Regulation of the human α2(1) procollagen gene by sequences adjacent to the CCAAT box

Malcolm COLLINS, Virna D. LEANER, Mziwandile MADIKIZELA and M. Iqbal PARKER*

Department of Medical Biochemistry, University of Cape Town Medical School, Observatory, 7925 Cape Town, South Africa

The human, rat, mouse and chicken α2(I) procollagen promoters analysed to date all contain an inverted CCAAT box at -80. In this study we have examined the binding of nuclear proteins to the proximal promoter of the human α2(I) procollagen gene, where an inverted CCAAT box is flanked by a downstream GGAGG sequence and its inverted counterpart (CCTCC) on the upstream end. Each of the GGAGG sequences is separated from the inverted CCAAT box by a single pyrimidine nucleotide (5′-CCTCCC**ATTGG**TGGAGGCCCTTTT-3′). Electrophoretic mobility-shift assays (EMSAs) revealed that two distinct DNA–protein complexes formed on this DNA sequence. Methylation interference analysis and *in vitro* mutagenesis studies revealed that the integrity of the sequence 5′-CCTCCC**ATTGG**-3′ (the GGAGG/CCAAT-binding element or G/CBE) was important for the binding of the CCAAT-binding factor (CBF) (complex I). Competition studies showed that complex formation on the human G/CBE could be competed by mouse CBE and nuclear factor-Y (NF-Y) oligonucleotides, suggesting that mouse CBE and human G/CBE-binding proteins belong to the same family of CCAAT box binding proteins. Furthermore, antibodies to mouse CBF specifically supershifted the G/CBE complex (complex I) in EMSAs. The downstream GGAGG and 3′-flanking sequences (5′-GGAGGCCCTTTT-3′) or collagen modulating element (CME), however, were important for the formation of a novel DNA–protein complex (complex III). The formation of this complex was not competed out by CBE or NF-Y oligonucleotides, nor was DNA–protein complex formation affected by the anti-CBF antibody. Functional analysis of G/CBE and CME elements subjected to mutagenesis, using promoter–chloroamphenicol acetyl transferase constructs in transient transfection assays, showed that both these elements were essential for activity of the human promoter. These experiments identified a novel regulatory element in the human α2(1) procollagen gene which is not present in the rodent gene.

INTRODUCTION

Type I collagen, a heterotrimer of two α1 and one α2 chains, is the major fibrillar collagen in bone, skin, tendons and ligaments [1]. The levels of this protein are carefully regulated during embryogenesis, development and in the adult organism [2]. The steady-state levels of the two different α-chains are also maintained at a stoichiometry of 2:1 [3]. Furthermore, type I collagen levels are altered during pathological conditions, such as wound healing, inflammatory conditions, cancer, arthritis and fibrosis [4]. The expression of the two type I collagen genes, therefore, has to be tightly and coordinately regulated in a cell- and tissue-specific manner during normal development of an organism and during pathological conditions. To achieve this, the spatial and/or temporal patterns of type I collagen gene expression are modulated by complex mechanisms, such as chromatin conformation [5], DNA methylation status [6–11], transcriptional [2,12,13] and post-transcriptional control [2,14].

Previous studies have shown that collagen gene expression is regulated primarily at the level of transcription. Several *cis*-acting elements involved in modulating the expression of the procollagen genes have been identified within the mouse α2(I) procollagen promoter [15–20] and first intron [15,21,22]. Except for the brain, all the essential regulatory elements within the mouse α2(I) procollagen promoter required for the temporal and spatial expression of reporter genes in transgenic mice have been shown to be located within the minimal DNA sequences between -350 and $+54$. These regulatory elements include a TATA box (-30 to -25), an inverted CCAAT box (-84 to -80), a CAGA box (-250 to -247), a nuclear factor-1 (NF-1)-like site (-312 to -300) and an inhibitory factor-1 (IF-1) site between -165 and -155 [17,23]. The heterotrimeric CCAAT-binding factor (CBF) and CCAAT transcription factor/NF-1 (or a related protein) have been shown to bind to their respective recognition sequences to stimulate *in vitro* transcription of the gene in NIH3T3 nuclear extracts [19,20,24–26]. Although the factor(s) which bind to the CAGA sequence has not been identified, *in vitro* mutagenesis of this motif to AAAG or ATAG almost totally abolishes DNA-binding activity and promoter-driven expression of a reporter-gene construct [17]. Since the overall activity of the minimal promoter is less than that of the -2000 bp promoter, additional upstream enhancer sequences are also required for high-level expression of the mouse α2(I) procollagen gene [18].

In the human α2(I) gene, the proximal 350 bp fragment is also essential for the cell-type-specific expression of reporter constructs in cultured cells [27]. There is a high degree of sequence similarity (86 %) between the mouse and human proximal promoters and, except for the IF-1- and NF-1-like sites, which each contain a two-base mismatch, some elements such as the TATA, CAGA and inverted CCAAT boxes are all conserved within the human promoter [28].

A previous study identified numerous DNA–protein complexes when gamma-radiation-transformed human embryonic lung WI-38 fibroblast (CT-1) nuclear extracts were assayed with the 350 bp human proximal promoter fragment [29]. Two major

Abbreviations used: IF-1, inhibitory factor-1; NF-1, nuclear factor-1; CAT, chloramphenicol acetyltransferase; CBF, CCAAT-binding factor; G/CBE, GGAGG/CCAAT-binding element; CME, collagen modulating element; NF-Y, nuclear factor-Y; C/EBP, CCAAT enhancer binding protein; EMSA, electrophoretic mobility-shift assay; TSP-1, thrombospondin-1.

* To whom correspondence should be addressed.

DNA–protein complexes, which form on the −107 to −50 bp promoter fragment containing an inverted CCAAT box, have been identified (complexes I and III). When using an equivalent fragment of the mouse promoter, however, only one DNA–protein complex, corresponding to complex I, was observed. The present study shows that mutation of sequences upstream of the inverted CCAAT box in the human α2(1) collagen promoter abolished complex I formation, whereas mutagenesis of sequences downstream of the CCAAT box abolished complex III formation. Mutation of the CCAAT box itself, however, resulted in only partial loss of both complex I and III formation. Moreover, the novel collagen modulating element (GGAGGC-CCTTTT) (CME)-binding proteins appear to act co-operatively with the G/CBE-binding proteins (G/CBE is the GGAGG/CCAAT-binding element) to regulate the human α2(I) procollagen promoter, since the changes in DNA-binding activity were reflected in the changes in the activity of promoter–chloramphenicol acetyl transferase (CAT) constructs in transient transfection assays. This study suggests that the GGAGG sequences flanking the human α2(I) CCAAT box are crucial for promoter activity. It is also clear that there is a major difference in the DNA-binding pattern between the mouse and human α2(I) procollagen genes, which would have important implications in cross-species transgenic studies.

