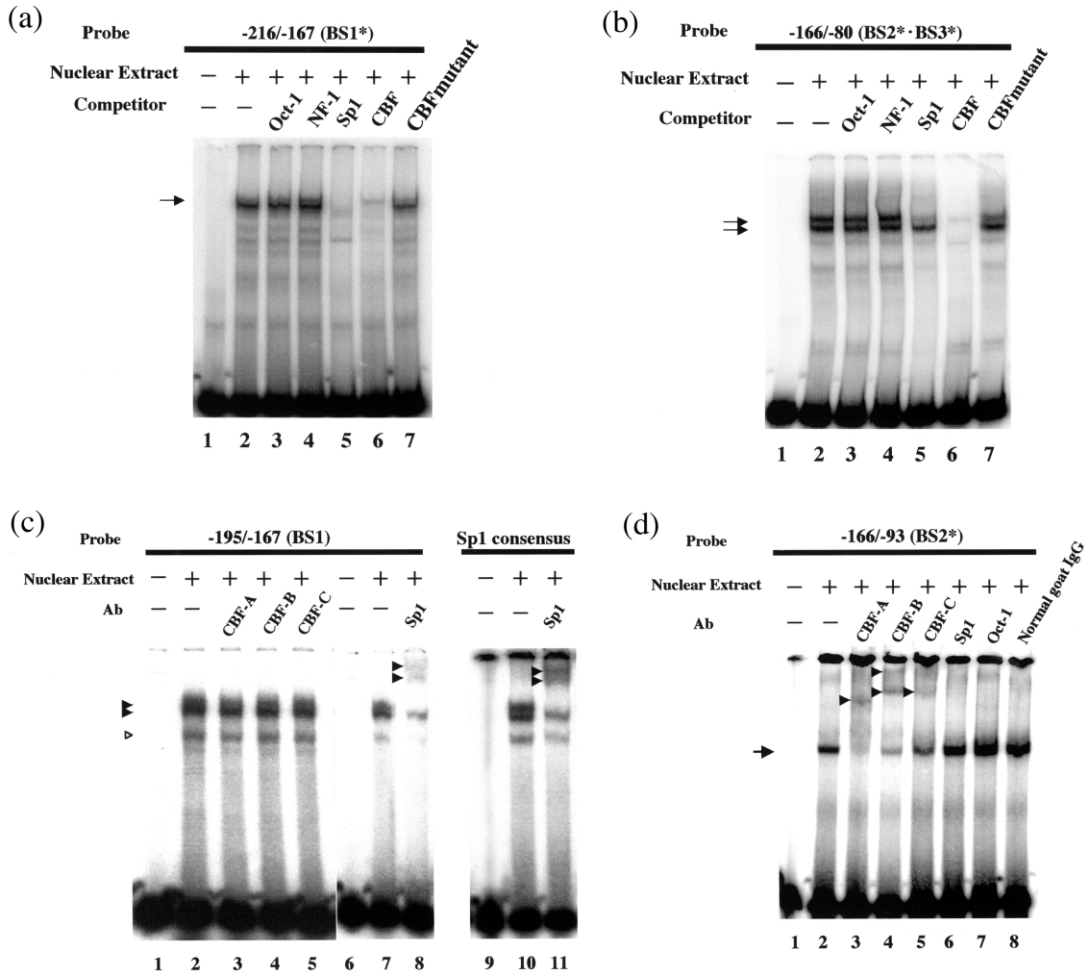

Fig. 5. Identification of nuclear factors binding to the *Col5a1* promoter. The competition assay with labeled probes corresponding to the regions −216 to −167 (a) and −166 to −80 (b) using oligonucleotides containing consensus binding sites for known nuclear factors, Oct-1 (lane 3), NF-1 (lane 4), Sp1 (lane 5), CBF (lane 6), and mutant CBF (lane 7), as competitors. Control assays were performed without nuclear extract or competitor (lane 1), and with nuclear extract but without competitor (lane 2). BS2*BS3* indicates the region containing BS2 and BS3. For the supershift assay, nuclear extract was incubated with labeled probes, −195/−167 and Sp1 consensus oligonucleotides (c) and −166/−93 oligonucleotide (d) after the addition of antibodies against CBF-A (lanes 3C and 3D), CBF-B (lanes 4C and 4D), CBF-C (lanes 5C and 5D), Sp1 (lanes 8C, 11C, and 6D), and Oct-1 (lane 7D). Control assays were performed without nuclear extract or antibody (lanes 1C, 6C, 9C, and 1D), with nuclear extract but without competitor (lanes 2C, 7C, 10C, and 2D), and with nuclear extract and normal goat IgG (lane 8D). Specific protein–DNA complexes and supershifted bands are indicated by arrows and arrowheads, respectively.

inhibited by CBF, whereas the slower one was competitively inhibited by both Sp1 and CBF in BS3 (Fig. 5b, lanes 5 and 6). We then evaluated supershift assays with antibodies against known factors (Fig. 5c,d). The complex in BS1 was supershifted with anti-Sp1 antibodies, but not with anti-CBFs antibodies (Fig. 5c). The complexes in BS3 were not markedly supershifted with anti-Sp1 antibodies (data not shown). The discrepancy between the oligonucleotide competition and supershift assays regarding the involvement of CBF in BS1 and BS3 is due to a cross-reaction between the consensus CBF-binding oligonucleotides and the Sp1 DNA-binding domain (data not shown). The complex in BS2 was supershifted by anti-CBF-A, -B, and -C antibodies, but by neither anti-Sp1 nor anti-Oct-1 antibodies, which is consistent with the data from the oligonucleotide competition assay (Fig. 5d).

