MOLECULAR AND CELLULAR BIOLOGY, Nov. 1992, p. 5024-5032
0270-7306/92/115024-09$02.00/0

Vol. 12, No. 11

A 40-Kilodalton Protein Binds Specifically to an Upstream Sequence Element Essential for Muscle-Specific Transcription of the Human Myoglobin Promoter

RHONDA BASSEL-DUBY,* MARIA D. HERNANDEZ, MARK A. GONZALEZ, JOANNA K. KRUEGER, AND R. SANDERS WILLIAMS

Departments of Internal Medicine and Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75235

Received 18 May 1992/Returned for modification 15 July 1992/Accepted 10 August 1992

To define transcriptional control elements responsible for muscle-specific expression of the human myoglobin gene, we performed mutational analysis of upstream sequences (nucleotide positions −373 to +7 relative to the transcriptional start site) linked to a firefly luciferase gene. Transient expression assays in avian and mammalian cells indicated that a CCCACCCCC (CCAC box) sequence (−223 to −204) is necessary for muscle-specific transcription directed either by the native myoglobin promoter or by a heterologous minimal promoter linked to the myoglobin upstream enhancer region. A putative MEF2-like site (−160 to −169) was likewise necessary for full transcriptional activity in myotubes. Mutations within either of two CANNTG (E-box) motifs (−176 to −148) had only minimal effects on promoter function. We identified and partially purified from nuclear extracts a 40-kDa protein (CBF40) that binds specifically to oligonucleotides containing the CCAC box sequence. A mutation of the CCAC box that disrupted promoter function in vivo also impaired binding of CBF40 in vitro. These data suggest that cooperative interactions between CBF40 and other factors including MEF-2 are required for expression of the human myoglobin gene in skeletal muscle.

Myoglobin is a cytoplasmic hemoprotein that is expressed selectively in skeletal and cardiac myocytes. The temporal and spatial pattern of myoglobin expression during development differs from that of other muscle-specific proteins in several respects: myoglobin transcription in skeletal muscle is activated late in the myogenic program in comparison with genes encoding contractile proteins (12), and its expression is low during fetal life relative to adult levels (37). Furthermore, myoglobin expression is spatially heterogeneous within the embryonic heart and skeletal muscles, even at late developmental stages when other muscle-specific genes are uniformly expressed (26). In adult skeletal muscles, expression of myoglobin is exquisitely sensitive to physiological inputs and undergoes up-regulation in response to stimuli that increase demand for mitochondrial respiration (32). Our long-term goal is to clarify the molecular basis for these distinctive regulatory characteristics.

Commitment and differentiation of skeletal myocytes is driven by a variety of muscle-specific regulatory proteins, several of which are members of the basic helix-loop-helix (bHLH) family of transcription factors (35). These proteins form heterodimers with ubiquitous bHLH proteins, bind DNA at CANNTG (E-box) motifs, and *trans* activate promoters of genes that encode several muscle-specific proteins (18, 27, 28, 38). This role for myogenic bHLH proteins as transcriptional activators has been established in skeletal muscle, but to date little is known about factors that serve an analogous function in cardiac development. In fact, despite extensive efforts, myogenic bHLH proteins have not been found in cardiac tissue (13, 29). Furthermore, studies with certain muscle-specific genes suggest the existence, even in skeletal muscle, of E-box-independent pathways controlling muscle-specific gene transcription (24, 31). Possible participants in such alternative pathways are the recently described MEF-2 (6, 11, 24), M-CAT (20), MAPF (33), and Hox (17, 21) factors.

In this study, we identify sequence elements from the 5′ flanking region of the human myoglobin gene that are required for transcriptional activation during myogenesis. Previously, we observed that the proximal 2-kbp upstream region of 5′ flanking region of the human myoglobin gene contains transcriptional control elements necessary and sufficient to direct muscle-specific expression in cultured avian skeletal myotubes (9) and during prenatal development of transgenic mice (26). A 380-bp upstream region (−373 to +7) functioned identically to the 2-kbp fragment in transient expression assays, thus directing attention to this region for the more detailed mutational analysis of the current study.

The myoglobin upstream region (−373 to +7) includes a pair of E-box motifs, spaced in a manner consistent with functional binding sites for bHLH proteins (18, 27), as well as a putative MEF-2 binding site (6). Previous deletional analysis in our laboratory (9) also directed attention to the upstream region between −261 and −205 as the location of a positive regulatory element and established a requirement for a canonical TATA element (34). The detailed base substitution analysis of the present study indicates that the cytosine-rich region, CCCACCCCC (−219 to −210), and the MEF-2-like sites are required for full activity of the myoglobin promoter in skeletal myotubes. In addition, we have identified and partially purified a 40-kDa protein (CBF40) that binds specifically to the CCCACCCCC (CCAC box) motif. A 3-base substitution mutation that disrupts DNA binding of CBF40 in vitro also compromises transcription in vivo. These data suggest that cooperative interactions between CBF40 and other factors such as MEF-2 are required for muscle-specific transcription of the human myoglobin promoter.

* Corresponding author.

MATERIALS AND METHODS

Cell culture. Sol8 myogenic cells were derived from a primary culture of mouse soleus muscle (7) and were obtained from Vijak Mahdavi. The sol8 monolayers were grown in Dulbecco's modified Eagle medium (DMEM) supplemented with 20% fetal bovine serum. Differentiation was induced by growing the cells to 80% confluence and switching the medium to differentiation medium (DMEM supplemented with 2% horse serum, 10 μg of insulin [GIBCO/BRL, Gaithersburg, Md.] per ml, and 10 μg of transferrin [GIBCO/BRL] per ml). Monolayers of mouse NIH 3T3 cells were grown in DMEM containing 10% fetal bovine serum. Primary cultures of breast muscle cells were isolated from 12-day chicken embryos as previously described (9).

Immunoperoxidase staining of sol8 myoblasts and myotubes. Sol8 cells (10^6) were plated in gelatin-coated 60-mm-diameter plates and grown overnight in growth medium. The cells were washed and fed with differentiation medium. At 0, 24, 48, and 72 h postdifferentiation, the cells were fixed for 3 min with −20°C acetone-ethanol (1:1), washed with cold saline, and incubated for 30 min with either goat antibody against myoglobin (Cappel Laboratories) or monoclonal antibody MF-20 to myosin heavy chain (1). The cells were washed with saline and incubated for 30 min with biotinylated secondary antibody. The Vectastain ABC peroxidase system and 3',3'-diaminobenzidine substrate kit (Vector Laboratories) were used to stain the cells.

