MOLECULAR AND CELLULAR BIOLOGY, Jan. 1993, p. 44–56
0270-7306/93/010044-13$02.00/0

Vol. 13, No. 1

Discrimination among Potential Activators of the β-Globin CACCC Element by Correlation of Binding and Transcriptional Properties

GRANT A. HARTZOG[1] AND RICHARD M. MYERS[1,2]*

Department of Biochemistry and Biophysics[1] and Department of Physiology,[2] University of California, San Francisco, San Francisco, California 94143-0444

Received 2 June 1992/Returned for modification 1 August 1992/Accepted 1 October 1992

Adult β-globin-like promoters contain a *cis*-acting element, CCACACCC, that is conserved across species and is required for wild-type levels of transcription. We have studied the contribution of this element and proteins that interact with it to activate β-globin transcription. We found that an erythroid-like cell line, MEL, contains several proteins that specifically bind the CACCC element. By comparing the DNA-binding properties of promoters with mutations in the CACCC element with the transcriptional activities of these mutant promoters, we found that two CACCC-binding proteins did not bind to mutant promoters that direct decreased levels of transcription. One of these proteins is the transcriptional activator Sp1, and the other we have designated CACD (CACCC-binding species D). We subjected CACD to a binding site selection procedure and obtained high-affinity CACD binding sites that are identical to that of the β-globin CACCC element. This result, combined with our finding that CACD binds the CACCC element with a higher affinity than does Sp1, argues that the CACCC element is a target of CACD rather than Sp1. The strategy of correlating the results of a binding site selection experiment with those of in vivo expression and in vitro binding studies may allow evaluation of the relative potential of different proteins to activate transcription through a single *cis*-acting site.

The adult β-globin-like promoters contain several elements whose sequences and locations are highly conserved (14). One of these is the CACCC element, CCA/TCACCC (Table 1). A large body of work has demonstrated that this element is required for wild-type levels of transcription of transfected β-globin-like genes in HeLa, L, and 3T6 cells (12, 39, 54). Similarly, this element is also required for transcription of transfected β-globin genes in erythroid cells (7, 49). More striking, several different naturally occurring single base substitution (point) mutations in the CACCC element of the human β-globin promoter that are known to decrease transcription cause β-thalassemia (Table 1) (18, 32, 40, 41, 43, 51). Also, there are indications that the β-globin promoter CACCC element may be involved in mediating the effects of the far-upstream locus control region (1).

Transcription of several other genes depends upon sequences that resemble or are identical to the β-globin CACCC site. These include the simian virus 40 enhancer, the erythroid cell-specific GATA-1 transcriptional activator gene, the tryptophan oxidase gene, and the porphobilinogen deaminase gene (16, 46, 52, 55). These genes, and other CACCC element-containing genes, are active in many different cell types. In some cases, it is known that the CACCC element is required for the activity of an additional nearby *cis* element. For example, in the tryptophan oxidase promoter, a CACCC element and a glucocorticoid response element (GRE) occur within 10 nucleotides of one another. Transcriptional induction of tryptophan oxidase by glucocorticoids occurs only if both the GRE and the CACCC element are intact (46). Additionally, the tryptophan oxidase GRE and CACCC element must be spaced an integral number of 10 nucleotides apart for maximal activity, suggesting that proteins binding these elements interact in a specific spatial orientation to activate transcription cooperatively. A similar interaction appears to occur in the prophobilinogen deaminase promoter between a CACCC element and a GATA-1 binding site (16).

The data presented above suggest that the CACCC element serves as a binding site for a factor or factors, widely distributed across tissues, that activate transcription and that likely cooperate with other transcriptional regulatory proteins. For these reasons, many laboratories have been interested in characterizing potential transcriptional regulatory proteins that bind to this sequence element. A number of such putative factors have been identified in many different cell types (4, 8, 9, 17, 21, 34, 35, 38, 47, 55). Because we are interested in the regulation of β-globin gene transcription, we have studied CACCC-binding proteins from MEL cells, a mouse erythroid precursor immortalized at the pre-erythroblast stage (36).

When treated with any of a variety of chemical agents, MEL cells differentiate into an erythrocyte-like cell (36). During this process, adult globin genes and other erythroid cell-specific genes are transcriptionally activated (36). We have found that several CACCC-binding activities occur in MEL nuclear extracts, in agreement with previous reports (9, 22, 35). Although these DNA-binding activities can be distinguished by their mobilities in native gels, they share many other properties. We sought to differentiate between these binding activities and deduce which activity(ies) is most likely to activate β-globin transcription from the CACCC element in vivo.

Given the rapid expansion in the number of known DNA-binding transcriptional regulatory proteins, many of which bind to similar sites, our problem of differentiating between several different activities binding a single site is not unusual (24, 26). To address this problem, we first looked for a correlation between the ability of CACCC-binding proteins to bind to β-globin promoters with mutations in or around

* Corresponding author.

TABLE 1. Alignment of CACCC elements from different genes

Gene	Sequence[a]	Reference
Adult β-globin	101	
Human β[b]	C A G A C C T C A C C C T G	12
	A G A G C C A C A C C C T A	
	88 87	
Mouse β-major	A G A G C C A C A C C C T G	12
Mouse β-minor	G A A G C C T C A C C C T G	12
Rabbit β1	C A G A C C T C A C C C T G	12
	A G A G C C A C A C C C T G	
Goat βA and βC	C A A G C C T C A C C C T G	12
Chicken β	T G A G C C C C A C C C T G	12
Others with CACCC elements		
Tryptophan oxidase	A G A G C C A C A C C C A G	46
Porphobilinogen deaminase	C A G G C C C C A C C C T C	12
Simian virus 40 GT-1C motif	C T T T C C A C A C C C T A	55
GATA-1	G G C G G C A C A C C C C C	53
	A C T G C C C C A C C C A C	

[a] The boxed sequences are those that are most highly conserved and required for transcription.
[b] The human CACCC element is duplicated. Naturally occurring, individual C-to-T point mutations at positions −101, −88, and −87 or a C-to-G mutation at position −87 result in β-thalassemia.

the CACCC element and the ability of these mutant promoters to activate transcription. Three of the four CACCC-binding species bound less well to β-globin promoters bearing mutations in the CACCC element that decreased transcription; this is behavior that might be expected of a transcriptional regulatory protein acting through this site. Two of these potential activators were examined in greater detail. We demonstrated that one CACCC-binding protein is the transcriptional activator Sp1. A second candidate activator was subjected to a site selection procedure to find its preferred high-affinity binding site. The results of this experiment lead us to believe that it is this activity, rather than Sp1, that is most likely the transcriptional regulator that binds the CACCC element and activates transcription of the adult β-globin genes. Finally, we suggest that this combination of binding, transfection, and site selection experiments may provide a general approach for differentiating between multiple proteins binding to a particular *cis* element.

MATERIALS AND METHODS

DNAs. The β-globin promoter was assayed in the context of a fusion of a minimal β-globin promoter (nucleotides −106 to +26) to the mouse metallothionein-1 gene (nucleotides +65 to +2200). Most of the mutant β-globin promoters, as well as the histone H4 reference gene, were derived from previous work in this laboratory; however, several new mutations were generated for this work (39, 49). These mutations, −95A, −94T, −92C, −92T, −90T and −92C/−93T, were all constructed by site-directed mutagenesis (33).

Transfections. MEL cells (*aprt* mutant [10]) were grown in Dulbecco's modified Eagle medium (Irvine Scientific) supplemented with 10% fetal calf serum (HyClone Laboratories, Inc.) and penicillin G (100 U/ml)–streptomycin sulfate (100 μg/ml). Cells were transfected by the DEAE-dextran method and, where indicated, induced to differentiate exactly as described previously (7).

RNA analysis. Cytoplasmic RNA was isolated from the transfected cells by the method of Favalaro et al. (15). The β-globin–metallothionein and H4-metallothionein fusion RNAs were quantitated by an S1 nuclease protection assay as described previously (7). Single-stranded, end-labeled probe DNAs complementary to the expected products of the transfected genes were made as described previously (7, 49) (Fig. 1A). Signals were visualized by autoradiography and were quantitated on a Phosphorimager (Molecular Dynamics). An example of the data collected is shown in Fig. 1B.

DNA binding assays. (i) Probes. The CAC1 probe was generated by cloning an oligonucleotide (Fig. 1C) into the *Xho*I site of plasmid pSP72 (31). This cloned product was excised from the plasmid by digestion with *Eco*RI and *Xba*I and was ^{32}P labeled by filling in the overhanging ends created by the restriction enzymes with Klenow enzyme in the presence of [α-^{32}P]dATP (3,000 Ci/mmol; Amersham). The labeled DNA was then gel purified in 6% native polyacrylamide gels, eluted in crush-soak buffer overnight at 37°C, and ethanol precipitated (37). Probes were stored in 10 mM Tris-HCl (pH 8)–1 mM EDTA–100 mM NaCl. β-globin promoter probe DNAs were prepared as described above, using *Bgl*II and *Cla*I to remove the 132-nucleotide β-globin promoter (−106 to +26) fragment from plasmid pβMTF5 and its mutant derivatives.

For DNase I footprinting, CAC1 probes labeled at a specific end were generated by cutting the DNA with either *Eco*RI or *Xba*I alone, treating it with calf intestinal alkaline phosphatase (Boehringer), and end labeling it with [γ-^{32}P]-ATP (7,000 Ci/mmol; ICN) and polynucleotide kinase (New England Biolabs). The DNA was subsequently digested with the other restriction enzyme to release the end-labeled probe and then gel purified.

The Sp1 probe was generated by using [γ-^{32}P]ATP and polynucleotide kinase to end label a double-stranded oligonucleotide bearing a high-affinity Sp1 binding site (29) (Fig. 1D).

The sequences of the two strands of the NF1 binding site oligonucleotide used as a competitor in Fig. 2 are 5′-GATCCTTTTGGATTGAAGCCAATATGAT and 5′-CTAG ATCATATTGGCTTCAATCCAAAAG (27). Annealing of these oligonucleotides produces a 5′-GATC overhang at each end.