To determine whether the Sp1 family or CBF activate the basal promoter of *Col5a1*, we examined the BS1, BS2, and BS3 regions using a luciferase assay, by introducing mutations into these regions (Fig. 6a). The BS2 mutation, m2-Luc, caused a decrease in promoter activity in NIH3T3 and A204 cells, as shown in Fig. 6b. However, the BS1 and the BS3 mutations, m1-Luc and m3-Luc, respectively, produced different results in the two cell lines. Overexpression of Sp1 also did not affect on wild, m1-Luc and m3-Luc promoter (data not shown).

Because the above experiment showed that CBF binds in BS2 where the CCAAT-like motif, 5′-ATTTG-3′

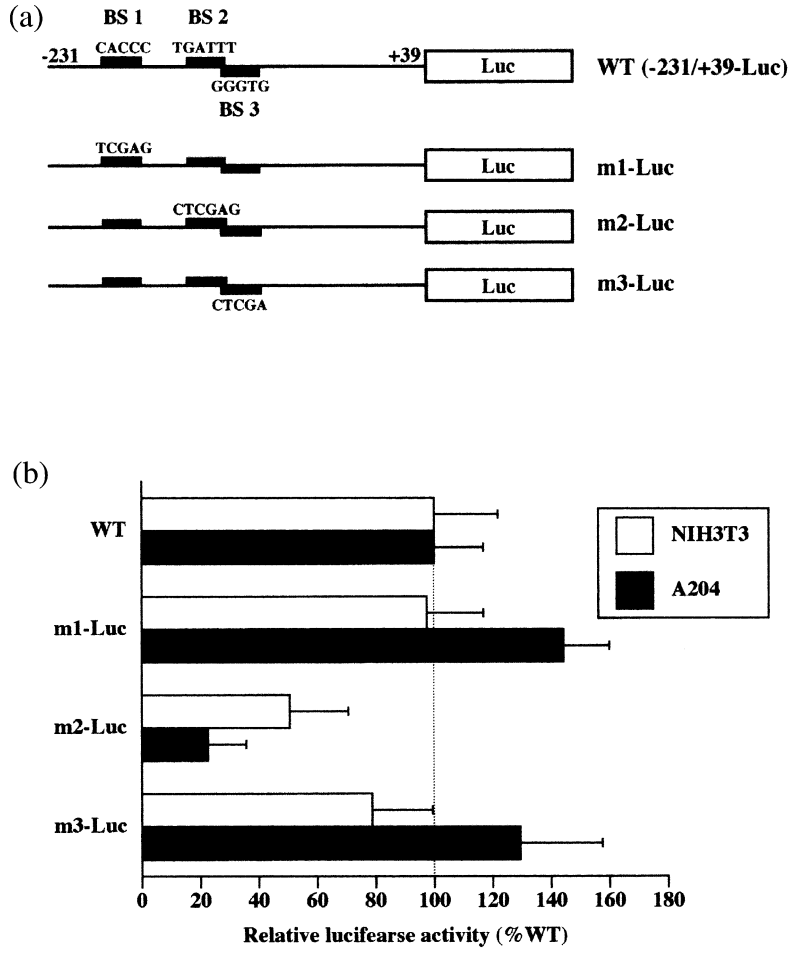

Fig. 6. Functional analysis of the *Col5a1* promoter. (a) Schematic representation of the mutant firefly luciferase reporter constructs, m1-Luc-m3-Luc. (b) Transient transfections with the wild-type construct from −231 to +39 and the mutant construct were carried out in NIH3T3 (white histogram) and A204 (black histogram) cells. A firefly luciferase reporter construct was cotransfected with the *Renilla* vector, pRL-TK, used as an internal control for transfection efficiency. The histograms indicate the percentage activity normalized to the internal control and shown relative to the activity of the wild-type construct transfected into NIH3T3 cells. The data represent the mean of three independent tests ± S.D.

(reverse CAAAT), is located, we examined the binding affinity of the CCAAT-like motif, including the prototype ATTGG (reversed CCAAT), using EMSA. DNA containing ATTTG in the *Col5a1* collagen promoter, or the mutant ATTCG, bound to CBF to similar degrees, but with a slightly weaker affinity than that of the CCAAT sequence. However, DNA with the TTTGG mutation completely failed to bind to CBF (Fig. 7).