Construction of myoglobin-luciferase expression vectors and site-directed mutagenesis of the myoglobin upstream region. Plasmid 380-luc contains a 380-bp *Hin*dIII fragment (−373 to +7) upstream of exon 1 of the human myoglobin gene ligated directly to a *Hin*dIII-*Bam*HI fragment encoding the firefly luciferase gene from plasmid pJD204 (10) and carried in a Bluescript plasmid (Stratagene, La Jolla, Calif.). Plasmid UR2-luc contains the luciferase gene controlled by upstream positions −373 to −205 of the human myoglobin gene ligated to a minimal promoter containing positions −47 to +1 from the human *hsp70* gene. Constructs $(CCAC)_1$-luc, $(CCAC)_2$-luc, $(CCAC)_3$-luc, $(CCAC)_4$-luc, and $(CCAC)_5$-luc also contain the luciferase gene controlled by minimal promoter elements from the human *hsp70* gene but incorporate one through five copies, respectively, of a synthetic oligonucleotide containing the CCAC box region. These $(CCAC)_n$ plasmids were constructed from complementary synthetic oligonucleotides designed to include a *Bam*HI overhang at the 5' end, an internal *Bgl*II site (shown in bold), and a 3' *Sma*I blunt end (shown in italic), as indicated below:

$(CCAC)_n$ sense, 5'-GATCGAACCACCCCACCCCCTGTGGAGATCT*CCC*-3'
$(CCAC)_n$ antisense, 5'-GGGAGATCTCCACAGGGGGTGGGGTGGTTG-3'

To construct $(CCAC)_1$-luc, the *Bam*HI-CCAC box-*Bgl*II-*Sma*I DNA fragment was ligated into complementary sites of vector containing the *hsp70* minimal promoter linked to the luciferase gene. $(CCAC)_2$-luc through $(CCAC)_5$-luc were generated by restricting construct $(CCAC)_1$-luc with *Bgl*II and *Sma*I and inserting, stepwise, one through four additional copies of the CCAC box oligonucleotide. All constructions included a simian virus 40 polyadenylation signal placed upstream of the cloning sites for promoter/enhancer fragments. This segment was included to eliminate low background levels of luciferase activity resulting from cryptic initiation sites within the plasmid vector sequences.

Site-directed mutagenesis was performed by using the Mutator kit (Stratagene) or the oligonucleotide-directed in vitro mutagenesis system (Amersham Corp., Arlington Heights, Ill.). Single-stranded template DNA was generated from the Bluescript or pUC118 phagemid vector containing a *Hin*dIII fragment with the proximal 373 nucleotides of the 5' flanking region of the human myoglobin gene. Oligonucleotides were prepared by using an Applied Biosystems DNA synthesizer and are listed below, with the mutated bases underlined:

CCAC mut 1, 5'-TTCGACACGCACAAGGTACCGACCCCCTGTGG-3'
CCAC mut 2, 5'-TTCGACACGCACAACCAGGTCACCCCCTGTGGCCT-3'
CCAC mut 3, 5'-ACGCACAACCACCCCGGTACCTGTGGCCTGAGC-3'
CCAC mut 4, 5'-ACGCACAACCACCCCACCGGTACCGGCCTGAGCTGTCC-3'
CCAC mut 5, 5'-ACGCACAACCACCCCACCCCGGTACCTGAGCTGTCCTGC-3'
E box mut 1, 5'-CTCGCCACAATGGTACCTGCCCTA-3'
E box mut 2, 5'-CTAAAATAGCTTCGCATGCAGGGCTAGA-3'
E box 1+2, 5'-CTCGCCACAATGGTACCTGCCCTAAAATAGCTTCGCATGCAGGGCTAGA-3'
MEF-2 mut, 5'-CACCTGCCCGCGCGCTTCTTCCCATGTG-3'

These oligonucleotides were designed to introduce a novel restriction site within mutant constructs, so as to facilitate distinction from nonmutant clones during screening procedures. *Hin*dIII fragments containing mutated myoglobin promoter sequences were excised and ligated to Bluescript containing the luciferase gene. The resulting constructs were used to transform *Escherichia coli* DH5α cells. Mutant sequences in each construct were confirmed by DNA sequencing.

Transfection and luciferase assay. Sol8 and NIH 3T3 cells were transfected by electroporation according to Chu et al. (4). Briefly, 5×10^6 cells were resuspended in 0.5 ml of growth medium and electroporated at room temperature in the presence of 10 μg of plasmid and 100 μg of sonicated salmon sperm DNA. Electroporation was performed with the Gene Pulser apparatus (Bio-Rad Laboratories, Richmond, Calif.), using a capacitance of 960 μF at 300 V with a discharge time of 18 to 20 ms. The cells were plated into two gelatin-coated 60-mm-diameter plates containing growth medium with 5 mM butyrate. Following 18 h of incubation at 37°C, the cells were washed and overlaid with differentiation medium. Three days after the addition of differentiation medium, the cells were harvested and lysed in 300 μl of 50 mM Tris-HCl (pH 8)–1% Nonidet P-40–100 μg of aprotinin per ml. Fifty microliters of lysate was added to 300 μl of luciferase reaction buffer containing 25 mM glycylglycine (pH 7.8), 15 mM $MgCl_2$, 2.5 mM ATP, 1 mM dithiothreitol, and 400 μg of bovine serum albumin per ml and assayed in the presence of 1 mM luciferin (Sigma Co.). Light emissions were integrated for the initial 10 s of emission at 25°C in a Berthold LB9500C luminometer. Reporter gene activity for each construct was determined in four or more independent experiments.

The efficiency of transfection was determined by electroporating cells with a β-galactosidase expression vector, pCMV-lacZ (from Radek Skoda, Boston, Mass.). Two methods were used to monitor β-galactosidase expression. In method 1, prior to the addition of differentiation medium, the transfected cells were stained with X-Gal (5-bromo-4-chloro-3-indolyl-β-galactoside), and the percentage of blue cells was calculated. In method 2, 30 μl of a 1:10 dilution of cellular lysate from transfected cells was added to a final concentration of 0.8 mg of *o*-nitrophenyl-β-D-galactopyranoside per ml in a buffer containing 0.1 M sodium phosphate (pH 7.9), 1 mM $MgCl_2$, and 45 mM 2-mercaptoethanol at a final volume of 300 μl. The samples were incubated at 37°C for 30 min, and then the A_{420} was measured with a spectrophotometer. Constructs containing the luciferase reporter gene were cotransfected with pCMV-lacZ expression vector to normalize for variations in transfection efficiency. The nonrestrictive human β-actin promoter was used to assess transfection of NIH 3T3 cells. This construct contained a 4.3-kb pro-

moter fragment from the human β-actin gene (19) linked to the luciferase gene.