(ii) Mobility shift assay. For mobility shift assays, protein samples were mixed with 1 to 12 fmol of probe DNA in a buffer consisting of 2 mM $MgCl_2$, 10% Ficoll 400 (Pharmacia), 20 mM Tris-HCl (pH 8), and 0.4 mM EDTA in a final reaction volume of 20 μl; 2 μg of nonspecific competitor

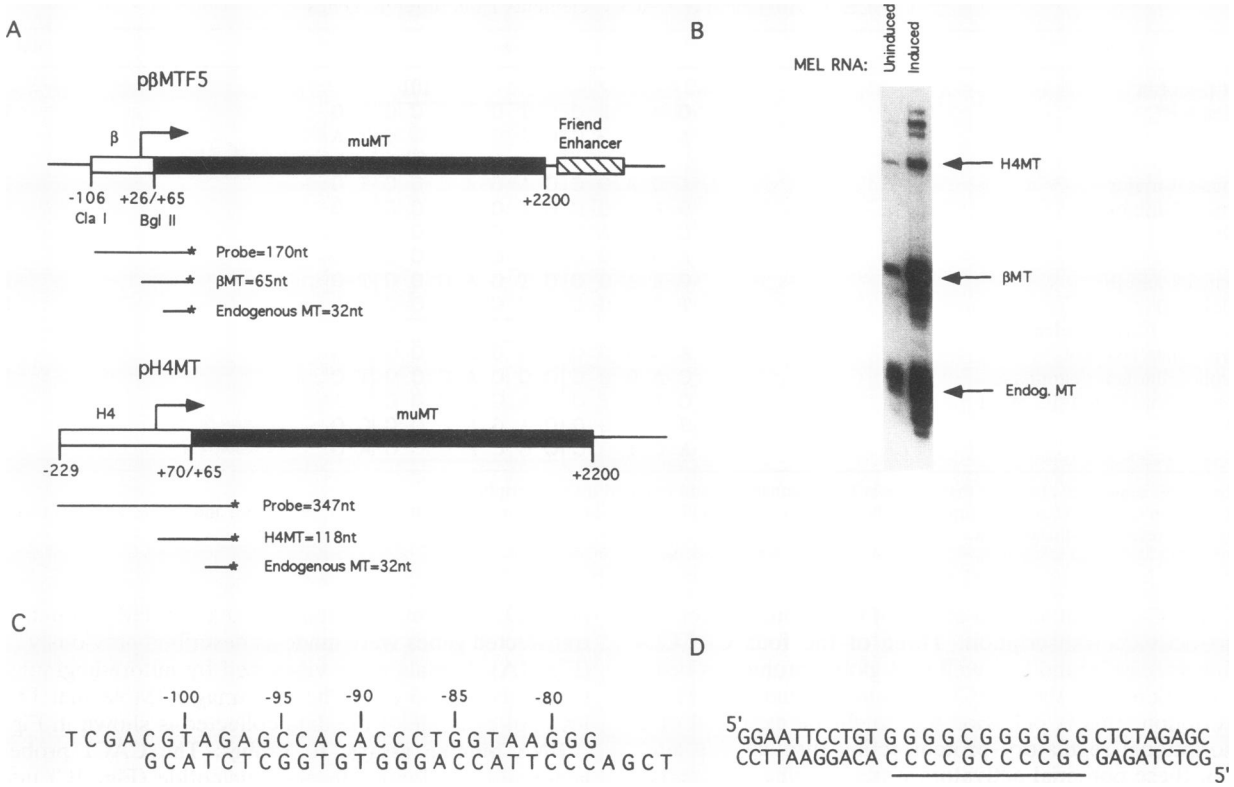


FIG. 1. Plasmid DNAs used in transfection and mobility shift experiments. (A) Chimeric β-globin–mouse metallothionein-1 (BMT) and histone H4–mouse metallothionein-1 (H4MT) constructs (7, 48) that were transfected into MEL cells. The single-stranded DNA probes used in the S1 nuclease protection assay are shown below each construct. In addition to detecting the transcript to which they are directed, each probe detects a small portion of the transcript from the other transfected construct in the region that is homologous to the mouse metallothionein-1 transcript as well as the endogenous metallothionein-1 transcript itself. In both cases, this extra signal is detected as a 32-nucleotide (nt) DNA fragment. (B) An example of the S1 nuclease assay. Cells were transfected with pβMTF5 and pH4MT, and a portion of the cells was subsequently induced to differentiate. RNA was isolated and subjected to an S1 nuclease protection assay in the presence of both the β-globin–metallothionein and H4-metallothionein probes. The products of this assay were fractionated on a 50% urea–polyacrylamide gel and are shown. The bands above the H4-metallothionein signal are due to upstream start sites in plasmid pH4MT. (C) Sequence of the CACCC oligonucleotide used to generate probe CAC1 and also used in competition assays. This oligonucleotide contains the mouse β-major-globin promoter sequence between nucleotides −78 and −102 and is numbered accordingly. (D) Sequence of the oligonucleotide with a high-affinity Sp1 binding site used as a competitor and, when end labeled with ^{32}P, as a probe in mobility shift assays. The Sp1 binding site is underlined.

DNA [poly(dI-dC); Sigma] was included when crude protein preparations were used, whereas no competitor was used in reactions with partially purified CACCC-binding protein species D (CACD) or with pure Sp1. Assays of purified Sp1 or partially pure CACD included 0.5 mg of bovine serum albumin (BSA; nuclease and protease free; Calbiochem) per ml as a carrier protein; in the case of partially purified CACD, no mobility shift was observed without carrier protein. Binding reaction mixtures were incubated at room temperature for 5 to 20 min and loaded onto a 6% 39:1 (acrylamide/bisacrylamide) cross-linked polyacrylamide gel. The gel was run at room temperature at approximately 6 V/cm in a buffer consisting of 50 mM Tris-HCl (pH 8.5), 190 mM glycine (Bio-Rad), and 1 mM EDTA (2). For assays with specific competitor DNAs, the protein sample was added after the probe and competitor DNAs were mixed; the reaction mixture was then incubated for at least 20 min at room temperature before being loaded on the gel.

The proteolytic clipping band shift assay (45) was performed by combining 50 μl of purified Sp1 or partially pure CACD with 10 μg of BSA carrier protein in 1 mM $CaCl_2$. Fifty nanograms of proteinase K (Boehringer Mannheim) was added, and the proteins were allowed to digest at room temperature. At intervals of 2, 5, and 10 min, 10-μl aliquots were withdrawn and treated with phenylmethylsulfonyl fluoride to inhibit the proteinase K. Four microliters each of these proteolyzed samples was then subjected to mobility shift analysis.

(iii) DNase I footprinting. DNase I footprinting assays were carried out in conjunction with the mobility shift assay. The mobility shift assay allowed the separation of the different CACCC-binding species in crude extracts as well as separation of bound and unbound probe. For the DNase I reactions, a typical mobility shift reaction was scaled up in size three to five times. After binding for 10 min at room temperature, the reaction mixture was brought to 1.25 mM in $CaCl_2$ and immediately treated with 5 to 100 ng of DNase I (Worthington) for 1 min at room temperature. EGTA was added to 1.25 mM, and the samples were immediately loaded onto a mobility shift gel. After electrophoresis, the free and shifted probe bands were excised and purified as described previously (2). The eluted DNAs were fractionated on 10% polyacrylamide–50% urea gels and visualized by autoradiography.

Proteins. MEL cells were grown in suspension at 0.5×10^6 to 1.5×10^6 cells per ml in Joklik's minimal essential medium (Irvine Scientific) supplemented with 5% newborn calf serum (Gemini Biosciences) and penicillin-streptomycin.

Nuclear extracts were prepared as described by Dignam et al. (13). All buffers used in the preparation of nuclear extracts and in the purification of proteins included 0.1% Nonidet P-40 (Sigma), 1 mM dithiothreitol, 0.5 mM phenylmethylsulfonyl fluoride (Sigma), 1 mM benzamidine (Sigma), 1 μg of aprotinin (Boehringer) per ml, 1 μg of leupeptin (Boehringer) per ml, and 1 μg of pepstatin A (Sigma) per ml. All manipulations of protein were carried out at 0 to 4°C. Proteins were stored at −70°C.

CACD was purified from MEL cell nuclear extracts by a combination of conventional and site-specific DNA affinity chromatography. These data will be presented in detail elsewhere (23a).

Sp1 was purified from MEL cells by a combination of wheat germ agglutinin agarose and site-specific DNA affinity chromatography as described by Jackson and Tjian (25). The protein produced by this procedure was 95% pure Sp1 protein as judged by silver staining and was authentic Sp1 as judged by immunoblot analysis.

Site selection assay. The site selection assay was performed as described by Blackwell and Weintraub (3). Briefly, an oligonucleotide, 5′-TCCGAATTCCTACAGN_{21}TGCAATGG ATCCGTCT, with 21 random internal nucleotides (N) and flanking primer sites was synthesized. The oligonucleotide was made double stranded by annealing with primer A (5′-AGACGGATCCATTGCA) and primer extension with Klenow enzyme in the presence of 60 μCi of [α-^{32}P]dATP and 0.5 mM each of the three other unlabeled deoxynucleoside triphosphates (dNTPs). The probe was gel purified as described above and, because it was internally labeled, was used as soon as possible after gel purification. The amount of probe was quantitated by counting the amount of incorporated ^{32}P and assuming a random distribution of adenines in the internal, random sequences. Mobility shift reactions were performed with 4 fmol of the labeled probe and the most highly purified fractions of CACD (see below). Marker lanes consisting of CACD binding reactions with the CAC1 probe were also run. After electrophoresis, gels were dried and exposed to film. The region of the gel containing probe DNA migrating at the same mobility as the CACD shift with the CAC1 probe was excised, and the DNA was eluted by the crush-soak method and collected by ethanol precipitation.