Finally, we tested the effect using a wild-type and mutant forms of CBF-B subunit, which interacts with the CBF-A and CBF-C subunits to form CBF as previously described (Mantovani et al., 1994). Overexpression of wild type significantly increased promoter activity using −231/+39-Luc in NIH3T3 and A204 cells. However, in cotransfection with the mutant type, the promoter activity decreased approximately 20% and 30% in NIH3T3 and A204 cells, respectively (Fig. 8).

3. Discussion

The cloned 1.7-kb genomic fragment from the 5′ flanking region of *Col5a1* had strong transcriptional activity. The promoter lacks the canonical TATA and CCAAT boxes, but has a high GC content; these are features of housekeeping genes. Because deletion analysis showed that the promoter activity is located within the −231 to +39 region and that removal of the sequences between −231 and −84 reduced transcriptional activity to near basal levels, we concentrated on characterizing the −231 to −84 region. Using EMSA, we identified three specific regions that clearly bind to nuclear proteins. EMSA with unlabelled competitor oligonucleotides and antibodies showed that Sp1 bind to the −195 to −167 region, where the CACCC sequence is critical. Although its complementary

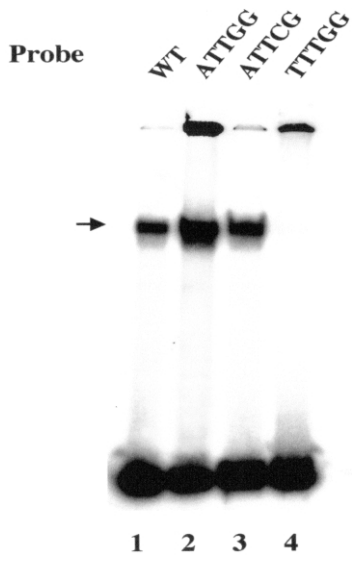

Fig. 7. Comparison of the binding affinity of CBF to different oligonucleotides containing CCAAT-like motifs. Electrophoretic mobility shift assays were performed using labeled −134/−106 oligonucleotides containing ATTTG (wild-type, lane 1), ATTGG (prototype, lane 2), ATTCG (lane 3), and TTTGG (lane 4).

sequence, GGGTG, in the −110 to −80 region, showed the same results as the CACCC sequence by in competition assays, we could not confirm the presence of supershifted bands in this region using Sp1 antibodies. A factor related to Sp1 might be involved in this region. However, this region is lacking in the human gene (Fig. 1b). Taken together with the data of the luciferase assay using mutations (Fig. 6b), these results suggest that Sp1/Sp3 or related factors do not greatly affect basic promoter function, or that their involvement is dependent upon cell type, wherein their contributions may differ. In contrast, despite the lack of a canonical CCAAT site or the reversed ATTGG site, CBF bound to the CCAAT-like motif (reversed ATTTG) in the −134 to −106 region, and activated the *Col5a1* promoter. Moreover, overexpression of a wild and mutant types of CBF-B subunit altered this promoter activity.

The CCAAT box is the most common element in eukaryotic promoters, found between 50 to 110 bp upstream from the start of transcription in both orientations. The mammalian promoters containing the CCAAT box are divided into three groups based on the effect on promoter activity of mutations to the motif (Maity and de Crombrugghe, 1998). In group I, mutation of the CCAAT motif in the promoter decreases basal promoter activity. In groups II and III, on the other hand, the CCAAT motif is necessary for the induction of promoter activity by various agents, and the alteration of promoter activity during cell growth, respectively. In the collagen gene family, CBF binds to the CCAAT motif in the α1(I) and α2(I) collagen promoters, resulting in an elevation of basal promoter activity. Similarly, we found that CBF activates basal promoter activity in the mouse α1(V) collagen gene. CBF requires the perfect pentanucleotide, CCAAT, except in a few genes (Mantovani, 1998). The globin r3′E gene and the ApoA-I gene (Wandersee et al., 1996; Papazafiri et al., 1991) contain a CAAAT sequence to which CBF can bind. For CBF to bind to the CCAAT motif, three 5′ (C, Pu, Pu) and five 3′ (C/G, A/G, G, A/C, G) flanking nucleotides also appear to be important (Fig. 9a) (Mantovani, 1998). These 5′ and 3′ flanking sequences of *Col5a1* perfectly match those of *Col1a1*.