Primary cultures of breast muscle cells were transfected by using the Lipofectin reagent (GIBCO/BRL). Briefly, 6×10^5 cells were plated in gelatin-coated 60-mm-diameter culture plates (Costar) and incubated at 37°C in growth medium for 24 h. The cells were transfected in serum-free medium with 1 to 2 μg of plasmid DNA. Pilot studies were performed with each plasmid to establish the optimal ratio of DNA to Lipofectin. After 18 h of incubation, the cells were washed and fed with fresh growth medium. Four days after transfection, cells were harvested and lysed in 600 μl of 25 mM Tris-phosphate (pH 7.4)–8 mM $MgCl_2$–1 mM EDTA–1% Triton X-100–1% bovine serum albumin–15% glycerol–1 mM dithiothreitol. Fifty microliters of lysate was added to 300 μl of luciferase reaction buffer, and luciferase activity was determined as described above. Reporter gene activity for each construct was determined in four or more independent experiments.

Preparation of nuclear proteins. Nuclear extracts from sol8 myoblasts and myotubes were prepared by washing the cell monolayers three times with isotonic-buffered saline and then adding to each 150-mm-diameter plate of cells 1 ml of lysis buffer (20 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid [HEPES; pH 7.6], 10 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 20% glycerol, 0.1% Triton X-100, 10 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride, 10 μg of leupeptin per ml, 10 μg of pepstatin per ml, 100 μg of aprotinin per ml). Nuclei were harvested by scraping the plates with a rubber policeman and then centrifuged for 1 min at 2,000 rpm at 4°C. The pellet was resuspended in nuclear extraction buffer (lysis buffer with 500 mM NaCl), rocked for 1 h at 4°C, and centrifuged at 12,000 rpm for 10 min at 4°C. The supernatant was collected and stored at 4°C. Protein concentration was determined by the Bio-Rad protein assay based on the Bradford dye-binding procedure (Bio-Rad, Richmond, Calif.).

DNA probes and competitors. Oligonucleotides corresponding to both strands of the CCAC box region (−226 to −205 upstream region of myoglobin) were synthesized with an added GATC nucleotide overhang at the 5'-terminal end of each oligonucleotide. The complementary strands were annealed, purified from a 12% polyacrylamide gel, and either end labeled with Klenow fragment of DNA polymerase I (Promega) and [α-^{32}P]dATP (3,000 Ci/mmol) or used as unlabeled competitors in protein binding experiments. In a similar manner, oligonucleotides corresponding to CCAC mut 3 or to a documented binding site for Sp1 (15) were prepared and used as unlabeled competitors in gel mobility shift assays. The sequences of these oligonucleotides were as follows:

CCAC sense,	5'-GATCACGGCACAACCACCCCACCCCCTGTG-3'
CCAC antisense,	5'-GATCCACAGGGGGTGGGGTGGTTGTGCGT-3'
CCAC mut3 sense,	5'-ACGCACAACCACCCCGGTACCTGTGGCCTGAGC-3'
CCAC mut3 antisense,	5'-GTGCTCAGGCCACAGGTACCGGGGTGGTTGT-3'
Sp1 sense,	5'-AGGGAGGCGTGGCCTGGGCGGGACTGGGG-3'
Sp1 antisense,	5'-CCCCAGTCCCGCCCAGGCCACGCCTCCCT-3'

The CCAC box probe used for binding to renatured protein blots (Southwestern [DNA-protein] blot analysis) and for biochemical isolation of the CCAC-box-binding protein was generated by excising a *Bam*HI-*Bgl*II fragment from plasmid $(CCAC)_3$-luc. For use in Southwestern blot analysis, the fragment was end labeled with the Klenow fragment of DNA polymerase I (Promega) in the presence of [α-^{32}P]dATP (3,000 Ci/mmol). For biochemical isolation of the CCAC-box-binding protein, biotin was incorporated into the ends of the *Bam*HI-*Bgl*II $(CCAC)_3$ fragment by using the Klenow fragment of DNA polymerase I (Promega) in the presence of biotin-14-dATP (GIBCO/BRL).

Gel mobility shift assays. Mobility shift reactions were prepared in a final volume of 12 μl by combining 1 ng of end-labeled double-stranded DNA (~10^4 cpm) with nuclear extract (~3 μg) and 1 μg of poly(dI-dC) (Boehringer Mannheim Biochemicals, Indianapolis, Ind.) in a binding buffer containing 25 mM HEPES (pH 7.6), 100 mM KCl, 5 mM $MgCl_2$, 5 mM dithiothreitol, 10% glycerol, and 0.2% Nonidet P-40. The amount of unlabeled competitor DNA added is indicated in the figure legends. After 20 min of incubation at room temperature, the reactions were loaded onto a 4% native polyacrylamide gel equilibrated in 45 mM Tris–45 mM boric acid–1 mM EDTA. Subsequent to electrophoresis at 10 V/cm for 40 min, the gels were dried and exposed to Amersham Hyperfilm-ECL film for 20 h.

Southwestern blot analysis. The protocol was modified from that of Miskimins et al. (22). Nuclear extract (~3 μg of protein) was mixed with sample buffer (62.5 mM Tris-HCl [pH 6.8], 3% sodium dodecyl sulfate [SDS], 10% glycerol, 0.6 M 2-mercaptoethanol, 0.01% bromphenol blue) and boiled for 3 min. Proteins were resolved by SDS-polyacrylamide gel electrophoresis (PAGE). Following electrophoresis at 15 V/cm for 2 h in 10 or 12% gels, the proteins were transferred to nitrocellulose paper overnight at 4°C, using the Mini Trans-Blot Cell (Bio-Rad) at 120 mA in transfer buffer (50 mM Tris, 40 mM glycine, 0.04% SDS, 20% methanol). The nitrocellulose paper was air dried at room temperature, immersed in binding buffer (25 mM HEPES [pH 7.6], 60 mM KCl, 1 mM EDTA, 1 mM dithiothreitol) supplemented with 6 M guanidine hydrochloride and gently rocked for 10 min at 4°C. The nitrocellulose paper was transferred to binding buffer containing 3 M guanidine hydrochloride and gently rocked for 10 min at 4°C. This procedure was repeated eight times, with each subsequent wash containing a concentration of guanidine hydrochloride twofold less than that in the previous wash. The final wash step consisted of binding buffer with no guanidine hydrochloride. The nitrocellulose was incubated in BLOTTO buffer (binding buffer containing 5% Carnation nonfat milk powder and 5 μg of sonicated salmon sperm DNA per ml) for 1 h at room temperature. Next, the nitrocellulose filter was immersed in BLOTTO consisting of 0.25% Carnation nonfat milk powder and incubated for 30 min. ^{32}P-labeled DNA probe was added to BLOTTO containing 0.25% Carnation nonfat milk powder and incubated with the nitrocellulose filter for 2 h at room temperature. The filter was washed four times with binding buffer for 7 min at room temperature. The filter was air dried and exposed to Amersham Hyperfilm-ECL film with an intensifying screen.