The recovered probe was immediately amplified by polymerase chain reaction (PCR) by using primer A and primer B (5′-TCCGAATTCCTACAG). PCR conditions were 2.5 nM each primer, 0.1 mM dNTPs, 1.5 mM Mg^{2+}, 50 mM KCl, 10 mM Tris-HCl (pH 8.4), 100 μg of gelatin per ml, and 0.5 μl of *Taq* DNA polymerase (Cetus-Perkin Elmer) in a total volume of 100 μl. Samples were put through 30 cycles of 94°C for 20 s, 55°C for 20 s, and 72°C for 30 s. After PCR, the products were separated from the unincorporated nucleotides and primers on a Sephadex G-25 (Pharmacia) spin column and concentrated by ethanol precipitation in the presence of 10 μg of carrier tRNA. These products were labeled with ^{32}P and made double stranded as described above except that both primers were now included since the starting template was double stranded. Subsequent mobility shift reactions were performed with lower amounts of probe and protein to select for higher-affinity protein-DNA interactions. After five rounds of selection, the PCR products were digested with *Eco*RI and *Bam*HI and purified on a G-25 spin column. These products were ligated into *Eco*RI/*Bam*HI-digested pSP72. Sequences of individual clones were determined by the dideoxy-chain termination method.

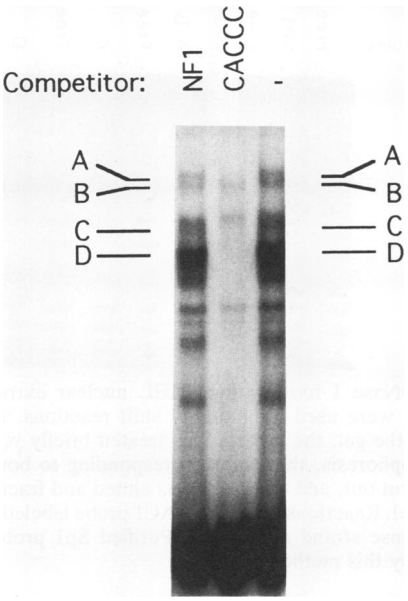


FIG. 2. CACCC mobility shift assay. When crude MEL nuclear extracts are mixed with the CAC1 probe and separated on a native gel, several bands are observed. These bands are competed for by a 100-fold molar excess of a specific (CACCC) competitor but not by a nonspecific (NF1) competitor. The four upper bands that are routinely observed with crude MEL nuclear extracts are labeled A to D. The species migrating faster than species D are variable. Band C and D sometimes appear to be monomeric and at other times appear to contain more than one species (see Fig. 6).

Antibodies. The anti-Sp1 antibodies used in this work were kindly provided by the Tjian laboratory. One, 2892, is a polyclonal rabbit antiserum produced against Sp1 protein made in *Escherichia coli*. The other, 2990, is a rabbit polyclonal antiserum produced against a peptide comprising most of Sp1 zinc finger 2.

Immunoblot analysis. Immunoblots were performed by using standard protocols (23). Proteins were separated on 10% polyacrylamide–sodium dodecyl sulfate (SDS) gels and electrophoretically transferred to 0.2-μm-pore-size nitrocellulose (Schleicher & Schuell). The filter was blocked with 2% gelatin (EIA grade; Bio-Rad) and was probed with antibody 2892 at a 1:1,000 dilution. The filter was subsequently probed with an anti-rabbit alkaline phosphatase-conjugated antibody (Promega) and was developed by using the Promega Protoblot reagents as instructed by the supplier.

RESULTS

DNA binding studies. To identify potential transactivating proteins that bind the CACCC element, we used the gel mobility shift assay. When a ^{32}P-labeled, double-stranded DNA probe (CAC1; Fig. 1C) containing nucleotides −101 to −78 of the mouse β-globin promoter was mixed with MEL nuclear extracts and electrophoresed on a native gel, four shifted species, A to D, were observed (Fig. 2). Occasionally, we observed species that migrated with a greater mobility than did species D. These occurred more commonly in extracts made without protease inhibitors and in extracts

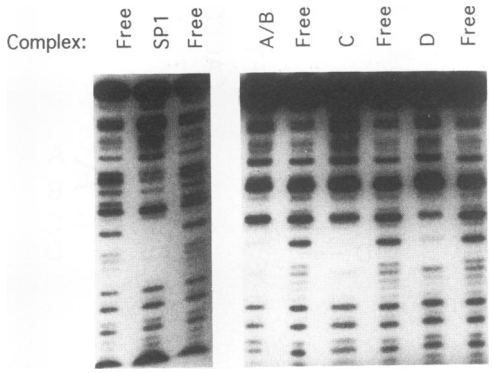


FIG. 3. DNase I footprinting. MEL nuclear extracts and the CAC1 probe were used for mobility shift reactions. Before being loaded onto the gel, the sample was treated briefly with DNase I. After electrophoresis, the bands corresponding to bound and free probe were cut out, and the DNA was eluted and fractionated on a denaturing gel. Reactions using the CAC1 probe labeled at the 5′ end of the antisense strand are shown. Purified Sp1 protein was also footprinted by this method.

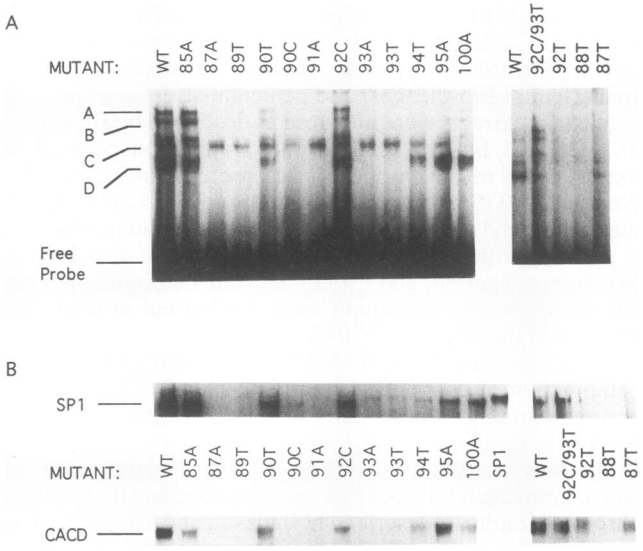


FIG. 4. Mobility shift analysis of point mutant promoter probes. The 132- nucleotide *Cla*I-*Bgl*II fragments of the wild-type (WT) and mutant β-globin promoters were labeled with ^{32}P to similar specific activities. Three femtomoles of each of these probes was used in the mobility shift assay with crude nuclear extracts of MEL cells (A) and pure Sp1 protein and partially purified CACD protein (B). In panel B, only a single shifted species of invariant mobility was observed for either protein on any of the probes, and only this species is shown. Figure 6 shows an entire mobility shift with the protein preparations used for this figure. Mutants −92C/93T, −92T, −88T, and −87T were assayed in a separate experiment and are shown alongside a reaction with the wild-type probe that was performed at the same time for comparison. In panel B, Sp1 binding to the Sp1 probe, in addition to the point mutant promoter probes, was assayed.

that were several months old. Thus, we assume that these higher-mobility bands represent proteolytic artifacts (see also Fig. 7A). Also, both species C and D occasionally appeared to be composed of two or more bands rather than one band (compare species C and D in Fig. 2 and 6). In the case of species C, when a second band was present, it competed specifically and appeared to bind to mutant promoters with properties similar to those of species A and B (see below). Therefore, this band may represent a proteolytic fragment of either of these activities. In the case of species D, the higher-mobility bands may represent a proteolyzed fragment of species D or a differentially phosphorylated form of species D (23a). Because these other bands did not appear reproducibly, even when the same sample of protein was used from experiment to experiment, we believe that they are experimental artifacts and will not consider them further.

Competition with an excess of the unlabeled, double-stranded CACCC oligonucleotide abolished each of the four CACCC-binding species. None of the shifted species was affected by competition with a nonspecific DNA fragment (NF1; Fig. 2). When the entire mouse β-globin promoter-proximal region, from nucleotides −106 to +26, was used as a probe, essentially the same pattern of shifts occurred (see below). Each shifted species observed with the promoter probe behaved in competition assays like the species of similar mobility observed with the CAC1 probe (data not shown). Therefore, under our mobility shift conditions with both the promoter and CAC1 probes, four CACCC-binding activities were detected.

The four CACCC-binding activities were further analyzed by DNase I footprinting. Binding reactions with crude MEL nuclear extracts and end-labeled CAC1 probe were assembled and treated briefly with DNase I. These reactions were electrophoresed on mobility shift gels to separate species A to D. The free and bound probe DNAs were eluted and further separated on sequencing gels (Fig. 3). Because species A and B run near one another on the mobility shift gels, these species could not be separated and were eluted as a single entity. Thus, the footprint for species A and B represents the composite of two individual footprints. The data from these experiments reveal that species A to D have nearly identical interactions with the CACCC binding site.

Point mutant binding studies. Because it appears that multiple factors with similar properties bind the CACCC element, we needed some way to differentiate between these potential transactivators. We reasoned that if a particular CACCC-binding species activates β-globin transcription, its ability to bind a mutant β-globin promoter fragment might parallel the ability of that mutant promoter to activate transcription. While mutations in the CACCC element that decrease transcription need not necessarily decrease binding of a CACCC transactivator protein, we expected that at least some mutations that reduce β-globin transcription would exert their effects by decreasing the affinity for a required transcription factor. Also, we expected that mutations that did not affect transcription would not disrupt DNA-activator protein interactions. Using a fragment of the β-globin promoter including nucleotides −106 to +26 as a probe in the mobility shift assay, we examined whether promoters with point mutations in or around the CACCC sequence could bind proteins with the same relative affinity as did the wild-type sequence.

