MOLECULAR AND CELLULAR BIOLOGY, Nov. 1993, p. 6969–6983 Vol. 13, No. 11
0270-7306/93/116969-15$02.00/0
Copyright © 1993, American Society for Microbiology

Activation of the β-Globin Promoter by the Locus Control Region Correlates with Binding of a Novel Factor to the CAAT Box in Murine Erythroleukemia Cells but Not in K562 Cells

N. L. DELVOYE,[1] N. M. DESTROISMAISONS,[1] AND L. A. WALL[1,2]*

Institut du Cancer de Montréal[1] and Department of Medicine, Université de Montréal,[2] 1560 Sherbrooke Street East, Montreal, Quebec, Canada H2L 4M1

Received 12 March 1993/Returned for modification 12 May 1993/Accepted 16 August 1993

Four distinct factors in extracts from murine erythroleukemia (MEL) cells interacted with the human β-globin gene promoter CAAT box: CP1, GATA-1, and two novel factors, denoted *a* and *b*, one of which is highly inducible in the MEL system. GATA-1 binding to the CAAT element was very unstable (half-life < 1 min), whereas bindings of *a*, *b*, and CP1 were comparatively stable, with half-lives of 18, 19, and 3.5 min, respectively. Stable transfections of MEL cells showed that in the presence of the β-globin locus control region (LCR), the wild-type CAAT box, a mutant which bound to GATA-1 with increased stability over the normal sequences, and a mutant which bound *a*, *b*, and CP1 specifically could all stimulate transcription greater than ninefold over that induced by a null CAAT mutation in both uninduced and terminally differentiated MEL cells. A mutant which bound the *a* and *b* factors specifically gave only a twofold stimulation of promoter activity, and this lower activity correlated with a decrease in the stability of binding of the *b* protein. On the other hand, CP1 binding alone did not stimulate transcription. Taken together, these results suggest that in the context of the wild-type β-globin CAAT element the *b* factor stimulates transcription directed by the LCR in MEL cells, although the LCR can also function through more stable GATA-1-binding sequences. However, in K562 cells, the wild-type β-globin CAAT box alone was unable to stimulate gene expression directed by the LCR and high levels of transcription were obtained only upon inclusion of more upstream β-globin promoter sequences. In contrast, a construct containing only the Aγ-globin CAAT box region did give high expression levels in K562 cells. Thus, there is a fundamental difference in the way the LCR functions in these two model systems in terms of its requirements at the promoter level.

Because of their complex regulation, the globin genes are a good model system for understanding the overall control of gene expression. In higher animals, there exist two independent loci which code separately for the α and β subunits of the tetrameric hemoglobin protein (reviewed in reference 25). These loci are active exclusively in erythrocytes, in which globin transcripts can represent greater than 1% of the total poly(A)$^+$ RNA. Moreover, each globin locus contains several genes which are expressed at different times during development (termed globin switching). For example, the human β-globin locus contains five active genes (ε, Gγ, Aγ, δ, and β). During development, the ε gene is expressed only during embryonic life in the yolk sac. By comparison, the two γ-globin genes are transcribed specifically in the fetal liver throughout fetal life. In the adult, when erythropoiesis occurs in the bone marrow, it is the β-globin gene which is expressed at high levels.

A milestone in our understanding of how globin gene expression is controlled was the discovery of a locus control region (LCR; previously called DCR and LAR [reviewed in reference 22]) in the human β-globin domain (16). The LCR is situated between 10 and 30 kbp upstream from the ε-globin gene (~50 kbp from the adult β-globin gene; see Fig. 1). The LCR enhances the levels of transcription of globin transgenes such that they are expressed per copy at the same levels as the endogenous globins and the levels of expression are independent of where the transgene is integrated into the genome (16). The LCR functions erythroid specifically in transgenic mice and in cell cultures (5, 10, 16, 26, 30).

When the LCR is present along with the β-globin gene, in transgenic mice the β-globin gene is expressed at all stages of development, although the gene alone had been shown to be adult stage specific (2, 12). Moreover, whereas the β-globin gene alone is expressed in adult-stage murine erythroleukemia (MEL) cells but not in human embryonic-fetal K562 cells, the gene is expressed in both cell types when the LCR is also present (5). However, when the human γ-globin gene is included on the same DNA fragment as the β-globin gene and LCR, the β-globin gene is then expressed developmental stage specifically in transgenic mice. These observations have suggested that the developmental specificity of the β-globin gene may be partially determined by competition between it and the γ-globin gene(s) for enhancement by the LCR (2, 12). A similar hypothesis was originally put forth for the chicken β-globin system (8). Understanding what gene-flanking sequences may be required for high-level expression to be directed by the LCR would help define how the LCR functions and how it might be able to choose between different genes in *cis*.

Antoniou and Grosveld (1) demonstrated that if the β-globin structural gene and 3′ flanking sequences are replaced by an *H-2K*^{k} reporter gene and the upstream promoter is deleted to −103, the LCR will still direct high levels of inducible transcription in MEL cells. However, if the promoter is further deleted to remove the CACC box and CAAT box sequences so that only the TATA box remains (see Fig. 1), expression of the transgene is low both before and after differentiation of MEL cells. Addition of the CAAT and/or

* Corresponding author.

TABLE 1. Mutant β-CAAT oligonucleotides

Oligonucleotide	Sequence[a]	Binding[b] to:			
		CP1	*a*	*b*	GATA-1
Wild type	CCTAGGGTTGGCCAATCTACTCCCAGGAGC	+	+	+	+
Mutant 1	-----------A-T---AC-----------	−	−	−	+
Mutant 2	----------------GA------------	+	+	+	−
Mutant 3	---------------C--------------	−	+	+	−
Mutant 4	-----TT-----------------------	+	±	±	+
Mutant 5	-----TT----A-T-CGAC-----------	−	−	−	−
Mutant 6	---------A------GA-A----------	+	±	±	−
Mutant 7	-------CCA------GAG-----------	+	−	−	−

[a] A dash indicates no change from the wild-type sequence.
[b] Data from Fig. 7. +, factor binds to the oligonucleotide; −, factor no longer recognizes the oligonucleotide; ±, factor still binds, but with reduced affinity.

CACC sequence(s) will restore high expression (1). Thus, full activity of the β-globin LCR may require transcription factor(s) that bind to the CAAT and/or CAAC box region. Defining these factors is important to an understanding of how the LCR controls transcription and to begin to define how competition between globin genes occurs.

In the present report, a detailed analysis of protein factors binding to the β-globin CAAT box region was performed. It is shown that with the LCR, high levels of inducible transcription in MEL cells are directed through a novel factor binding the CAAT element which is described. It is also demonstrated that the β-globin CAAT box alone does not stimulate transcription in K562 cells.

MATERIALS AND METHODS

Cell cultures, transfections, and inductions. The MEL cells used are an adenine phosphoribosyltransferase-negative strain (C88) which was grown in α-minimal essential medium containing 10% fetal calf serum, 20 to 50 μg of diaminopurine per ml, and antibiotics. To induce differentiation, cultures were diluted at 0.5×10^6 cells per ml in the same medium containing 2% dimethyl sulfoxide (DMSO) and grown for 2 days for protein extracts or for 3 to 4 days for RNA extracts. In some cases, nuclear extracts were also prepared from a human chromosome 11 hybrid (H11A) MEL line, which gave similar results. K562 cells were grown in 10% fetal calf serum in α-minimal essential medium and were induced by growing in the presence of 80 μM hemin for 3 to 4 days.

Transfection of cells was performed by the lipofection procedure with a kit purchased from GIBCO-BRL Life Technologies, Inc. Cells (10^7) were transfected with 5 μg of μ-locus plasmid DNA linearized with *Pvu*I. The treated cells were grown overnight in fresh medium and then were selected by growth in medium containing 0.8 (MEL cells) or 1.0 (K562 cells) mg of G418 sulfate per ml. This medium was changed after 3 days, and then stably transfected populations selected after approximately 2 (MEL cells) or 3 (K562 cells) weeks were maintained in 0.6 to 0.8 mg of G418 per ml.

Protein extraction, methylation interference, and GMS assays. Making of protein extracts from nuclei that were isolated from tissues or cell cultures by centrifugation through 2.2 M sucrose solutions, methylation interference assays, and gel mobility shift (GMS) assays were all performed exactly as described previously (31), except that 1 mg of bovine serum albumin (BSA) per ml was used in the GMS assay mixes, unless stated otherwise, and 0.8 ng of ^{32}P-labeled double-stranded oligonucleotides (oligonucleotides) was used in each 10 μl of assay mix. For off-time analyses, the GMS assay mixes were scaled up to 100 μl and incubated at 22°C for 30 min, after which time a 1,000-fold excess of unlabeled competitor oligonucleotide was added. At the times indicated after addition of the competitor, 10-μl aliquots were loaded onto a gel that was continuously running at 4°C. To keep the assay conditions consistent, the wild-type CAAT oligonucleotide was used as the competitor in all cases except where noted otherwise, although similar results were obtained when the same competitor as the labeled probe was used.

The DNA sequences of synthetic oligonucleotides used (both strands are shown 5′ to 3′) were as follows: β-CAAT, CGCGGGATCCTAGGGTTGGCCAATCTACTCCCAG GAGCTT/GCTCCTGGGAGTAGATTGGCCAACCCTA GG; γ-CAAT, TTGCCTTGACCAATAGCCTTGACAAGG CAAAC/AAGTTTGCCTTGTCAAGGCTATTGGTCAAG GC; β-CACC, CCTAGGGTGTGGCTCCACAGGGTGAG GTC/AGACCTCACCCTGTGGAGCCACACCCTAGG; C, CGTCAGGAGCTTTAAGATTAGCATTCAGGAAG/CGC TTCCTGAATGCTAATCTTAAACATCCTGA; B, CGTCT TATTACCCTATCATAGGCCCACC/CGGGTGGGCCTA TGATAGGGTAATAAGA; FBG-CP1, AATGACCAGTTC CAGCCACTCTT/AAGAGTGGCTGGAACTGGTCA; and PBGD-NFE2, CACTGGGGAACCTGTGCTGAGTCACTG GAGG/TGCCTCCAGTGACTCAGCACAGGTTCCCCAG.

Note that the β-CAAT oligonucleotide has a *Bam*HI overhang which was used for cloning of polymerase chain reaction (PCR) products (see below). The sequence of mutant oligonucleotides is given in Table 1.

Northern and Southern blots and RNase protection assays. Northern (RNA) blot and Southern blot analyses were performed by common procedures (24) using GeneScreen Plus membranes and the hybridization and washing procedures recommended by the manufacturer. To measure copy number in transfected MEL cells, Southern blots of genomic DNA that had been digested with *Xba*I were probed with a 3-kb *Xba*I fragment from the LCR of the μ locus and the same blots were probed for a 2-kb fragment from the dt locus on mouse chromosome 1 (probe kindly provided by Rashmi Kothary) as a loading control. Probes were labeled by random oligonucleotide priming to approximately equal specific activities. Copy numbers were estimated by comparison with 1 and 10 equivalent copies of the μ-locus plasmid digested with *Xba*I and by comparison with a MEL population that had been transfected with the vector alone and was deemed to have only 1 copy. Copy numbers for K562 populations were estimated on Southern blots of *Sac*I digests by using a *Pvu*II fragment from the Neo gene as a probe for the transgene and a probe for the endogenous β-actin gene.