Biochemical isolation of CCAC-box-binding proteins. Biotinylated $(CCAC)_3$ DNA (~500 μg) was incubated with nuclear extract (~200 μg) for 30 min at room temperature. One-half milliliter of streptavidin-agarose (Sigma) was equilibrated with buffer F [10 mM HEPES (pH 7.6), 10 mM KCl, 1.5 mM $MgCl_2$, 1.0 mM dithiothreitol, 0.2% Nonidet P-40, 0.1 mM phenylmethylsulfonyl fluoride, 30 μg of poly(dI-dC) per ml] and added to the mixture of DNA and nuclear extract. The mixture was incubated overnight at room temperature and then poured into a 1-ml column. The column was washed five times with 1 ml of buffer N (10 mM HEPES [pH 7.6], 10 mM KCl, 1.5 mM $MgCl_2$, 0.1 mM EDTA, 1.0 mM dithiothreitol, 0.1 mM phenylmethylsulfonyl fluoride, 10% glycerol). The CCAC-box-binding proteins were eluted in two steps: first by adding 0.5 ml of buffer B (10 mM

TABLE 1. Expression patterns of endogenous myoglobin and myosin in myoblasts and myotubes

Cell line	Differentiation stage	Myoglobin protein	Myosin heavy-chain protein
Sol8	Myoblast	−	−
Sol8	Day 1 postdifferentiation	−	+
Sol8	Day 2 postdifferentiation	+	+
C2C12	Myoblast	−	−
C2C12	Day 1 postdifferentiation	−	+
C2C12	Day 2 postdifferentiation	+/−	+

Tris-HCl [pH 6.8], 150 mM KCl, 1.0 mM dithiothreitol, 0.1 mM EDTA, 0.2% Nonidet P-40, 0.1 mM phenylmethylsulfonyl fluoride, 0.25 μg of bovine serum albumin per ml, 10% glycerol) and collecting 50-μl fractions and then by adding 0.5 ml of buffer BI (buffer B containing 500 mM KCl) and collecting 50-μl fractions. The eluted proteins were loaded onto an SDS–8 to 25% gradient polyacrylamide gel and electrophoresed in the PhastSystem (Pharmacia LKB, Alameda, Calif.). The proteins were visualized by silver staining, using the PhastGel silver kit.

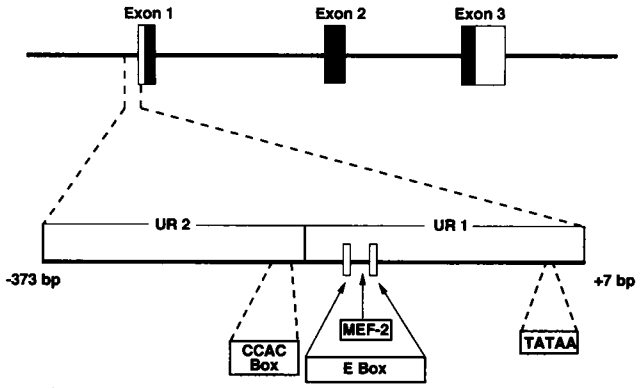

FIG. 1. Schematic diagram of the human myoglobin gene and the upstream promoter region. The 10.4-kb myoglobin gene consists of three exons, as illustrated by boxes in which the nonshaded part denotes the noncoding region and the shaded section represents coding sequence. A 380-bp upstream region (−373 to +7) has been enlarged to highlight the positions of sequence motifs, including the CCAC box, E boxes, putative MEF-2 binding site, and TATAA element.

RESULTS

Sol8 myotubes express myoglobin. Differentiation of sol8 myogenic cells was monitored by morphological changes (formation of multinucleated myotubes) and by immunohistochemistry, using antibodies directed against myoglobin or myosin heavy-chain protein (Table 1). Undifferentiated myoblasts expressed neither myoglobin nor myosin heavy-chain protein. Upon growth to confluence and withdrawal of growth factors associated with a shift to differentiation medium, myosin heavy-chain protein was detected within 1 day, whereas myoglobin protein was not detectable at this early stage of differentiation. Two days postdifferentiation, sol8 myotubes uniformly expressed both endogenous myoglobin and myosin heavy-chain protein. Myoglobin mRNA also was undetectable by Northern (RNA) analysis (not shown) in sol8 myoblasts but was present in myotubes maintained in differentiation medium for 2 or more days. A second murine myogenic cell line, C2C12, expressed myoglobin only in a subpopulation of myotubes (indicated by +/− in Table 1) even after 2 or more days in differentiation medium, when all C2C12 myotubes expressed myosin heavy-chain protein (Table 1). These results demonstrated that the endogenous myoglobin gene is transcriptionally active in sol8 myotubes.

Myoglobin upstream regions direct muscle-specific expression of luciferase when linked either to the myoglobin core promoter or to a heterologous minimal promoter. Previously, we reported that a 380-bp upstream fragment (−373 to +7) from the human myoglobin gene (Fig. 1) can direct muscle-specific expression of a reporter gene in primary cultures established from embryonic chicken myoblasts. Transcripts generated by using this fragment as a promoter are initiated at the same site as is the endogenous human myoglobin gene and accumulate in parallel to the activity of the reporter enzyme (9, 34). In the present study, we observed that this 380-bp fragment also directs cell-specific and developmentally regulated expression of a reporter gene (the firefly luciferase gene) in mammalian myogenic cells.

A construct containing the 380-bp myoglobin upstream fragment exhibited 23-fold-greater activity than did a promoterless construct in sol8 myotubes (Fig. 2A). Similar results were seen when this construct was introduced into primary cultures of skeletal myotubes derived from chicken embryos (data not shown). However, this myoglobin upstream fragment was devoid of promoter activity in a murine nonmuscle cell line, NIH 3T3 (Fig. 2B). The absence of promoter activity directed by the myoglobin upstream fragment in NIH 3T3 fibroblast cells was attributable neither to low efficiencies of transfection nor to instability of reporter gene products; luciferase expression data were internally controlled by cotransfection of pCMV-lacZ, and the nonrestricted β-actin promoter drove high levels of luciferase activity in all cell types (Fig. 2B).

To define functional domains within the myoglobin upstream region, the 380-bp fragment was divided into smaller fragments, UR1 and UR2 (Fig. 1). The UR2 fragment encompassing nucleotide residues −373 through −205 was placed upstream of a minimal promoter derived from the human *hsp70* gene linked to the luciferase coding region. This construct also was expressed selectively in differentiated myotubes, indicating that sequence elements capable of directing muscle-specific transcription reside within UR2 (Fig. 2).