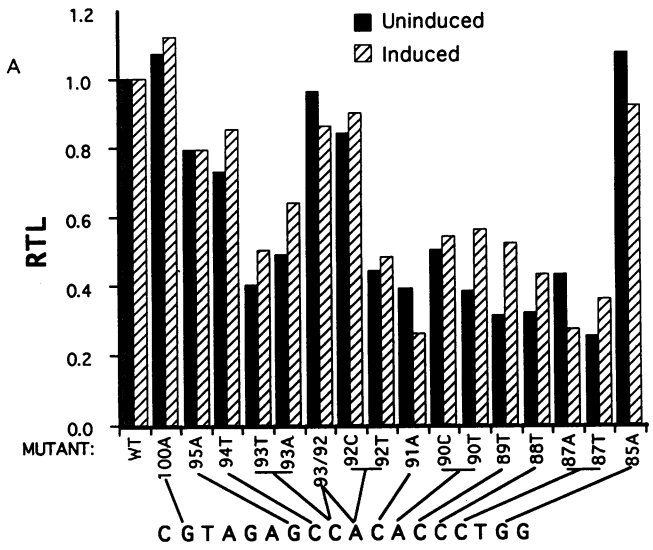
With crude nuclear extracts, the mobility shifts observed with the mutant promoter probes could be placed into several different classes (Fig. 4A; the sequence of the wild-type mouse β-globin promoter in the region of the CACCC element is shown in Fig. 1C and 5A). One class of mutants gave rise to shifted species that were identical to those observed with the wild-type probe (Fig. 4A, lane WT)

or to a pattern that had only slightly decreased levels of species A, B, and D. This class of mutations includes −100A, −95A, and −85A, which flank the CACCC site, as well as −92C and −90T, which fall within the conserved sequence, CCACACCC. The second class of mutant promoters bound low or negligible amounts of species A, B, and D. This class includes mutations −87A, −87T, −88T, −89T, −91A, −92T, −93A, and −93T. Three mutants that do not fit into these two classes are −90C, −94T, and −92C/−93T. Mutation −90C bound species B and C at reduced levels and did not bind species A or D. In contrast, mutant −94T bound species C and D but not species A or B. Mutation −92C/−93T increases binding of species A, B, and C and decreases binding of species D. Finally, only mutations −92C/−93T and −92T seemed to affect the affinity of the promoter probes for species C. Overall, these results are reminiscent of those for experiments that tested binding of HeLa and F9 cell proteins to the simian virus 40 GT-1C motif, which is similar to the CACCC element (55) (Table 1). However, in contrast to this previous study, several mutations that we examined, including −90T, −94T, −95A, and −92C/−93T, helped us to differentiate between the binding specificities of the four CACCC-binding activities.

Transfections. Although previous studies of mutations in the β-globin CACCC element have examined the transcriptional effects of point mutations at nearly every nucleotide of this conserved sequence, these studies were carried out primarily in nonerythroid cells (6, 7, 12, 20, 39). To confirm the extent of the CACCC element and its nature as a *cis* sequence required for β-globin transcription in erythroid cells, mutant β-globin promoters linked to the mouse metallothionein-1 gene were transiently cotransfected into MEL cells along with an internal reference gene (Fig. 1A). A portion of the transfected cells was subsequently induced to differentiate with dimethyl sulfoxide. The amount of β-globin–metallothionein mRNA produced by these transfected cells was measured with an S1 nuclease assay, quantitated, and expressed as the level of transcription relative to that of the wild type (RTL). The levels of transcripts measured in this assay are known to reflect accurately the transcriptional activity of the template that produced them (5).

The results of these experiments, summarized in Fig. 5, are in good agreement with previous results for HeLa, MEL, and 3T6 cells with the exception of those regarding the −94T and −95A mutations (7, 12, 39). The −94T mutant gave an RTL of 0.25 when transfected into HeLa cells in one study and an RTL of 1.0 when transfected into 3T6 cells in another (12, 39). In the present study, the −94T mutant resulted in only a slight decrease in transcription. The −95A mutant previously gave RTLs of 0.17 in HeLa cells and of 0.34 in uninduced and 0.64 in induced MEL cells (7, 39). In this study, mutant −95A gave nearly wild-type levels of transcription. It is possible that these differences can be explained in part by different CACCC-binding activities in different cell types. Alternatively, the discrepancy between the results of this and previous studies could be due to some inadvertent mutation occurring elsewhere in this construct. All point mutant promoters used in this study were resequenced and shown to have the expected sequence in the −106 to +1 region.

Comparison of the transfection data with the point mutant binding data shows that the binding behavior of CACCC species A, B, and D on the point mutant promoters correlates well with the transcriptional activity of those promoters. CACCC species C DNA binding does not correlate with the transcriptional activity of the point mutant templates,



B

MUTANT	RTL UNINDUCED	RTL INDUCED
WT	1.00 (0.11)	1.00 (0.02)
100A	1.07 (0.35)	1.12 (0.10)
95A	0.79 (0.17)	0.79 (0.21)
94T	0.73 (0.34)	0.85 (0.07)
93T	0.40 (0.13)	0.50 (0.10)
93A	0.49 (0.15)	0.64 (0.11)
92C/93T	0.96 (0.21)	0.86 (0.18)
92C	0.84 (0.29)	0.90 (0.21)
92T	0.44 (0.10)	0.48 (0.07)

MUTANT	RTL UNINDUCED	RTL INDUCED
91A	0.39 (0.27)	0.26 (0.06)
90C	0.50 (0.21)	0.54 (0.23)
90T	0.38 (0.11)	0.56 (0.38)
89T	0.31 (0.10)	0.52 (0.25)
88T	0.32 (0.14)	0.43 (0.20)
87A	0.43 (0.20)	0.27 (0.04)
87T	0.25	0.36
85A	1.07 (0.20)	0.92 (0.09)

FIG. 5. Quantitated results of transfection assays. All data represent averages of a least three separate transfections except for the induced value of mutant −89T, which is the average of two experiments. The data for mutant −87T are from reference 7 and are shown for comparison. RTLs were calculated as the ratio of the βMT signal to the H4MT signal for the test plasmid relative to the same ratio for the wild-type βMT construct. RTLs of constructs in induced cells were calculated relative to the wild-type βMT signal in induced cells; in uninduced cells, the RTLs were calculated relative to the wild-type βMT signal in uninduced cells. In the experiments reported here, dimethyl sulfoxide induction resulted in an average increase in wild-type β-globin–metallothionein fusion gene transcript levels of 6.5-fold. (A) Data presented graphically. The sequence of the wild-type (WT) promoter in this region and the positions of the mutations are shown at the bottom. (B) Average RTL for each construct in both induced and uninduced cells. The standard deviation for each value is given in parentheses.

indicating that this activity is less likely to activate the CACCC element in vivo.

CACCC species B is Sp1. Other groups have suggested that the transcriptional activator Sp1 can bind to the CACCC element (17, 21, 34, 47, 55, 56). If this is true, it is possible that Sp1 is responsible for activating transcription through the β-globin CACCC element. The results of five experiments described below showed that Sp1 can bind to the CACCC element.

First, we compared the abilities of unlabeled oligonucleotides bearing either a high-affinity Sp1 binding site or a CACCC binding site to compete for binding to the CAC1 probe (Fig. 6A). The CACCC oligonucleotide competed for species D binding (Fig. 6A, lanes 2 to 6) at concentrations four- to fivefold below those required for equal levels of competition by the Sp1 oligonucleotide (lanes 7 to 11).

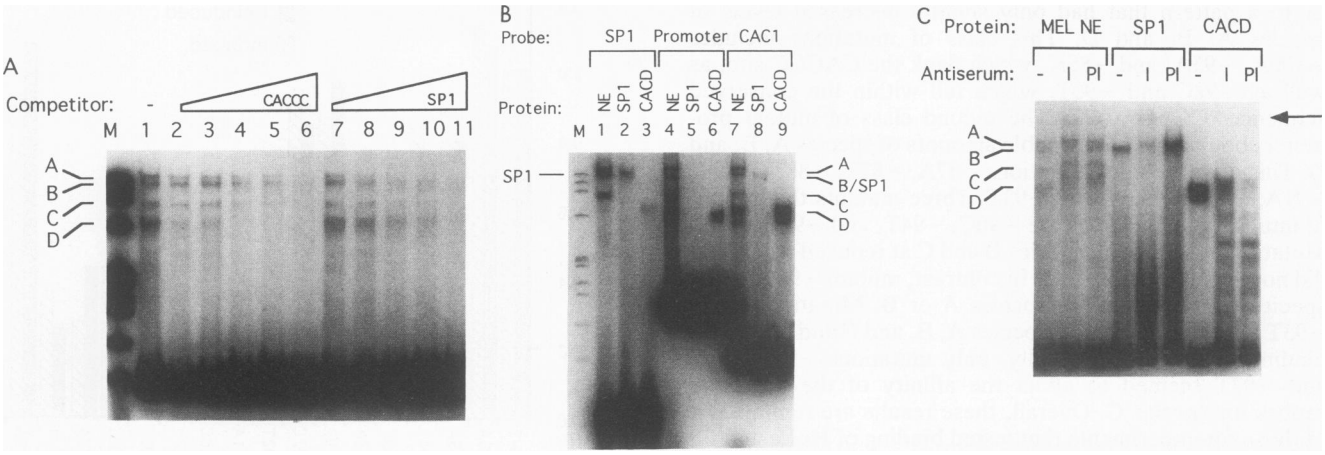


FIG. 6. Comparison of CACD and Sp1. (A) Competition assays with CACCC and Sp1 oligonucleotides. Mobility shift reactions with crude nuclear extracts, 12 fmol of the CAC1 probe, and competitor DNA were assembled. These reactions included no competitor (lane 1) or a 4 (lanes 2 and 7)-, 8 (lanes 3 and 8)-, 20 (lanes 4 and 9)-, 40 (lanes 5 and 10)-, or 80 (lanes 6 and 11)-fold molar excess of competitor oligonucleotide. (B) Mobility shift assays using the Sp1, β-globin promoter, and CAC1 probes. Mobility shift reactions were assembled with each probe and either crude nuclear extracts of MEL cells (NE), pure Sp1 protein, or partially pure CACD protein, as indicated. Each reaction contained 3 fmol of probe, and the amount of each protein used was kept constant for the three probes. Note that in lane 5, a very faint shift comigrating with CACCC species B was seen on the original autoradiograph. (C) The indicated protein was mixed with the CAC1 probe in the presence of no antiserum (−) or an immune (I) or preimmune (PI) antiserum directed against Sp1. These reactions were subjected to mobility shift analysis. MEL NE, MEL cell nuclear extract. The arrow indicates the mobility of the new, low-mobility complex observed when an anti-Sp1 antiserum was included with either pure Sp1 or crude MEL nuclear extracts.