EXPERIMENTAL

Cell culture and transient transfection

CT-1 cells are WI-38 human embryonic lung fibroblasts transformed by gamma-radiation [30], in which collagen synthesis is not significantly affected. Cells were cultured as previously described [31] and transiently transfected using the calcium phosphate–DNA precipitation method [32]. Cells were co-transfected with CMV-βGAL (the bacterial β-galactosidase gene driven by the cytomegalovirus promoter) to control for variation in transfection efficiency. CAT activity was measured using the protocol of Seed and Sheen [33]. β-Galactosidase activity was measured using *O*-nitrophenyl-β-D-galactopyranoside as substrate [34].

Site-directed mutagenesis

Site-directed mutagenesis was performed as described by Sayers et al. [35] using the Amersham oligonucleotide-directed *in vitro* mutagenesis system. The *Pst*I–*Sph*I fragment (−343 to +54) of the human α2(I) promoter was subcloned into M13mp18 and mutagenesis was performed using the oligonucleotides indicated in Figure 4. Recombinant plaques were subjected to sequence analysis and recloned into the promoterless p8CAT vector for CAT assays.

Preparation of nuclear proteins

Nuclear proteins were isolated using the method of Dignam et al. [36], except that 1 μg/μl each of leupeptin and pepstatin A was added to all the buffers. The protein concentration and DNA-binding activity of each preparation were determined by the Bradford method [37] and the electrophoretic mobility shift assay (EMSA) [29] respectively.

EMSA

Oligonucleotides were synthesized on a Beckman model 1000A DNA synthesizer. The double-stranded oligonucleotides were prepared by denaturing complementary single-stranded oligonucleotides at 90 °C for 5 min and allowing them to anneal at 37 °C for 60 min before cooling to room temperature. Oligonucleotides were end-labelled with [γ-^{32}P]ATP and polynucleotide kinase, as recommended by the manufacturers. DNA–protein complexes were identified using double-stranded oligonucleotides or the *Sma*I–*Bst*NI fragment (−107 to −50) of the human α2(I) procollagen promoter as probes in the EMSA, which was performed as previously described [29]. In the competition assays, unlabelled double-stranded competing oligonucleotides were incubated with the nuclear proteins before the addition of radioactive probe. In the antibody supershift assay, polyclonal rabbit anti-(mouse CBF-B) IgG [38] was added after DNA–protein complex formation and left on ice for a further 30 min before analysis on non-denaturing polyacrylamide gels.

Methylation interference assay

DNA was end-labelled at the 5′-terminus of either the coding or non-coding strand using [γ-^{32}P]ATP and polynucleotide kinase, followed by partial methylation with dimethyl sulphate. The modified DNA probe (2.5×10^5 c.p.m.) was incubated with 80 μg of crude nuclear extract and 20 μg of poly(dI-dC)·poly(dI-dC) in a final volume of 80 μl and electrophoresed on non-denaturing polyacrylamide gels as described for EMSA. The resolved complexes and free probe were electrophoretically transferred from the gel onto DEAE membranes at 500 mA for 1 h at room temperature in TBE (90 mM Tris, 90 mM boric acid, 2.5 mM EDTA, pH 8.0). The DEAE membranes were sealed in plastic bags and the complexes visualized by autoradiography (1 h at 4 °C). The complexed and free probe bands were excised, placed in microcentrifuge tubes and eluted in 200 μl of 10 mM Tris/HC1, pH 8.0/1 mM EDTA/1 M NaC1/1 % (w/v) SDS for 60 min at 65 °C. The supernatants were transferred to separate tubes, the membranes washed in an equal volume of TE (10 mM Tris/HC1, pH 7.5/1 mM EDTA) and the supernatants pooled. The residual debris was removed by centrifugation and the supernatants transferred to fresh tubes. The DNA samples were precipitated by the addition of 2.5 vol. of ethanol, pelleted, dried, cleaved with piperidine, dissolved in formamide buffer and analysed on 12 % polyacrylamide/urea sequencing gels as previously described [39].

RESULTS

Analysis of DNA-binding sites

Previous studies have shown that incubation of the 59 bp *Sma*I–*Bst*NI fragment (−107 to −50) of the human α2(1) procollagen promoter with crude nuclear extracts, prepared from collagen-producing WI-38 fibroblasts and its gamma-radiation transformed counterpart (CT-1 fibroblasts), resulted in two distinct DNA–protein complexes (complexes I and III) in EMSAs [29]. The CT-1 cells are transformed, as measured by their growth in soft agar and production of tumours in nude mice, but their collagen synthetic capability is only slightly reduced (by 20 %) when compared with the parental WI-38 cell line.

Methylation interference assays were used to identify the purine bases which are crucial for the binding of the protein to the DNA [40]. Nuclear extracts prepared from the CT-1 cell line indicated that these two complexes bind to overlapping or adjacent DNA elements within the proximal promoter (Figure 1A). These results also suggested that the formation of complex I involved the inverted GGAGG sequence upstream of the inverted CCAAT box, the CCAAT box itself, and the downstream GGAGG sequence (−92 to −72; summarized in Figure