Mutations within the CCAC box impair muscle-specific expression directed by myoglobin upstream sequences. The myoglobin UR2 segment contains a cytosine-rich region characterized by the sequence CCACCCCACCCCC (boxed in Fig. 3A). On the basis of previous deletional analysis (9), this region was suspected to contain a positive regulatory element, but the specific sequences responsible for this activity had not been identified. We constructed a series of nucleotide substitution mutations to scan this region (Fig. 3A). Mutations of bases flanking the CCCACCCCC sequence (CCAC mut 1 or 5) had minimal (<2-fold) or no effects on expression of the reporter gene. By contrast, disruption of bases within the CCCACCCCC sequence (CCAC mut 2, 3, or 4) impaired expression more severely; the mutation present in CCAC mut 3 reduced promoter function more than 20-fold.

Mutations of the E-box motifs exert little or no effect on the function of the myoglobin promoter, while the putative MEF-2

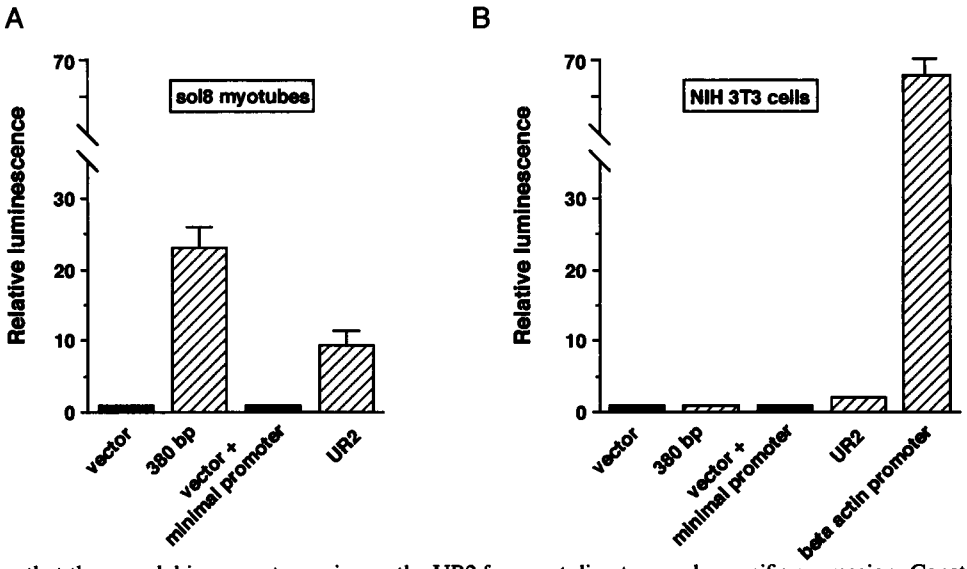

FIG. 2. Evidence that the myoglobin promoter region or the UR2 fragment directs muscle-specific expression. Constructs bearing either the 380-bp upstream fragment (−373 to +7) of the myoglobin gene, the UR2 fragment (−373 to −205) coupled to a minimal promoter (from human *hsp70*), or the human β-actin promoter were linked to the luciferase coding region and transfected into sol8 myogenic cells (A) or NIH 3T3 fibroblasts (B). Luciferase activity was measured in cell lysates 3 days after differentiation of sol8 cells or 1 day after transfection of NIH 3T3 cells. Results were expressed relative to the activity of a promoterless luciferase gene (vector) and represent mean values (± standard error of the mean) from four or more independent observations. All results were normalized by cotransfection of pCMV-lacZ expression vector.

motif is a positive regulatory element. Disruption of one or both of the E-box motifs (boxed in Fig. 4A) within the myoglobin 5′ flanking region (see Fig. 1) resulted in constructs that were expressed in a manner similar to the wild-type promoter in both mammalian (Fig. 4B) and avian (data not shown) myotubes. Thus, although less extensive mutational analysis was directed at this region than to the CCAC box, it appears that canonical E-box elements comprising putative binding sites for myogenic bHLH proteins are not essential for muscle-specific transcription directed by the myoglobin promoter. By contrast, mutations within a putative MEF-2 binding site (dashed box in Fig. 4A) reduced expression of the reporter gene to less than 20% of control levels (Fig. 4).

The CCAC box element alone is not sufficient for muscle-specific transcription. The finding that the CCAC box region was necessary for muscle-specific transcription led us to determine whether the CCAC box element, acting alone, was sufficient to augment transcription from a minimal promoter and whether such activating function, if present,

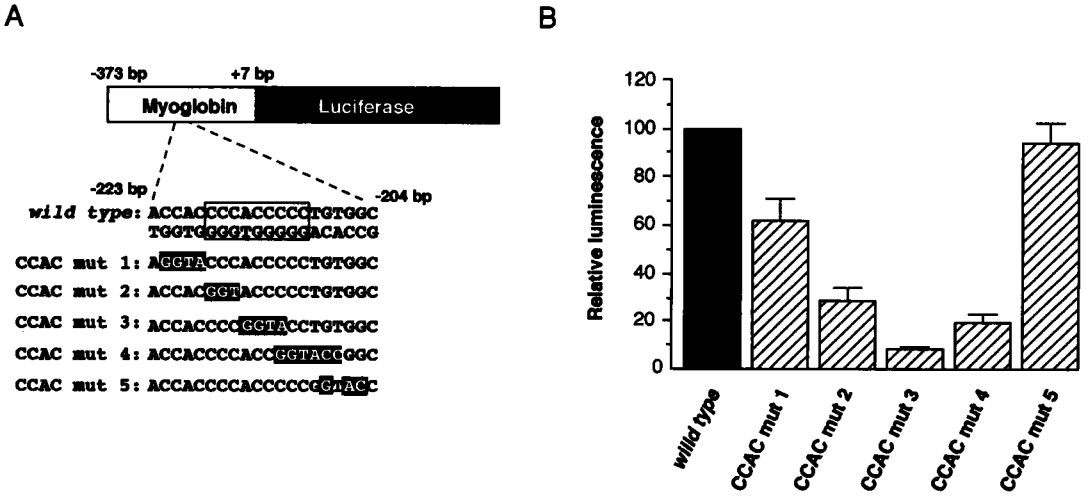

FIG. 3. Mutations of the CCAC box motif within the myoglobin promoter region. (A) Within the UR2 region (between bp −204 through −223) lies a CCCACCCCC sequence outlined in a box. Base substitutions in mutant constructs are highlighted. (B) The CCAC box mutants within the 380-bp fragment were ligated to the luciferase gene and transfected into sol8 muscle cells. Three days postdifferentiation, cell lysates were assayed for luciferase activity. The results were expressed as percent activity with respect to the luciferase expression controlled by the wild-type myoglobin 380-bp sequence. All results were normalized by cotransfections of pCMV-lacZ expression vector and represent mean values (± standard error of the mean) of at least four independent observations.

