Journal of Cellular Biochemistry 70:455–467 (1998)

Characterization of Two Distinct Families of Transcription Factors That Bind to the CCAAT Box Region of the Human COL1A2 Gene

Malcolm Collins, Arlene A. Smith, and M. Iqbal Parker*

Department of Medical Biochemistry, Faculty of Health Sciences, University of Cape Town, Observatory, Cape Town, South Africa

Abstract Both the mouse and human α2(I) procollagen promoters contain an inverted CCAAT box at −80, but only the human promoter contains an additional regulatory element, the collagen modulating element (CME), immediately downstream of the CCAAT box [Collins et al. (1997): Biochem J 322:199–206]. In this study, the transcription factors that bind to the G/CBE and CME within the human promoter were characterized in SVWI-38 and CT-1 nuclear extracts. Two distinct proteins bind to the CME, and both were identified as heat-labile factors that were sensitive to high ionic strengths and required Zn^{2+} for DNA-binding activity. These proteins had Stokes radii of 4.12 and 3.15 nm, sedimentation coefficients of 3.9 and 3.2 S and native molecular weights of 66 and 41 kDa, respectively. On the basis of biochemical and DNA-binding properties, the CME binding proteins are probably novel factors involved in the regulation of the human α2(I) procollagen gene. By contrast, the G/CBE binding proteins were more resistant to heat, ionic strength, and divalent metal ion chelators, demonstrating that the G/CBE and CME binding proteins had distinct DNA-binding properties. The above properties suggest that this factor is a member of the previously characterized family of CCAAT box-binding factors, CBF, NF-Y, CP-1 and α-CP1. Taken together, these physicochemical properties of the COL1A2 CCAAT box and CME-binding proteins demonstrated that they were distinct unrelated transcription factors. These results also suggest that there is a distinct difference in the DNA-binding activity between the equivalent region of the mouse and human α2(I) procollagen promoters. J. Cell. Biochem. 70:455–467, 1998.

Key words: collagen; gene regulation; DNA-binding proteins

Type I collagen, the most abundant protein in the extracellular matrix, is a heterotrimer consisting of two α1 and one α2 chains [for review, see van der Rest and Garrone, 1991]. Besides its well-characterized structural properties, the protein can also function as an adhesive substratum for mesenchymal cells and influence cellular functions through its interactions with integrins [Damsky et al., 1993]. Although type I collagen is synthesized and secreted predominantly by fibroblasts, it can also be produced by a variety of other differentiated cells [Bornstein and Sage, 1980, 1989]. The genes that encode the human α1(I) and α2(I) chains have been mapped to chromosomes 17 and 7, respectively [Huerre et al., 1982] and are coordinately regulated in a strict spatial and temporal manner during normal development and in the adult organism [Bornstein and Sage, 1989]. Since type I collagen levels are altered during pathological conditions such as wound healing, inflammatory conditions, fibrosis, arthritis, scleroderma, and cancer [Bornstein and Sage, 1980], the mechanisms that regulate the levels of these proteins have been extensively studied [reviewed in de Crombrugghe et al., 1991; Bornstein and Sage, 1989; Slack et al., 1993; Majack and Bornstein, 1984; Adams, 1989]. In spite of the progress made to date, the precise mechanisms still remain poorly understood.

Boast et al. [1990] have shown that the proximal human COL1A2 promoter (350 bp) is sufficient for high levels of cell-type-specific expression of reporter constructs in cultured cells. Subsequently, both positive and negative regu-

Contract grant sponsor: South African Medical Research Council; Contract grant sponsor: Cancer Association of South Africa; Contract grant sponsor: University of Cape Town.

*Correspondence to: M. Iqbal Parker, Department of Medical Biochemistry, Faculty of Health Sciences, University of Cape Town, Observatory 7925, Cape Town, South Africa. E-mail: mparker@physio.uct.ac.za

Received 22 January 1998; Accepted 26 March 1998

latory elements, as well as cytokine-responsive elements, have been identified within the proximal promoter [Parker et al., 1992; Collins et al., 1997; Tamaki et al., 1995; Ihn et al., 1996; Chung et al., 1996; Inagaki et al., 1994, 1995b]. The proximal promoter has also been implicated in type I collagen synthesis during hepatofibrogenesis [Inagaki et al., 1995a]. The positive regulatory elements within the proximal human COL1A2 promoter include an inverted CCAAT box (−84 to −80) and GC-rich regions [Collins et al., 1997; Ihn et al., 1996; Inagaki et al., 1994]. One of these GC-rich regions is located upstream of the CCAAT box between nucleotides −133 to −119, while the second region has been mapped to a distal element containing three clustered GC-boxes between nucleotides −303 and −271 [Inagaki et al., 1994; Ihn et al., 1996]. The ubiquitous transcription factor, Sp1, has been shown to bind to the distal element were it functions as a strong activator [Inagaki et al., 1994; Greenwel et al., 1997]. An unknown transcriptional activator(s) binds to the downstream GC-rich region, where it contributes significantly to the basal promoter activity [Ihn et al., 1996]. Both Sp1 and the tissue-specific zinc finger protein, c-Krox, have been shown to bind to an equivalent region within the mouse promoter [Galera et al., 1996; Hasegawa et al., 1996]. The inverted CCAAT box at −80 is conserved within the α2(I) procollagen promoter across all the species analyzed to date and an analogue of the extensively studied mouse heterotrimeric CCAAT-binding factor (CBF) is believed to bind to this element in the human gene [Collins et al., 1997; Ihn et al., 1996]. The CCAAT box is also conserved in the COL1A1 gene, where it, like the c-Krox binding motif, probably plays an important role in the coordinated expression of the α1(I) and α2(I) procollagen genes [Karsenty and de Crombrugghe, 1990].

A pyrimidine-rich repressor element, which binds to an unknown factor(s) has been identified between nucleotides −164 and −159 of the human α2(I) procollagen promoter [Ihn et al., 1996]. A homologous region has also been identified within the mouse Col1A2 and Col1A1 promoters and shown to bind the transcriptional repressor IF1 [Karsenty and de Crombrugghe, 1991; Karsenty et al., 1991]. Transcriptional activators, such as, Sp1, c-Krox, other GC-box binding proteins, and more recently, BFCOL1, the mouse homologue of the human transcription factor htβ, have also been shown to bind to this region within the mouse Col1A2 promoter [Galera et al., 1996; Hasegawa et al., 1996, 1997].

Studies on the TGF-β-responsive elements in the human promoter, together with those in the highly homologous mouse proximal α2(I) procollagen promoter suggest that species-specific mechanisms operate in regulating the expression of this gene [Inagaki et al., 1994, 1995b; Chung et al., 1996; Rossi et al., 1988]. Species-specific mechanisms in the regulation of the α2(I) procollagen gene were also suggested in studies where a novel element, which was not present in the mouse Col1A2 promoter, was identified within the human promoter immediately downstream of the inverted CCAAT box between nucleotides −78 and −67 [Parker et al., 1992; Collins et al., 1997]. In type I collagen producing human embryonic lung WI-38 fibroblasts, a transcriptional activator binds to this element and probably functions cooperatively with the CCAAT box binding factors in the regulation of the human COL1A2 gene [Collins et al., 1997]. In SV40-transformed human embryonic lung WI-38 fibroblasts (SVWI-38), in which the gene is inactive, a second complex forms on this downstream element [Parker et al., 1992]. Our previous studies also suggested that this second factor functions as a transcriptional repressor and is responsible for the inactivation of the COL1A2 gene in SVWI-38 fibroblasts [Parker et al., 1990, 1992]. Since this novel element is able to bind both a transcriptional activator and a repressor, it has been termed the collagen modulating element (CME). In this study, the transcription factors in SVWI-38 and CT-1 nuclear extracts which bind to the GGAGG/CCAAT box (G/CBE) and CME were characterized.