Species A, B, and C, on the other hand, were competed for by concentrations of the Sp1 oligonucleotide four- to fivefold below those required for equal levels of competition by the CACCC oligonucleotide. Thus, species D binds with higher affinity to the CACCC element than to a high-affinity Sp1 binding site, whereas species A, B, and C bind with a higher affinity to a high-affinity Sp1 binding site than to a CACCC element.

Second, we purified Sp1 from MEL cells and directly examined its ability to bind the CACCC probe in the mobility shift assay. These experiments demonstrated that Sp1 shifts the probe to a position identical to that of CACCC species B (Fig. 6B; compare lanes 7 and 8).

Third, when a polyclonal anti-Sp1 antibody was included in a mobility shift binding reaction with pure Sp1, the Sp1-binding activity was diminished and a new, lower-mobility binding species appeared (Fig. 6C). Preimmune serum did not affect Sp1 binding in the mobility shift assay. When the same experiment was performed with a crude MEL nuclear extract, species B was clearly diminished and a new, lower-mobility species appeared in the gel, presumably representing a supershifted DNA-antibody-species B complex (Fig. 6C).

Fourth, when pure Sp1 was used in the gel mobility shift assay with the point mutant probes, results were similar to those observed with CACCC species B in crude MEL nuclear extracts (Fig. 4B). Finally, when purified Sp1 was DNase I footprinted on the CAC1 probe, it gave a footprint almost identical to that of the CACCC species A/B complex (Fig. 3). In aggregate, the results of these experiments argue that Sp1 in the crude MEL nuclear extract binds and retards the mobility shift probe to give rise to CACCC species B. While species A and C bind with higher affinity to an Sp1 site than to a CACCC site, they appear to be unrelated to Sp1 by the criteria of the supershift experiment.

Characterization of species D. Since species D is by far the most prominent activity observed in a mobility shift of the CAC1 probe with crude MEL nuclear extracts, we decided to characterize this activity in more detail. By a combination of conventional and site-specific DNA affinity chromatography, we purified CACCC-binding species D approximately 3,000 fold to ~1% purity and named this activity CACD.

We observed no other CACCC-binding activities in this preparation of CACD. Nonspecific competitor DNA was not required in mobility shifts of partially pure CACD, presumably because the bulk of nonspecific DNA-binding activities were removed during the purification procedure. The mobility of CACD in the gel shift assay, its relative binding affinity for the point mutant probes, and its DNase I footprint were the same as those observed for species D in the crude extract (Fig. 4B and 6B and data not shown). The partially pure CACD generated in this preparation was used for the remainder of this work.

Comparison of CACD with species B/Sp1. The similar behaviors of CACD and species B/Sp1 in the footprinting and point mutant binding assays raised the possibility that CACD is related to Sp1, perhaps as an alternately spliced or proteolytic product of Sp1. Alternatively, CACD and Sp1 may have distinct but related DNA binding domains. We compared CACD and Sp1 in several different assays, and in each case, CACD appeared to differ from Sp1.

First, we examined the ability of CACD to bind to an Sp1 probe. Consistent with the competition experiments of Fig. 6A, we found that CACD could bind to an Sp1 probe, although less efficiently than to the CAC1 probe (Fig. 6B, lanes 3 and 9; Fig. 7A, lanes 9 and 13). Conversely, Sp1 protein bound to the CAC1 probe, but less efficiently than it bound to the Sp1 probe (Fig. 6B, lanes 2 and 8; Fig. 7A, lanes 1 and 5).

Second, immunoblot analysis of the most purified fractions of CACD did not show cross-reactivity to either of two polyclonal anti-Sp1 antibodies (one of which is an antipeptide antibody directed against finger 2 of the Sp1 DNA binding domain; Fig. 7B and data not shown). Third, these

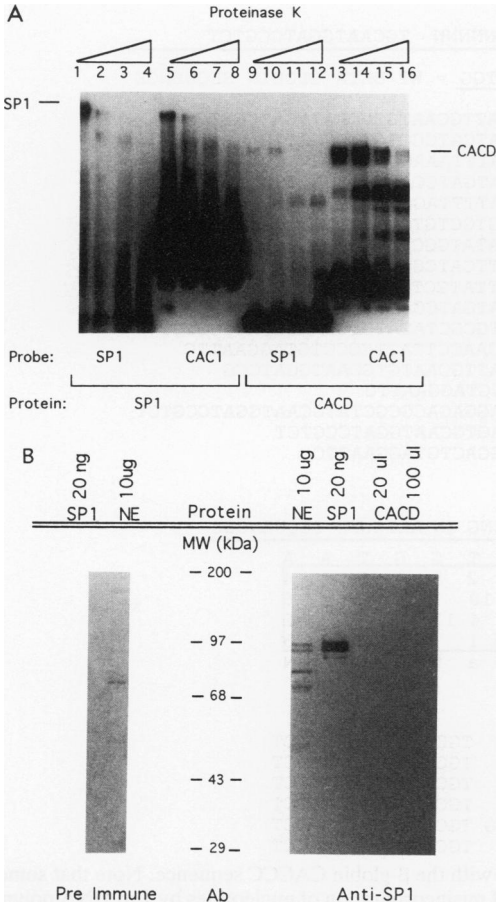


FIG. 7. (A) Proteolytic clipping band shift assay of Sp1 and CACD protein. Mobility shift reactions using the indicated probes were assembled with either pure Sp1 (lanes 1 to 8) or partially pure CACD (lanes 9 to 16). The protein in these reactions had been treated with proteinase K for 0 (lanes 1, 5, 9, and 13), 2 (lanes 2, 6, 10, and 14), 5 (lanes 3, 7, 11, and 15), or 10 (lanes 4, 8, 12, and 16) min prior to being assayed by mobility shift. Similar amounts of protein were assayed at each time point except for the no-treatment samples, which contained approximately 60% as much protein as did the other samples. (B) Immunoblot analysis of Sp1 and CACD. The indicated amounts of MEL nuclear extracts (NE), Sp1 protein, and CACD protein were separated on an SDS–10% gel, transferred to nitrocellulose, and probed with an antiserum (2892) directed against Sp1 or a preimmune serum (Pre Immune Ab). Given the amount of CACD-binding activity present in the fractions blotted and our estimate that CACD is a 44-kDa protein (23a), a minimum of 15 ng of CACD was used in this experiment. Immunoblots using antiserum 2990 gave similar results (data not shown).

anti-Sp1 antisera did not supershift CACD when included in the mobility shift assay (Fig. 6C).

Our fourth approach for comparing CACD and Sp1 was to use the proteolytic clipping band shift assay (45). In this assay, proteins are subjected to gentle protease treatment before being used in the mobility shift assay. If portions of the protein not essential for DNA binding are removed, then the mobility of the resulting DNA-protein complex will be greater than that of the intact protein. By subjecting both Sp1 and CACD to proteinase K for various times and then observing their mobility shifts, we examined whether these proteins showed similar sensitivities to protease and whether their patterns of proteolytic products still active for DNA binding were similar. Sp1 and CACD differed substantially in this assay (Fig. 7A). Sp1 was relatively sensitive to proteinase K treatment and produced only one fragment of higher mobility. CACD gave rise to several fragments of higher mobility, and the relative abundance of these fragments suggests that they were produced stoichiometrically from the intact protein.

Selection of binding sites for CACD. Because we had tested CACD binding only on the Sp1, CAC1, and promoter probes, we did not know whether sites with a higher affinity for the CACD protein existed. Therefore, CACD was subjected to a site selection procedure (3). An oligonucleotide containing 21 random internal nucleotides and defined flanking sequences that serve as primer binding sites was used to make a pool of random, double-stranded ^{32}P-labeled probe molecules. This random probe pool was used in mobility shift assays with the most purified fractions of CACD. DNA-protein complexes migrating with the same mobility as the CACD-CAC1 complex were eluted from the gel, PCR amplified, ^{32}P labeled, and subjected to subsequent rounds of selection. By using low concentrations of CACD and standard CACD binding conditions and by decreasing the amount of PCR-amplified DNA probe used in later rounds of selection, we hoped to select for high-affinity CACD binding sites. After five rounds of selection, the PCR products were cloned and the sequences of 22 individual clones were determined.

Of the 22 clones, 16 contained sequences that matched the sequence CCACACCC at six or more of these eight nucleotides (Fig. 8). These sequences were aligned by their CACCC element homology, and the frequency of individual nucleotides occurring at each position of this alignment was determined. Overall, a consensus sequence, ANCCACACCCAT (N = any nucleotide), emerges from these data. Of the other six clones sequenced, three of them, 1, 8, and 22, have a five-of-eight-base match to the sequence CCAC ACCC and show more extensive homology among themselves. The other three clones do not match the consensus and may represent contaminating PCR products or sequences with a low but greater than nonspecific affinity for the CACD activity.

The results of the selection experiment suggest that the primary determinant of CACD binding affinity is the conserved sequence CCACACCC. In addition, an AT dinucleotide (corresponding to positions −86 and −85) was commonly recovered at the 3′ end of the consensus sequence and might also contribute to binding affinity. The −85A point mutant probe was bound by CACCC species A to D as efficiently as was the wild-type promoter (Fig. 4), but mutations −85T and −86A, which change the wild-type promoter to the consensus sequence, had not been tested. We therefore made a −86A/−85T double-mutant oligonucleotide that was otherwise identical to the CACCC oligonucleotide used in the competition assays. We found that it competed for CACD binding at half the concentration required by the wild-type oligonucleotide, indicating that the −86A/−85T mutation results in higher-affinity CACD DNA binding (data not shown). Finally, site selection experiments similar to those described above, in which the starting oligonucleotide was fixed as NNNNNCNNCACCCNNN, returned the sequence cNNNaCCACACCCa/ttt (lowercase letters indicate a lower but still nonrandom frequency of occurrence and N indicates a random nucleotide; data not shown), again consistent with the idea that the primary determinant of CACD binding affinity is the sequence of the core conserved sequence found in the β-globin promoters.