Lee and Greenspan reported that the human *COL5A1* minimal promoter region contains GAGA boxes, which have been found in the TATA-less promoter regions of

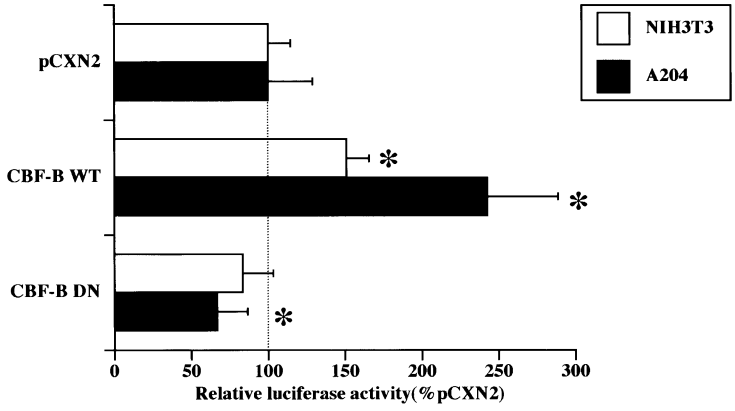

Fig. 8. Overexpression of a wild-type (WT) and mutant (DN) forms of CBF-B. NIH3T3 and A204 cells were transfected with the −231/+39-Luc construct together with plasmid DNA expressing a wild-type, or mutant form of CBF-B, or control vector with no insert (pCXN2). Transfections were normalized by cotransfection with the Renilla vector, pRL-TK. The data represent an average of five independent tests ±S.D. Asterisks indicate statistically significant compared to control vector ($P < 0.05$).

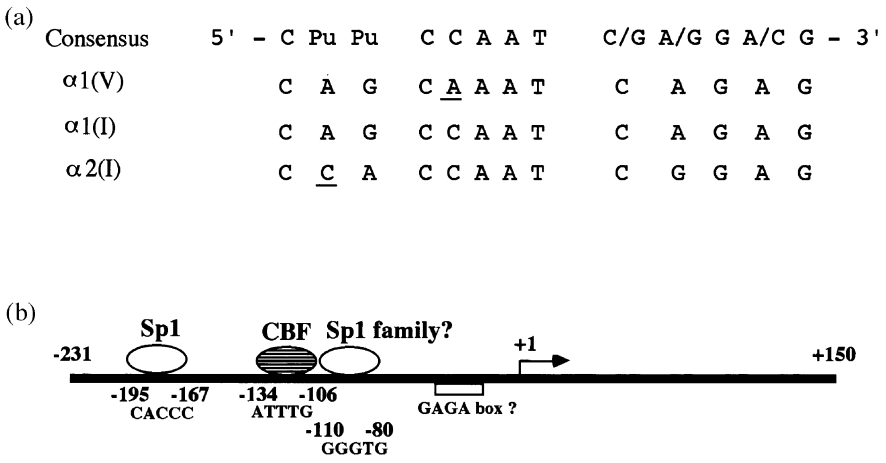

Fig. 9. The flanking sequence of CCAAT motif in the collagen promoters and schema of the core promoter region of *Col5a1*. (a) The 5′ and 3′ flanking sequences of the CCAAT motif in α1(V), α1(I), and α2(I) collagen promoters with reverse orientation. The nucleotides that differ from the consensus sequence are indicated with underlining. (b) Schematic summary of the binding of nuclear factors to the core promoter region of *Col5a1*. Sp1 bind to a CACCC box in the −195 to −167 region. CBF binds to a CCAAT-like motif in the −134 to −106 region. The Sp1 family may bind in the −110 to −80 region.

a number of extracellular matrix genes, immediately upstream and downstream from the major transcription start site (Lee and Greenspan, 1995). The region containing the downstream 56-bp GAGA box is sensitive to cleavage with S1 nuclease and has promoter activity in *COL5A1*, although the upstream region does not. They speculated that the binding factor at the GAGA box affects chromatin structure. However, we observed no promoter activity in the corresponding region of the mouse gene. Because the sequence is not similar in both species, it may only influence the human gene, and not the mouse gene. Our preliminary experiment showed a faint band in the upstream GAGA box of the mouse promoter by EMSA (data not shown). Although we could not identify a factor that binds to the upstream GAGA boxes, one may contribute to promoter activity (Fig. 1b and Fig. 9b).

Type V collagen is found in most connective tissue matrices, usually co-expressed with type I collagen. In the proximal promoter of α1(I) and α2(I) collagen genes, Sp1, which binds at several sites, and CBF are responsible for basal promoter activity. CBF is especially required for the coordinated expression of both type I collagen genes. Type III collagen is also co-expressed with type I collagen, but CBF cannot bind to its proximal promoter, suggesting that other factors are involved in the mechanism underlying the co-expression of type I and III collagen genes (Ruteshouser and de Crombrugghe, 1989). Since CBF activates the basal promoter of the α1(V) collagen gene, as shown in our experiment, CBF may be a key factor in the co-expression of type I and type V collagens. For collagen fibril formation, two collagen molecules must be produced simultaneously, even if the ratio of their production is different spatially and temporally. Sox9 and related factors are necessary for the specific production of type II and XI collagens, which form collagen fibrils in cartilaginous tissue (Lefebvre et al., 1997; Bridgewater et al., 1998). Similarly, type I and V collagens co-polymerize to form fibrils in non-cartilaginous tissues. Because CBF itself is the ubiquitous factor, it may be utilized for the broad expression of type I and V collagens. A specific factor, such as cbfa1, which is critical for osteoblast differentiation (Ducy et al., 1997), probably binds to some region of the α1(V) gene. The identification of factors other than CBF that affect cell-specific expression should provide further information on the expression of the type V collagen gene.