MATERIALS AND METHODS

Cell Culture and Preparation of Nuclear Proteins

SVWI-38 and CT-1 cells are WI-38 human embryonic lung fibroblasts transformed by the SV40 virus and γ-radiation, respectively [Parker et al., 1986; Namba et al., 1980]. Type I collagen synthesis is not significantly affected in CT-1 fibroblasts, while the α2(I) procollagen gene is not expressed in SVWI-38 cells [Parker et al., 1989]. The latter cell line therefore synthesizes low amounts of an overmodified homotrimer of the α1(I) chain. Cells were cultured and nuclear extracts were prepared as previously described [de Haan et al., 1986; Collins et al., 1997].

Electrophoretic Mobility-Shift Assays

The *Sma* I/*Sph* I (−107 to +54) and *Sma* I/*Bst* NI (−107 to −50) fragments of the human α2(I) procollagen promoter were radiolabeled with the Klenow fragment of DNA polymerase and used as probes in electrophoretic mobility-shift assays (EMSA), which were performed as previously described [Parker et al., 1992]. The divalent cation chelators used in EMSA were prepared as previously described [Hooft van Huijsduijnen et al., 1987]. To study the effects of ionic strength on DNA–protein complex formation, NaCl was added to the final concentrations indicated in Figure 3.

For the temperature sensitivity assays, aliquots of nuclear extracts were incubated at the desired temperature (see Fig. 4) for 5 min, the samples placed on ice and 4 µg of nuclear extract was assayed by EMSA.

Analytical Gel Filtration Chromatography

A total of 5–18 mg of crude nuclear extract in a final volume of 2 ml of chromatography buffer (50 mM Tris–Cl (pH 7.9); 0.1 mM EDTA; 0.1 M KCl; 0.5 mM DTT; 0.5 mM PMSF; 1 µg/ml leupeptin; and 1 µg/ml pepstatin A) containing 30% glycerol and 0.1% Nonidet P-40 was applied to a Sephacryl S-300 column (1.6 cm × 100 cm) and eluted with chromatography buffer containing 0.1 M KCl and 20% glycerol at a flow rate of 6 ml/h. One-ml fractions were collected and 2–4 µg of protein from each fraction was used in EMSAs containing 2 µg poly(dI − dC)/poly(dI − dC) per reaction as described above. The columns were calibrated using protein standards of known Stokes radii, consisting of 2 mg thyroglobin, 8.50 nm; 2 mg ferritin, 6.10 nm; 4 mg catalase, 5.22 nm; 2 mg aldolase, 4.81 nm (Pharmacia, Uppsala, Sweden); 4 mg bovine serum albumin (BSA), 3.55 nm (Miles-Seravac, Cape Town, S. Africa) and 7 mg ovalbumin, 3.05 nm (Sigma Chemical Co., St. Louis, MO), in 0.5 to 1 ml of chromatography buffer containing 30% glycerol. The elution volumes (V_e) of the protein standards were monitored by absorbance at 280 nm (LKB 2158 Uvicord SD UV-monitor). The void (V_o) and total (V_t) volumes of the columns were determined using 0.2 mg blue dextran 2000 (Pharmacia) and 2 mg thymidine (Sigma), respectively [Kim and Sheffery, 1990]. A calibration curve was prepared by plotting the log of the Stokes radius versus k_{av}, where $k_{av} = (V_e - V_o)/(V_t - V_o)$ [Siegel and Monty, 1966].

Glycerol Gradient Centrifugation

Linear 15–40% glycerol gradients (4.6 ml) in 20 mM Hepes (pH 7.9), 0.1 M KCl, 0.2 mM EDTA, 0.5 mM PMSF, 0.5 mM DTT, 1 µg/ml leupeptin and 1 µg/ml pepstatin A were prepared at 4°C as previously described [Martin and Ames, 1961]. Parallel gradients were calibrated with 0.5–0.75 mg each of catalase (11.3 S), BSA (4.4 S), ovalbumin (3.7 S), carbonic anhydrase (3.2 S) and lysozyme (1.9 S). A calibration curve was prepared by plotting percentage glycerol in the fractions versus sedimentation coefficient (S). Approximately 0.5 mg of crude nuclear extracts and 0.5 mg of ovalbumin (internal standard) in a final volume of 100 µl was layered on top of the gradients and centrifuged for 16 h at 53 000 rpm and 4°C in a Beckman SW65Ti rotor with slow deceleration. Three-drop fractions were collected from the bottom of each tube and the DNA-binding activity in 10 µl of each fraction was quantitated using the EMSA as described above.

UV Cross-Linking

Oligonucleotides were synthesized on a Beckman system 1000A DNA synthesizer. A 37-mer oligonucleotide containing the G/CBE and CME was annealed with a complementary primer as previously described [Collins et al., 1997]. Bromodeoxyuridine (Br-du) and ^{32}P-deoxycytidine monophosphate (^{32}P-dCMP) were incorporated into the annealed oligonucleotides using the Klenow fragment of DNA polymerase. The modified DNA probe (10^5 cpm) was incubated with 80 µg of crude nuclear extract 20 µg of poly(dI−dC)/poly(dI−dC) in a final volume of 80 µl and electrophoresed on nondenaturing polyacrylamide gels as described for EMSA. The wet gels were irradiated for 30 min with 305 nm UV light at 7,000 µW/cm^2 (Fotodyne transilluminator) [Chodosh et al., 1986; Wu et al., 1987]. The UV cross-linked complexes were visualized by autoradiography, excised from the gel, chopped into slices, and eluted with 1 vol buffer containing 50 mM Tris (pH 8.0), 0.1 mM EDTA, 150 mM NaCl, 0.1% sodium dodecyl sulafte (SDS), 5 mM DTT, 1 µg/ml leupeptin and 1 µg/ml pepstatin A at room temperature for 3 h with gentle shaking. The supernatants were transferred to fresh tubes, the gel slices were washed and the pooled supernatants clarified by centrifugation. Carrier BSA was added to the supernatants to a final concentration of 0.1 mg/ml, followed by 5 vol of cold (−20°C)

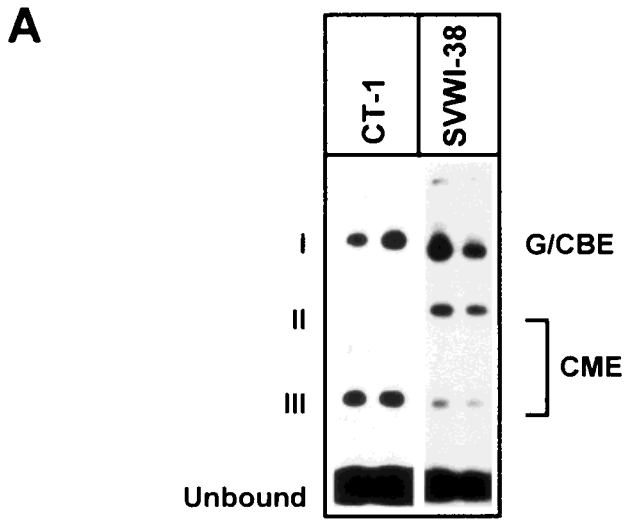
acetone, and the proteins were precipitated on ice for 30 min [Hager and Burgess, 1980]. The precipitates were washed with 100% ethanol, dried, and resolved by SDS-polyacrylamide gel electrophoresis (PAGE)[Laemmli, 1970], the gels dried, and the bands visualized by autoradiography.