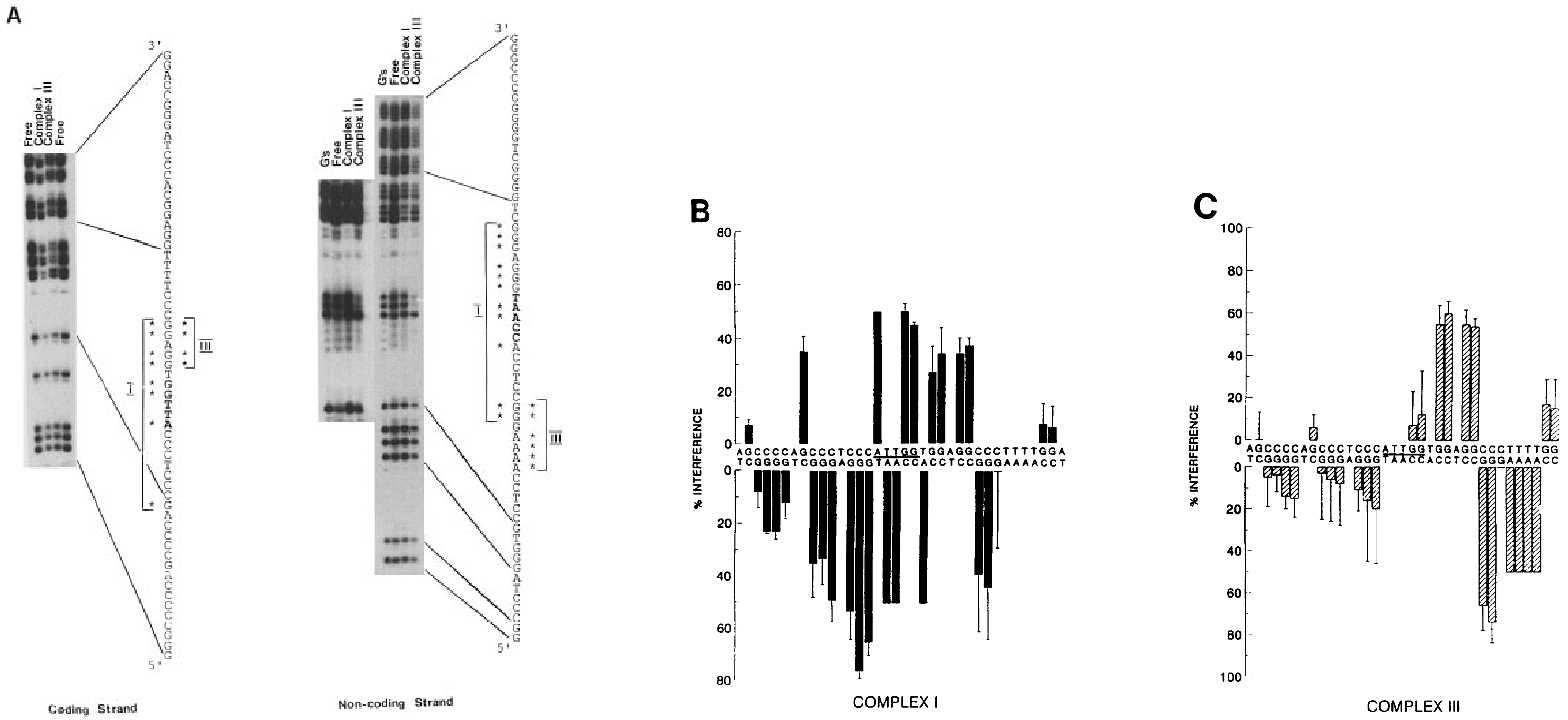


Figure 1 Methylation interference analysis of the human α2(1) procollagen promoter

(**A**) DNA was end-labelled with [γ-^{32}P]ATP, partially methylated with dimethyl sulphate and incubated with crude CT-1 nuclear extract as described in the Materials and methods section. DNA–protein complexes (complexes I and III) were resolved on 5% non-denaturing polyacrylamide gels, the free and complexed DNA bands were extracted, cleaved with piperidine, resolved on 12% sequencing gels, dried and exposed to X-ray film for at least 16 h. The DNA sequence of the coding (−107 to −47) and the non-coding (−107 to −50) strands are shown with the CCAAT box in bold. Purines required for complex I and III formation are indicated with asterisks. The protected adenines on the non-coding strand are shown in the longer exposure, shown to the left of the non-coding strand. The lanes marked G's represent naked DNA modified with dimethyl sulphate and cleaved with piperidine to localize the guanines in the sequence. Histograms of the methylation interference data of complexes I and III are shown in (**B**) and (**C**) respectively. The autoradiograms were scanned and the average percentage interference and S.D.s for each purine residue were determined in four separate experiments. Both the coding (top) and non-coding (bottom) strands are shown. The inverted CCAAT box is underlined.

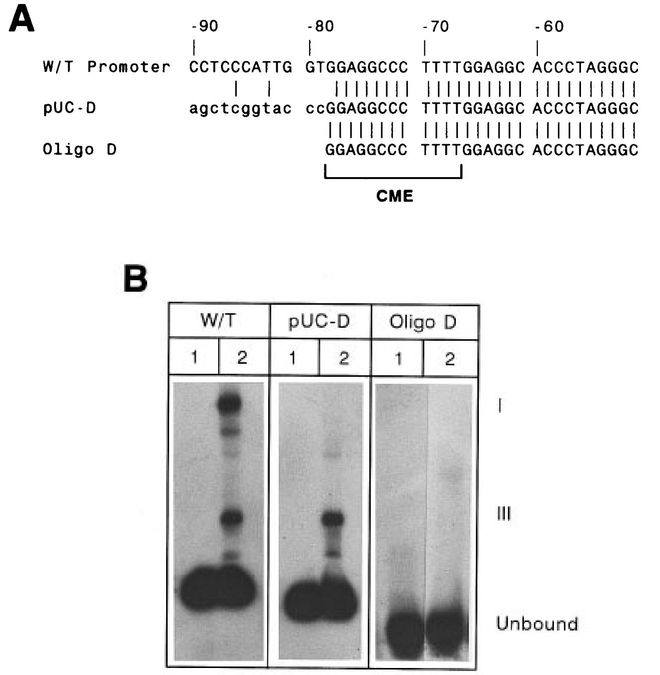

A
-90
-80
-70
-60
W/T Promoter CCTCCCATTG GTGGAGGCCC TTTTGGAGGC ACCCTAGGGC
pUC-D agctcggtac ccGGAGGCCC TTTTGGAGGC ACCCTAGGGC
Oligo D GGAGGCCC TTTTGGAGGC ACCCTAGGGC
CME
B
W/T
pUC-D
Oligo D
1
2
1
2
1
2
I
III
Unbound

Figure 2 Upstream sequences do not play a role in complex III formation

(**A**) Double-stranded wild-type promoter (−90 to −50) or oligo(D) containing the CME and downstream sequences were cloned into the *Sma*I site of pUC-19. DNA fragments were excised by double digestion with *Eco*RI and *Hin*dIII for use in EMSAs. (**B**) The DNA fragment released from pUC-D [oligo(D) cloned in pUC-19], the wild-type promoter and the double-stranded CME oligonucleotide were used to detect DNA–protein complex formation using crude nuclear extracts. Lanes 1 contain no nuclear proteins, showing the position of the free probe. Lanes 2 contain 4 μg of CT-1 nuclear extract. The free DNA in the oligo D lanes migrated faster than the free DNA in the wild type (W/T) and pUC-D lanes, since the latter fragments are larger due to the presence of the additional pUC-19 sequences.