A

OLIGO	GAATTCCTACAG	NNNNNNNNNNNNNNNNNNNN	TGCAATGGATCCGTCT
		AGAG CCACACCC TGG = WT BETA GLOBIN SEQUENCE	
CLONE[a]			
2	GAATTCCTACAGCGAA	CCACCGCC	ATTGCAATGTGCAATGGATCCGTCT
3	AGACGGATCCATTGCATCGGGT	CCGCGCCC	ATCGTCCCTGTAGGAATTC
5	GAATTCCTACAGG	CCACACCC	ATTTAAAGGTAGTGCAATGGATCCGTCT
6	GAATTCCTACAGGC	CAACACCC	ATGATCCATCATGCAATGGATCCGTCT
7	GAATTCCTACAGC	CCACACCT	ATTTTAGGGAGGTGCAATGGATCCGTCT
9	AGACGGATCCATTGCACAGAGACCCG	CCACACCC	GTGCTGTAGGAATTC
11	AGACGGATCCATTGCATGGGAGA	ACACACCC	ATATGGCTGTAGGAATTC
13	AGACGGATCCATTGCACAACGG	CCACGCCC	TTCATCGCTGTAGGAATTC
16	AGACGGATCCATTGCAGATTGTGT	ACCACGCC	TTATGCTGTAGGAATTC
20	GAATTCCTACAGGC	CAACACCC	ATGATCCATCATGCAATGGATCCGTCT
22	GAATTCCTACAGC	CAACGCCC	TGCGCTACATGTGCAATGGATCCGTCT
23	AGACGGATCCATTGCA	CCACGCCC	CAACCTTATTGGCCTGTAGGAATTC
25	GAATTCCTACAGCGAA	CCACGCCC	ATTGCAATGTGCAATGGATCCG
27	AGACGGATCCATTGCATAACCATGATATGT	CCACGCCC	TGTAGGAATTC
28	GAATTCCTACAG	CCCCACCC	AGGAGACGCGCTATGCAATGGATCCGTCT
29	GAATTCCTACAGTCCGGGTGGGG	CCGCATCC	AGTGCAATGGATCCGTCT
29(rev.)	AGACGGATCCATTGCACTGGATGCGG	CCCCACCC	GGACTGTAGGAATTC

FREQUENCY OF INDIVIDUAL NUCLEOTIDES OCCURRING IN ALIGNED SEQUENCES

	A	G	A	G	C	C	A	C	A	C	C	C	T	G	G	T	A	A
G	5	7	12	6			2		6	2			2	5	4	5	4	4
A	5	4	3	4	2	3	12	1	9				10	1	4	5	1	6
T	1	2		3						1		1	4	11	6	4	7	2
C	6	4	2	4	15	14	3	16	2	14	17	16	1		3	3	5	5
CONSENSUS	N	N	a	N	C	C	a	C	a	C	C	C	a	t	N	N	t	N

B

31	GAATTCCTACAG	CACAGTTAACGGCCATGTTAG	TGCAATGGATCCGTCT
33	GAATTCCTACAG	GCAAGCAGAGGCATGTCCTTA	TGCAATGGATCCGTCT
1	GAATTCCTACAG	TGGGCGTATCCCGATCCCTTG	TGCAATGGATCCGTCT
8	GAATTCCTACAG	GCCATGTAGAGGGGCGTATCC	TGCAATGGATCCGTCT
21	GAATTCCTACAG	GCCGAGTTTGGGCGTTACCTGG	TGCAATGGATCCGTCT
24	GAATTCCTACAG	CCCAGTTGATCGTTATCCGAC	TGCAATGGATCCGTCT

FIG. 8. Site selection data. (A) Sequences aligned according to their homology with the β-globin CACCC sequence. Note that some of the sequences vary in length from the starting oligonucleotide, presumably, because of misincorporation of nucleotides by *Taq* DNA polymerase. (B) Sequences of site selection products with a less than six-of-eight-base match to the sequence CCACACCC.

DISCUSSION

Several common activities can bind the CACCC element. To identify potential transcription factors acting through the CACCC element, we employed a gel shift assay using MEL nuclear extracts and a short probe containing the conserved CACCC sequence. This assay identified at least four activities that bind the CACCC element in a specific manner, species A, B, C, and D, with species D binding most tightly. Furthermore, these four activities exhibited extremely similar DNase I footprints (Fig. 3) and, in a dimethyl sulfate interference assay, identical binding characteristics (23a).

CACCC species B is Sp1. With use of purified Sp1 protein in gel shifts with the CAC1 probe, the Sp1-CAC1 complex migrated with the same mobility as did CACCC species B in native gels. Species B and Sp1 also behaved similarly in binding competition experiments. When either of two anti-Sp1 antibodies was included in mobility shift binding reactions with crude MEL extracts, species B was specifically abolished and a new, lower-mobility band appeared. Therefore, we conclude that CACCC species B contains Sp1.

Although species A and C behaved similarly to species B in competition assays and exhibited DNase I footprints nearly identical to that of Sp1, they were not reproducibly perturbed by anti-Sp1 antisera in the supershift assay. Occasionally, CACCC species A, C, and D were perturbed when Sp1 antisera were included in mobility shift reactions (Fig. 6C). However, these other interactions also occurred with preimmune serum, suggesting that they were not the result of specific antigen-antibody interactions.

What activity is responsible for transcriptional activation of β-globin through the CACCC element? One phenotype expected of mutations in the binding site of a transcriptional activator is that those mutations that prevent the activator from binding DNA should decrease the activity of the linked promoter. Although it is possible that mutations in the CACCC element can decrease β-globin transcription by some other mechanism, we felt that at least some mutations were likely to exert their effects in this manner. Starting with the assumption that the mobility shift assay, as performed here, allows examination of any CACCC-binding activity relevant to the activity of the β-globin promoter, we develop the following model of CACCC element function.

The transcription levels observed for the most deleterious mutations are approximately the same as those seen when the entire CACCC element is deleted (12) or when a series of four clustered point mutations is introduced into the CACCC element (49a). Therefore, the data in Fig. 5 are consistent with the idea that a severe, single point mutation in the CACCC element can abolish all activity normally directed by this *cis* sequence. Because none of the point mutations in the CACCC element that decreased transcription also decreased binding of CACCC species C, this activity is unlikely to be responsible for activation of β-globin transcription. Sp1 and CACCC species A and D are all candidate

activators since their relative binding affinities for the point mutant promoters correlate well with the transcriptional activities of these promoters in the transfection assay. However, the few exceptions to this last statement are potentially instructive.

Mutants −94T and −95A, which bound wild-type levels of species D and low levels of Sp1 and species A, were transcribed at nearly wild-type levels. One model that would account for this observation is that species D normally binds the CACCC element in vivo and activates transcription of the β-globin gene. Therefore, CACCC element mutations that do not affect species D binding, such as −94T and −95A, would not affect β-globin transcription. Mutant −92C/−93T, which was transcribed at nearly wild-type levels, bound low levels of species D, wild-type levels of Sp1, and increased amounts of species A and C. Thus, this mutant CACCC element would be less likely to use CACD as a transactivator in vivo. These data can be incorporated into our model by assuming that some other factor, perhaps Sp1 or species A, compensates for the loss of affinity for species D in the −92C/93T mutant by binding the mutant CACCC element and activating transcription. Indeed, this type of CACCC element has been shown to confer Sp1 responsiveness on a linked promoter in vivo (21). Since the −92C/−93T mutation creates a CACCC element of the type found in fetal and embryonic globin genes as well as the α- and β-globin locus control regions, this result suggests that more than one CACCC-binding protein may regulate transcription of the different globin genes in vivo (42, 48).

Finally, this model is consistent with the data derived from various cases of β-thalassemia in humans. Mutations −87 C to T (equivalent to −88T in the mouse promoter) and −88 C to T (−89T in the mouse promoter) both result in reduced β-globin transcription in vivo and hence cause β-thalassemia (32, 40). Each of these mutations bound reduced amounts of CACD in vitro (Fig. 4B) and was transcribed at reduced levels in MEL cells (Fig. 5). This finding is in accordance with our original assumption that mutations in the CACCC element that affect β-globin transcription would exert their affects by disrupting protein-DNA interactions. Two other human mutations, −101 C to T and −87 C to G, that cause β-thalassemia were not tested in this study (18, 51).

Because CACCC species A bound promoter probes only when both CACD and Sp1 also bound, we cannot say that CACCC species A is sufficient for transcription. However, because several mutations appear to abolish CACCC species A binding without affecting transcription, species A must not be necessary for transcription. Because it is less abundant than CACD and because we have not characterized it extensively, we shall not consider it further in this discussion.

One mutation, −90T, results in decreased β-globin transcription, yet it binds significant amounts of Sp1, CACD, and species A. It is possible that this mutation affects transcription by altering interactions of an activator with the rest of the transcriptional machinery rather than by preventing the activator from binding to its DNA site.

Finally, if our initial assumption is incorrect and our in vitro assay does not detect all relevant protein-DNA interactions at the CACCC element, then the model presented above may be incorrect. To test this possibility, any candidate activator of the CACCC element must be specifically depleted from a cell or cell extract and be shown to be required for β-globin transcription.

The CACCC element is a high-affinity target of CACD. From the analysis presented above, we conclude that Sp1, a known activator of transcription, can probably stimulate transcription from the β-globin promoter but that it is neither the only nor the primary activity that provides this function. To further assess the role of CACD in β-globin transcription, a site selection experiment to find high-affinity CACD binding sites was performed. These experiments were performed with a partially purified preparation of CACD. Thus, it is possible that multiple entities rather than a single entity was assayed in the selection experiment. By separating free and bound DNA on a mobility shift gel, we were able to select for DNA sequences that were bound and shifted to the same mobility as the CAC1 probe is shifted by the CACD activity. Furthermore, because the site selection experiment was performed with binding conditions identical to those used to assay CACD throughout this study, we selected for high-affinity DNA-protein interactions under conditions in which CACD normally binds the CAC1 probe with high affinity. Thus, although CACD has not been purified to homogeneity, the site selection experiment was performed under conditions in which the CACD activity appears to be homogeneous. Even if our CACD preparation is composed of multiple species, the following minimal statement can still be made about the results of the site selection experiment: the proteins that make up the activity that we have named CACD share a common high-affinity binding sequence, ANCCACACCCAT. This site is nearly identical to the site that naturally occurs in the mouse β-globin promoter. The site selection experiment also showed that the −86A/85T mutation slightly increases affinity for CACD. Whether or not this mutation has any functional consequence remains to be determined.