RESULTS

Effects of Divalent Cations on Complex Formation

In previous studies, we identified a novel regulatory element, the CME, which is located immediately downstream of the inverted CCAAT box, within the human α2(I) procollagen promoter between nucleotides −78 and −67 [Parker et al., 1992; Collins et al., 1997]. Two distinct DNA–protein complexes (complexes II and III) form on the CME with crude nuclear extracts prepared from SV40-transformed WI-38 human embryonic lung fibroblasts (SVWI-38), which do not express the COL1A2 gene (Fig. 1), while only one of these complexes (complex III) is present in nuclear extracts prepared from CT-1 and WI-38 collagen producing fibroblasts. A single distinct complex (complex I), however, forms on the inverted CCAAT box (the G/CBE) with nuclear extracts prepared from all the cell lines tested [Parker et al., 1992; Collins et al., 1997].

Since many transcription factors require the presence of metal ions, such as zinc for maximum DNA binding activity, nuclear extracts were assayed in the presence of various metal ion chelators to determine whether the G/CBE and CME binding factors were metalloproteins. The divalent cation chelator ortho-phenanthroline (1,10-phenanthroline) (OP) inhibited the formation of complexes II and III at low concen-



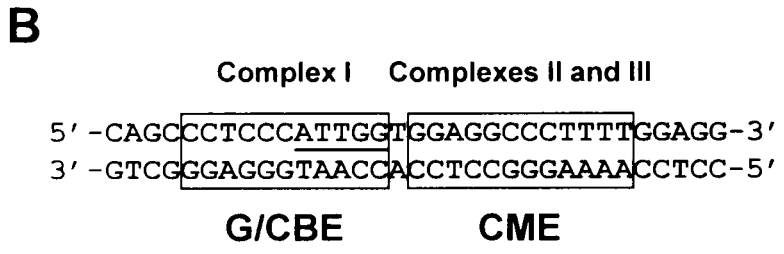


Fig. 1. The G/CBE and CME DNA-binding elements within the proximal human α2(I) procollagen promoter. **A:** A fragment of the proximal human α2(I) procollagen promoter (−107 to −50) was end-labeled with ^{32}P-dCTP and used to detect DNA–protein complex formation using crude CT-1 and SVWI-38 nuclear extracts as described under Materials and Methods. The positions of complexes I, II, and III, as well as the free probe, are indicated. Complex I forms with the G/CBE, while complexes II and III form with the CME. **B:** Summary of the transcription factor binding sites on the CCAAT box region of the human α2(I) procollagen promoter. The G/CBE containing an inverted CCAAT box (underlined) and CME are *boxed*. The G/CBE also corresponds to the CBF-binding site within the mouse promoter.

trations, while they were only mildly inhibited by EDTA and even less by 8-hydroxyquinoline (Fig. 2). Complex I formation was not abrogated in the presence of EDTA and 8-hydroxyquinoline, while ortho-phenanthroline showed inhibition at higher concentrations only (8 mM). A possible explanation for the reduced sensitivity of complex formation in the presence of EDTA and 8-hydroxyquinoline is that, unlike ortho-phenanthroline, EDTA, and 8-hydroxyquinoline are unable to chelate metal ions that are already complexed internally with the active centres of proteins [Hooft van Huijsduijnen et al., 1987]. These results, nevertheless, demonstrate that the CME binding proteins require divalent cations for DNA binding.

Since similar concentrations of ortho-phenanthroline have previously been used to demonstrate the presence of divalent cations in other DNA-binding proteins [Hooft van Huijsduijnen et al., 1987], the ability of Zn^{2+}, Co^{2+}, Cu^{2+} and Mg^{2+} to restore complex formation in the presence of 4 mM ortho-phenanthroline was examined. Excess Zn^{2+} restored the formation of complexes II and III, suggesting that the CME binding proteins require zinc (Fig. 3). Other divalent cations, such as Co^{2+}, Cu^{2+} and Mg^{2+}, were unable to restore the formation of these complexes. Excess Co^{2+} and Cu^{2+} and higher concentrations of Zn^{2+} and Mg^{2+} inhibited the formation of complex I. This is probably due to the phenomenon of "metal poisoning" by excess metal ions.

Effects of Ionic Strength and Temperature on Complex Formation

Complex I formation was much more resistant than complexes II and III to variations in ionic strength when SVWI-38 nuclear extracts were assayed for DNA-binding activity in the presence of increasing concentrations of NaCl (Fig. 4A). Although there was a significant decrease in complex I formation in the presence of low concentrations of NaCl, complex formation was totally abolished at 0.8 M NaCl and higher. The CME binding proteins (complexes II and III), on the other hand, were much more sensitive to NaCl, since NaCl concentrations greater than 0.2 M completely abolished formation of these complexes. There was no significant difference in the effects of ionic strength between complex II and III formation.

To test the heat sensitivity of these proteins, aliquots of SVWI-38 nuclear extracts were heated at the indicated temperatures (Fig. 4B) for 5 min before assaying for DNA-binding activ-

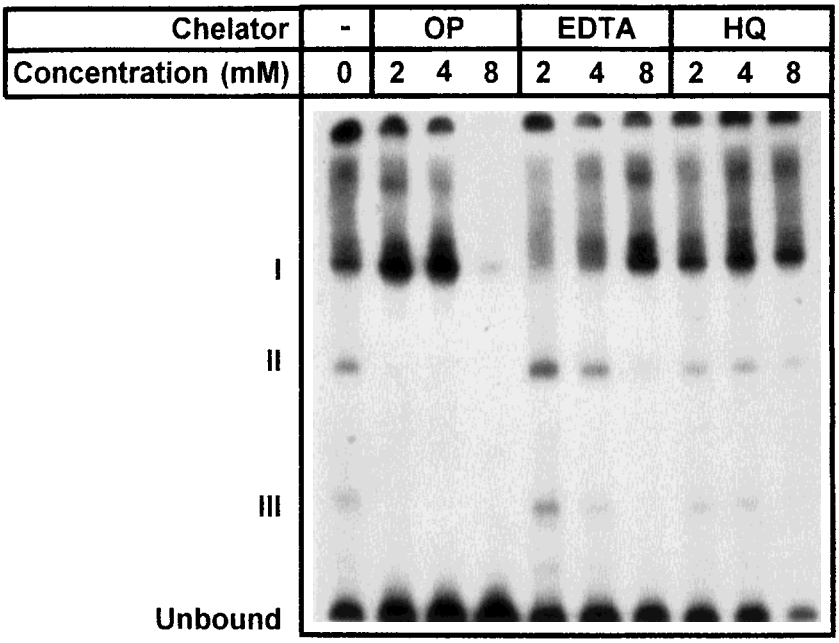


Fig. 2. Dependence of DNA–protein complex formation on divalent metal ions. SVWI-38 nuclear extracts were assayed for DNA binding activity using the *Sma* I/*Sph* I (−107 to +54) fragment of the human α2(I) procollagen promoter in EMSA, except that the reaction buffer was supplemented with the indicated final concentration of metal ion chelators. The samples were analyzed on 5% nondenaturing polyacrylamide gels and the gels dried and exposed to X–ray film for at least 16 h. The positions of the DNA–protein complexes I, II and III, as well as the free probe, are indicated. OP, ortho-phenanthroline; EDTA, ethylenediaminetetra-acetic acid; HQ, 8-hydroxyquinoline.