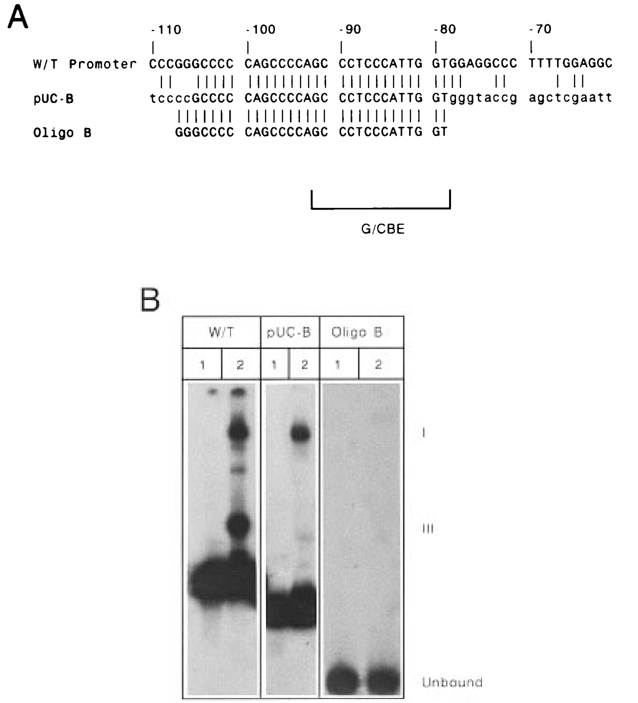

A
-110
-100
-90
-80
-70
W/T Promoter CCCGGGCCCC CAGCCCCAGC CCTCCCATTG GTGGAGGCCC TTTTGGAGGC
pUC-B tccccGCCCC CAGCCCCAGC CCTCCCATTG GTgggtaccg agctcgaatt
Oligo B GGGCCCC CAGCCCCAGC CCTCCCATTG GT
G/CBE
B
W/T
pUC-B
Oligo B
1
2
1
2
1
2
I
III
Unbound

Figure 3 Downstream sequences are not involved in complex I formation

(**A**) Double-stranded wild-type promoter (−90 to −50) or oligo(B) containing the G/CBE and downstream sequences were cloned into the *Sma*I site of pUC-19. DNA fragments were excised by double digestion with *Eco*RI and *Hin*dIII and used in EMSAs. (**B**) The DNA fragment released from pUC-B [oligo (B) cloned into pUC-19], the wild-type promoter and the double-stranded G/CBE oligonucleotide were used to detect DNA–protein complex formation using crude nuclear extracts. Lanes 1 contain no nuclear proteins to show the position of the free probe. Lanes 2 contain 4 μg of CT-1 nuclear extract. The free DNA in the oligo B lanes migrated faster than the free DNA in the wild-type (W/T) and pUC-B lanes, since the latter fragments are larger due to the presence of the additional pUC-19 sequences.

1B). Since the upstream GGAGG sequence is important in the formation of the CCAAT box DNA–protein complex (see also Figures 2 and 3), this element has been named the GGAGG/CCAAT binding element or G/CBE. Complex III formation, on the other hand, involved sequences mainly downstream of the CCAAT box (−78 to −67; summarized in Figure 1C) and has been termed the CME, since it would appear that an inhibitor of α2(1) collagen gene transcription also binds to this sequence [29]. It is possible that the upstream and downstream GGAGG sequences (CCTCCCATTGGTGGAGG) are involved in the formation of cruciform structures.

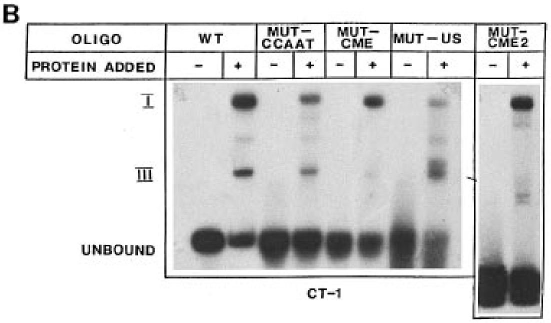
Since the methylation interference data indicated that the two sites overlap each other, it is possible that the proteins either bind in a mutually exclusive manner, or that these sites are adjacent to each other and that the factors bind in a co-operative manner. In order, therefore, to clearly define the boundaries of the G/CBE and CME, double-stranded oligonucleotides were used in EMSAs. A short double-stranded oligonucleotide containing the CME and 3′-flanking sequences (oligo D in Figure 2A) failed to form any DNA–protein complexes (Figure 2B). When non-specific DNA from the polylinker of pUC-19 was added onto the 5′-flanking region of the oligonucleotide, however, complex III formation occurred. This pUC-19 sequence has no similarity to the sequence flanking the CME in the human α2(1) procollagen promoter. Similarly, an oligonucleotide containing the G/CBE and 5′-flanking sequences failed to form any DNA–protein complex [oligo(B) in Figure 3A], but the addition of non-specific DNA on the 3′-end resulted in complex I formation. Both the cloned oligonucleotides pUC-B and pUC-D [oligo(B) and oligo(D) cloned in pUC-19 respectively] had the same specificity of DNA–protein complex formation as the full-length wild-type oligonucleotide. These experiments clearly defined the G/CBE and CME as two distinct but adjacent DNA elements. The contact points outside each element exhibited by methylation interference analysis (Figure 1) therefore appear to be non-specific, and these non-specific sequences are probably required for the proteins to ‘grip’ onto the DNA. This property is similar to that established for the binding of restriction endonucleases to DNA.