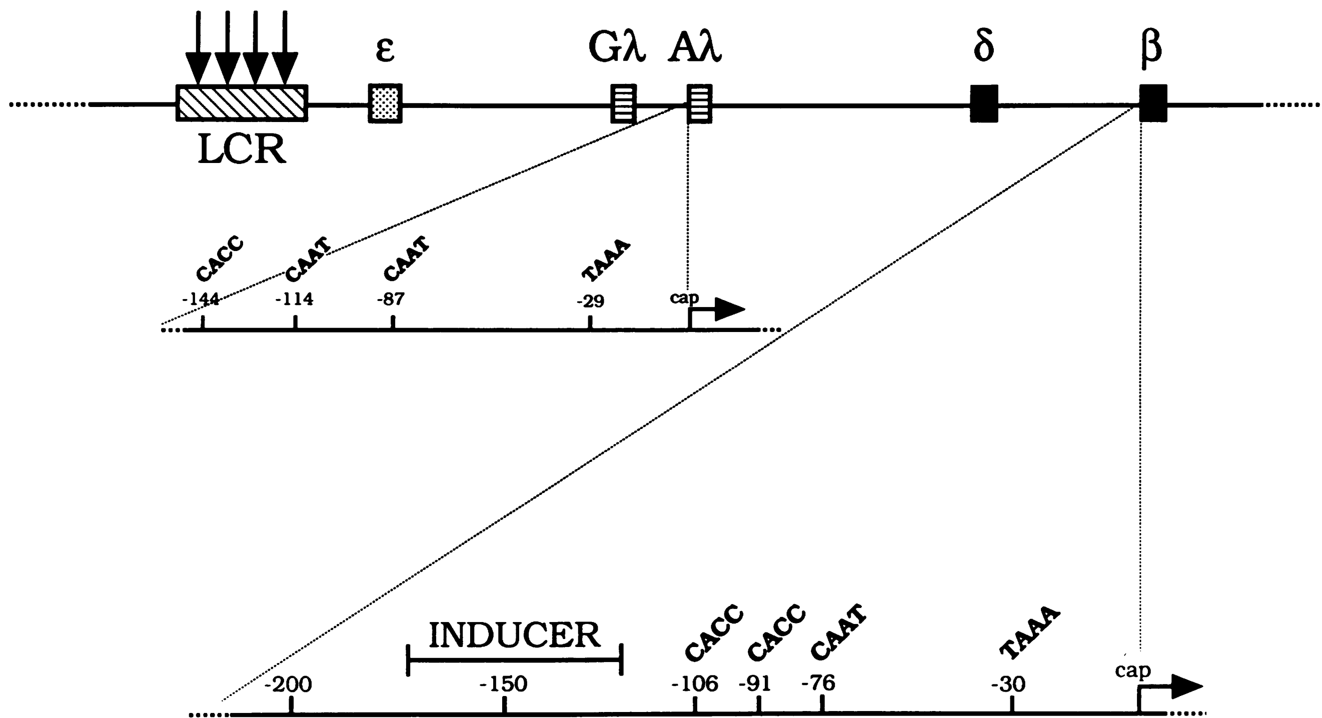


FIG. 1. Structure of the human β-globin locus and the promoter regions of the $^{A}\gamma$-globin and β-globin genes.

RNase protection assays were done by the procedures in reference 24. Ten micrograms of RNA was hybridized at 51°C to 3×10^5 cpm of [^{32}P]UTP-labeled RNA probe. RNase digestion was done at 34°C. For the β-*H-2K*^{k} RNA probe, a 900-bp *Cla*I-*Not*I DNA fragment from the wild-type β-globin promoter-*H-2K*^{k} gene hybrid described by deBoer et al. (11), which contains −800 to +32 of the β-globin gene and the first 55 bp of the *H-2K*^{k} gene, was cloned into the same restriction enzyme sites in the Stratagene BlueScript SK+ vector. This construct was linearized with *Cla*I, and antisense RNA was synthesized with T7 polymerase. For the GATA-1 RNA probe, a DNA fragment corresponding to nucleotides 1 to 323 in the mouse GATA-1 cDNA (28) was synthesized by PCR (see below) and cloned in the opposite orientation into the *Xba*I and *Eco*RI sites of the SK+ vector. The plasmid was linearized with *Not*I, and RNA was synthesized with T3 polymerase. The human β-actin probe, which was kindly provided by Benoit Houle, was a 299-bp fragment, comprising 212 bp of exon 3 and 87 bp of intron 3, which was cloned by PCR into the *Sac*I and *Bam*HI sites of pGem (Promega). The plasmid was linearized with *Hin*dIII, and RNA was synthesized with T7 polymerase.

PCR synthesis and μ-locus constructs. To synthesize wild-type and β-CAAT mutants in the μ locus, which are described in the legend to Fig. 9 and Table 1, the sense strand of each oligonucleotide pair, which spans from −89 to −58 in the β-globin promoter and includes a *Bam*HI overhang on the 5′ end, was used in a PCR with the 900-bp wild-type β-*H-2K*^{k} fragment cloned in the SK+ vector (see above) and an oligonucleotide corresponding to the T7 promoter in the SK+ vector. After 30 PCR cycles (93°C for 2 min, 57°C for 2 min, and 72°C for 2 min for each cycle) the synthesized DNA fragments were cut with *Not*I and *Bam*HI and were cloned into the same sites of the SK+ vector. Clones were sequenced to ensure that only the mutation of interest was present. The β-globin promoter fragment was removed from one of these constructs by digestion with *Hin*dIII and was replaced by an Aγ-globin gene promoter fragment spanning the *Bal*I site at −130 to the *Nco*I site at +54, which includes the two upstream CAAT boxes but not the CACC box (Fig. 1). The cloning was done by filling in the *Hin*dIII and *Nco*I overhangs and results in +56 of the γ-globin gene being fused to the first 55 bp of *H-2K*^{k}.

The μ locus containing the β-globin LCR and unique *Cla*I and *Kpn*I cloning sites has been described previously (9). The single *Not*I site in this plasmid was removed by digestion with *Not*I and filling in with T4 polymerase. The *H-2K*^{k} gene was then cloned as a 5.5-kbp *Kpn*I fragment into the unique *Kpn*I site (this introduces a new *Not*I site which is located at the 5′ end of the *H-2K*^{k} gene [11]). The β-globin- and γ-globin-*H-2K*^{k} fragments in the SK+ vector were cloned into this μ-locus *H-2K*^{k} construct as *Cla*I-*Not*I fragments.

Quantitation assays. Protein concentrations were measured with the dye-binding kit from Bio-Rad Laboratories with BSA as a standard. The relative level of expression of β-*H-2K*^{k} (Table 2) in transfected MEL cells and Southern blots for copy number determination were quantitated by scanning autoradiograms using and LKB Ultra-Scan laser densitometer coupled to an LKB 2200 integrator. Autoradiograms exposed for different times were scanned to ensure the linearity of the response and the proper background substraction.

The different affinities of the *a*, *b*, CP1, and GATA-1 proteins for the wild-type and mutated CAAT sequences were estimated via GMS assays by using the oligonucleotide for which the affinity was to be determined as the unlabeled competitor, with either the mutant itself or the wild-type CAAT as the labeled probe, and varying the amounts of competitor over at least a 300-fold range with 1.5- to 2-fold serial dilutions. The amount of label associated with each protein band was determined by laser densitometry.

TABLE 2. Expression of CAAT mutants in MEL cells

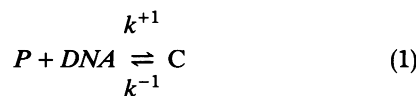
Construct	Relative expression[a] (copy no.) of population:			Avg expression[b] ± mean variation	Inducibility[c]
	A	B	C		
Wild type	9.3 (7.0)	16.0 (5.4)	4.4 (3.0)	9.9 ± 4.1 (9.4 ± 3.4)	16.3 ± 4.3
Mutant 1	10.2 (4.6)	8.5 (3.0)	6.5 (3.6)	8.4 ± 1.3 (10.9 ± 1.7)	12.5 ± 4.9
Mutant 2	13.8 (2.2)	10.4 (3.6)	7.4 (3.9)	10.5 ± 2.2 (17.9 ± 8.4)	13.4 ± 3.9
Mutant 3	9.8 (16.1)	5.8 (14.9)	3.6 (7.5)	6.4 ± 2.3 (2.4 ± 0.4)	15.2 ± 1.7
Mutant 5	1.6 (5.4)	0.8 (6.6)	0.6 (2.7)	1.0 ± 0.4 (1.0 ± 0.3)	12.9 ± 2.3
Mutant 7	1.2 (4.9)	0.6 (20.0)	1.0 (3.3)	0.9 ± 0.2 (0.9 ± 0.5)	15.8 ± 6.2

[a] Expression for each population in induced MEL cells, which was determined as the ratio of β-*H-2K*^{k} RNA to GATA-1 RNA by RNase protection assays (see Fig. 9). All values are relative to the average value of the three mutant 5 populations, which was arbitrarily given a value of 1.

[b] The value not corrected for copy number is given first, and the corrected value is given in parentheses. Again, all values are relative to the average of the mutant 5 populations.

[c] Average ratio for the three populations (± mean variation) of β-*H-2K*^{k} expression in DMSO-induced MEL cells to uninduced cells as determined by RNase protection assays (bottom panel in Fig. 9).

The reaction between a DNA-binding protein (P) and its cognate sequence to form a complex (C) can be written as



$$P + DNA \underset{k^{-1}}{\overset{k^{+1}}{\rightleftharpoons}} C \quad (1)$$

where k^{+1} and k^{-1} are the on-time and off-time rate constants, respectively. K_{dis}, the dissociation constant for such a reaction, is

$$K_{dis} = \frac{[C]}{[P]\,[DNA]} = \frac{k^{+1}}{k^{-1}} \quad (2)$$

and the affinity constant (K_{aff}) is defined as $1/K_{dis}$. The [P] equals $[P]_0 - [C]$, and under conditions where $[DNA] \gg [P]$, [DNA] equals $[DNA]_0$ at all times. Equation 2 can be rearranged to give

$$\frac{1}{[DNA]} = \frac{K\,[P]_0}{[C]} - K_{dis} \quad (3)$$

By varying the amount of DNA as the cold competitor and measuring [C] by scanning of the GMS gels, one can make plots of 1/[C] versus 1/[DNA] (Lineweaver-Burk plots). As can be seen in equation 3, when 1/[C] tends to 0 (i.e., the x-axis intercept), 1/[DNA] equals $-K_{dis}$ or $-1/K_{aff}$. In our hands, such plots were linear, with correlation factors in the 0.99 range. However, when working with crude extracts, excess amounts of nonspecific competitor must be added and corrections for the binding of nonspecific as well as other specific proteins cannot be made. Thus, the values obtained are likely much higher (that is, show lower affinity) than the actual values for K_{aff} expected for purified proteins. The relative values obtained under fixed assay conditions, however, should be valid with respect to one another.

The half-life of DNA-protein complexes were determined by plots of log [C] versus time from off-time analyses as previously described (18).