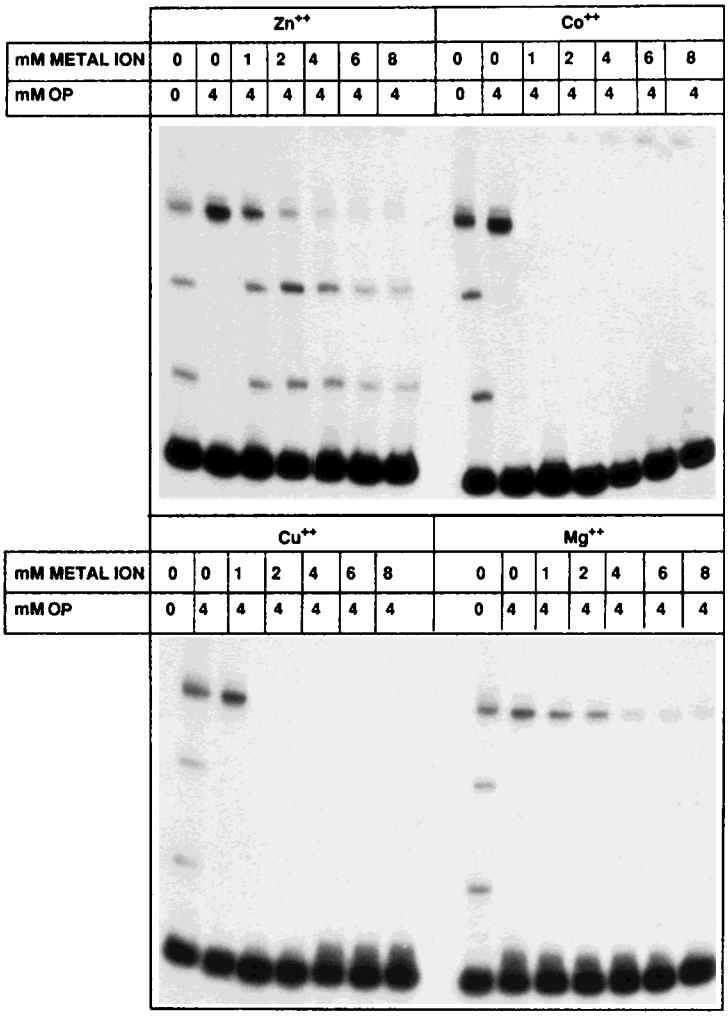


Fig. 3. Zn^{2+} requirement for DNA–protein complex formation. SVWI-38 nuclear extracts were assayed for DNA binding activity using the *Sma* I/*Sph* I (−107 to +54) fragment of the human α2(I) procollagen promoter in EMSA, except that the reaction buffer was supplemented with 4 mM ortho-phenanthroline and the indicated final concentrations of divalent cations. The samples were analyzed on 5% nondenaturing polyacrylamide gels, the gels were dried and exposed to X-ray film for at least 16 h.

ity. The CME binding proteins were more heat labile than the G/CBE binding proteins. Complex I formation was stable after heating at 55°C, while complex III formation was totally abolished when nuclear extracts were heated at 50°C. Although the bulk of the complex II activity was lost when heated at 50°C, some residual activity still remained, even when heated at 75°C. Similar results were obtained when the effects of ionic strength and temperature on complex formation were determined for CT-1 nuclear extracts (data not shown). These results indicated that the G/CBE binding proteins were more resistant to both temperature and ionic strength than the CME binding proteins. These findings also serve to confirm that these were indeed different factors that bind to the G/CBE and CME.

Native Molecular Weights of G/CBE and the CME Binding Proteins

The effects of divalent cations, ionic strength, and temperature clearly demonstrated that the G/CBE and CME binding proteins have distinct properties. However, there were no significant differences between complexes II and III formation under these conditions. Analytical gel filtration and glycerol gradient sedimentation analysis were therefore performed to determine the Stokes radii and the sedimentation coefficients of the G/CBE and CME binding proteins in crude SVWI-38 nuclear extracts under nondena-

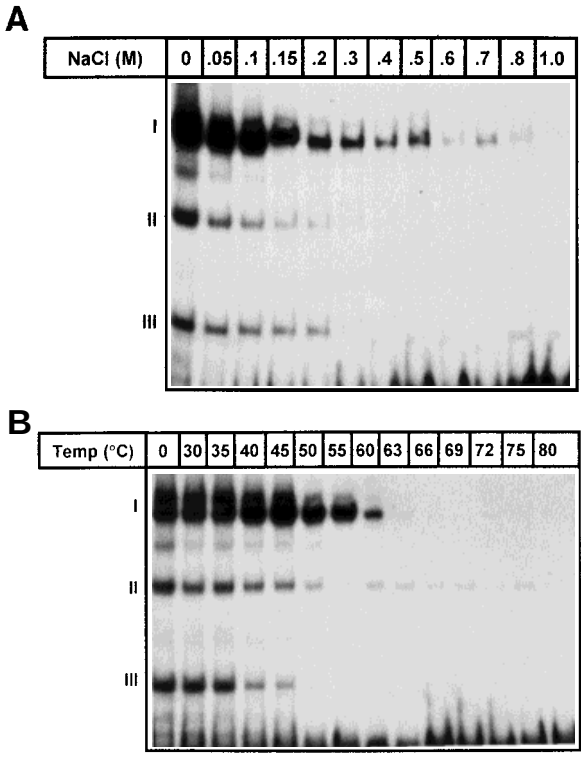


Fig. 4. Effects of ionic strength and temperature on the binding of transcription factors to the G/CBE and CME. **A:** SVWI-38 nuclear extracts were assayed for DNA-binding activity with the *Sma*I/*Sph*I (−107 to +54) fragment of the human α2(I) procollagen promoter as described in Figure 1, except that the reaction buffer was supplemented with the indicated final concentrations of NaCl. **B:** Aliquots of SVWI-38 nuclear extracts were incubated at the indicated temperature for 5 minutes, cooled on ice and assayed for DNA binding activity with the *Sma* I/*Sph* I (−107 to +54) fragment of the human α2(I) procollagen promoter as described above. The samples were analyzed on 5% nondenaturing polyacrylamide gels, the gels were dried and exposed to X-ray film for at least 16 h. The positions of the DNA–protein complexes are indicated.

turing conditions. As expected, the larger G/CBE binding proteins (complex I) eluted first, followed by complex II and finally complex III proteins when crude nuclear extracts were fractionated on Sephacryl S-300 (Fig. 5A). Under nondenaturing conditions, complex I, II, and III proteins eluted from the column with Stokes radii of 5.22 ± 0.04 nm, 4.12 ± 0.03 nm and 3.15 ±0.01 nm, respectively (Fig. 5B). The Stokes radii of complex I, II, and III proteins were comparable to spherical proteins of 232, 107, and 50 kDa, respectively. The stokes radii of the CME binding proteins (complexes II and III) were not significantly altered when nuclear extracts were fractionated on a sephadex G-200 column in the presence of 0.8 M KCl, suggesting that these proteins were either monomers or that noncolvant protein–protein interactions were not disrupted under these conditions (data not shown). The noncovalent interactions of the α- and β-subunits of α-CP1, for example, has been shown to be stable under similar conditions [Kim and Sheffery, 1990]. The stokes radius of the G/CBE binding proteins (complex I) decreased from 5.22 nm to 3.67 nm in the presence of 0.8 M KCl, suggesting that the factor is multimeric (data not shown). Under these conditions, the various components of the factor probably eluted independently and G/CBE binding activity was detected only in those fractions where the components co-eluted.

As shown in Figure 6, the sedimentation coefficients ($S_{20,W}$) of the three complexes were 4.9 ± 0.4, 3.9 ± 0.1 and 3.1 ± 0.3 for the protein components of complexes I, II, and III, respectively (Fig. 6B). Using BSA as a standard, the G/CBE binding protein was calculated to have a native molecular weight of 80 kDa, while the CME binding proteins (complex II and III) had molecular weights of 57 and 40 kDa, respectively, as estimated from the sedimentation data alone [Martin and Ames, 1961]. Identical Stokes radii and sedimentation coefficients were obtained for the proteins common to SVWI-38 and CT-1 cells (data not shown).