4. Materials and methods

4.1. Isolation and characterization of genomic clones

To isolate the 5′ flanking region of the mouse α1(V) collagen gene, a partially *Sau*3A1-restricted genomic library constructed in the Lambda FIX II vector (Stratagene, La Jolla, CA) was screened using the product of a rapid amplification of cDNA ends (RACE) as a probe (Yoshioka et al., 1995). RACE was carried out according to the method described by Frohman et al. (1988). Briefly, 20 μg of total RNA from 18-day-old mouse lung was reverse-transcribed using a *Col5a1*-specific primer, RACE1 (Fig. 1b). After tailing, the cDNA molecules were amplified by the polymerase chain reaction (PCR) using a 5′ adaptor primer and a 3′ *Col5a1*-specific primer, RACE2. Amplifications were carried out for 35 cycles of 94 °C for 1 min, 55 °C for 1 min, and 72 °C for 2 min, followed by a final

extension at 72 °C for 8 min. The RACE products were subcloned using the TA Cloning Kit (Promega, Madison, WI). Positive genomic clones were characterized as described previously (Yoshioka et al., 1995). Sequences were analyzed with an ABI 310 sequencer (Perkin–Elmer Life Sciences and Applied Biosystems).

4.2. Determination of the transcription start site

For this purpose, RACE and ribonuclease protection assays were used. In addition to the primers described above, another set of *Col5a1*-specific primers, RACE3 and RACE4, was used for the RACE procedure (Fig. 1b). Ribonuclease protection assays were performed using an RPA III Kit (Ambion, Austin, TX) according to the manufacturer's instructions. The DNA template for generating the riboprobe was derived from mgKS3 genomic DNA (Fig. 1a), using PCR with the RNase2 and RACE3 primers. The product was subcloned into the TA vector and linearized at the *Hin*dIII site. Total RNA (10 μg) was hybridized overnight at 42 °C to 10^5 counts per minute (cpm) of radiolabeled antisense riboprobe produced with an in vitro transcription kit (Ambion). Following hybridization, the reactions were incubated with single-strand-specific RNase, and the protected fragments were analyzed on a 5% denaturing polyacrylamide gel (Yoshioka et al., 1995).

4.3. Construction of chimeric plasmids for luciferase assay

Luciferase reporter constructs were generated using the promoterless firefly luciferase reporter vector, pGL3-Basic (Promega, Madison, WI). *Sac*I-site-linked 5′ and *Xho*I-site-linked 3′ primers specific for the *Col5a1* genomic sequence determined above were used to amplify the mgKS3 genomic fragment by PCR. Initially, the −811/+39 fragment was amplified and cloned into the pGEM-T Easy vector. The −468/+39-Luc and −231/+39-Luc constructs were produced by digesting this fragment with *Sac*I/*Xho*I and *Sma*I/*Xho*I, respectively, and were ligated into the corresponding sites in pGL3. The −84/+39-Luc, −1/+39-Luc, and −231/+156-Luc fragments were generated by PCR amplification, cloned into the pGEM-T vector and ligated into the *Sac*I/*Xho*I sites of pGL3. To construct the fragments −1700/+39-Luc, −1095/+39-Luc, and −811/+39-Luc, the upstream fragments −1700/−469, −1095/−469, and −811/−469 were amplified by PCR and ligated into the *Sac*I site in −468/+39-Luc in the correct orientation. To generate the −1700/−232-Luc fragment, the −1700/−469 fragment was ligated to −468/−232-Luc. To generate the −468/−232-Luc fragment, −468/+39-Luc was digested with *Sac*I/*Sma*I to remove the 3′ section. For the −1700/+156-Luc fragment, the −1700/+39-Luc fragment was cut with *Sma*I/*Xho*I and ligated to the −231/+156 fragment derived from the original genomic fragment. The *Xho*I site, CTCGAG, was introduced into the −231/+39-Luc fragment to generate the mutant constructs, m1-Luc, m2-Luc, and m3-Luc. The upstream fragments were amplified by PCR using *Sma*I-site-linked 5′ and *Xho*I-site-linked 3′ primers, cloned into the pGEM-T vector, transferred into the *Sma*I/*Xho*I sites of pGL3, and ligated to the downstream fragments linked to *Xho*I sites. All PCR-derived constructs were fully sequenced and compared with the original genomic clone.

4.4. Construction of CBF-B mutant construct

The CBF-B construct was generated by RT-PCR using RNA extracted from A204 cells. A set of primers were
sense: 5′-GTCGACGGAGGGACCATGGAGCAGTA-TA-3′; and
antisense: 5′-TCTAGAGGGTTAGGACACTCGGATG-ATC-3′.