Mutation analysis

In order to confirm the data obtained by methylation interference and mobility-shift assays, the mobility-shift assays were performed with mutated oligonucleotides. In these experiments, either the CCAAT box or the flanking GGAGG sequences were mutated as indicated in Figure 4(A). Mutations in the CCAAT box (MUT-CCAAT) resulted in a reduction in, but not total abolition of, complex I formation (80% reduction), while a minimal reduction (30%) in complex III formation was observed (Figure 4B). Mutation of the upstream inverted GGAGG sequence (MUT-US,) on the other hand, totally abolished complex I formation without any significant effect on complex III formation. EMSAs with mutated oligonucleotides

A

```
            -107                                              -59
             |                                                 |
WT           GGGCCCCCAGCCCCAGCCCTCCCATTGGTGGAGGCCCTTTTGGAGGC
MUT-CCAAT    -----------------------C---C-------------------
MUT-CME      -----------------------------TT-TT-------------
MUT-CME2     ----------------------------------AA-----------
MUT-US       -----------------AA-AA-------------------------
```



C

G/CBE　　CME

5'-CCCTCCCATTGGTGGAGGCCCTTTTGGAGG-3'
3'-GGGAGGGTAACCACCTCCGGGAAAACCTCC-5'

Figure 4 ***In vitro* mutagenesis of DNA-binding sites**

(**A**) Double-stranded oligonucleotides, containing mutations in the GGAGG sites (MUT-CME, MUT-US), CCAAT box (MUT-CCAAT) and the cytosine dinucleotides at −72 and −73 (MUT-CME2), were used in EMSAs. WT, wild type. (**B**) ^{32}P-labelled double-stranded oligonucleotide (1 ng; 10^4 c.p.m.) was incubated with 3–4 μg of crude nuclear extract (+ lanes) or without nuclear extract (− lanes) and analysed on 5% non-denaturing polyacrylamide gels, as described in the Materials and methods section. The gels were dried and exposed to X-ray film for at least 16 h. The positions of complexes I and III are indicated. (**C**) Summary of transcription factor binding sites on the human α2(I) procollagen promoter. The sequence of the proximal promoter shows the G/CBE containing an inverted CCAAT box, which also corresponds to the CBF-binding site in the mouse promoter. The CME is the binding site for a novel factor, forming complex III in the human α2(1) procollagen promoter.

A

Oligonucleotide/fragment	Sequence
NF-Y	5'-AACATTTTTCTGATTGGTTAAAAGTTG-3' 3'-TTGTAAAAAGACTAACCAATTTTCAAC-5'
TSP 1	5'-GCCCGGCCGCCGCCCATTGGCCGGAGGAAT-3 3'-CGGGCCGGCGGCGGGTAACCGGCCTCCTTA-5
C/EBP	5'-AATTCAATTGGGCAATCAGG-3' 3'-GTTAACCCGTTAGTCCTTAA-5'
G/CBE	5'-GATCCGCCCTCCCATTGGTG-3' 3'-GCGGGAGGGTAACCACCTAG-5'
CME	5'-AATTCGGAGGCCCTTTTGGAGG-3' 3'-GCCTCCGGGAAAACCTCCTTAA-5'
CBE	-110 to +54, murine Col1A2 gene containing an inverted CCAAT box

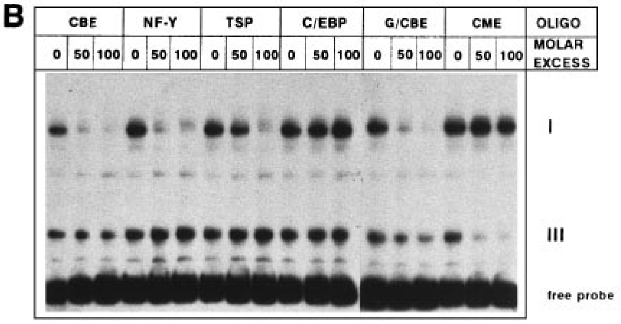


Figure 5 **Characterization of the human G/CBE**

(**A**) Complementary oligonucleotides for NF-Y, TSP-1, C/EBP, G/CBE, CME and CBF were synthesized on a Beckman Oligo 1000 DNA synthesizer. (**B**) EMSAs using the human α2(1) proximal procollagen promoter (−107 to −50) as probe. CT-1 nuclear extracts were incubated with the indicated amount of unlabelled double-stranded competitor oligonucleotide before the addition of ^{32}P-labelled probe. DNA–protein complexes were analysed on a non-denaturing 5% polyacrylamide gel and exposed to X-ray film for 16 h.

showed that the downstream GGAGG sequence and the adjacent CC dinucleotide are both important for complex III formation, confirming the results obtained by methylation interference (Figure 1). Mutations within these sequences resulted in < 95% inhibition of complex III formation while complex I formation was largely unaffected. The sequences which are important for complex I formation could clearly be localized to the CCAAT box and upstream inverted GGAGG sequence, while the sequence crucial for complex III formation was located downstream of the CCAAT box. These data also support the results obtained in Figures 2 and 3 and confirm that the G/CBE and CME are indeed independent elements. A summary of the binding data showing the adjacent G/CBE and CME elements is presented in Figure 4(C).

In order to determine whether the G/CBE-binding proteins are related to any of the known CCAAT box binding factors, gel-shift competition analysis was performed using double-stranded nuclear factor-Y (NF-Y), CCAAT enhancer binding protein (C/EBP), CBF or thrombospondin-1 (TSP-1) oligonucleotides [41], as indicated in Figure 5(A). The G/CBE competitor should compete out complex I formation very efficiently, since it is identical with the G/CBE in the probe. CBE is the mouse homologue of the human G/CBE, with which it shares 100% sequence identity. The TSP-1 promoter also contains high sequence similarity to the human G/CBE, whereas NF-Y is a well-characterized CCAAT box element. C/EBP, on the other hand, is an enhancer binding protein which does not bind to these CCAAT elements and was used as a negative control. The CBE [mouse α2(1) CCAAT box] and NF-Y oligonucleotides competed very strongly for complex I formation (i.e. on the G/CBE), whereas the C/EBP and CME oligonucleotide did not compete at all. The TSP-1 oligonucleotide, on the other hand, was a much weaker competitor than either CBE or NF-Y (Figure 5B). It is also clear that none of the CCAAT box oligonucleotides interfered with complex III (CME) formation. These competition studies using the CME oligonucleotide provided further evidence that the G/CBE and CME are two independent elements, since the CME oligonucleotide was unable to compete out complex I formation (Figure 5B). It is even more interesting that the mouse oligonucleotide, which contains all the 3′-flanking sequences down to +54, was unable to compete out complex III formation. Since some elements may be located at different positions in different species [42], a search of the mouse promoter and first intron was performed, which revealed that the CME is not present elsewhere in the rodent gene.