As there was a discrepancy when the molecular weights of the protein components in the three complexes were calculated from the gel filtration and sedimentation data, the native molecular weights of the factors were determined from the Stokes radii and sedimentation coefficients using the following equation [Siegel and Monty, 1966]:

$$MW = 6\pi nNas/(1 - vp)$$

where MW is the molecular weight, π is 3.14, n is the viscosity of H_2O at 20°C, N is Avogadro's number, a and s are the calculated Stokes radii and sedimentation coefficients, respectively, and p is the density of H_2O at 20°C [Resnick et al., 1993]. If the partial specific volume, v, is assumed to be 0.725 cm^3/g [Kim and Sheffery, 1990], then the native molecular weights of complex I, II, and III proteins under nondenaturing conditions correspond to 105, 66, and 40 kDa, respectively.

UV CrossLinking Experiments

An independent determination of the molecular weights of the G/CBE and CME binding proteins was performed by substituting a double-stranded 37-mer oligonucleotide (containing the G/CBE and CME) with bromodeoxy-

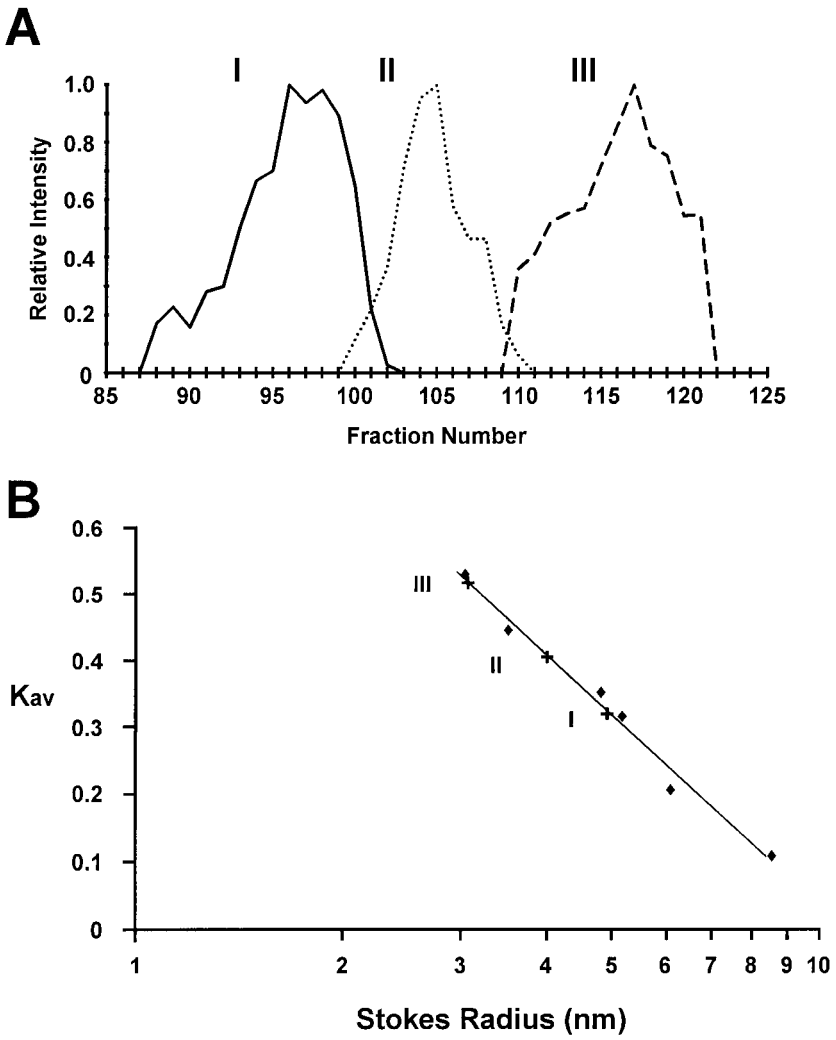


Fig. 5. Stokes radii of the G/CBE and CME binding proteins under nondenaturing conditions. **A:** Crude SVWI-38 nuclear extracts were applied to a Sephacryl S-300 column (1.6 cm X 100 cm) in chromatography buffer containing 30% glycerol and 0.1% Nonident P-40 and eluted with chromatography buffer containing 20% glycerol at a flow rate of 3 ml/cm^2/h, as described under Materials and Methods. One ml fractions were collected and assayed for DNA-binding activity with the *Sma*I/*Bst*NI (−107 to −50) fragment of the human α2(I) procollagen promoter as described in Figure 1. The autoradiograms were scanned densitometrically and the results for complexes I, II, and III were expressed as the relative intensity of each fraction. **B:** The gel filtration column was calibrated using protein standards (2 mg thyroglobin, 8.50 nm; 2 mg ferritin, 6.10 nm; 4 mg catalase, 5.22 nm; 2 mg aldolase, 4.81 nm; 4 mg BSA, 3.55 nm and/or 7 mg ovalbumin, 3.05 nm) and the log Stokes radius versus the k_{av} for each protein was plotted, where $k_{av}=(V_e-V_o)/(V_t-V_o)$ and V_e, V_o, and V_t are the elution, void and total volumes of the column, respectively. The Stokes radii of complexes I, II, and III were determined from the curve. The positions of the marker proteins and complexes are indicated by ◆ and +, respectively.

uridine and ^{32}P-deoxycytidine monophosphate before using it as a probe in UV cross-linking experiments as described under Materials and Methods. The replacement of thymidine with bromodeoxyuridine did not affect the ability of the oligonucleotides to form any of the DNA–protein complexes (data not shown).

Complex I comprised proteins with a molecular weight of 120 kDa in both SVWI-38 and CT-1 extracts (Fig. 7). Occasionally, two faint bands of 240 and 50 kDa were present. Bands of 78 and 69 kDa were cross-linked in complex II, while identical proteins within the range 30–46 kDa were identified for complex III in extracts prepared from both cell lines.

On average the size of the major bands generated by the UV cross-linking experiments were larger (12–16 kDa) than those calculated from the gel filtration and glycerol gradient sedimentation analysis. This increase is molecular