RESULTS

Novel proteins bind to the β-globin CAAT box. Previously, deBoer et al. (11) had shown that an oligonucleotide that encompasses the CAAT region of the human β-globin gene promoter interacts strongly with the ubiquitous CAAT-binding factor CP1 and weakly with the erythrocyte-megakaryocyte-specific factor GATA-1, which has been shown to be involved in the expression of a variety of erythrocyte-specific genes (6, 13, 19, 23, 29, 33). To confirm this result, GMS assays were performed using the β-CAAT oligonucleotide and nuclear extracts isolated from uninduced MEL cultures or from cells treated with DMSO for 48 to 72 h (henceforth referred to as induced cells). As controls for extract integrity and for the position of GATA-1 and CP1 on the gels, GMS assays were also performed with the strong GATA-1 (also weak OCT-1)-binding C oligonucleotide from the β-globin gene 3′ enhancer (31) (Fig. 2a, lanes 1 and 2) and with the strong CP1 (also weak GATA-1)-binding γ-CAAT oligonucleotide from the human γ-globin gene (lanes 3 and 4)

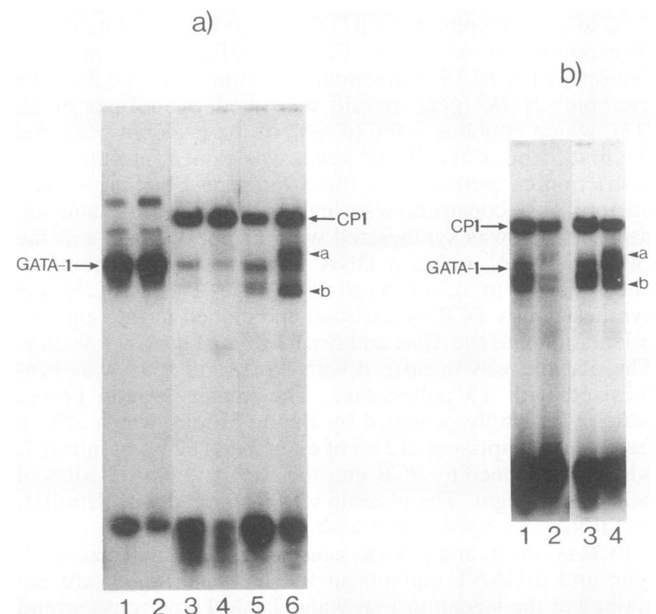


FIG. 2. Four different factors interact with the β-globin CAAT box region. (a) Nuclear extracts from untreated MEL cells (odd-numbered lanes) or from cells treated with DMSO for 48 h (even-numbered lanes) were assayed by GMS with ^{32}P-labeled C oligonucleotide (lanes 1 and 2), γ-CAAT oligonucleotide (lanes 3 and 4), or β-CAAT oligonucleotide (lanes 5 and 6). The extract concentration was 0.4 mg of protein per ml in the assay, and the position of DNA-protein complexes described in the text are indicated on the sides of the radiogram. (b) Uninduced (lanes 1 and 3) or induced (lanes 2 and 4) MEL extracts were assayed as in panel a with the β-CAAT oligonucleotide, except that BSA was omitted from the reaction mix in lanes 1 and 2.

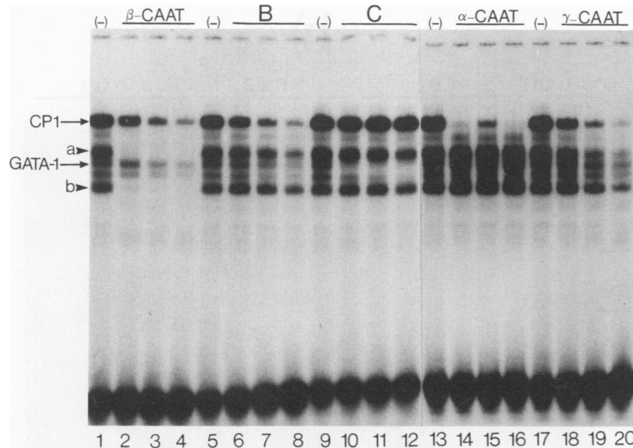


FIG. 3. Factors *a* and *b* are not related to GATA-1 or CP1. GMS assays were performed as for Fig. 2a with extracts from induced MEL cells and with the β-CAAT oligonucleotide as the ^{32}P-labeled probe, without additions [lanes (−)] or in the presence, from right to left, of a 10-, 30-, or 100-fold excess of the unlabeled oligonucleotide competitor indicated on the top of each set of lanes. The positions of specific DNA-protein complexes seen on the radiogram are indicated. Note that the band seen migrating between the position of GATA-1 and *b* shows the same competition pattern as GATA-1; hence, it may be related to this protein.

(3). Not only were CP1 and GATA-1 detectable with the β-CAAT oligonucleotide (Fig. 2a, lanes 5 and 6), two additional labeled DNA-protein complexes were observed (denoted *a* and *b*). Complex *a* was particularly intriguing, as it was easily detectable with induced MEL extracts (lane 6) but was just visible when extracts from untreated cells were used (lane 5; also see Fig. 4 and 7a). As expected, the γ-CAAT oligonucleotide gave a strong CP1 band and a weak GATA-1 band in the GMS assays. However, this probe also gave a very weakly labeled band migrating at the *b* position, although no band at the *a* position was observed. On the other hand, the C oligonucleotide, which gave rise to an intense GATA-1 band and weak OCT-1 band as expected, did not give rise to any detectable bands at the *a* or *b* position.

Why were the *a* and *b* bands not detected in previous results (11)? One alteration that had been made in the GMS assay protocol since this earlier report was that 0.1% BSA was now included in the reaction buffer. In Fig. 2b it is demonstrated that the inclusion of BSA was at least partially responsible for the detection of complexes *a* and *b*. If BSA was omitted, although band *b* could be observed with the uninduced extract (lane 1), with induced extracts only faint smears were seen at the positions of the *a* and *b* bands (lane 2), whereas distinct bands were seen when BSA was included (lanes 3 and 4). In later experiments, it was found that BSA was not required if the assays were done at a high extract concentration (>0.8 mg of protein per ml). However, BSA did always improve the signals in terms of clarity and reproducibility from one extract to the next.

Note that with the extract used in Fig. 2b, a band migrating between the positions of GATA-1 and *b* was also observed. This band was seen only with some extract preparations, and competition assays (see below) suggested that it may be related to GATA-1 (Fig. 3).

***a* and *b* represent sequence-specific DNA-binding proteins.** To characterize the *a* and *b* proteins, competitions in GMS assays were performed. Both the *a* and *b* complexes were easily inhibited by addition of an excess of unlabeled β-CAAT oligonucleotide (Fig. 3, lanes 1 to 4). However, the strongly GATA-1-binding C oligonucleotide and the strongly CP1-binding α-CAAT oligonucleotide (7) inhibited only GATA-1 and CP1, respectively (lanes 9 to 16). They did not inhibit the *a* and *b* DNA-protein complexes even when they were present at a 100-fold molar excess. On the other hand, the B oligonucleotide from the human β-globin gene 3′ enhancer (31) and the γ-CAAT oligonucleotide, both of which bind to CP1 strongly and to GATA-1 weakly, did partially compete against the interaction between the *a* and *b* proteins and the β-CAAT probe (lanes 5 to 8 and lanes 17 to 20). However, this competition was at least 10-fold weaker than that seen with the β-CAAT oligonucleotide itself, since a 100-fold excess of either the B or γ-CAAT oligonucleotide reduced binding of the *a* and *b* proteins only by approximately 50% (lanes 8 and 20), whereas a 10-fold excess of the β-CAAT oligonucleotide was sufficient to reduce both bands to near background levels (lane 2).

Several ubiquitous proteins have been described to interact with CAAT box sequences: CP1, CP2, and NF-1 (7). C/EBP has also been shown to bind CAAT motifs; however, the protein is liver specific (4). There have been at least two transcription factors which have been observed to increase during chemically induced differentiation of MEL cells: the CAAT-binding factor CP2 (17) and NF-E2, an erythroid-specific factor with the same DNA-binding sequence specificity as AP1 (20, 21). In Fig. 4, we investigated any relationships between these proteins and complexes *a* and *b*.

Whereas the β-CAAT oligonucleotide again inhibited binding of the *a* and *b* proteins to itself (Fig. 4a, lanes 1 to 4), a CP2-binding oligonucleotide from the rat fibrinogen gene (7) (lanes 5 and 6) or an NF-E2-binding oligonucleotide from the human porphobilinogen deaminase gene (18) (lanes 7 and 8) did not inhibit any complex formed with the β-CAAT oligonucleotide. On the other hand, when either of these oligonucleotides was used as the radiolabeled probe in GMS assays, it efficiently inhibited its own binding but was not at all inhibited by the β-CAAT oligonucleotide (Fig. 4b, lanes 3 to 10). In similar experiments, specific DNA sequences which bind NF-1, SP-1, or OCT-1 did not inhibit binding of any proteins to the β-CAAT sequences (not shown). Thus, the proteins responsible for formation of the *a* and *b* complexes are highly sequence specific in their interaction with DNA, and they show no similarity to the factors described above (CP1, CP2, GATA-1, NF-E2, AP1, NF-1, OCT-1, and SP-1).

On the other hand, a 100-fold excess of unlabeled β-CACC oligonucleotide reduced the β-CAAT *a* complex considerably and also inhibited the β-CAAT complex *b* slightly, but it had no effect on GATA-1 or CP1 (Fig. 4a, lane 10). Moreover, when the β-CACC oligonucleotide was used as the probe in GMS assays, besides several labeled bands that are collectively referred to as CACC-binding factors (11, 32) (note that the most slowly migrating band may be SP-1), there was also a labeled band in the induced MEL sample, migrating near the same position as complex *a*, that was not obvious in the uninduced-MEL lane (Fig. 4b, compare lanes 11 and 12). This band was specifically inhibited by the β-CAAT oligonucleotide (lane 13) as well as by the β-CACC oligonucleotide (lane 14). Thus, the proteins in complex *a* also bind the β-CACC region. However, no band near the position of complex *b* was observed with the β-CACC probe, although the competition assays did suggest that β-CACC does bind the complex *b* protein(s) weakly.

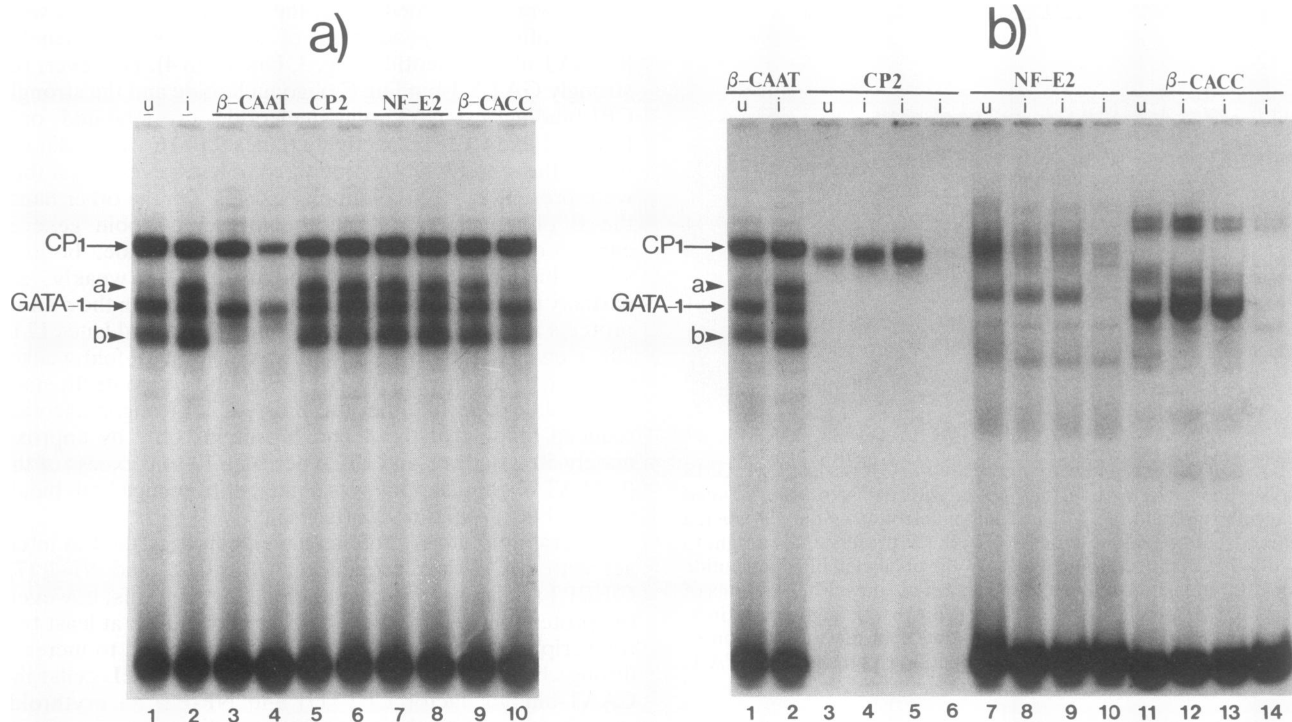


FIG. 4. *a* and *b* are not related to CP2 or NF-E2. (a) GMS assays were performed with uninduced (lane 1) or induced (all other lanes) MEL extracts with the β-CAAT oligonucleotide as the ^{32}P-labeled probe, with no additions (lanes 1 and 2) or with a 10-fold (first lane of each pair) or 100-fold (second lane of each pair) excess of the unlabeled oligonucleotide indicated at the top of each pair of lanes. Lane u, uninduced control, and lane i, induced control. (b) Uninduced (lanes u) or induced (lanes i) MEL extracts were used in GMS assays with the oligonucleotide indicated at the top of each set of lanes as the ^{32}P-labeled DNA probe. In lanes 5, 9, and 13, a 100-fold excess of unlabeled β-CAAT oligonucleotide was added as the competitor, while in lanes 6, 10, and 14, a 100-fold excess of CP2, NF-E2, or β-CACC oligonucleotide (i.e., the same oligonucleotide as was used as the probe), respectively, was added as the competitor.