That complex I is indeed the CCAAT-binding factor was shown by electrophoretic mobility supershifting using an antibody to the mouse CBF-B component. Addition of rabbit anti-(mouse CBF-B) to the DNA–protein complexes resulted in

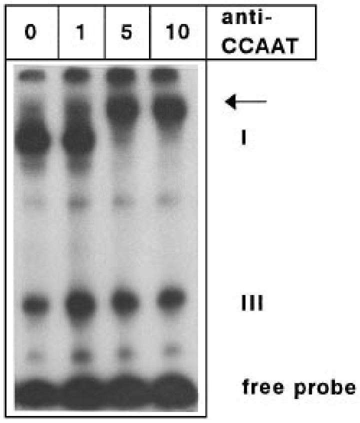
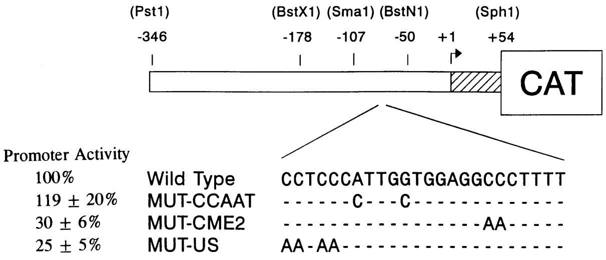


Figure 6 Identification of complex I as a CCAAT box binding factor

The proximal human α2(1) procollagen promoter fragment (−107 to −50) was end-labelled with *[^{32}P]dCTP and incubated with crude nuclear extract as described in the legend to Figure 5, followed by the addition of either 1, 5 or 10 μl of a 1:100 dilution of rabbit anti-(mouse CBF-B) antibody. The antibody was allowed to react for 30 min on ice before analysis of DNA–protein complexes on non-denaturing 5% polyacrylamide gels. The gels were dried and exposed to X-ray film for 16 h.



Promoter Activity		
100%	Wild Type	CCTCCCATTGGTGGAGGCCCTTTT
119 ± 20%	MUT-CCAAT	------C---C-------------
30 ± 6%	MUT-CME2	-----------------AA-----
25 ± 5%	MUT-US	AA-AA-------------------

Figure 7 Promoter activity of mutated proximal promoter constructs

The *PstI–SphI* fragment of the human α2(I) procollagen promoter was subjected to *in vitro* mutagenesis to generate the indicated mutants, as described in the Materials and methods section. The mutant promoters were cloned into the promoterless p8CAT vector. The promoter/CAT constructs were co-transfected with CMV-βGAL (the bacterial β-galactosidase gene driven by the cytomegalovirus promoter) into CT-1 fibroblasts and assayed for CAT and β-GAL activities as described in Materials and methods section. The CAT activity is expressed relative to that of β-GAL in order to correct for transfection efficiency ($n = 8$). The activity of the wild-type promoter is set as 100%.

supershift of only complex I and not complex III (Figure 6). This also clearly shows that complex I protein is indeed the CCAAT-binding factor and that complex III protein is not related to CBF.

Promoter activity

In view of the fact that mutations in the G/CBE and CME totally abolished complex I and III formation respectively (Figure 4), it was essential to determine whether these mutations affected the functional activity of the human α2(I) procollagen promoter. The promoter constructs subjected to *in vitro* mutagenesis were cloned into the promoterless p8CAT vector, as described in the Materials and methods section, and tested for promoter activity in transient transfection assays using the type I collagen producing CT-1 fibroblast cell line. Mutations in the upstream inverted GGAGG sequence (MUT-US) or the downstream CC dinucleotide (MUT-CME2) resulted in drastic decreases in promoter activity of 70 % and 80 % respectively (Figure 7). The binding of transcription factors to the CME immediately downstream of the CCAAT box is therefore crucial for activity of the human α2(1) procollagen promoter. These findings suggest that regulation of the human α2(I) promoter may be different from that of the mouse promoter, and that the G/CBE (complex I) and the CME (complex III) proteins are probably co-activators in the human promoter.

DISCUSSION

Two DNA–protein complexes (complexes I and III), which form on a 59 bp fragment (−107 to −50) of the human α2(I) procollagen promoter, have previously been identified in nuclear extracts prepared from CT-1 fibroblasts [29]. In this study we have identified these as two distinct adjacent DNA-recognition elements, both of which are essential for activity of the human α2(1) procollagen promoter.

The minimal sequence element required for the formation of both complexes I and III was located within the 26 bp DNA fragment between −92 and −67, i.e. the CCAAT box and downstream sequences (5′-GCCCTCCC**ATTGG**TGGAGG CCCTTTT-3′). Methylation interference analysis, mutation analysis and transcription factor binding studies indicated that the CCAAT box and upstream inverted GGAGG sequence were the crucial motifs for complex I formation. Although the methylation interference patterns showed that guanines in the downstream GGAGG box were also involved in complex I formation, mutation or replacement of these guanines with non-specific sequences did not significantly reduce complex I formation. This suggested that non-specific 3′-flanking sequences were required in order for the proteins to hold or 'grip' onto the DNA.

The specificity of a DNA-binding factor for a particular *cis*-element may be influenced by the sequences flanking the consensus site. In this study, the nucleotides downstream of the CCAAT box motif were shown not to be important for complex I formation, whereas the upstream sequences were crucial. The upstream sequences involved in G/CBE protein binding are 100 % conserved between the human and mouse α2(1) collagen promoters (Figure 8A). This high level of sequence similarity, the competition and antibody supershift data (Figures 5 and 6) and other protein characterization data, such as native molecular mass, Stokes radius, sedimentation coefficient, etc. (results not shown), strongly suggest that the G/CBE-binding protein(s) (complex I) is a member of the family of the structurally related CCAAT box binding factors, which include CBF [17,24–26], NF-Y [43], α-CP1 and CP1 [44]. There are, however, significant differences between the sequences flanking the human α2(I) procollagen CCAAT box when compared with sequences flanking the other CCAAT boxes, such as NF-Y and others mentioned above. It is not clear why an NF-Y oligonucleotide should compete out complex I formation, since the upstream sequences do not bear any similarity to the G/CBE. The NF-Y oligonucleotide was also able to compete out binding of CBF to the CBE in the mouse promoter (results not shown), confirming the similarity between the human and mouse CCAAT-binding factors.