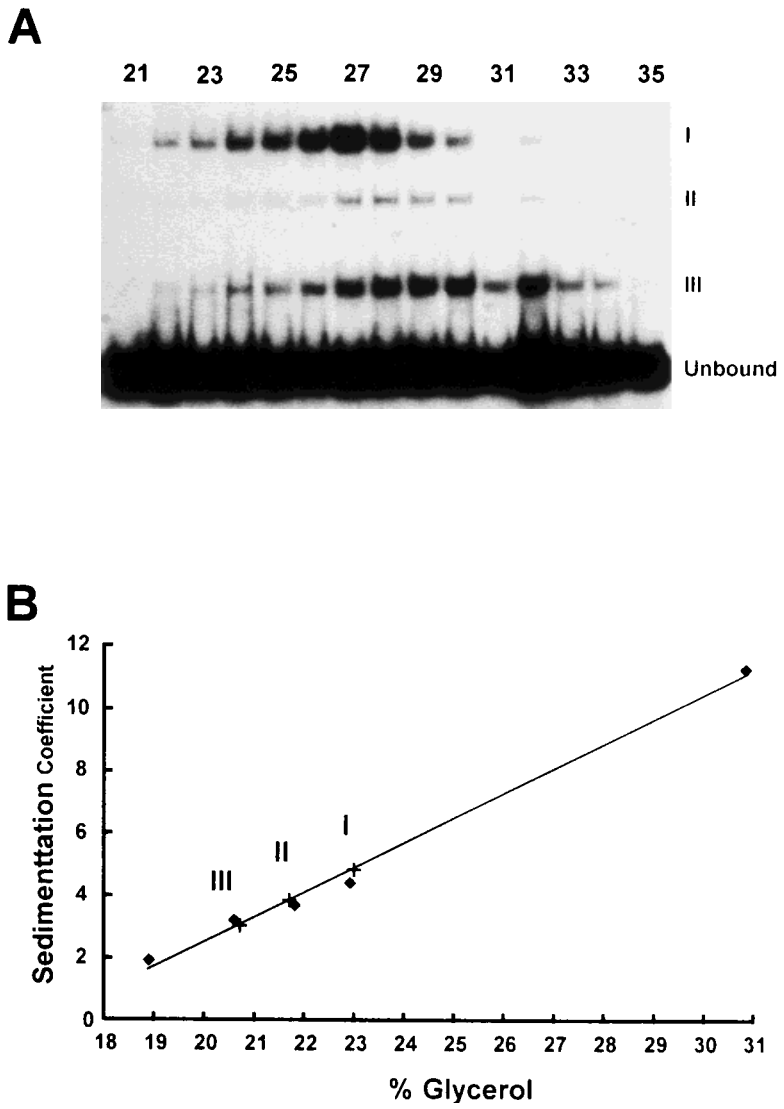


Fig. 6. Sedimentation coefficients of the G/CBE and CME binding proteins under nondenaturing conditions. **A:** SVWI-38 crude nuclear extracts (0.5 mg) and ovalbumin (0.5 mg) (internal standard) in a final volume of 100 µl were layered onto 15–40% linear glycerol gradients and centrifuged for 16 hours at 53,000 rpm at 4°C in a Beckman SW65 rotor. Three-drop fractions were collected from the bottom of each centrifuge tube and DNA-binding activity in 10 µl of each fraction was assayed using the *Sma* I/*Bst* NI (−107 to −50) fragment of the human α2(I) procollagen promoter. The positions of complexes I, II, and III are indicated. **B:** The percentage glycerol in each fraction was calculated and a glycerol gradient calibration curve prepared by plotting the percentage glycerol in the fractions where the standard proteins sedimented versus sedimentation coefficient. The sedimentation coefficients of the three complexes were determined from the curve. Parallel tubes containing catalase (11.3 S), BSA (4.4 S), ovalbumin (3.7 S), carbonic anhydrase (3.2 S), and lysozyme (1.9 S) were centrifuged at the same time. The positions of the standard proteins and complexes are indicated with ◆ and +, respectively.

weight could probably be due to the cross-linking with the oligonucleotide. Taking the contribution of the DNA fragments into account, the sizes of the bands in the UV cross-linking data were closer to the native molecular weights calculated from the gel filtration and glycerol gradient sedimentation analysis. The physicochemical characterization of these factors suggests that distinct unrelated transcription factors bind to the G/CBE and the adjacent downstream CME within the proximal human α2(I) procollagen promoter.

DISCUSSION

The physicochemical properties of the G/CBE (complex I) and CME (complexes II and III) binding proteins on the α2(I) procollagen promoter were investigated in order to determine

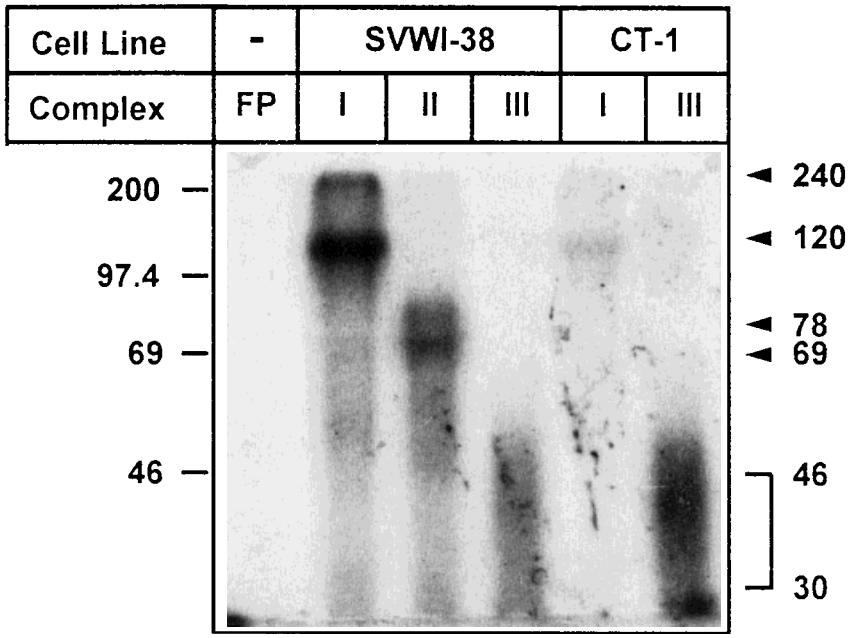


Fig. 7. Molecular weights of UV cross-linked DNA–protein complexes. SVWI-38 and CT-1 nuclear extracts were incubated with Br-dUTP and ^{32}P-dCTP-labeled probe and resolved on 5% nondenaturing polyacrylamide gels as described under Materials and Methods. The DNA–protein complexes were UV cross-linked in situ, visualized by autoradiography, the bands excised, the DNA–protein complexes eluted from the gel slices and resolved on 10% SDS-polyacrylamide gels. The gels were dried and exposed to X-ray film for at least 16 h. The molecular weight markers and cross-linked products are indicated in kDa. The lane labeled FP indicates the free probe.

whether these factors are related to previously characterized DNA-binding proteins.

The CME binding proteins (complexes II and III) are heat-labile factors and are sensitive to high ionic strengths. Binding to the CME was also abolished in the presence of ortho-phenanthroline, a chelator of protein-bound Zn^{2+} cations, and only Zn^{2+} was able to restore complex II and III formation in the presence of this chelator. DNA-binding proteins that require Zn^{2+} for binding activity include, amongst others, Sp1 [Kadonaga et al., 1987] and TFIIIA [Miller et al., 1985]. These proteins interact with DNA via their DNA-binding domains that contain one or more zinc fingers, and the data suggest that the DNA-binding motifs of the CME binding proteins could consist of such zinc fingers.

Under nondenaturing conditions, the CME binding proteins (complexes II and III) had Stokes radii of 4.12 and 3.15 nm, sedimentation coefficients of 3.9 and 3.2 S, and native molecular weights of 66 and 41 kDa, respectively. The Stokes radii of these factors did not significantly differ when fractionated on gel filtration columns in the presence of 0.8 M KCl, suggesting that the CME binding proteins were either monomers or that protein–protein interactions between the various components are stable under these denaturing conditions. Two bands of 78 and 69 kDa were, however, identified when complex II proteins were cross-linked to an oligonucleotide containing the CME. In addition, no SVWI-38 specific bands were identified during Southwestern blotting experiments suggesting that complex II is heteromultimeric (M. Collins and M.I. Parker, unpublished data). If the complex II protein(s) were a monomer or homomultimer then the subunit(s) would migrate as a single band which would be able to bind with the probe during Southwestern blotting experiments. If, on the other hand, the factor is heteromultimeric (and if the subunits are of different sizes) and if all the subunits were required for DNA-binding activity, the probe would not bind during these assays. Although the UV cross-linking and Southwestern blotting experiments implied that complex II is heteromultimeric, it is possible that the proteins did not renature correctly after electrophoresis during the Southwestern blotting experiments. This is unlikely, however, since complex III formation was not inhibited. It is also possible that the 69-kDa band produced during the UV cross-linking experiments is a specific degradation product of the 78-kDa protein, but

subsequent affinity purification has shown these two proteins to be integral constituents of complex II (A. Masemola and M.I. Parker, unpublished data).

The identification of cross-linked bands with molecular weights similar to the calculated native molecular weight of 41 kDa in SWVI-38 and CT-1 nuclear extracts suggested that the complex III proteins were monomers. Southwestern blot data also suggested that the complex III proteins were monomeric (M. Collins and M.I. Parker, unpublished data).