(Underlining indicates tagged *Sal*I and *Xba*I restriction sites for cloning). This PCR products was cloned into the pGEM-T Easy vector and subcloned into the *Sal*I/*Xba*I site in the empty vector.

The CBF-B mutant construct was generated by replacing Arg-Gly-Glu (amino acids 312–314) with Ala-Ala-Ala (Mantovani et al., 1994). Two sets of primers for the 5′ and 3′ CBF-B mutant fragments were:
sense: 5′-GTCGACGGAGGGACCATGGAGCAGTA-ATA-3′;
antisense: 5′-*GCGGCCGC*CTTCCGTGCCATGGCAT-GAC-3′;
sense: 5′-*GCGGCCGC*AGGTGGACGATTTTTCTC-TC-3′; and
antisense: 5′-TCTAGAGGGTTAGGACACTCGGATG-ATC-3′.

(Underlining indicates tagged *Sal*I and *Xba*I restriction sites for cloning, and italicized sequences are *Not*I sites, which introduce mutation). These PCR products were cloned into the pGEM-T Easy vector and subcloned into the *Sal*I/*Not*I sites for the 5′-CBF-B mutant, followed by the *Not*I/*Xba*I sites for the 3′-CBF-B mutant into the empty vector.

Finally, wild and mutant forms were cloned into the mammalian expression vector, pCXN2.

4.5. Cells and culture conditions

The cell lines used in this study were NIH3T3 mouse fibroblasts and human rhabdomyosarcoma A204 (HTB82) cells. These cells were grown and maintained in Dulbecco's modified Eagle's medium supplemented with 10% heat-inactivated fetal bovine serum (Sanko-Junyaku, Tokyo, Japan).

4.6. Transient transfections and luciferase assays

Approximately 2×10^5 cells were seeded in 20 mm dishes 24 h before transfection using 5 μg of various *Col5a1* promoter reporter constructs co-transfected with 0.25 μg of *Renilla* pRL-TK plasmid. Transient transfections were performed using the calcium-phosphate precipitation method (Yoshioka et al., 1995; Graham and van der Eb, 1973). After transfection, the cells were incubated at 37 °C for 3 h, shocked for 1 min in a 15% (v/v) glycerol solution, washed with phosphate-buffered saline, and incubated with fresh medium. After 48 h, the cells were washed with phosphate-buffered saline and lysed in Passive Lysis Buffer (Promega) at room temperature for 5 min. Firefly and *Renilla* luciferase activities were determined using the Dual-Luciferase Reporter Assay System (Promega), according to the manufacturer's instructions using a luminometer (Lumat LB 9507, Perkin–Elmer Life Sciences). Transfection experiments were independently repeated 3–5 times.

For co-transfection experiments with a wild and mutant types of CBF-B expression vector, 1.25 μg of *Col5a1* reporter plasmid DNA was transfected together with 3.75 μg of wild or mutant types of CBF-B cDNA, or pCXN2 vector DNA (Mantovani et al., 1994). Transfections were normalized by co-transfection with 0.25 μg of the *Renilla* vector pRL-TK.

4.7. Preparation of nuclear extracts

Nuclear extracts were prepared from 1×10^8 NIH3T3 cells (Yoshioka et al., 1995; Morris et al., 1986). Cells were scraped in phosphate-buffered saline and centrifuged for 5 min at 1500 rev./min. The pellets were resuspended in 10 ml hypotonic buffer (10 mM HEPES [pH 7.8], 10 mM KCl, 0.1 mM EDTA-2Na, 0.1% NP40) and homogenized for 10 min on ice. The homogenate was centrifuged for 10 min at 3000 rev./min at 4 °C. The supernatant was removed, and the nuclei in the pellet were resuspended in 400 μl of low-salt buffer (20 mM HEPES [pH 7.8], 420 mM KCl, 0.1 mM EDTA-2Na, 20% glycerol, 5 mM $MgCl_2$) by stirring with a pipette tip. Nuclear proteins were extracted for 30 min at 4 °C with continuous gentle mixing with a rotator, and centrifuged for 15 min at 15 000 rev./min. The supernatants were subsequently centrifuged for 30 min at $24\,000\times g$ at 4 °C. The protein concentration of the extracts was determined using Bio-Rad reagent. Solutions containing the nuclear proteins were frozen immediately at −80 °C. The samples were thawed gently on ice just before electrophoretic mobility-shift assays were performed.

4.8. Electrophoretic mobility-shift assay (EMSA)

Probes for EMSA were generated by PCR using *Hin*dIII-site-linked *Col5a1*-specific primers. The mutant oligonucleotides were generated using mutant luciferase constructs as DNA templates. DNA fragments were digested with *Hin*dIII and end-labeled with [α-^{32}P]dCTP using Klenow fragment (Amersham Biosciences). EMSA was carried out by incubating 5–10 μg of nuclear extract, 3 μg of poly(dI-dC), 5× reaction buffer containing 50 mM HEPES-KOH (pH 7.8), 250 mM KCl, 5 mM EDTA-2Na (pH 8.0), 50% glycerol, and 25 mM $MgCl_2$ (Yoshioka et al., 1995). Approximately 30 000 cpm of radiolabeled probe was added to the preincubation mixture. After 30 min incubation at 25 °C, DNA–protein complexes were resolved on non-denaturing 4.5% acrylamide gels in 0.25 TBE/2.5% glycerol at 4 °C and visualized by autoradiography. For the competition assays, double-stranded oligonucleotides containing a consensus Sp1, Oct-1, NF-1, CBF, or CBF mutant binding site were generated by annealing equimolar complementary oligonucleotides at 100 °C.