The methylation interference patterns showed that complex III proteins bound to a 12 bp element between −78 and −67, the CME, which contains the downstream GGAGG sequence and 3′-flanking sequences (5′-GGAGGCCCTTTT-3′). Mutations in this GGAGG sequence and the downstream cytosine dinucleotide resulted in a drastic reduction, or abolition, of complex III formation (on the CME). This element is adjacent

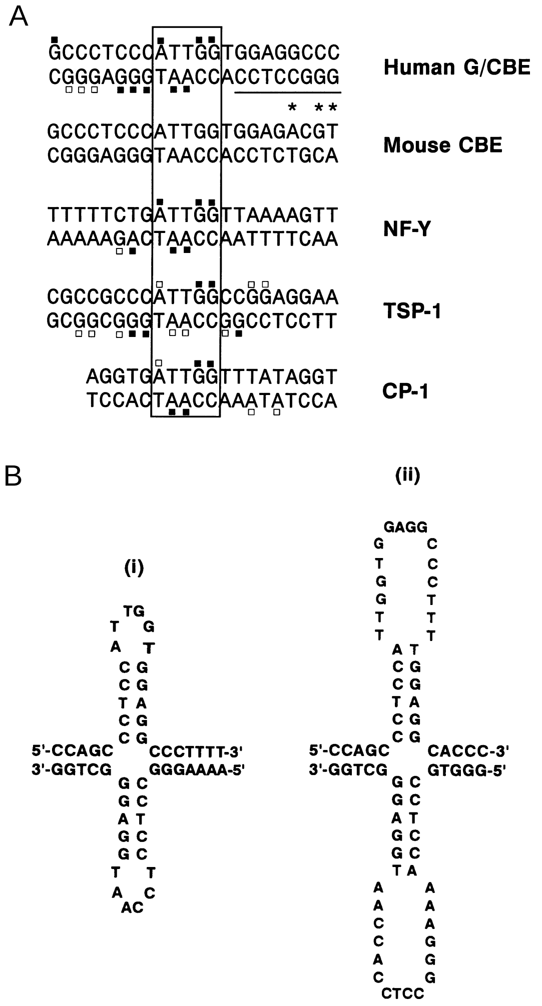


Figure 8 Comparison of the human G/CBE with other CCAAT box elements

(**A**) Alignment of a 50 bp region (nucleotides −110 to −61) shows a very high sequence identity between the human and mouse CCAAT box elements. The inverted CCAAT boxes are boxed, while the human CME is underlined. Mismatches between the human and mouse sequences in the region of the CME are indicated by asterisks. Full and partial methylation interference of purines in the G/CBE, NF-Y [43], TSP-1 [41] and CP-1 [44] are indicated by (■) and (□) respectively. (**B**) Possible secondary structures of the human α2(1) procollagen promoter involving the upstream inverted GGAGG sequence and one of the two downstream GGAGG sequences shown in Figure 4(**C**).

to the G/CBE (Figure 4C) and does not bear any sequence similarity to any of the known transcription factor binding sites published to date. Similar GGGAGGG boxes have recently been reported to footprint in the promoters of other genes, such as the human α1(XI) procollagen gene [45], but no information on its significance is available. Comparison of the human CME with the mouse sequence revealed a 3 bp mismatch within this region, involving bases which were crucial for complex III formation in the human promoter (Figure 8A), resulting essentially in the absence of a CME in the mouse α2(I) promoter. Only one complex was obtained when NIH 3T3 nuclear extracts were assayed with an equivalent region of the mouse α2(I) procollagen promoter [16]. It is therefore more than likely that species-specific mechanisms operate in regulating the expression of the α2(I) procollagen gene, as has been suggested by studies on the transforming growth factor-β responsive elements in the human and mouse α2(I) promoters [20,46] and reviewed by Bornstein [42].

The methylation interference data in Figure 8 also indicate that the purine bases within the CCAAT box are crucial for binding of the CBF. Moreover, the upstream sequences of the human α2(1) procollagen and TSP-1 genes are very similar, as are the methylation interference patterns.

Transfection studies revealed that both an intact G/CBE and CME were essential for α2(I) procollagen promoter activity. It would appear that the binding of both factors was essential for promoter activity, since mutations which resulted in a total loss of DNA–protein complex formation on either the G/CBE or the CME resulted in an associated loss of promoter activity. These results imply that the two factors bind co-operatively and that they possibly are co-activators of the human α2(I) procollagen gene. A similar synergistic relationship exists between the HNF-1 and NF-Y binding sites in the albumin promoter. NF-Y is a CCAAT-binding protein, whereas HNF1 binds to an element adjacent to and immediately downstream of the CCAAT box [47,48]. Mutation analysis showed that binding of both these factors are required for gene activity. The affinity of HNF-1 for its binding site is greatly enhanced by co-operative interaction with other factors which bind to an adjacent element. Our study shows that a similar situation exists in the human α2(1) procollagen gene.

One interesting feature of the proximal human α2(I) procollagen promoter is that this region could potentially form cruciform structures involving the upstream inverted and one of the two downstream GGAGG sequences. The model in Figure 8(B) (i) would place the CCAAT box on the tip of the loop, while the CME would be the hairpin double-stranded structure. In the second configuration, both the CCAAT box and CME would be in the loop. Such structures may facilitate the co-operative binding of the different proteins to the promoter, or could even be induced upon binding of proteins to the promoter.

In summary, two distinct adjacent regulatory elements, the G/CBE and CME, have been identified in the human α2(I) proximal procollagen promoter. Two types of *trans*-acting factors, the G/CBE proteins (complex I) and CME proteins (complex III) bind to these elements to regulate the expression of the α2(I) procollagen gene in normal type I collagen producing cells. Moreover, the mouse and human promoters do not have the same DNA-binding sites, i.e. the CME, suggesting an important difference in the regulation of the α2(1) collagen gene between these species. These findings also have important implications in the use of transgenic mice to study the regulation of the human collagen gene.

This work was supported by research grants from the South African Medical Research Council, the Cancer Association of South Africa and the University of Cape Town. We are grateful to Dr. F. Ramirez for providing us with the original genomic clones, and to Dr. B. de Crombrugghe for the anti-CBF antibodies.