It was unclear from the data how the CME binding proteins are related to each other. Complex II and III proteins could contain a common DNA-binding domain or they may be two distinct transcription factors that bind to the CME. One example of the latter is the transcriptional activator, interferon regulatory factor-1 (IRF-1), and repressor, IRF-2, which bind to the same regulatory elements in the type I interferon (IFN) and IFN-inducible genes [Harada et al., 1989].

Characterization of the binding activity of the G/CBE binding proteins (complex I) revealed a less stringent requirement for divalent metal cations. The binding activity was fairly resistant to temperature and was abolished at around 63°C. The ubiquitous CCAAT box-binding factor, NF-Y, was originally identified in nuclear extracts of B-lymphoid cells and shown to bind to the Y-box of the major histocompatibility complex (MHC) class II gene, E_α [Dorn et al., 1987a; Dorn et al., 1987b; Hooft van Huijsduijnen et al., 1987]. NF-Y has subsequently been shown to be relatively heat resistant and displayed binding activity which persisted at 69°C [Hooft van Huijsduijnen et al., 1987] (Table I). Like the G/CBE binding proteins, the binding of NF-Y to its cognate DNA-binding motif was also fairly resistant to high ionic strengths, where a substantial degree of DNA–protein complex formation was detected in the presence of 0.5 M NaCl [Hooft van Huijsduijnen et al., 1987].

Under nondenaturing conditions, the G/CBE binding proteins had a Stokes radius and sedimentation coefficient of 5.22 ± 0.04 nm and 4.9 ± 0.4 S, respectively, a calculated native molecular weight of 105 kDa and a frictional ratio of 1.66. Another member of the family of CCAAT box-binding proteins, α-CP1, originally purified from murine erythroleukemia (MEL) cells, has a native molecular weight of 101 kDa, sedimentation coefficient of 4.3 S, a Stokes radius of 5.7 nm, and a frictional ratio of 1.78 [Kim and Sheffery, 1990] (Table I). NF-Y also has a similar sedimentation coefficient on glycerol gradients and the estimated molecular weight of the NF-Y/E_α oligo complex is 250–300 kDa [Hooft van Huijsduijnen et al., 1987]. The complex I protein–G/CBE oligonucleotide complex, had a calculated molecular weight of 299 kDa (M. Collins and M.I. Parker, unpublished data). The gel filtration data also showed that the G/CBE-binding protein is multicomponent, since the various subunits of the factor eluted independently on Sephadex G-200 in the presence of 0.8 M NaCl and as an intact complex in the absence of high salt concentrations on Sephacryl S-300. NF-Y, CBF and other members of this protein family are also heterotrimers [Hooft van Huijsduijnen et al., 1987; Hatamochi et al., 1988; Maity et al., 1992], and this observation is consistent with the proposed heterotrimeric structure of the G/CBE binding protein.

TABLE I. Comparison of the Physicochemical Properties of the G/CBE Binding Proteins (Complex I) with the Published Properties for NF-Y, α-CP1 and CBF

Parameter	Complex I	NF-Y	α-CP1	CBF
Apparent or native molecular weight	105 kDa	—	101 kDa [1]	113 kDa[2]
Stokes radius	5.2 nm	—	5.7 nm[1]	—
Frictional ratio	1.66	—	1.78[1]	—
Sedimentation coefficient	4.9 S	—	4.3 S[1]	—
DNA–protein complex molecular weight	299 kDa	250–300 kDa[3]	—	—
Heat lability[4]	63°C	69°C[3]	—	—
Ionic strength[4]	0.7 M NaCl	0.5 M NaCl[3]	—	—
Polypeptide composition	Multimeric	Heterotrimer	Heterotrimer	Heterotrimer

[1]Kim and Sheffery (1987).
[2]Maity et al. (1992).
[3]Hooft van Huijsduijnen et al. (1987).
[4]Temperature and ionic strength at which complex formation is abolished.

A major band of 120 kDa was identified when complex I protein(s) were UV cross-linked to a 37-mer oligonucleotide containing the G/CBE. Since the factor in this complex had a native molecular weight of 109 kDa, the 120 kDa band was probably produced when a single heterotypic complex was UV cross-linked to the oligonucleotide, with the cross-linked oligonucleotide probably causing the apparent increase in the size of the protein. The occasional minor 50 kDa band was probably a degradation product, a finding supported by the data of Hooft van Huijsduijnen et al. [1987]. The weak 240 kDa band suggests that the stoichiometry of the G/CBE binding protein in the DNA–protein complex may be more complex. Although the three polypeptides making up CBF associate with one another to form a functional factor with an apparent molecular weight of 113 kDa [Maity et al., 1992], a band with an apparent molecular weight of 170 kDa was produced when purified CBF was cross-linked with glutaraldehyde, suggesting that the structure of CBF may also be more complex [Maity and de Crombrugghe, 1992].

Taken together, the physicochemical properties of the G/CBE binding proteins strongly suggest that this factor belongs to the family of previously identified heterotrimeric CCAAT box-binding proteins, which include, among others, CBF [Hatamochi et al., 1988; Maity et al., 1992], NF-Y [Hooft van Huijsduijnen et al., 1987], α-CP1 [Kim and Sheffery, 1990] and CP-1 [Chodosh et al., 1988]. Supershift experiments with antibody to CBF, as well as DNA binding experiments, add additional support to this structural relationship [Collins et al., 1997]. Since the CME binding proteins are not, to our knowledge, related to any previously characterized DNA-binding factor, they are probably novel factors involved in the expression of the human α2(I) procollagen gene and have distinct biochemical and DNA-binding properties as compared with the G/CBE binding proteins. The data also demonstrate that, unlike the equivalent region of the mouse promoter where only one transcription factor, CBF, binds to the inverted CCAAT box, two distinct unrelated families of transcription factors bind to the CCAAT box region of the human α2(I) procollagen promoter. This implies that important species-specific mechanisms operate in the expression of the human and mouse genes.

REFERENCES

Adams SL (1989): Collagen gene expression. Am J Respir Cell Mol Biol 1:161–168.

Boast S, Su M, Ramirez F, Sanchez M, Avvedimento EV (1990): Functional analysis of cis-acting DNA sequences controlling transcription of the human type I collagen genes. J Biol Chem 265:13351–13356.

Bornstein P, Sage H (1980): Structurally distinct collagen types. Annu Rev Biochem 49:957–1003.

Bornstein P, Sage H (1989): Regulation of collagen gene expression. Prog Nucleic Acid Res Mol Biol 37:67–106.

Chodosh LA, Carthew RW, Sharp PA (1986): A single polypeptide possesses the binding and transcription activities of the adenovirus major late transcription factor. Mol Cell Biol 6:4723–4733.

Chodosh LA, Baldwin AS, Carthew RW, Sharp PA (1988): Human CCAAT–binding proteins have heterologous subunits. Cell 53:11–24.

Chung K, Agarwal A, Uitto J, Mauviel A (1996): An AP-1 binding sequence is essential for regulation of the human α2(I) collagen (COL1A2) promoter activity by transforming growth factor-β. J Biol Chem 271:3272–3278.

Collins M, Leaner VD, Madikizela M, Parker MI (1997): Regulation of the human α2(I) procollagen gene by sequences adjacent to the CCAAT box. Biochem J 322:199–206.

Damsky C, Sutherland A, Fisher S (1993): Extracellular matrix 5: Adhesive interactions in early mammalian embryogenesis, implantation, and placentation. FASEB J 7:1320–1329.

de Crombrugghe B, Vuorio T, Karsenty G, Maity S, Rutheshouser EC, Goldberg H (1991): Transcriptional control mechanisms for the expression of type I collagen genes. Ann Rheum Dis 50:872–876.

de Haan JB, Gevers W, Parker MI (1986): Effects of sodium butyrate on the synthesis and methylation of DNA in normal cells and their transformed counterparts. Cancer Res 46:713–716.