GATA-1, CP1, and the *a* and *b* factors bind overlapping sequences in the CAAT box. Since the above results demonstrated that the proteins responsible for complexes *a* and *b* bind strongly to the β-CAAT region, weakly to the β-CAAT sequences, and even more weakly to the B and γ-CAAT oligonucleotides, we did a comparison between these four sequences to determine whether they are related. From this comparison we derived a consensus sequence shared among these oligonucleotides (Fig. 5). Emphasis was put on the β-CAAT sequence in determining this common sequence, as it was the only probe which clearly showed both the *a* and *b* bands in GMS assays. To determine whether this shared sequence may be related to binding of the *a* and/or *b* protein, methylation interference assays were performed to determine where each of the factors binds on this DNA fragment.

As expected, binding of the CP1 factor was strongly inhibited by methylation of residues centered around the CAAT sequence (Fig. 6, lanes 1 and circles). Binding of the GATA-1 protein (lanes 3 and squares) was centered near CCAATCTA, a sequence which matches 7 of 8 bases of the GATA-1 consensus (C/APyT/AATCT/APy [31]). These results showed that even though the methylation interference technique was made difficult by the close proximity of the various bands on the gels and the weak intensity of some of the bands, the data were still adequate to distinguish the areas of binding of the factors. Both complex *a* (lanes 2 and triangles) and complex *b* (lanes 4 and diamonds) binding was inhibited most strongly by methylation of some G and A residues within the derived consensus sequence (boxed area), although binding did extend slightly upstream of this region. These results establish that the consensus derived in Fig. 5 is related to the binding specificity of factors *a* and *b*. However, this must be considered with caution since the data are presently limited to one strong, one weak, and two feeble binding sites for the *a* and *b* factors. The methylation interference results also show that although all four factors (*a*, *b*, GATA-1, and CP1) bound overlapping sequences, there were also some differences in the methylation interference pattern between each of them.

β-CAAT: A**GGGTTG**G**CCAAT**CTACTCCCAGGAGC

γ-CAAT: T**G**CC**TTGA**CC**AAT**AGCCTTGACAAGGC

β-CACC: G**GTGTGG**C**TCCAC**AGGGTGAGG

B-OLIGO: C**GGGTGG**GC**CTAT**GATAGGGTAATAAG

CONSENSUS: **G N G T T G N C C N A T**
G T C

FIG. 5. Consensus for binding of *a* and *b*. The residues used in determining the consensus shown on the bottom of the figure are indicated in boldface in the top part of the figure. N, any residue.

The LCR can function with GATA-1 or complexes *a* and *b* but not with CP1 alone. On the basis of the consensus data, methylation interference data, and the known DNA sequence specificities of CP1 and GATA-1 (7, 31), mutations in the β-CAAT sequences were derived. Each mutant sequence was tested with extracts from induced MEL cells by

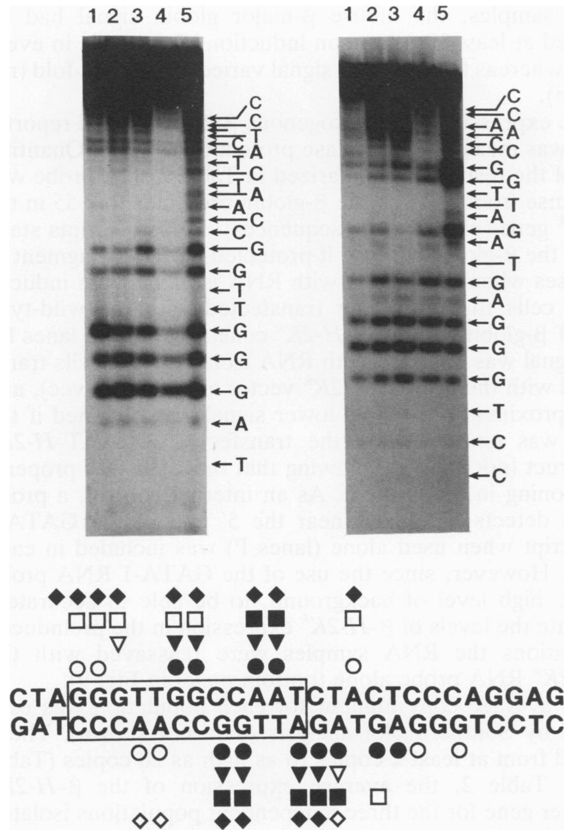


FIG. 6. Methylation interference assays. The β-CAAT oligonucleotide sense strand (left panel) or antisense strand (right panel) was ^{32}P labeled, methylated, annealed to its partner, and then used in GMS assays with nuclear extracts from induced MEL cells. The radioactive CP1 (lanes 1), *a* (lanes 2), GATA-1 (lanes 3), and *b* (lanes 4) DNA-protein complexes and unbound probe (lanes 5) were excised from the gel, and the DNA was cleaved at methylated G and A residues and then run on a sequencing gel, a radiogram of which is shown at the top of the figure. At the right of each panel the DNA sequence is shown. The data are summarized in the bottom part of the figure. Circles, CP1; triangles, *a*; squares, GATA-1; diamonds, *b*. Closed and open symbols represent strong and weak interactions, respectively. The consensus sequence derived in Fig. 5 is boxed.

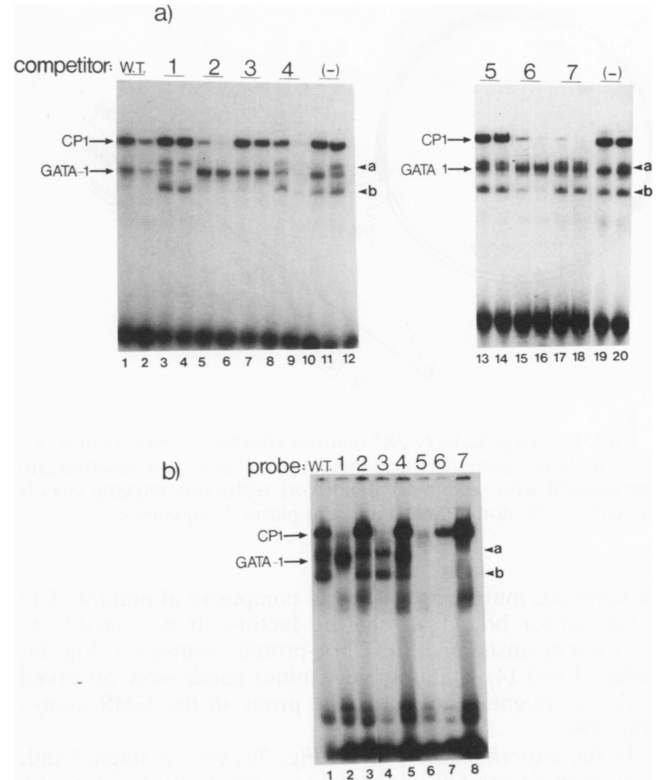


FIG. 7. Mutations in the β-globin CAAT box and their effect on protein binding. (a) Uninduced (lanes 11 and 19) or induced (all remaining lanes) MEL extracts were used in GMS assays with the ^{32}P-labeled β-CAAT oligonucleotide probe with no additions [lanes (−)] or with a 10-fold (first lane of each pair) or 100-fold (second lane of each pair) excess of the unlabeled competitor oligonucleotide indicated at the top of each pair of lanes. The sequences of the wild-type (W.T.) and mutant β-CAAT oligonucleotides are shown in Table 1. (b) The oligonucleotide indicated at the top of each lane was used as the ^{32}P-labeled probe in GMS assays with extracts from induced MEL cells.

competition assays against the wild-type β-CAAT oligonucleotide (Fig. 7a), as well as by using each directly as the ^{32}P-labeled probe in the GMS assays (Fig. 7b). The results are summarized in Table 1. Identical results were obtained with noninduced MEL cells, except for the low level of the *a* factor binding activity with such extracts (not shown).

In mutant 1, four nucleotides were altered to retain the GATA-1 consensus and to remove the binding of CP1 and both complexes *a* and *b* on the basis of their methylation interference patterns. This mutant competed only against GATA-1 binding, and only a single, intense band, migrating at the GATA-1 position, was seen with this mutant as the probe (Fig. 7b). Although the intensity of the GATA-1 band seen with mutant 1 versus wild-type CAAT in Fig. 7b would suggest the affinity of GATA-1 to be vastly stronger for mutant 1 than for wild-type CAAT, with some less active extracts the difference was not as great as in Fig. 7b and measurement of relative affinity constants gave only a four- to fivefold-stronger affinity for mutant 1 (data not shown). With the extract used in Fig. 7b, most of the wild-type CAAT probe is bound by CP1, *a*, and *b*, which inhibit GATA-1 binding.

The conversion of CT to GA in mutant 2 destroys the GATA-1 consensus, but it is outside of the consensus derived for the *a* and *b* factors (Fig. 5) and retains the CP1 consensus as is found in the human α-globin gene promoter CAAT region (7). The assays (Fig. 7) did show that this mutant binds CP1 and the *a* and *b* factors but does not interact with GATA-1. The T-to-C conversion in mutant 3 (Table 1) destroys the consensus for both CP1 and GATA-1. However, the consensus derived for the *a* and *b* factors suggested that this position can be either a T or a C. In agreement, only bands corresponding to factors *a* and *b* were clearly evident when mutant 3 was used as the labeled probe in GMS assays (Fig. 7b). However, in some experiments, a very slight trace of CP1 binding to mutant 3 was still evident (see Fig. 10e for an example), and when this mutant was used at high levels of competitor against the wild-type CAAT probe, it was found to retain some binding capacity for CP1, but with an at least 10-fold-reduced affinity over the wild-type sequences. Thus, it is indicated as not binding CP1.

The mutant 4 oligonucleotide still bound to all four proteins, but it did appear to have a reduced affinity for factors *a* and *b* (compare competition in lane 9 to lane 1 of Fig. 7a).