The consensus sequences are:

Sp1: 5′-ATTCGATCGGGGCGGGGCGAGC-3′,
Oct-1: 5′-TGTCGAATGCAAATCACTAGAA-3′,
NF-1: 5′-TTTTGGATTGAAGCCAATGATA-3′,
CBF: 5′-AGACCGTACGTGATTGGTTAATCTCTT-3′,
CBF mutant: 5′-AGACCGTACGAAATACGGGAATCTCTT-3′.

For the supershift assay, 2 μg of polyclonal antibody directed against Sp1, Oct-1, or CBF subunits A, B or C, or preimmune goat polyclonal IgG (Santa Cruz Biotechnology, CA) were added to the reactions and incubated at 4 °C for 1 h before the addition of probe.

Acknowledgments

We thank Dr F. Ramirez for reviewing the paper, Dr T. Kumamoto and Dr K. Sugisaki for support during this work, and the staff of Division of Biomolecular Medicine and Medical Imaging, and Division of Radioisotope Research, Institute of Scientific Research, Oita University where we performed some experiments. This work was supported by Grants-in-Aid for Scientific Research (11470312 and 14370468 to H.Y.) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

Adachi, E., Hayashi, T., 1986. In vitro formation of hybrid fibrils of type V collagen and type I collagen into thick fibrils by type V collagen. Connect. Tissue Res. 14, 257–266.

Andrikopoulos, K., Liu, X., Keene, D.R., Jaenisch, R., Ramirez, F., 1995. Targeted mutation in the *col5a2* gene reveals a regulatory role for type V collagen during matrix assembly. Nat. Genet. 9, 31–39.

Barsky, S.H., Rao, C.N., Grotendorst, G.R., Liotta, L.A., 1982. Increased content of type V collagen in desmoplasia of human breast carcinoma. Am. J. Pathol. 108, 276–283.

Bridgewater, L.C., Lefebvre, V., de Crombrugghe, B., 1998. Chondrocyte-specific enhancer elements in the *Col11a2* gene resemble the *Col2a1* tissue-specific enhancer. J. Biol. Chem. 273, 14 998–15 006.

Brown, J.C., Timpl, R., 1995. The collagen superfamily. Int. Arch. Allergy Immunol. 107, 484–490.

Burgeson, R.E., El Adli, F.A., Kaitila, I.I., Hollister, D.W., 1976. Fetal membrane collagens: identification of two new collagen alpha chains. Proc. Natl. Acad. Sci. USA 73, 2579–2583.

Coustry, F., Maity, S.N., de Crombrugghe, B., 1995. Studies on transcription activation by the multimeric CCAAT-binding factor CBF. J. Biol. Chem. 270, 468–475.

Ducy, P., Zhang, R., Geoffroy, V., Ridall, A.L., Karsenty, G., 1997. Osf2/cbfa1: a transcriptional activator of osteoblast differentiation. Cell 89, 747–754.

Frohman, M.A., Dush, M.K., Martin, G.R., 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA 85, 8998–9002.

Graham, F., van der Eb, A.J., 1973. Transformation of rat cells by DNA of human adenovirus 5. Virology 54, 536–539.

Haralson, M.A., Michell, W.M., Rhodes, R.K., Kresina, T.F., Gay, R., Miller, E.J., 1980. Chinese hamster lung cells synthesize and confine to the cellular domain a collagen composed solely of B chains. Proc. Natl. Acad. Sci. USA 77, 5206–5210.

Hatamochi, A., Golumbek, P.T., van Schaftingen, E., de Crombrugghe, B., 1988. A CCAAT DNA binding factor consisting of two different components that are both required for DNA binding. J. Biol. Chem. 263, 5940–5947.

Ihn, H., Ohnishi, K., Tawaki, T., LeRoy, E.C., Trojanowska, M., 1996. Transcriptional regulation of the human α2(I) collagen gene. J. Biol. Chem. 271, 26 717–26 723.

Inagaki, Y., Truter, S., Ramirez, F., 1994. Transforming growth factor-β stimulates α2(I) collagen gene expression through a *cis*-acting element that contains an Sp1 binding site. J. Biol. Chem. 269, 14 828–14 834.

Karsenty, G., Golumbek, P.T., de Crombrugghe, B., 1988. Point mutations and small substitution mutations in three different upstream elements inhibit the activity of the mouse α2(I) collagen promoter. J. Biol. Chem. 263, 13 909–13 915.