Dorn A, Bollekens J, Staub A, Benoist C, Mathius D (1987a): A multiplicity of CCAAT box-binding proteins. Cell 50: 863–872.

Dorn A, Durand B, Marfing C, Le Meur M, Benoist C, Mathius D (1987b): Conserved major histocampatibility complex class II boxes—X and Y—are transcriptional control elements and specifically bind nuclear proteins. Proc Natl Acad Sci USA 84:6249–6253.

Galera P, Park R, Ducy P, Mattei M, Karsenty G (1996): c-Krox binds to several sites in the promoter of both mouse type I collagen genes. J Biol Chem 271:21331–21339.

Greenwel P, Inagaki Y, Hu W, Walsh M, Ramirez F (1997): Sp1 is required for the early response of α2(I) collagen to transforming growth factor-β1. J Biol Chem 272:19738–19745.

Hager DA, Burgess RR (1980): Elution of proteins from sodium dodecyl sulphate-polyacrylamide gels, removal of sodium dodecyl sulfate, and renaturation of enzymatic activity: Results with sigma subunits of *Escherichia coli* RNA polymerase, wheat germ DNA topoisomerase, and other enzymes. Anal Biochem 109:76–86.

Harada H, Fujita T, Miyamoto M, Kimura Y, Maruyama M, Furia A, Miyata T, Taniguchi T (1989): Structurally similar but functionally distinct factors, IRF-1 and IRF-2, bind to the same regulatory elements of IFN and IFN-inducible genes. Cell 58:729–739.

Hasegawa T, Zhou X, Garrett LA, Ruteshouser EC, Maity SN, de Crombrugghe B (1996): Evidence for three major transcription activation elements in the proximal mouse proα2(I) collagen promoter. Nucleic Acids Res 24:3253–3260.

Hasegawa T, Takeuchi A, Miyaishi O, Isobe K, de Crombrugghe B (1997): Cloning and characterization of a transcription factor that binds to the proximal promoters of the mouse type I collagen genes. J Biol Chem 272:4915–4923.

Hatamochi A, Golumbek PT, 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.

Hooft van Huijsduijnen RAM, Bollekens J, Dorn A, Benoist C, Mathis D (1987): Properties of a CCAAT box-binding protein. Nucleic Acids Res 15:7265–7282.

Huerre C, Junien C, Weil D, Chu M, Morabito M, van Cong N, Myers JC, Foubert C, Gross M, Prockop DJ, Boue A, Kaplan J, de la Chapelle A, Ramirez F (1982): Human type I procollagen genes are located on different chromosomes. Proc Natl Acad Sci USA 79:6627–6630.

Ihn H, Ohnishi K, Tamaki T, LeRoy EC, Trojanowska M (1996): Transcriptional regulation of the human α2(I) collagen gene. J Biol Chem 271:26717–26723.

Inagaki Y, Truter S, Ramirez F (1994): Transforming growth factor-beta stimulates alpha 2(I) collagen gene expression through a cis-acting element that contains an Sp1-binding site. J Biol Chem 269:14828–14834.

Inagaki Y, Truter S, Greenwel P, Rojkind M, Unoura R, Kobayashi K, Ramirez F (1995a): Regulation of the alpha 2(I) collagen gene transcription in fat-storing cells derived from a cirrhotic liver. Hepatology 22:573–579.

Inagaki Y, Truter S, Tanaka S, Di Liberto M, Ramirez F (1995b): Overlapping pathways mediate the opposing actions of tumor necrosis factor-α and transforming growth factor-β on α2(I) collagen gene transcription. J Biol Chem 270:3353–3358.

Kadonaga JT, Carner KR, Masiarz FR, Tjian R (1987): Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell 51:1079–1090.

Karsenty G, de Crombrugghe B (1990): Two different negative and one positive regulatory factors interact with a short promoter segment of the alpha 1 (I) collagen gene. J Biol Chem 265:9934–9942.

Karsenty G, Ravazzolo R, de Crombrugghe B (1991): Purification and functional characterization of a DNA-binding protein that interacts with a negative element in the mouse alpha 1(I) collagen promoter. J Biol Chem 266: 24842–24848.

Karsenty G, de Crombrugghe B (1991): Conservation of binding sites for regulatory factors in the coordinately expressed alpha 1 (I) and alpha 2 (I) collagen promoters. Biochem Biophys Res Commun 177:538–544.

Kim CG, Sheffery M (1990): physical characterization of the purified CCAAT transcription factor, α-CP1. J Cell Biol 265:13362–13369.

Laemmli UK (1970): Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685.

Maity SN, de Crombrugghe B (1992): Biochemical analysis of the B subunit of the heteromeric CCAAT-binding factor. A DNA-binding domain and a subunit interaction domain are specified by two separate segments. J Biol Chem 267:8286–8292.

Maity SN, Sinha S, Ruteshouser EC, de Crombrugghe B (1992): Three different polypeptides are necessary for DNA binding of the mammalian heteromeric CCAAT binding factor. J Biol Chem 267:16574–16580.

Majack RA, Bornstein P (1984): Heparin and related glycosaminoglycans modulate the secretory phenotype of vascular smooth muscle cells. J Cell Biol 99:1688–1695.

Martin RG, Ames BN (1961): A method for determining the sedimentation behavior of enzymes: Application to protein mixtures. J Biol Chem 236:1372–1379.

Miller J, McLachlan AD, Klug, A (1985): Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. EMBO J 4:1609–1614.

Namba M, Nishitani K, Kimoto T (1980): Characteristics of WI-38 cells (WI-38 CT-1) transformed by treatment with Co-60 gamma rays. Gann 71:300–307.

Parker MI, de Haan JB and Gevers W (1986): DNA hypermethylation in sodium butyrate-treated WI-38 fibroblasts. J Biol Chem 261:2786–2790.

Parker MI, Smith AA, Gevers W (1989): Absence of alpha 2(1) procollagen synthesis in a clone of SV40-transformed WI-38 human fibroblasts. J Biol Chem 264:7147–7152.

Parker MI, Smith AA, Boast S, Su M, Ramirez F (1990): Regulation of the human α2(I) procollagen gene by a trans-acting factor. Ann NY Acad Sci 580:451–453.

Parker MI, Smith AA, Mundell K, Collins M, Boast S, Ramirez F (1992): The abolition of collagen gene expression in SV40-transformed fibroblasts is associated with trans-acting factor switching. Nucleic Acids Res 20:5825–5830.

Resnick D, Freedman NJ, Xu S, Krieger M (1993): Secreted extracellular domains of macrophage scavenger receptors form elongated trimers which specifically bind crocidolite asbestos. J Biol Chem 268:3538–3545.

Rossi P, Karsenty G, Roberts AB, Roche NS, Sporn MB, de Crombrugghe B (1988): A nuclear factor 1 binding site mediates the transcriptional activation of a type I collagen promoter by transforming growth factor-beta. Cell 52:405–414.

Siegel LM, Monty KJ (1966): Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. Application to crude preparations of sulphite and hydroxylamine reductases. Biochim Biophys Acta 112: 346–362.

Slack JL, Liska DJ, Bornstein P (1993): Regulation of expression of the type I collagen genes. Am J Med Genet 45:140–151.

Tamaki T, Ohnishi K, Hartl C, LeRoy EC, 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.

Wu C, Wilson S, Walker B, Dawid I, Paisley T, Zimarino V, Ueda H (1987): Purification and properties of *Drosophila* heat shock activator protein. Science 238:1247–1253.