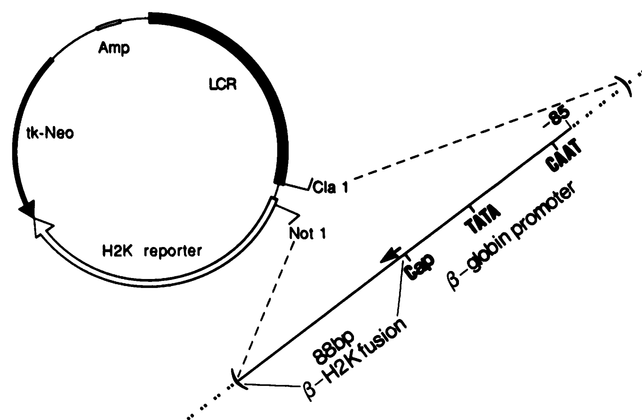


FIG. 8. The μ-locus *H-2K*^{k} reporter construct. The way in which the mutated β-promoter fragments (in parentheses) are inserted into the plasmid with unique *Cla*I and *Not*I restriction enzyme sites is indicated. The dotted lines represent plasmid sequences.

In contrast, mutant 5, which is a composite of mutants 1 to 4, no longer bound any of the factors. It was unable to compete against any of the DNA-protein complexes (Fig. 7a, lanes 13 and 14), and only very minor bands were observed with this oligonucleotide as the probe in the GMS assays (Fig. 7b).

In the experiment shown in Fig. 7b, only a single band, corresponding to CP1, is clearly evident with the mutant 6 oligonucleotide. However, in some experiments bands at the *a* and *b* positions were also visible with this mutant probe, and the competition assays suggested that this mutant still retains some affinity for the *a* and *b* factors (Fig. 7a, lanes 15 and 16). On the other hand, mutant 7, which was designed on the basis of the strongly CP1-binding α-globin CAAT sequence (see Fig. 3), was clearly able to compete only against CP1 binding to the wild-type β-CAAT sequence (Fig. 7a, lanes 17 and 18), and it gave rise to only a single, strong CP1 band when it was used as the probe in the assay system (Fig. 7b).

Thus, it was possible to derive mutations which bound either none of the factors (mutant 5), only GATA-1 (mutant 1), only CP1 (mutant 7), or only complexes *a* and *b* (mutant 3), results which again demonstrated that factors *a* and *b* are distinct from GATA-1 and CP1. However, we were unable to design a mutant which bound to complex *a* or to complex *b* separately. In fact, mutant 4 was an attempt to do this since the two G residues which were converted to T residues in this mutant had been shown to interact strongly with only complex *b* by the methylation interference assays (Fig. 6). However, binding of both *a* and *b* was reduced significantly (Fig. 7).

We next investigated whether some of the mutant sequences, in conjunction with the LCR, could direct high levels of inducible expression in MEL cells. The positive strand of each oligonucleotide pair was used in a PCR to synthesize from the 5′ CAAT promoter region to +32 in the β-globin gene, fused to the 5′ end of the mouse *H-2K*^{k} structural gene. The PCR products were transferred in front of the complete *H-2K*^{k} reporter gene within the μ-locus construct (Fig. 8) (9). Each construct was transfected into MEL cells in triplicate, and independent, stable populations were selected for resistance to G418. For each population, total RNA was isolated from untreated cells and from cultures induced for 3 to 4 days. On Northern blots of the RNA samples, the mouse β-major globin signal had increased at least 10-fold upon induction of the cells in every case, whereas the GATA-1 signal varied less than 2-fold (not shown).

The expression of the exogenous β–*H-2K*^{k} fused reporter gene was measured by RNase protection (Fig. 9). Quantitation of the results is summarized in Table 2. The probe was antisense from −800 in the β-globin promoter to +55 in the *H-2K*^{k} gene (plus plasmid sequences). For transcripts starting at the β-globin cap site it protected an RNA fragment of 87 bases when used alone with RNA isolated from induced MEL cells that had been transfected with the wild-type CAAT β-globin promoter–*H-2K*^{k} construct (Fig. 9, lanes P). No signal was detected with RNA isolated from cells transfected with the μ-locus *H-2K*^{k} vector alone (lanes vec), and an approximately 100-fold-lower signal was obtained if the LCR was omitted from the transfected β-CAAT–*H-2K*^{k} construct (not shown), showing that the LCR was properly functioning in this system. As an internal control, a probe which detects 323 bases near the 5′ end of the GATA-1 transcript when used alone (lanes P) was included in each assay. However, since the use of the GATA-1 RNA probe gave a high level of background, to be able to accurately estimate the levels of β–*H-2K*^{k} expression in the preinduced populations the RNA samples were reassayed with the β–*H-2K*^{k} RNA probe alone (bottom panel in Fig. 9).

The average copy number for each population was estimated by Southern blot analysis (data not shown). These ranged from at least 2 copies to as high as 20 copies (Table 2). In Table 2, the average expression of the β–*H-2K*^{k} reporter gene for the three independent populations isolated for each construct is given both uncorrected and corrected for copy number. As previously noted using the MEL system (27), we found that populations with the highest copy numbers usually also had the lowest expression per copy. A comparison of population B with populations A and C for mutant 7 in Table 2 is a good example of this phenomenon. The interpretation of the results below is based on expression levels per integrated copy. However, where the conclusions may be partially influenced by the actual copy number is also discussed below.

In MEL populations carrying the wild-type CAAT-LCR construct, there was a low level of β–*H-2K*^{k} RNA expression in pretreated cultures that was increased about 16-fold with DMSO treatment (Fig. 9 and Table 2). In contrast, populations harboring mutant 5, which does not bind CP1, GATA-1, or the *a* and *b* factors, gave about ninefold-lower levels of β–*H-2K*^{k} RNA expression per copy of the plasmid in both preinduced and induced cultures. Note, however, that mutant 5 and the wild-type construct appeared to give the same relative level of inducibility of expression with DMSO treatment (13- to 16-fold). Thus, factor(s) binding to the β-globin promoter CAAT box region can substantially increase the levels of transcription driven by the LCR, and this effect appears to be equally high in both preinduced and induced MEL cultures.

Mutant 1, which binds to GATA-1 alone, and mutant 2, which binds to CP1, *a*, and *b* but not GATA-1, both gave levels of expression per gene copy that were at least as high as that seen with the wild-type CAAT construct (Fig. 9 and Table 2). In each case, β–*H-2K*^{k} RNA expression was at least 10-fold higher than that observed with the mutant 5 construct, and expression increased 12- to 14-fold with DMSO treatment. Although mutant 2 appears to have close to twofold-higher levels of expression than either the wild-type or mutant 1 construct, it should be noted that the three

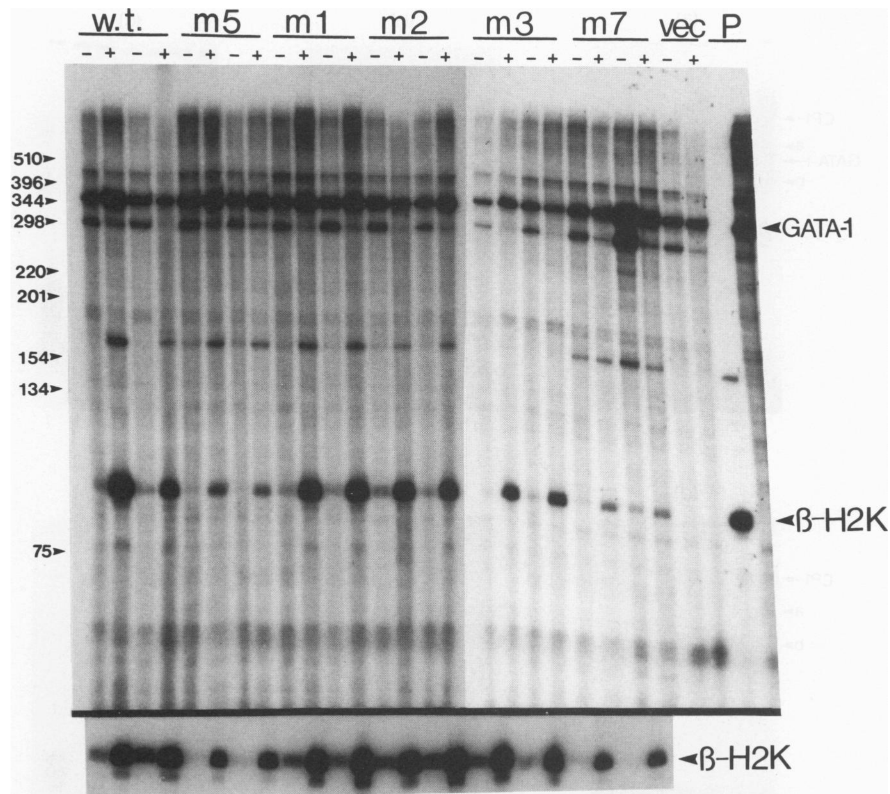


FIG. 9. RNase protection assays of transfected MEL cells. Total RNA from uninduced (−) or induced (+) G418-resistant populations of MEL cells that had been transfected with the μ-locus–*H-2K*^{k} reporter construct, with the β-globin promoter containing the CAAT box mutation (Table 1) indicated at the top of each set of lanes, was assayed by RNase protection with a mixture of probes that detect a 330-base GATA-1 RNA fragment and an 87-base β–*H-2K*^{k} fusion RNA fragment. Both probes were labeled with a specific activity of 200 Ci of [α-^{32}P]UTP per mmol. In lanes P, each of the probes was used separately with RNA isolated from induced cells that had been transfected with the wild-type β-CAAT globin promoter construct. The signal detected with each of the probes is indicated by the arrowheads. In the bottom panel, RNA samples that were estimated to have the same amount of GATA-1 in independent experiments were assayed by RNase protection with the β–*H-2K*^{k} probe alone. The lanes are as indicated in the corresponding lanes of the top panel. Note that each pair of uninduced and induced lanes represents a completely independent population of transfected MEL cells.

mutant 2 populations had the lowest overall average copy number for all constructs tested. Thus, this difference may not be significant within the experimental variation. For example, if expression is considered without copy number correction, the wild-type, mutant 1, and mutant 2 populations all give approximately equivalent average levels of expression within error (Table 2).

The mutant 3 construct, which only binds the *a* and *b* factors in the CAAT region, also gave higher levels of expression per gene copy than mutant 5, although in this case the stimulation was only slightly higher than twofold. Since the mutant 3 populations had the highest overall average copy number of all constructs tested, this value may be arbitrarily low. Without copy number correction, the expression levels for mutant 3 are closer to the wild-type levels (Table 2). On the other hand, mutant 7, which retains only CP1 binding, resulted in low levels of expression of the β–*H-2K*^{k} mRNA that were close to equivalent to those resulting from mutant 5. Yet net inducibility was again retained.

Expression correlates with the stability of DNA binding. The expression results from MEL cells showed that mutant 3 is much less active than mutant 2. In terms of protein binding, the only difference observed between these two sequences was the lack of binding of CP1 to mutant 3 (Table 1). Yet the results with mutant 7 suggested that CP1 does not play a role in transcriptional activation. In seeking alternative explanations, the affinity constants of the *a*, *b*, CP1, and GATA-1 factors for the various sequences to which they bind were determined (see Material and Methods). However, the affinity constants for binding of *a*, *b*, and/or CP1 to the wild-type, mutant 2, mutant 3, and/or mutant 7 sequences were all found to be within twofold of one another. As was noted above, the only highly significant difference seen was the four- to fivefold-increased affinity GATA-1 has for mutant 1 versus wild-type CAAT. We next determined whether the off rates were similar or significantly different. In an off-time analysis, extracts are preincubated with labeled DNA, as in normal GMS assays, then a large excess of competitor DNA is added, and the disappearance of protein-DNA complexes is monitored over time by loading aliquots onto a continuously running gel (18). Thus, this analysis tests for the stability of specific DNA-protein complexes once they are formed.