The CACCC elements of the adult β-globin-like genes are most likely conserved across species as a result of an evolutionary selection for a high-affinity binding site for a protein that prefers to bind to the sequence CCACACCC. This argument predicts that positions that are less well conserved among CACCC elements should be less important both for binding of the activator protein and for transcription of the β-globin promoter; this is what we observed for binding of CACD to promoters with the −92C mutation. Although the −92T mutation reduces DNA binding and transcription, this finding is not inconsistent with our argument since the upstream, duplicated CACCC element of the human, which has a −92T type of sequence, makes only a modest contribution to β-globin transcription in comparison with the downstream CACCC element (Table 1) (12, 18, 43).

It is formally possible that Sp1 provides all transactivation functions through the CACCC element. However, although Sp1 appears to be able to function through the CACCC element, its binding to the mouse β-major CACCC element is weak compared with its affinity for a consensus Sp1 site or compared with the affinity of CACD for the β-globin CACCC site (compare the binding of these proteins to their cognate and heterologous sites in Fig. 6B and 7A). Site selection experiments similar to the ones described here have been performed with Sp1 and defined a consensus high-affinity binding site, GGGGCGGGGT, that is different from the CACCC element (50). If Sp1 normally functions through the CACCC element, it is hard to understand why higher-affinity consensus Sp1 sites are not found in the β-globin-like promoters. If low-affinity Sp1 binding at this position were important, then, by the arguments given above, lower levels of sequence conservation in the CACCC elements of the β-globin-like promoters would be expected. Another piece of evidence that argues that CACCC and Sp1 binding sites are functionally distinct, and therefore likely to be the

targets of different proteins, comes from the work of Wang and Gralla (53). When testing the ability of promoter-proximal elements to interact with remote enhancer-type elements, they found that CACCC and Sp1 sites behaved quite differently even when compared in nearly identical contexts. This finding suggests that CACCC and Sp1 sites are the targets of different transcriptional activators with distinct functions. On the basis of the arguments presented above, we believe that the CACCC element represents a high-affinity protein binding site and, on the strength of the site selection result and the correlation of CACD binding to and the transcriptional activity of the point mutant promoters, that this element is a target for the CACD protein.

The problem of multiple proteins binding to a single DNA sequence element is one of general significance (26). Many DNA-binding proteins are members of families that have similar DNA-binding domains and sequence specificities. For example, different homeodomain proteins of *Drosophila melanogaster* can bind to very similar sites in vitro, yet genetic data show that they act at different sites in vivo (24). Nevertheless, the relative affinity differences between homeodomain proteins for a particular binding site can be a good predictor of which protein(s) will act on the binding site in vivo (11). The site selection procedure as used here provides an avenue for inferring what protein or proteins are most likely to operate through a particular site in vivo.

A potential problem with this approach is that the DNA binding specificities of some proteins are influenced by cooperative interactions with other proteins. This is true for the yeast homeodomain protein α2 (19, 30). The α2 protein's DNA binding specificity is determined in part by its ability to interact cooperatively with either of two proteins, MCM1 or a1. In this situation, the site selection procedure would be of use only if the protein-protein interaction that occurred in vivo was recognized and reproduced in vitro. Mutations near a protein's DNA binding site that decrease transcription but do not affect its binding in vitro may be indicative of such an in vivo protein-protein cooperative interaction. Thus, a prerequisite for using the site selection assay to discriminate between potential activators is a thorough understanding of the binding site of interest and the functional consequences of mutations in or around the site. In the case of the β-globin CACCC element, mutations at positions flanking the CACCC element have no effect on transcription (7, 39). Also, CACD has similar affinities for the CAC1 and β-globin promoter probes (Fig. 6B; compare lanes 4 and 7). These results are consistent with the model that the protein that binds the CACCC element does so independently of proteins binding other *cis* sequences in the β-globin promoter.

Role of the CACCC element in transcription. In the transfection experiments presented in Fig. 5, the effects of CACCC element mutations were the same in uninduced and induced MEL cells. In contrast to a previous study, we found no evidence for an inducible CACCC-binding activity in extracts of induced MEL cells (23a, 35). Also, when nuclear extracts were prepared from MEL cells that had been induced to differentiate for various periods of time, the relative amount of CACCC-binding activities increased, but only slightly (23a). These data are consistent with the idea that the CACCC element mediates the effects of a factor that provides activation functions not specific to terminally differentiating MEL cells. Likewise, the presence of CACCC-binding factors and the activity of CACCC elements in nonerythroid cells suggest that CACCC-binding transcription factors are widely distributed among tissues (9, 22, 23a, 35, 55).

CACD. Several groups have described DNA-binding activities that are similar in sequence specificity and mobility in the mobility shift assay to CACD. However, these activities, which include TEF-2 and AP3 (8, 38, 55), have not been described in detail. The proteolytic clipping band shift assay (Fig. 7A) shows that CACD has a protein component that can be partially degraded and still retain its DNA-binding activity. We have found that the DNA-binding activity of CACD is inhibited by chelators of divalent cations and that low concentrations of zinc can restore binding activity (23a). The zinc-dependent DNA-binding activity of CACD is reminiscent of that demonstrated for several zinc finger proteins, including Sp1 and an activity, likely to be Sp1, that binds the chicken CACCC element (28, 34).

Given the similarity of Sp1 and CACD biochemical and DNA-binding properties, it seemed possible that CACD is related to Sp1, possibly via an alternate splice of the Sp1 mRNA or by proteolysis of the 92.5-kDa Sp1 protein to produce the smaller CACD protein. However, alternate forms of the Sp1 cDNA have not been reported, its mRNA appears to be a single species on Northern (RNA) blots, and low-stringency Southern blots fail to reveal Sp1 homologs (28, 44). Furthermore, by four experimental criteria presented here, relative affinities for their DNA binding sites, immunoblot analysis, supershift analysis, and protease clipping, Sp1 and CACD appear to be unrelated. Further evidence suggesting that Sp1 and CACD are different entities includes their divergent chromatographic behaviors on both wheat germ agglutinin and site-specific DNA affinity columns (23a). While these results do not rule out the possibility that CACD is related to Sp1 by an alternate splice or proteolytic event, they make it unlikely. We believe it likely that CACD, like Sp1, is a member of the Cys_2-His_2 zinc finger protein family but is only distantly related to Sp1.

In conclusion, we characterized four CACCC-binding activities that occur in MEL cell nuclear extracts. By correlating the ability of these activities to bind to mutant promoters with the ability of the mutant promoters to activate transcription, we found that two activities, CACD and Sp1, are likely activators of the β-globin CACCC element. We subsequently used a site selection experiment to determine the high-affinity sites for CACD. This experiment gave the surprising result that CACD's high-affinity sites are nearly identical to that of the β-globin CACCC element.

We suggest that these experiments illustrate a general methodology for discriminating between potential activators of a single *cis* element. Observation of the effects of mutations in and around a *cis* element allows determination of its extent in a functional sense. Binding studies using the same mutant sites allow one to search for proteins with binding properties consistent with those suggested by the functional studies. If more than one candidate activator remains at this point, the site selection assay can be used to determine the preferred high-affinity sites of the candidate activators. Additionally, this assay will reveal the variations in the high-affinity site that still allow DNA binding. These data can be used in conjunction with the functional studies to identify the activity(ies) that has the properties expected of the in vivo activator.

ACKNOWLEDGMENTS

This work was supported by a grant from the NIH to R.M.M. G.A.H. was supported by the Medical Scientist Training Program.

We thank the Tjian laboratory for the gift of antibodies. We thank Pablo Garcia and Andy Peterson for helpful conversations and

Laura Stuvé, Andy Peterson, Keith Yamamoto, Rudi Groschedl, and Billy Forrester for critical reading of the manuscript.