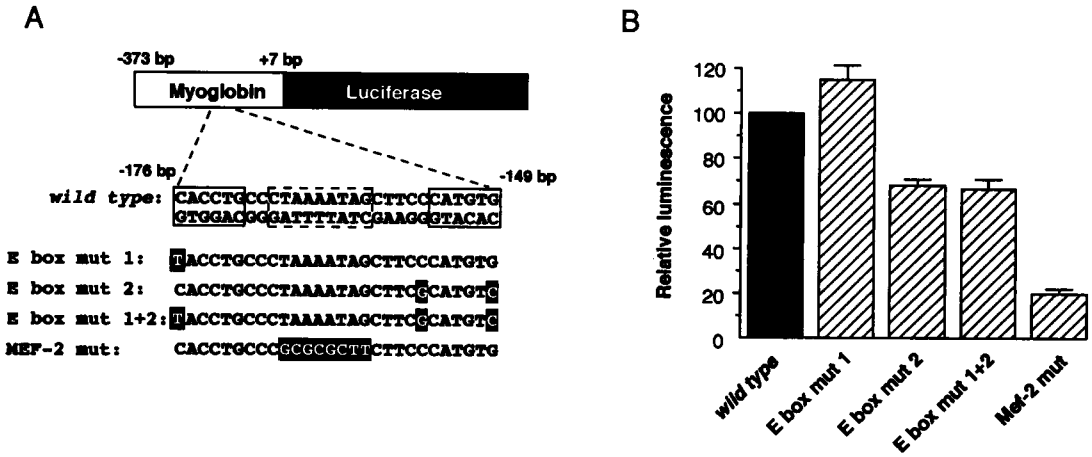

FIG. 4. Mutations of the E-box motifs or MEF-2-like motif within the myoglobin promoter region. (A) Within the myoglobin promoter between nucleotides residues −149 through −176 lie two E-box motifs (outlined in boxes) and a putative MEF-2 binding motif (outlined in a dashed box). Base substitutions that alter one or both E-box motifs or the putative MEF-2 binding site are highlighted. (B) The E-box or MEF-2 mutants contained in the 380-bp fragment were ligated to the luciferase gene and transfected into sol8 muscle cells. Three days postdifferentiation, the cell lysates were assayed for luciferase activity. Data were calculated and presented as described in the legend to Fig. 3.

was limited to differentiated myotubes. The results are shown in Fig. 5. A single copy of the CCAC box had no effect, but multiple copies increased the activity of the minimal promoter, producing up to a sixfold-activating effect in the case of the construct containing five copies (Fig. 5A). This effect, however, was not limited to myotubes; similar activating effects were observed in NIH 3T3 cells (Fig. 5B). Promoter activities of constructs containing multimeric CCCACCCCC elements remained well below those of constructs containing native myoglobin upstream fragments (note different scales on the vertical axes in Fig. 2 and 5).

The CCAC box element binds specifically to nuclear proteins. A single copy of the CCCACCCCC element was recognized by nuclear proteins present in sol8 myotubes (Fig. 6A, lane 2), myoblasts (Fig. 6B, lane 2), and 3T3 cells (data not shown) in gel mobility shift assays. Binding of nuclear protein factors was sequence specific, as assessed by competitive binding of unlabeled oligonucleotides bearing either the wild-type CCCACCCCC sequence or the functionally inactive CCAC mut 3 sequence (Fig. 6). Other competitive binding studies (not shown) indicated that the CCAC-box-binding factor in sol8 myotubes is distinct from the transcription factor Sp1, which recognizes a similar but nonidentical DNA sequence (CCGCCC) (16).

Characterization and partial purification of a CCAC-box-binding factor. To identify nuclear proteins that bind to the CCAC box sequence, we used a protein-blotting technique (Southwestern blotting). Nuclear extracts from sol8 myotubes were subjected to SDS-PAGE, electroblotted onto nitrocellulose paper, renatured, and probed with ^{32}P-labeled

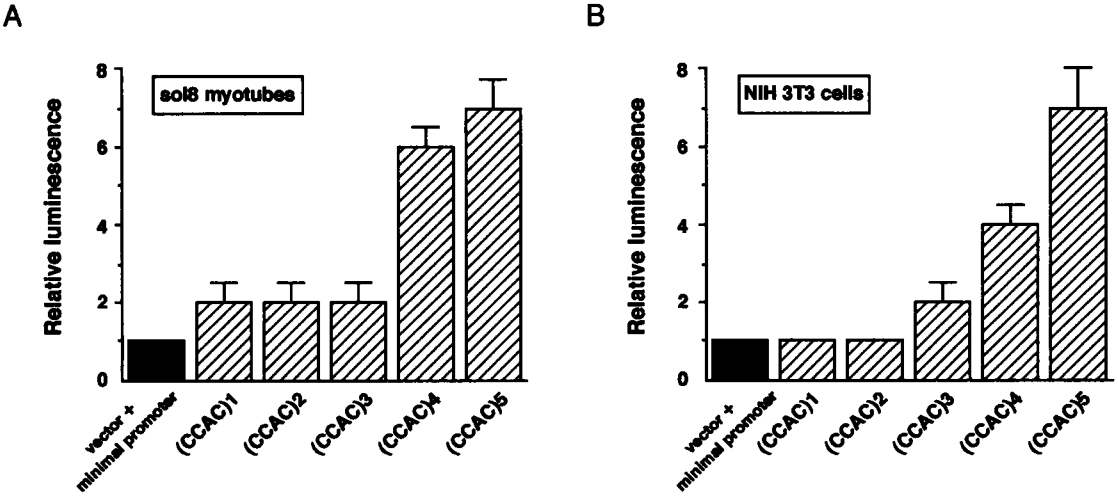

FIG. 5. Functional analysis of multiple copies of the CCAC box element placed upstream of the *hsp70* minimal promoter region. One through five copies of the CCCACCCCC box element (5′-TTCCACACGCACAACCACCCCACCCCCTGTGGCCT-3′) were placed upstream of the minimal promoter linked to the luciferase gene. The constructs were used to transfect sol8 muscle cells (A) or NIH 3T3 fibroblast cells (B). Three days after differentiation of sol8 cells or 1 day after transfection of NIH 3T3 cells, the cell lysates were assayed for luciferase activity. Data were calculated and presented as described in the legend to Fig. 2.