REFERENCES

1 Van der Rest, M. and Garrone, R. (1991) FASEB J. **5**, 2814–2823
2 Bornstein, P. and Sage, H. (1989) Progr. Nucleic Acid Res. Mol. Biol. **37**, 67–106
3 Vuust, J., Sobel, M. E. and Martin, G. R. (1985) Eur. J. Biochem. **151**, 449–453
4 Bornstein, P. and Sage, H. (1980) Annu. Rev. Biochem. **49**, 957–1003
5 Raghow, R. and Thompson, J. P. (1989) Mol. Cell. Biochem. **86**, 5–18
6 Guenette, D. K., Ritzenthaler, J. D., Foley, J., Jackson, J. D. and Smith, B. D. (1992) Biochem. J. **283**, 699–703
7 McKeon, C., Ohkubo, H., Pastan, I. and de Crombrugghe, B. (1982) Cell **29**, 203–210
8 Parker, M. I., de Haan, J. B. and Gevers, W. (1986) J. Biol. Chem. **261**, 2786–2790

9 Parker, M. I., Smith, A. A. and Gevers, W. (1989) J. Biol. Chem. **264**, 7147–7152

10 Smith, B. D. and Marsilo, E. (1988) Biochem. J. **253**, 269–273

11 Thompson, J. P., Simkevich, C. P., Holness, M. A., Kang, A. H. and Raghow, R. (1991) J. Biol. Chem. **266**, 2549–2556

12 de Crombrugghe, B., Karsenty, G., Maity, S., Vuorio, T., Rossi, P., Ruteshouser, E. C., McKinney, S. H. and Lozano, G. (1990) Ann. N.Y. Acad. Sci. **580**, 88–96

13 Slack, J. L., Liska, D. J. and Bornstein, P. (1993) Am. J. Med. Genet. **45**, 140–151

14 Prockop, D. J., Kivirikko, K. I., Tuderman, L. and Guzman, N. A. (1979) New Engl. J. Med. **301**, 13–23

15 Goldberg, H., Helaakoski, T., Garrett, L. A., Karsenty, G., Pellegrino, A., Lozano, G., Maity, S. and de Crombrugghe, B. (1992) J. Biol. Chem. **267**, 19622–19630

16 Hatamochi, A., Paterson, B. and de Crombrugghe, B. (1986) J. Biol. Chem. **261**, 11310–11314

17 Karsenty, G., Golumbek, P. and de Crombrugghe, B. (1988) J. Biol. Chem. **263**, 13909–13915

18 Niederreither, K., D'Souza, R. N. and de Crombrugghe, B. (1992) J. Cell Biol. **119**, 1361–1370

19 Oikarinen, J., Hatamochi, A. and de Crombrugghe, B. (1987) J. Biol. Chem. **262**, 11064–11070

20 Rossi, P., Karsenty, G., Roberts, A. B., Roche, N. S., Sporn, M. B. and de Crombrugghe, B. (1988) Cell **52**, 405–414

21 Pogulis, R. J. and Freytag, S. O. (1993) J. Biol. Chem. **268**, 2493–2499

22 Rossi, P. and de Crombrugghe, B. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 5590–5594

23 Karsenty, G. and de Crombrugghe, B. (1991) Biochem. Biophys. Res. Commun. **177**, 538–544

24 Hatamochi, A., Golumbek, P. T., Van Schaftingen, E. and de Crombrugghe, B. (1988) J. Biol. Chem. **263**, 5940–5947

25 Maity, S. M., Golumbek, P. T., Karsenty, G. and de Crombrugghe, B. (1988) Science **241**, 582–585

26 Maity, S. N. and de Crombrugghe, B. (1992) J. Biol. Chem. **267**, 8286–8292

Received 30 July 1996/23 September 1996; accepted 2 October 1996

27 Boast, S., Su, M.-W., Ramirez, M. and Avvedimento, E. V. (1990) J. Biol. Chem. **265**, 13351–13356

28 Dickson, L. A., de Wet, W., Liberto, M. D., Weil, D. and Ramirez, F. (1985) Nucleic Acids Res. **13**, 3427–3438

29 Parker, M. I., Smith, A. A., Mundell, K., Collins, M., Boast, S. and Ramirez, F. (1993) Nucleic Acids Res. **20**, 5825–5830

30 Namba, M., Nishitani, K. and Kimoto, T. (1980) Jpn. J. Cancer Res. **71**, 300–307

31 de Haan, J. B., Gevers, W. and Parker, M. I. (1986) Cancer Res. **46**, 713–716

32 Graham, F. L. and van der Eb, A. J. (1973) Virology **52**, 456–467

33 Seed, B. and Sheen, J.-Y. (1988) Gene **67**, 271–277

34 Herbomel, P., Bourachot, B. and Yaniv, M. (1984) Cell **39**, 653–662

35 Sayers, J. R., Schmidt, W. and Eckstein, F. (1988) Nucleic Acids Res. **16**, 791–802

36 Dignam, J. D., Lebovitz, R. M. and Roeder, R. G. (1983) Nucleic Acids Res. **11**, 1475–1489

37 Bradford, M. M. (1976) Anal. Biochem. **72**, 248–254

38 Maity, S. N. and de Crombrugghe, B. (1992) J. Biol. Chem. **267**, 8286–8292

39 Perbal, B. (1988) A Practical Guide to Molecular Cloning, 2nd edn., John Wiley & Sons, New York

40 Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1993) Current Protocols in Molecular Biology, Greene Publishing Associates, Inc and John Wiley & Sons, Inc. New York

41 Framson, P. and Bornstein, P. (1992) J. Biol. Chem. **268**, 4989–4996

42 Bornstein, P. (1996) Matrix Biol. **15**, 3–10

43 Chodosh, L. A., Baldwin, A. S., Carthew, R. W. and Sharp, R. A. (1988) Cell **53**, 11–24

44 Kim, C. G. and Sheffery, M. (1990) J. Biol. Chem. **265**, 13362–13369

45 Yoshioka, H., Greenwel, P., Inoguchi, K., Truter, S., Inagaki, Y., Ninomiya, Y. and Ramirez, F. (1995) J. Biol. Chem. **270**, 418–424

46 Inagaki, Y., Truter, S. L. and Ramirez, F. (1994) J. Biol. Chem. **269**, 14828–14838

47 Lichtsteiner, S. and Schibler, U. (1989) Cell **57**, 1179–1187

48 Tronche, F., Rollier, A., Bach, I., Weiss, M. C. and Yaniv, M. (1989) Mol. Cell. Biol. **9**, 4758–4766