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. **Ausubel, F. K., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1991. Current protocols in molecular biology. Greene Publishing Associates/ Wiley-Interscience, New York.
3. **Blackwell, T. K., and H. Weintraub.** 1990. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science **250**:1104–1110.
4. **Catala, F., E. de Boer, G. Habets, and F. Grosveld.** 1989. Nuclear protein factors and erythroid transcription of the human A gamma-globin gene. Nucleic Acids Res. **17**:3811–3827.
5. **Chao, M. V., P. Mellon, P. Charnay, T. Maniatis, and R. Axel.** 1983. The regulated expression of beta-globin genes introduced into mouse erythroleukemia cells. Cell **32**:483–493.
6. **Charnay, P., P. Mellon, and T. Maniatis.** 1985. Linker scanning mutagenisis of the 5′-flanking region of the mouse β-major-globin gene: sequence requirements for transcription in erythroid and nonerythroid cells. Mol. Cell. Biol. **5**:1498–1511.
7. **Cowie, A., and R. M. Myers.** 1988. DNA sequences involved in transcriptional regulation of the mouse β-globin promoter in murine erythroleukemia cells. Mol. Cell. Biol. **8**:3122–3128.
8. **Davidson, I., J. H. Xiao, R. Rosales, A. Staub, and P. Chambon.** 1988. The HeLa cell protein TEF-1 binds specifically and cooperatively to two SV40 enhancer motifs of unrelated sequence. Cell **54**:931–942.
9. **de Boer, E., M. Antoniou, V. Mignotte, L. Wall, and F. Grosveld.** 1988. The human β-globin promoter: nuclear protein factors and erythroid specific induction of transcription. EMBO J. **7**:4203–4212.
10. **Deisseroth, A., and D. Hendrick.** 1978. Human alpha-globin gene expression following chromosomal dependent gene transfer into mouse erythroleukemia cells. Cell **15**:55–63.
11. **Dessain, S., C. T. Gross, M. A. Kuziora, and W. McGinnis.** 1992. Antp-type homeodomains have distinct DNA binding specificities that correlate with their different regulatory functions in embryos. EMBO J. **11**:991–1002.
12. **Dierks, P., A. van Ooyen, M. D. Cochran, C. Dobkin, J. Reiser, and C. Weissman.** 1983. Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell **32**:695–706.
13. **Dignam, J. D., R. M. Lebovitz, and R. G. Roeder.** 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. **11**:1475–1489.
14. **Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Spritz, J. K. DeRial, B. G. Forget, S. M. Weissman, J. L. Slightom, A. E. Blechl, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proudfoot.** 1980. The structure and evolution of the human β-globin gene family. Cell **21**:653–668.
15. **Favalaro, J., R. Treisman, and R. Kamen.** 1980. Transcription maps of polyoma virus-specific RNA: analysis by two dimensional nuclease S1 gel mapping. Methods Enzymol. **65**:718–749.
16. **Frampton, J., M. Walker, M. Plumb, and P. R. Harrison.** 1990. Synergy between the NF-E1 erythroid-specific transcription factor and the CACCC factor in the erythroid-specific promoter of the human porphobilinogen deaminase gene. Mol. Cell. Biol. **10**:3838–3842.
17. **Giglioni, B., P. Comi, A. Ronchi, R. Mantovani, and S. Ottolenghi.** 1989. The same nuclear proteins bind the proximal CACCC box of the human beta-globin promoter and a similar sequence in the enhancer. Biochem. Biophys. Res. Commun. **164**:149–155.
18. **Gonzalez-Redondo, J. M., T. A. Stoming, A. Kutlar, F. Kutlar, K. D. Lanclos, E. F. Howard, Y. J. Fei, M. Aksoy, C. Altay, and A. Gurgey.** 1989. A C----T substitution at nt–101 in a conserved DNA sequence of the promoter region of the beta-globin gene is associated with "silent" beta-thalassemia. Blood **73**:1705–1711.
19. **Goutte, C., and A. D. Johnson.** 1988. a1 protein alters the DNA binding specificity of alpha2 repressor. Cell **52**:875–882.
20. **Grosveld, G. C., E. de Boer, C. K. Shewmaker, and R. A. Flavell.** 1982. DNA sequences necessary for transcription of the rabbit β-globin gene *in vivo*. Nature (London) **295**:120–126.
21. **Gumucio, D. L., K. L. Rood, K. L. Blanchard-McQuate, T. A. Gray, A. Saulino, and F. S. Collins.** 1991. Interaction of Sp1 with the human gamma globin promoter: binding and transactivation of normal and mutant promoters. Blood **78**:1853–1863.
22. **Gumucio, D. L., K. L. Rood, T. A. Gray, M. F. Riordan, C. I. Sartor, and F. S. Collins.** 1988. Nuclear proteins that bind the human γ-globin gene promoter: alteration in binding produced by point mutations associated with hereditary persistence of fetal hemoglobin. Mol. Cell. Biol. **8**:5310–5322.
23. **Harlow, E., and D. Lane.** 1988. Antibodies: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

23a. **Hartzog, G. A., and R. M. Myers.** Unpublished data.

24. **Hayashi, S., and M. Scott.** 1990. What determines the specificity of action of Drosophila homeodomain proteins? Cell **63**:883–894.
25. **Jackson, S. P., and R. Tjian.** 1989. Purification and analysis of RNA polymerase II transcription factors by using wheat germ agglutinin affinity chromatography. Proc. Natl. Acad. Sci. USA **86**:1781–1785.
26. **Johnson, P. F., and S. L. McKnight.** 1989. Eukaryotic transcriptional regulatory proteins. Annu. Rev. Biochem. **58**:799–839.
27. **Jones, K. A., J. T. Kadanoga, P. Rosenfeld, T. Kelly, and R. Tjian.** 1987. A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell **48**:79–89.
28. **Kadonaga, J. T., K. R. Carner, F. R. Masiarz, and R. Tjian.** 1987. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell **51**:1079–1090.
29. **Kadonaga, J. T., K. A. Jones, and R. Tjian.** 1986. Promoter specific activation of RNA polymerase II transcription by Sp1. Trends Biochem. Sci. **11**:20–23.
30. **Keleher, K. A., C. Goutte, and A. D. Johnson.** 1988. The yeast cell-type-specific repressor alpha 2 acts cooperatively with a non-cell-type-specific protein. Cell **53**:927–936.
31. **Krieg, P. A., and D. A. Melton.** 1987. In vitro RNA synthesis with SP6 RNA polymerase. Methods Enzymol. **155**:397–415.
32. **Kulozik, A. E., A. Bellan-Koch, S. Bail, E. Kohne, and E. Kleihauer.** 1991. Thalassemia intermedia: moderate reduction of beta globin gene transcriptional activity by a novel mutation of the proximal CACCC promoter element. Blood **77**:2054–2058.
33. **Kunkel, T. A., J. D. Roberts, and R. A. Zakour.** 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol. **154**:367–382.
34. **Lewis, C. D., S. Clark, G. Felsenfeld, and H. Gould.** 1988. An erythrocyte-specific protein that binds to the poly(dG) region of the chicken β-globin promoter. Genes Dev. **2**:863–873.
35. **Mantovani, R., N. Malgaretti, S. Nicolis, B. Giglioni, P. Comi, N. Cappellini, M. T. Bertero, F. Caligaris-Cappio, and S. Ottolenghi.** 1988. An erythroid specific nuclear factor binding to the proximal CACCC box of the beta-globin gene promoter. Nucleic Acids Res. **16**:4299–4313.
36. **Marks, P. A., and R. A. Rifkind.** 1978. Erythroleukemic differentiation. Annu. Rev. Biochem. **47**:419–448.
37. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
38. **Mitchell, P. J., C. Wang, and R. Tjian.** 1987. Positive and negative regulation of transcription *in vitro*: enhancer binding protein AP-2 is inhibited by SV40 T antigen. Cell **50**:847–861.
39. **Myers, R. M., K. Tilly, and T. Maniatis.** 1986. Fine structure genetic analysis of a β-globin promoter. Science **232**:613–618.
40. **Orkin, S. H., S. E. Antonarakis, and H. H. J. Kazazian.** 1984. Base substitution at position −88 in a β-thalassemic globin gene. J. Biol. Chem. **259**:8679–8681.
41. **Orkin, S. H., H. H. Kazazian, S. E. Antonarakis, S. C. Goff, C. D. Boehm, J. P. Sexton, P. G. Waber, and G. P.J.V.** 1982.

Linkage of β-thalassaemia mutations and β-globin gene polymorphisms with DNA polymorphisms in human β-globin gene cluster. Nature (London) **296**:627–631.

42. **Philipsen, S., D. Talbot, P. Fraser, and F. Grosveld.** 1990. The β-globin dominant control region: hypersensitive site 2. EMBO J. **9**:2159–2167.
43. **Ristaldi, M. S., S. Murru, G. Loudianos, L. Casula, S. Porcu, D. Pigheddu, B. Fanni, G. V. Sciarratta, S. Agosti, and M. I. Parodi.** 1990. The C-T substitution in the distal CACCC box of the beta-globin gene promoter is a common cause of silent beta thalassaemia in the Italian population. Br. J. Haematol. **74**:480–486.
44. **Saffer, J. D., S. J. Thurston, M. B. Annarella, and J. G. Compton.** 1990. Localization of the gene for the trans-acting transcription factor Sp1 to the distal end of mouse chromosome 15. Genomics **8**:571–574.
45. **Schreiber, E., P. Matthias, M. M. Muller, and W. Schaffner.** 1988. Identification of a novel lymphoid specific octamer binding protein (OTF-2B) by proteolytic clipping bandshift assay (PCBA). EMBO J. **7**:4221–4229.
46. **Schule, R., M. Muller, H. Otsuka-Murakami, and R. Renkawitz.** 1988. Cooperativity of the glucocorticoid receptor and the CACCC-box binding factor. Nature (London) **332**:87–90.
47. **Spanopoulou, E., V. Giguere, and F. Grosveld.** 1991. The functional domains of the murine *Thy-1* gene promoter. Mol. Cell. Biol. **11**:2216–2228.
48. **Strauss, E. C., N. C. Andrews, D. R. Higgs, and S. H. Orkin.** 1992. In vivo footprinting of the human α-globin locus upstream regulatory element by guanine and adenine ligation-mediated polymerase chain reaction. Mol. Cell. Biol. **12**:2135–2142.
49. **Stuvé, L. L., and R. M. Myers.** 1990. A directly repeated sequence in the β-globin promoter regulates transcription in murine erythroleukemia cells. Mol. Cell. Biol. **10**:972–981.

49a. **Stuvé, L. L., and R. M. Myers.** Unpublished data.

50. **Thiesen, H., and C. Bach.** 1990. Target detection assay (TDA): a versatile procedure to determine DNA binding sites as demonstrated on Sp1 protein. Nucleic Acids Res. **18**:3203–3209.
51. **Treisman, R., S. Orkin, and T. Maniatis.** 1983. Specific transcription and RNA splicing defects in five cloned β-thalassemia genes. Nature (London) **302**:591–596.
52. **Tsai, S. F., E. Strauss, and S. H. Orkin.** 1991. Functional analysis and in vivo footprinting implicate the erythroid transcription factor GATA-1 as a positive regulator of its own promoter. Genes Dev. **5**:919–931.
53. **Wang, W. D., and J. D. Gralla.** 1991. Differential ability of proximal and remote element pairs to cooperate in activating RNA polymerase II transcription. Mol. Cell. Biol. **11**:4561–4571.
54. **Wright, S., A. Rosenthal, R. Flavell, and F. Grosveld.** 1984. DNA sequences required for regulated expression of β-globin genes in murine erythroleukemia cells. Cell **38**:265–273.
55. **Xiao, J. H., 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-I motif of the SV40 enhancer. Genes Dev. **1**:794–807.
56. **Yu, C. Y., K. Motamed, J. Chen, A. D. Bailey, and C. K. Shen.** 1991. The CACC box upstream of human embryonic epsilon globin gene binds Sp1 and is a functional promoter element *in vitro* and *in vivo*. J. Biol. Chem. **266**:8907–8915.