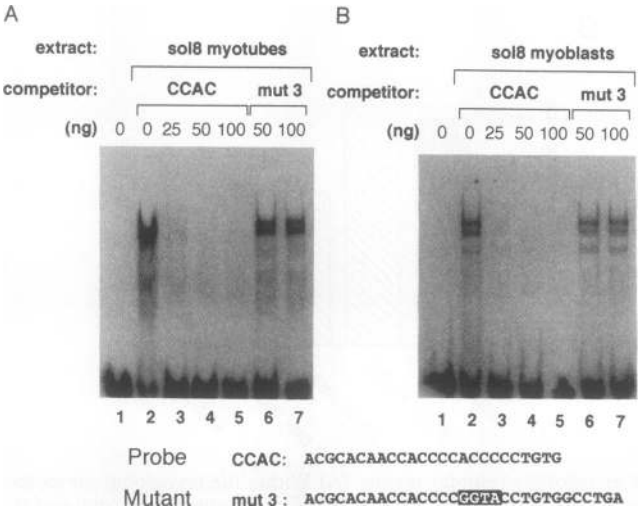

FIG. 6. Specific binding of the CCAC box element to nuclear factors. A ^{32}P-labeled CCAC box probe was used in a gel mobility shift assay to identify nuclear factors that bind to the CCAC box element. Competitive binding studies included 25 ng (lane 3), 50 ng (lane 4), or 100 ng (lane 5) of unlabeled CCAC box sequence. A functionally inactive CCAC box mutant, mut 3, was used in competition studies at an amount of 50 ng (lane 6) or 100 ng (lane 7). The DNA sequences of the CCAC probe and CCAC mut 3 are shown.

oligonucleotides consisting of one or more copies of the myoglobin CCAC box element. All of these probes identified a 40-kDa protein in renatured protein blots (Fig. 7A) that we termed CBF40. Binding of the probe varied in relation to the amount of nuclear extract loaded on the polyacrylamide gel (compare lanes 2 and 3) or the copy number of CCCACC CCC elements (compare lane 1 with lane 2 and lane 3 with lane 4). A ^{32}P-labeled oligonucleotide consisting of a binding site for Sp1 (15) bound a renatured protein with a molecular

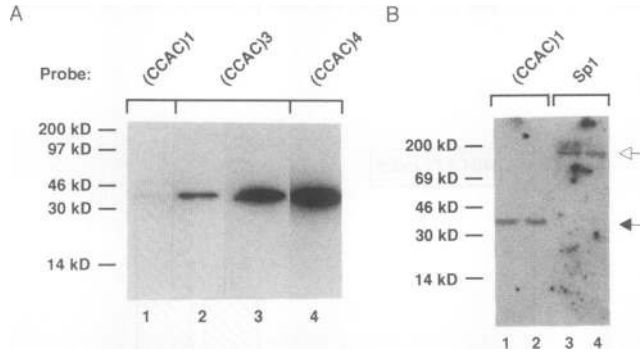

FIG. 7. Binding of ^{32}P-labeled oligonucleotides to renatured protein blots. (A) Nuclear proteins from sol8 myotubes were resolved by SDS-PAGE (12% polyacrylamide gel) after loading of 10 μg (lanes 1 and 2) or 25 μg (lanes 3 and 4) per lane and electrophoretically transferred to nitrocellulose. Proteins bound to nitrocellulose were renatured as described in Materials and Methods. The blots were probed with one (lane 1), three (lanes 2 and 3), or four (lane 4) copies of the ^{32}P-labeled CCAC box sequence motif. (B) Similar renatured protein blots (SDS-PAGE [10% polyacrylamide gel]) were probed with ^{32}P-labeled oligonucleotides containing either one copy of the myoglobin CCAC box motif (lanes 1 and 2) or a documented binding site for Sp1 (lanes 3 and 4). The open arrow points to the protein that binds Sp1 oligonucleotide, and the closed arrow points to the protein that binds the CCAC box sequence.

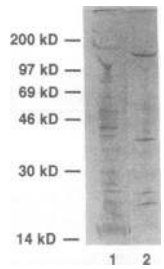

FIG. 8. Partial purification of the CCAC-box-binding factor. Shown is a silver stain of proteins separated by SDS-PAGE (8 to 25% gradient gel) loaded with unfractionated nuclear extract (lane 1) or affinity-purified fraction (lane 2). Proteins were eluted from streptavidin-agarose columns with elution buffer containing 0.5 M KCl. The arrow points to the 40-kDa protein (CBF40) corresponding in size to the protein binding to CCCACCCCC sequences in renatured protein blots (see Fig. 7).

weight consistent with the known size of Sp1 (Fig. 7B). The oligonucleotide containing the binding site for Sp1 failed to bind CBF40. This finding indicated that CBF40 protein is distinct from Sp1.

Partial purification of CBF40 was achieved by DNA affinity chromatography. Nuclear extracts were incubated with a biotinylated oligonucleotide containing the myoglobin CCCACCCCC element. Protein-DNA complexes were bound to streptavidin-agarose beads, and proteins were eluted stepwise with buffer containing increasing concentrations of KCl. After separation by SDS-PAGE, a limited number of protein bands was observed by silver staining in fractions eluted with 0.5 M KCl (Fig. 8). The protein band at approximately 40 kDa corresponds with the results of oligonucleotide binding to renatured protein blots (Fig. 7).

DISCUSSION

We have identified a sequence motif, CCCACCCCC (CCAC box), that is required for muscle-specific transcription directed by the human myoglobin promoter. This CCAC box element is not sufficient in itself to drive muscle-specific expression of a minimal promoter but requires cooperative interactions with other *cis*-acting elements. Such auxiliary elements include a putative MEF-2 binding site located downstream of the CCAC box (Fig. 1). As yet undefined elements located upstream of the CCAC box also appear to be able to complement the function of the CCAC box, at least when placed in close proximity to the TATA box of a minimal promoter, as in the UR2-luc construct (Fig. 1 and 2). In the transient expression assays used in this study, binding sites for myogenic bHLH proteins (E-box motifs) are not essential for activation of the myoglobin promoter during muscle differentiation.

A 40-kDa protein binds specifically to the myoglobin CCAC box and was termed CBF40. A mutation within the CCAC box motif that reduces binding of this factor in vitro also compromises transcription in vivo. A protein of similar size and binding activity is present both in differentiated myotubes, in which the myoglobin promoter is transcriptionally active, and in undifferentiated myoblasts or fibroblasts, in which myoglobin is not expressed. Thus, muscle-specific enhancer activity of regions containing a CCAC box does not appear to be attributable simply to the exclusive expression of CBF40 in muscle cells.

Sequences containing an adenine flanked by multiple

cytosines are present within regions upstream of other mammalian myoglobin genes (2, 3) as well as within muscle-specific enhancers from muscle creatine kinase (14), alpha-myosin heavy-chain (5), and slow-cardiac troponin C (25) genes. However, this motif seems to be required for muscle-specific enhancer activity only in a subset of the genes in which it appears. For example, disruption of a CCCCCAC CCCC region within an upstream enhancer of the muscle creatine kinase gene has little effect on enhancer function (14). In contrast, Parmacek et al. (25) observed a requirement for CCACCCCC motifs within both cardiac-specific and skeletal muscle-specific enhancer regions from the slow-cardiac troponin C gene. Among genes that do not exhibit muscle-specific expression, a tryptophan oxygenase gene requires a CCACACCCC sequence for induction by glucocorticoids (30).