Karsenty, G., de Crombrugghe, B., 1990. Two different negative and one positive regulatory factors interact with a short promoter segment of the α1(I) collagen gene. J. Biol. Chem. 265, 9934–9942.

Lee, S., Greenspan, D.S., 1995. Transcriptional promoter of the human α1(V) collagen gene (*COL5A1*). Biochem. J. 310, 15–22.

Lefebvre, V., Huang, W., Harley, V.R., Goodfellow, P.N., de-Crombrugghe, B., 1997. SOX9 is a potent activator of the chondrocyte-specific enhancer of the proα1(II) collagen gene. Mol. Cell. Biol. 17, 2336–2346.

Linsenmayer, T.F., Gibney, E., Igoe, F., Gordon, M.K., Fitch, J.M., Fessler, L.I., et al., 1993. Type V collagen: molecular structure and fibrillar organization of the chick α1(V) NH_2-terminal domain, a putative regulator of corneal fibrillogenesis. J. Cell Biol. 121, 1181–1189.

Maity, S.N., Golumbek, P.T., Karsenty, G., de Crombrugghe, B., 1988. Selective activation of transcription by a novel CCAAT binding factor. Science 241, 582–585.

Maity, S.N., Sinha, S., Ruteshouser, C., de Crombrugghe, B., 1992. Three different polypeptide are necessary for DNA binding of the mammalian heteromeric CCAAT binding factor. J. Biol. Chem. 267, 16 574–16 580.

Maity, S.N., de Crombrugghe, B., 1998. Role of the CCAAT-binding protein CBF/NF-Y in transcription. TIBS 23, 174–178.

Mantovani, R., Li, X.-Y., Pessara, U., van Huisjduijnen, R.H., Benoist, C., Mathis, D., 1994. Dominant negative analogs of NF-YA. J. Biol. Chem. 269, 20 340–20 346.

Mantovani, R., 1998. A survey of 178 NF-Y binding CCAAT boxes. Nucleic Acids Res. 26, 1135–1143.

Morris, G.F., Price, D.H., Marzluff, W.F., 1986. Synthesis of U1 RNA in a DNA-dependent system from sea urchin embryos. Proc. Natl. Acad. Sci. USA 83, 3674–3678.

Narayanan, A.S., Enge l, L.D., Page, R.C., 1983. The effect of chronic inflammation on the composition of collagen types in human connective tissue. Collagen Relat. Res. 3, 323–334.

Ooshima, A., 1981. Collagen alpha B chain: increase proportion in human atherosclerosis. Science 213, 666–668.

Papazafiri, P., Ogami, K., Nicosia, M., Moraci, P., Cladaras, C., Zannis, V.I., 1991. Promoter elements and factors involved in hepatic transcription of the human ApoA-I gene positive and negative regulators bind to overlapping sites. J. Biol. Chem. 266, 5790–5797.

Ramirez, F., Di Liberto, M., 1990. Complex and diversified regulatory programs control the expression of vertebrate collagen genes. FASEB J. 4, 16 116–16 123.

Ruteshouser, E.C., de Crombrugghe, B., 1989. Characterization of two distinct positive cis-acting elements in the mouse α1(III) collagen promoter. J. Biol. Chem. 264, 13 740–13 744.

Sage, H., Bornstein, P., 1979. Characterization of a novel collagen chain in human placenta and its relation to AB collagen. Biochemistry 18, 3815–3822.

Slack, J.L., Liska, D.J., Bornstein, P., 1993. Regulation of expression of the type I collagen genes. Am. J. Med. Genetics 45, 140–151.

Tamaki, T., Ohnishi, K., Hartl, C., LeRoy, C., Trojanowska, M., 1995. Characterization of a GC-rich region containing Sp1 binding site(s) as a constitutive responsive element of the α2(I) collagen gene in human fibroblasts. J. Biol. Chem. 270, 4299–4304.

van der Rest, M., Garrone, R., 1991. Collagen family of proteins. FASEB J. 5, 2814–2823.

Vuorio, E., de Crombrugghe, B., 1990. The family of collagen genes. Annu. Rev. Biochem. 59, 837–872.

Wandersee, N.J., Ferris, R.C., Ginder, G.D., 1996. Intronic and flanking sequences are required to silence enhancement of an embryonic beta-type globin gene. Mol. Cell. Biol. 16, 236–246.

Wu, Y.-L., Sumiyoshi, H., Khaleduzzaman, M., Ninomiya, Y., Yoshioka, H., 1998. cDNA sequence and expression of the mouse α1(V) collagen gene (*Col5a1*). Biochim. Biophys. Acta 1397, 275–284.

Yoshioka, H., Greenwel, P., Inoguchi, K., Truter, S., Inagaki, Y., Ninomiya, Y., et al., 1995. Structural and functional analysis of the promoter of the human α1(XI) collagen gene. J. Biol. Chem. 270, 418–424.