As is shown in Fig. 10a, binding of GATA-1 to the wild-type CAAT element was very unstable. In several experiments performed, no GATA-1 binding was discernible after more than 1 min of chase and with different exposures of the autoradiograms. In contrast, factors *a*, *b*, and CP1 bound to wild-type CAAT with a much higher stability than

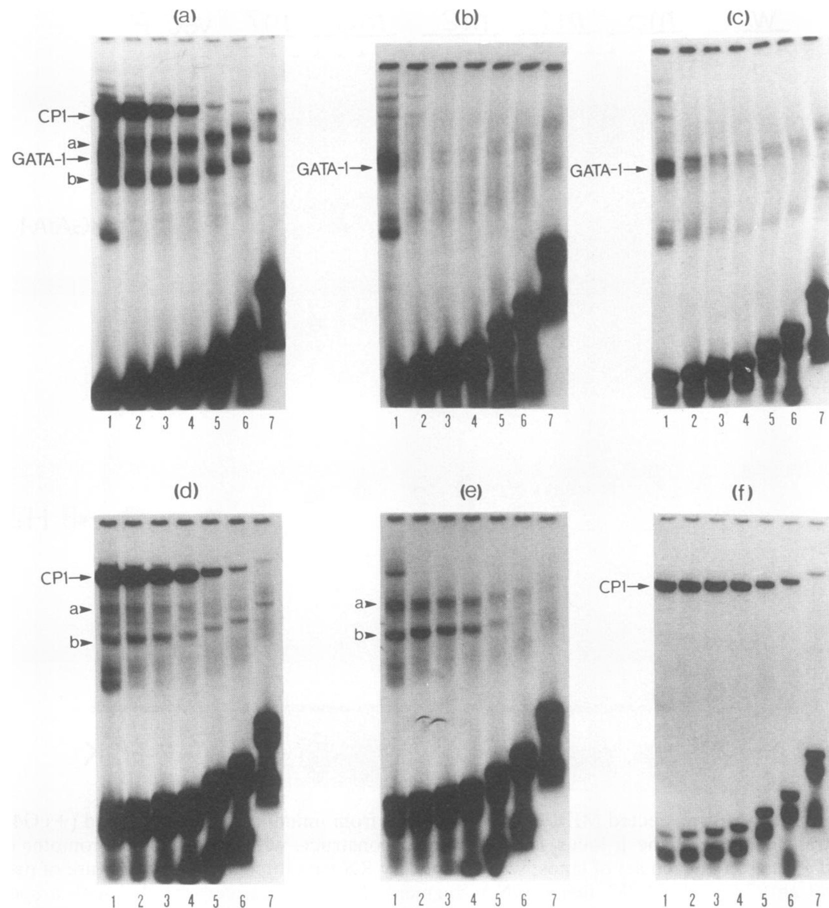


FIG. 10. Off-time analysis of DNA binding. Extracts were preincubated with ^{32}P-labeled probe [a] wild-type CAAT; [b and c] mutant 1; [d] mutant 2; [e] mutant 3; [f] mutant 7), then an excess of unlabeled probe was added, and aliquots were loaded onto a continuously running gel at different times (lanes 1, zero time before competitor was added; lanes 2, 3, 4, 5, 6, and 7, 1, 3, 6, 15, 25, and 50 min, respectively. In panel c, Mutant 3 was used as the unlabeled competitor, while in all other panels the wild-type CAAT sequence was used. Note that panel a shows an overexposed autoradiogram to demonstrate the lack of any detectable GATA-1 after 1 min of competition.

GATA-1, with a large percentage of the *a* and *b* complexes remaining even after 50 min of chase (Fig. 10a). The measured half-lives of these DNA-protein complexes, the time required to remove 50% of binding (18), were 18, 19, and 3.5 min for *a*, *b*, and CP1, respectively.

Binding of GATA-1 to mutant 1 was more stable than binding to wild-type CAAT (Fig. 10b). Although there was a large decrease in binding after only 1 min of competition, a small amount of GATA-1 remained bound after this time, and this binding actually remained stable throughout the 50 min of competition. In fact, the initial decrease in the GATA-1 band appeared to reflect nonspecific binding under the assay conditions of running the gels at 4°C, since nonspecific competitors also gave a similar decrease. As an example, a competition of the mutant 1 probe with mutant 3, which does not bind GATA-1 under normal assay conditions (Fig. 7), is shown in Fig. 10c. Thus, the specific portion of the interaction of GATA-1 with the mutant 1 probe appears to be very stable.

Binding of CP1 to mutants 2 and 7 was more stable than that to the wild-type CAAT, with half-lives of 8 and 13 min versus 3.5 min, respectively (compare Fig. 10d and f with panel a; also note the very low and very unstable binding of CP1 to mutant 3 in Fig. 10e). On the other hand, the stability of binding of the *a* factor to both mutants 2 and 3 was much less than that of binding to the wild-type sequences (compare Fig. 10d and e with panel a), with half-times of only 4 and 5 min versus 18 min, respectively. In contrast, binding of the *b* protein to mutant 2 gave a stability similar to that seen with the wild-type CAAT, with a large percentage of the *b* complex remaining after 50 min of chase in both cases, whereas binding of *b* to mutant 3 was relatively unstable (compare panel e in Fig. 10 with panels a and d), with an average half-life of only 3.5 min for the *b*-mutant 3 complex. Taken together, as is discussed below, the stability studies suggest that the lower activity of mutant 3 is due to a decrease in the stability of binding of the *b* protein and that it is this factor that functions in the context of the wild-type sequences.

The β-globin CAAT box does not function in K562 cells. Contrary to the situation with MEL cells, when β-*H-2K*^{k} expression was compared with human actin expression in transfected populations of K562 cells, the wild-type β-globin CAAT construct gave no increase in expression over that seen with the mutant 5 construct (Fig. 11). In both cases, the β-*H-2K*^{k} RNA levels were almost undetectable and these low levels of expression did not appear to be affected by hemin treatment of the cells. On Northern blots, all K562

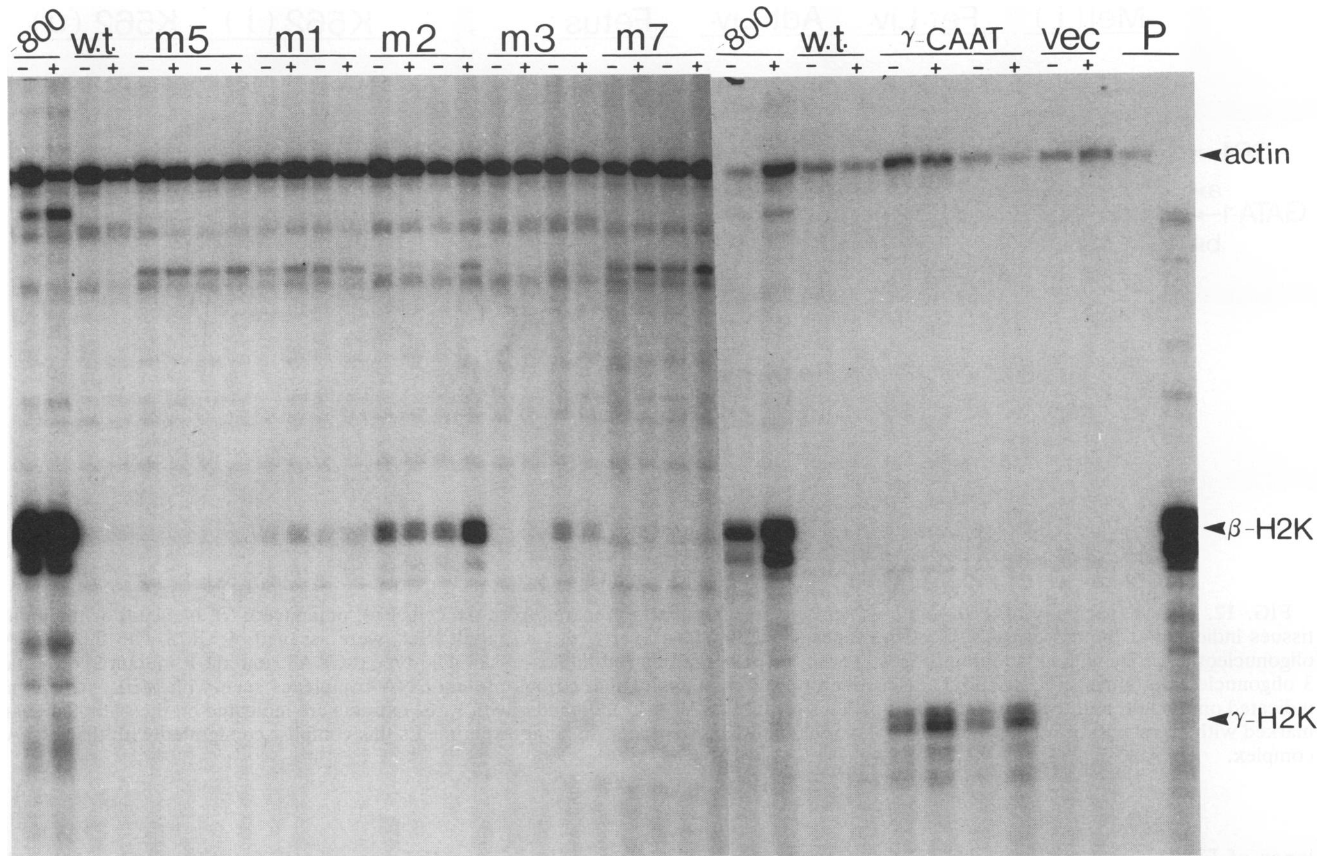


FIG. 11. RNase protection assays of K562 transfectants. RNA (20 μg) from K562 populations transfected with the μ-locus construct indicated at the top of each set of lanes was assayed by RNase protection with a mixture of probes for β-*H-2K*^{k} (labeled with [α-^{32}P]UTP at 3,000 Ci/mmol) and human β-actin (labeled with [α-^{32}P]UTP at 2 Ci/mmol). In lanes P, the actin probe was used alone with RNA from an induced −800 K562 population or the β-*H-2K*^{k} probe was used alone with an induced MEL population transfected with the wild-type β-CAAT construct. The position of the signal from each probe is indicated by the arrowheads. The copy numbers for all K562 populations shown were within twofold of one another, and cells transfected with vector alone gave no discernible signal with the β-*H-2K*^{k} probe (lanes vec). −, untreated cells; +, cells treated with 80 μM hemin for 3 days.

populations tested showed an approximately fivefold induction of the endogenous γ-globin mRNA with hemin treatment and all K562 populations had the same relative copy number within twofold (data not shown). In contrast, when the β-globin gene promoter was extended to −800, such that the CACC and further upstream sequences were included (Fig. 1), high levels of β-*H-2K*^{k} expression were obtained in K562 cells (Fig. 11). Moreover, a μ-locus-*H-2K*^{k} construct which contained the Aγ-globin gene promoter, deleted just upstream of the duplicated CAAT box region, thus excluding the CACC box for this promoter (Fig. 1), also gave high levels of expression in the K562 system (Fig. 11; note that the β-*H-2K*^{k} RNase protection probe detects only the homologous 55-base *H-2K*^{k} portion of the γ-globin-*H-2K*^{k} mRNA fusion).

Since the above results indicated that the β-globin CAAT box specifically is unable to support transcriptional activation by the LCR in K562 cells, we investigated whether some of the mutations might be able to rescue this nonfunctionality in the K562 system (Fig. 11). As was anticipated, mutant 7 did not stimulate transcription over the levels seen with the β-globin wild-type CAAT and mutant 5 constructs. On the other hand, mutants 1 and 2 and one of the mutant 3 populations, all of which gave full or partial activity in MEL cells (Table 2), did give some stimulation over the wild-type β-CAAT sequences in K562 cells, with mutant 2 appearing to give slightly higher stimulation than mutant 1 or 3. Because of the nearly undetectable levels of transcription observed with the wild-type, mutant 5, and mutant 7 β-globin sequences in K562 cells, it was not possible to quantitate the level of stimulation afforded by mutants 1, 2, and 3 in these cells. However, it should be realized that the β-*H-2K*^{k} RNase protection probes used for the assays of K562 transfectants in Fig. 11 were labeled at a 15-fold-higher specific radioactivity than those used to assay MEL transfectants in Fig. 9. By assaying samples of MEL transfectants and K562 transfectants together, we estimate that the full-length (−800) β-globin and the Aγ-CAAT constructs are expressed in K562 cells at levels about fivefold lower than that of the wild-type β-globin CAAT construct in induced MEL cells. For example, population B of MEL cells transfected with the wild-type β-CAAT construct (Fig. 9) was used for the β-*H-2K*^{k} probe lane in Fig. 11 (lanes P) for comparative purposes. On the other hand, the average expression of the K562 populations harboring the mutant 2 β-CAAT construct was at least 10-fold lower than that of either the −800 or γ-CAAT constructs in K562 cells, as was measured relative to the actin control. Note that the actin control in the first 24

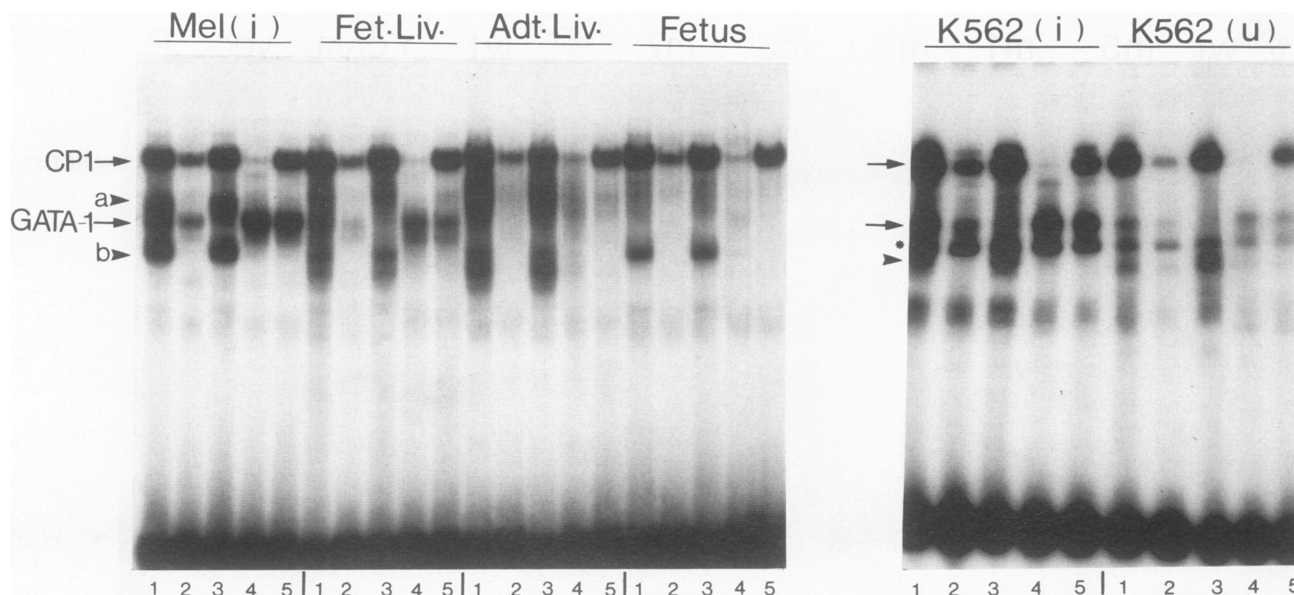


FIG. 12. Tissue specificity of the *a* and *b* proteins. Nuclear extracts prepared from the cells [(u), uninduced; (i), induced] or the mouse tissues indicated at the top of each set of five lanes (Fet.Liv., fetal liver; Adt.Liv., adult liver) were assayed by GMS with the β-CAAT oligonucleotide probe with no additions (lanes 1) or in the presence of a 100-fold excess of wild-type β-CAAT, mutant 1, mutant 2, or mutant 3 oligonucleotide (lanes 2, 3, 4, and 5, respectively). The positions of the various protein-DNA complexes seen with MEL extracts are indicated on the left, and the corresponding positions for CP1, GATA-1, and band *b* with K562 extracts are indicated on the right. The band marked with an asterisk is a nonspecific complex seen with K562 extracts. In some experiments this complex comigrated with the GATA-1 complex.

lanes of Fig. 11 was overexposed to demonstrate the very low levels of expression of the β-globin CAAT constructs in K562 cells. Thus, we suggest that the stimulation by mutants 1, 2, and 3 over the wild-type β-CAAT sequences in K562 cells does not result in full activation. In other words, the rescue is very partial, at best, compared with the full activity of the γ-CAAT and −800 β-globin promoters.

Tissue specificity of the β-CAAT *a* and *b* complexes. In GMS assays of nuclear extracts derived from HeLa cells, L cells, and neuronal N2A cells, bands corresponding to *a* and *b* were not evident, although CP1 could be easily detected (data not shown). Although this may have been related to extract quality (in our hands it is more difficult to get high-quality extracts from adherent cell lines) and/or a possible tissue specificity of the *a* and *b* proteins, since complex *a* was more abundant in terminally differentiated MEL cells, it was deemed that a comparison with undifferentiated, transformed cell lines may not be representative of the true situation. Thus, to compare terminally differentiated tissues, nuclear extracts were prepared from 14.5-day fetal mouse livers (hematopoietic and hepatic cells), 14.5-day mouse fetuses (with the livers removed), and adult mouse livers and were compared with induced MEL extracts by GMS assays. All three tissue extracts produced a protein-DNA complex that comigrated with complex *b* observed with the MEL extracts (Fig. 12). In all cases, these bands showed the same competition profile as was seen for the *b* complex in MEL extracts. Similar results were also obtained with extracts from the heart, kidney, and spleen (not shown). Thus, the protein(s) responsible for complex *b* is not tissue specific.

At the position of band *a* seen with induced MEL extracts, two faint labeled bands were observed with the fetal liver extracts. Although these bands are faint in comparison with those seen with MEL extracts, it should be realized that the fetal liver is a mixture of hepatic and hematopoietic cells. For example, the erythrocyte-megakaryocyte-specific GATA-1 band in the fetal liver lanes is also much weaker than GATA-1 in the MEL lanes. Again, the competition profile of fetal liver (Fig. 12) showed that these weak bands are related to the complex *a* proteins and not to GATA-1 or CP1. Two faint bands with extracts derived from fetuses and a smear of bands with extracts derived from adult liver that showed the same competition profile as those seen with MEL and fetal liver extracts were observed near the *a* position. With extracts made from mouse adult heart, kidney, and spleen, no bands at the *a* position were discernible. However, such extracts were of noted poorer quality than those of fetal tissues and adult liver (not shown). Thus, further characterization of the protein(s) responsible for complex *a* will be necessary before we can determine whether it has a limited tissue distribution and/or is modified differentially in various tissues.

As was expected, GMS assays of extracts from both untreated and hemin-treated K562 cells gave a binding activity corresponding to GATA-1 that was inhibited specifically by the wild-type and mutant 1 sequences (Fig. 12). The band seen migrating just below GATA-1 with K562 extracts (marked with an asterisk in the figure) appeared to be nonspecific, since it was barely inhibited by any of the oligonucleotides. A strong CP1 band was also seen with the K562 extracts. However, the K562 CP1 was slightly but obviously inhibited by a 100-fold molar excess of mutant 3 oligonucleotide, which was not the case with MEL extracts nor other mouse tissue extracts (compare lanes 5 in Fig. 12). Whether this reflects a difference in binding characteristics of human versus mouse CP1 is not known. In any case, K562 CP1 did evidently have a much-reduced affinity for mutant 3,

compared with that of wild-type and mutant 2 oligonucleotides, as was the case for the murine CP1 (compare competitions in lanes 3 and 4 with those in lanes 5).

A labeled band migrating at the *b* position and showing the appropriate competition profile was also clearly visible with extracts from both untreated and hemin-treated K562 cells. The difference in intensity between uninduced and induced K562 cells for the *b* complex and GATA-1 in Fig. 12 was not reproducible. On the other hand, no band at the *a* position was visible with either extract. Thus, K562 cells may lack the protein(s) corresponding to the *a* DNA binding activity, although the possibility that the protein(s) is present at levels too low to detect, the protein(s) is not extracted from these cells in active form, or its presence in GMS assays is obscured by other complexes cannot be ruled out.

DISCUSSION

This report describes two DNA-binding factors, *a* and *b*, which, in addition to CP1 and GATA-1, interact with the human β-globin gene promoter CAAT box region. The results suggested that the *a* and *b* factors are not directly related to and are not proteolytic fragments of CP1 or GATA-1. For example, mutants of the β-globin CAAT box that bound CP1, GATA-1, or the *a* and *b* proteins separately were derived. Thus, the binding specificities of *a* and *b* are different from those of CP1 or GATA-1. On the basis of these investigations and the comparative studies of other transcription factors, it may be that the *a* and *b* factors have not been described in the MEL system previously, although they could be related to proteins described in other systems, especially considering the ubiquitous expression of factor *b* activity. While this paper was being revised, a report described a protein (referred to as NF-E6) binding to the γ-globin CAAT box which we believe is the *b* factor described here (3).

It appears that factors *a* and *b* have similar DNA-binding specificities. Methylation interference assays demonstrated that both complexes bound strongly to the same core sequence within the β-CAAT region, although factor *b* did appear to interact with some nucleotides outside the region bound by factor *a*. Several mutations in the β-CAAT oligonucleotide that were studied affected binding of both factors in a nearly identical fashion. Moreover, in GMS assays with the β-CAAT, β-CACC, γ-CAAT, and B oligonucleotides as competitors, binding of both complexes was again affected in a nearly parallel manner. In relation to this similarity and since factor *a* DNA-binding activity was highly induced during differentiation of MEL cells, it is proposed that factor *a* be designated DSF-1 (differentiation-specific factor) and factor *b* be designated DSF-r (DSF related).

To evaluate the function of the proteins interacting with the β-globin promoter CAAT box, we derived a mutation of the CAAT region that was unable to bind any of the factors (mutant 5). The expression data demonstrated that this mutant results in approximately ninefold-lower levels of expression than does the wild-type CAAT box in MEL cells. These results are in close agreement with those reported by Antoniou and Grosveld (1), who showed a 10-fold-lower level of expression for a LCR μ-locus construct containing only the β-globin promoter TATA box (referred to as −77 deletion). Mutant 5 had the same relative inducibility (≈13-fold), albeit with low levels of expression, as all other constructs tested. This indicates that the transcriptional enhancement by the CAAT box and its cognate factors in the presence of the LCR is equally high in both uninduced and induced MEL populations. Thus, inducibility may be controlled solely by the LCR. This would be in agreement with the finding that the LCR can invoke induced expression of unrelated transgenes in the MEL system (1, 5, 26).

Recent results have shown that mutation of a GATA-1 site in the human ε-globin gene promoter greatly reduces transactivation of this promoter by hypersensitive site region 2 of the β-globin LCR (15) and that alteration in a GATA-1 binding site may be the cause of the Greek form of hereditary persistence of fetal hemoglobin (3), suggesting that the GATA-1 protein can play a role in directing enhancement by the LCR of globin gene promoters. In agreement, we found that conversion of the β-globin CAAT box to a comparatively stable GATA-1 binding sequence results in full activation by the LCR in MEL cells (mutant 1 in Fig. 9). These results might appear to indicate that GATA-1 has a transactivational role in binding to the wild-type β-globin CAAT box. However, GATA-1 binding to the wild-type CAAT box was observed to be extremely unstable, in contrast to the very stable binding of the specific portion of the GATA-1 interaction with mutant 1. Thus, it is unlikely that GATA-1 plays any activator role in the context of the wild-type β-globin CAAT box, although it is clear from the results presented here and elsewhere that strong GATA-1-binding sites found in the promoter and enhancer regions of globin genes (11, 13, 23, 31) could play a role in directing transcriptional enhancement by the LCR.

In contrast, the results do strongly indicate that it is DSF-r which is responsible for directing activation by the LCR through the wild-type CAAT box. This conclusion is based on several observations.

(i) In contrast to GATA-1, both DSF-r and DSF-1 and, to a lesser extent, CP1 formed complexes with the β-globin CAAT element that had high stability. However, since the β-globin CAAT box does not appear to play a role in the induction of promoter expression in MEL cells, the high inducibility of DSF-1 DNA-binding activity in these cells implies that this factor does not play a role in transcriptional activation in binding to this CAAT element. This agrees with the fact that the DSF-1 DNA complex formed with mutant 2 was about fivefold less stable than the DSF-1–wild-type CAAT complex yet mutant 2 was at least as active as the normal CAAT sequences in MEL cells. But the strong inducibility of DSF-1 activity during terminal differentiation of MEL cells and its apparent absence from K562 cells do suggest that this factor may have an important role in erythroid cell development.

(ii) The null activity of mutant 7, which was able to form a DNA-protein complex with CP1 that had an approximately fourfold-longer half-life than that formed with the wild-type CAAT motif, ruled out the possibility that CP1 directs the LCR activity through the β-globin CAAT box.

(iii) On the other hand, the four- to sevenfold-lower activity of mutant 3, compared with that of the wild-type and mutant 2 sequences, in MEL populations, as is based on copy number correction (Table 2), correlated with the much-decreased stability of the complex formed between DSF-r and mutant 3 DNA versus the stable binding of DSF-r to the wild-type and mutant 2 oligonucleotides. Thus, DSF-r is an important factor to globin gene expression, and its future identification should further our knowledge into how the LCR activity is directed through promoter regions.

Finally, the data demonstrated that there is a fundamental difference in how the LCR functions to drive globin gene expression in MEL cells versus K562 cells, in that the wild-type CAAT element was able to increase transcription

directed by the LCR greater than ninefold over the null mutation in MEL cells, but it was completely ineffective in K562 cells. If the difference in the two systems is directly related to the CAAT box, and assuming that GATA-1 is equivalent in the two cells types, it was expected that mutant 1 would fully rescue activity in K562 cells. Although mutant 1 did give some increase in expression, its activity in K562 cells was still very low compared with that of the γ-globin promoter CAAT element in these cells and even compared with that of the same mutant in uninduced MEL cells. Moreover, both mutants 2 and 3, which would be expected to function in K562 because of the presence of DSF-r in these cells (Fig. 12), also only gave very partial rescue of activity. Thus, it may be that the β-globin CAAT box itself is not defective in K562 cells. A second possibility worth considering is that it is the LCR that functions differently in the two cell types, in terms of its requirements at the promoter level. This possibility is particularly interesting with respect to the recent discovery that different regions of the LCR have some developmental and/or globin gene specificity (14). However, whether the finding of a difference between functioning of the LCR with the β-globin CAAT element in MEL and K562 cells reflects the different developmental profiles of the two cell types cannot be concluded. The difference could just as easily represent their different species of origin (mouse versus human), their differentiation statuses, and/or their transformed natures. We are in the process of doing a transgenic mouse analysis to address these questions.

ACKNOWLEDGMENTS

We are indebted to R. Gronostajski (University of Toronto) for providing us with an oligonucleotide probe for NF1, to Frank Grosveld's laboratory for providing the μ-locus plasmid, and to R. Kothary, D. Skup, and Nahum Sonenberg for critical reading of the manuscript. We also thank C. Nault for typing the manuscript and Roger Duclos for the photography.

This work was supported by a grant from the National Cancer Institute of Canada and funds from the Fonds de Recherche en Santé du Québec. L.W. is a chercheur-boursier of the FRSQ.

REFERENCES

1. **Antoniou, M., and F. Grosveld.** 1990. β-Globin dominant control region interacts differently with distal and proximal promoter elements. Genes Dev. **4**:1007–1013.
2. **Behringer, R. R., T. M. Ryan, R. D. Palmiter, R. L. Brinster, and T. M. Townes.** 1990. Human γ- to β-globin switching in transgenic mice. Genes Dev. **4**:380–389.
3. **Berry, M., F. Grosveld, and N. Dillon.** 1993. A single point mutation is the cause of the Greek form of hereditary persistence of fetal haemoglobin. Nature (London) **358**:499–502.
4. **Birkenmeier, E. H., B. Gwynn, S. Howard, J. Jerry, J. I. Gordon, W. H. Landschulz, and S. L. Knight.** 1989. Tissue-specific expression, developmental regulation and genetic mapping of the gene encoding CCAAT/enhancer binding protein. Genes Dev. **3**:1146–1156.
5. **Bloom van Assendelft, G., O. Hanscombe, F. Grosveld, and D. R. Greaves.** 1989. The β-globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner. Cell **56**:969–977.
6. **Chiba, T., Y. Ikawa, and K. Todokoro.** 1991. GATA-1 transactivates erythropoietin receptor gene, and erythropoietin receptor-mediated signals enhance GATA-1 gene expression. Nucleic Acids Res. **19**:3843–3848.
7. **Chodosh, L. A., A. S. Baldwin, R. W. Carthew, and P. A. Sharp.** 1988. Human CAAT-binding proteins have heterologous subunits. Cell **53**:11–24.
8. **Choi, O.-R. B., and J. D. Engel.** 1988. Developmental regulation of β-globin gene switching. Cell **55**:17–26.
9. **Collis, P., M. Antoniou, and F. Grosveld.** 1990. Definition of the minimal requirements within the human β-globin gene and the dominant control region for high level expression. EMBO J. **9**:233–240.
10. **Curtin, P. T., D. Liu, W. Liu, J. C. Chang, and Y. W. Kan.** 1989. Human β-globin gene expression in transgenic mice is enhanced by a distant DNase I hypersensitive site. Proc. Natl. Acad. Sci. USA **86**:7082–7086.
11. **de Boer, E., M. Antoniou, V. Mignotte, L. Wall, and F. Grosveld.** 1988. The human β-globin gene promoter: nuclear protein factors and erythroid specific induction of transcription. EMBO J. **7**:4203–4212.
12. **Enver, T., N. Raich, A. J. Ebens, T. Papayannopolous, F. Constantini, and G. Stamatoyannopolous.** 1990. Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. Nature (London) **344**:309–313.
13. **Evans, T., and G. Felsenfeld.** 1991. *trans*-activation of a globin promoter in nonerythroid cells. Mol. Cell. Biol. **11**:843–853.
14. **Fraser, P., S. Pruzina, M. Antonious, and F. Grosveld.** 1993. Each hypersensitive site of the human β-globin locus control region confers a different developmental pattern of expression on the globin genes. Genes Dev. **7**:106–113.
15. **Gong, Q., and A. Dean.** 1993. Enhancer-dependent transcription of the ε-globin promoter requires promoter-bound GATA-1 and enhancer-bound AP-1/NFE2. Mol. Cell. Biol. **13**:911–917.
16. **Grosveld, F., G. Bloom van Assendelft, D. Greaves, and G. Kollias.** 1987. Position-independent, high-level expression of the human β-globin gene in transgenic mice. Cell **57**:975–985.
17. **Kim, G. C., S. L. Swendeman, K. M. Barnhart, and M. Sheffery.** 1990. Promoter elements and erythroid cell nuclear factors that regulate α-globin gene transcription in vitro. Mol. Cell. Biol. **10**:5958–5966.
18. **Li, R., J. Knight, G. Bream, A. Stenlund, and M. Botchan.** 1989. Specific recognition nucleotides and their DNA context determine the affinity of E_2 protein for 17 binding sites in the BPV-1 genome. Genes Dev. **3**:510–526.
19. **Martin, D. I. K., and S. H. Orkin.** 1990. Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Erf1. Genes Dev. **4**:1886–1898.
20. **Mignotte, V., J. F. Eleouet, N. Raich, and P.-H. Romeo.** 1989. Cis- and trans-acting elements involved in the regulation of the human porphobilinogen deaminase gene. Proc. Natl. Acad. Sci. USA **86**:6548–6552.
21. **Mignotte, V., L. Wall, E. deBoer, F. Grosveld, and P.-H. Romeo.** 1989. Two tissue-specific factors bind the erythroid promoter of the human porphobilinogen deaminase gene. Nucleic Acids Res. **17**:37–54.
22. **Orkin, S. H.** 1990. Globin gene regulation and switching: circa 1990. Cell **63**:665–672.
23. **Plumb, M., J. Frampton, H. Wainwright, M. Walker, K. Macleod, G. Goodwin, and P. Harrison.** 1989. GATAAG: a *cis*-control region binding an erythroid-specific nuclear factor with a role in globin and non-globin gene expression. Nucleic Acids Res. **17**:73–92.
24. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
25. **Stamatyannopolous, G., and A. W. Neinhuis.** 1987. Developmental control of globin gene expression. Alan R. Liss, Inc., New York.
26. **Talbot, D., P. Collis, M. Antoniou, M. Vidal, F. Grosveld, and D. R. Greaves.** 1989. A dominant control region from the human β-globin locus conferring integration site-independent gene expression. Nature (London) **338**:352–355.
27. **Talbot, D., S. Philipsen, P. Fraser, and F. Grosveld.** 1990. Detailed analysis of the site 3 region of the human β-globin dominant control region. EMBO J. **9**:2169–2178.
28. **Tsai, S.-F., D. I. K. Martin, L. I. Zon, A. D. D'Andrea, G. G. Wong, and S. H. Orkin.** 1989. Cloning of the cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. Nature (London) **339**:446–451.
29. **Tsai, S.-F., E. Strauss, and S. H. Orkin.** 1991. Functional analysis and *in vivo* footprinting implicate the erythroid tran-

scription factor GATA-1 as a positive regulator of its own promoter. Genes Dev. **5**:919–931.
30. **Tuan, D. Y. H., W. B. Solomon, I. M. London, and I. P. Lee.** 1989. An erythroid-specific, developmental-stage-independent enhancer far upstream of the human "β-like globin" genes. Proc. Natl. Acad. Sci. USA **86**:2554–2558.
31. **Wall, L., E. deBoer, and F. Grosveld.** 1988. The human β-globin gene 3′ enhancer contains multiple binding sites for an erythroid-specific protein. Genes Dev. **2**:1089–1100.
32. **Xiao, J., I. Davidson, M. Macchi, R. Rosales, M. Vigneron, A. Staub, and P. Chambon.** 1987. *In vitro* binding of several cell-specific and ubiquitous nuclear proteins to the GT-1 motif of the SV-40 enhancer. Genes Dev. **1**:794–807.
33. **Yamamoto, M., L. J. Ko, M. W. Leonard, H. Beug, S. H. Orkin, and J. D. Engel.** 1990. Activity and tissue-specific expression of the transcription factor NF-E1 multigene family. Genes Dev. **4**:1650–1662.