Erythroid globin genes utilize a CCACCCC sequence as an important positive transcriptional control element (8). Because myoglobin and erythroid globins have a common evolutionary heritage and are similar in genomic organization (36), it is interesting to speculate that transcriptional control elements may have been retained in conjunction with exon sequences during genomic rearrangements that led to divergence of the globin gene family, but modified (on the basis of variations in flanking sequences or combinatorial effects with other regulatory elements) to participate in entirely different programs of gene activation.

The apparent paradox in which similar *cis*-acting DNA sequence motifs participate in differentiation of unrelated cell lineages has several possible explanations. All CCAC boxes may bind an identical and ubiquitous protein factor that activates transcription, but only in the context of cooperative interactions with other, lineage-specific, transcriptional regulators. Alternatively, the superficial similarity of cytosine-rich transcriptional control elements may be misleading, and such sites may bind to a diversity of proteins. Different members of a family of CCAC-box-binding proteins may discriminate among CCAC box motifs on the basis of differences in flanking sequence around a CAC core. It is conceivable that certain members of a family of CCAC-box-binding proteins may be restricted to a given cell lineage or developmental stage, in a manner analogous to E-box-binding proteins of the bHLH family (23).

Our data suggest that both explanations may be correct. Even multimers of the myoglobin CCCACCCCC motif are insufficient to confer muscle specificity to a minimal promoter, indicating a requirement for cooperative interactions between CCAC-box-binding factors and other transcriptional activators. Our data also support the existence of a family of CCAC-box-binding proteins. In the human embryonic ε-globin gene, a CCACCCC sequence is a functional promoter element in transient expression assays and binds Sp1 protein in vitro (39). Our present results, however, suggest that the nuclear factor that recognizes the CCCAC CCCCC element of the human myoglobin gene is distinct from Sp1, as determined by competitive binding studies and Southwestern analysis. Thus, at least two discrete proteins, Sp1 and CBF40, bind apparently similar DNA sequence motifs. Further studies and more complete characterization of CBF40 will be required to define the full repertoire of cognate factors that *trans* activate myoglobin and other genes by binding to CCAC box elements and the nature of cooperative interactions with other transcription factors.

ACKNOWLEDGMENTS

We are grateful to Carol Mendelson and Erwei Gao for helpful advice on protein renaturation and transcription factor purification. We thank Zohre German for skillful technique in preparing the primary chicken muscle cells, R. Gaynor for providing the Sp1 oligonucleotides, and W. Wright and J. Sambrook for critical reading of the manuscript.

This work was supported by grants from NIH (RO1 AR40849) and from the Perot Family Foundation. M.D.H. is supported by NIH training grant HL07360.

REFERENCES

1. **Bader, D., T. Masaki, and D. A. Fischman.** 1982. Immunochemical analysis of myosin heavy chain during avian myogenesis in vivo and in vitro. J. Cell Biol. **95**:763–770.
2. **Blanchetot, A., M. Price, and A. J. Jeffreys.** 1986. The mouse myoglobin gene: characterization and sequence comparison with other myoglobin genes. Eur. J. Biochem. **159**:469–474.
3. **Blanchetot, A., V. Wilson, D. Wood, and A. J. Jeffreys.** 1983. The seal myoglobin gene: an unusually long globin gene. Nature (London) **301**:732–734.
4. **Chu, G., H. Hayakawa, and P. Berg.** 1987. Electroporation for the efficient transfection of mammalian cells with DNA. Nucleic Acids Res. **15**:1311–1326.
5. **Cribbs, L. L., N. Shimizu, C. E. Yockey, J. E. Levin, S. Jakovcic, R. Zak, and P. K. Umeda.** 1989. Muscle-specific regulation of a transfected rabbit myosin heavy chain β gene promoter. J. Biol. Chem. **264**:10672–10678.
6. **Cserjesi, P., and E. N. Olson.** 1991. Myogenin induces the myocyte-specific enhancer binding factor MEF-2 independently of other muscle-specific gene products. Mol. Cell. Biol. **11**: 4854–4862.
7. **Daubas, P., A. Klarsfeld, I. Garner, C. Pinset, R. Cox, and M. Buckingham.** 1988. Functional activity of the two promoters of myosin alkali light chain gene in primary muscle cell cultures: comparison with other muscle gene promoters and other culture systems. Nucleic Acids Res. **16**:1251–1271.
8. **deBoer, E., M. Antoniou, V. Mignotte, L. Wall, and F. Grosveld.** 1988. The human beta-globin promoter; nuclear protein factors and erythroid specific induction of transcription. EMBO J. **7**:4203–4212.
9. **Devlin, B. H., F. C. Wefald, W. E. Kraus, T. S. Bernard, and R. S. Williams.** 1989. Identification of a muscle-specific enhancer within the 5′ flanking region of the human myoglobin gene. J. Biol. Chem. **264**:13896–13900.
10. **deWet, J. R., K. V. Wood, M. DeLuca, D. R. Helsinki, and S. Subramani.** 1987. Firefly luciferase gene: structure and expression in mammalian cells. Mol. Cell. Biol. **7**:725–737.
11. **Gosset, L. A., D. J. Kelvin, E. A. Sternberg, and E. N. Olson.** 1989. A new myocyte-specific enhancer-binding factor that recognizes a conserved element associated with multiple muscle-specific genes. Mol. Cell. Biol. **9**:5022–5033.
12. **Gunning, P., E. Hardeman, R. Wade, P. Ponte, W. Bains, H. M. Blau, and L. Kedes.** 1987. Differential patterns of transcript accumulation during human myogenesis. Mol. Cell. Biol. **7**:4100–4114.
13. **Hopwood, N. D., A. Pluck, and J. B. Gurdon.** 1989. MyoD expression in the forming somites is an early response to mesoderm induction in *Xenopus* embryos. EMBO J. **8**:3409–3417.
14. **Horlick, R. A., and P. A. Benfield.** 1989. The upstream muscle-specific enhancer of the muscle creatine kinase gene is composed of multiple elements. Mol. Cell. Biol. **9**:2396–2413.
15. **Jones, K. A., J. T. Kadonaga, P. A. Luciw, and R. Tjian.** 1986. Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1. Science **232**:755–759.
16. **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.
17. **Kessel, M., and P. Gruss.** 1990. Murine development control genes. Science **249**:374–379.

18. **Lassar, A. B., J. N. Buskin, D. Lockshon, R. L. Davis, S. Apone, S. D. Hauschka, and H. Weintraub.** 1989. MyoD is a sequence-specific DNA binding protein requiring a region of *myc* homology to bind to the muscle creatine kinase enhancer. Cell **58**:823–831.
19. **Leavitt, J., P. Gunning, S. Y. Ng, C. S. Lin, and L. Kedes.** 1984. Molecular cloning and characterization of mutant and wild-type human beta-actin genes. Mol. Cell. Biol. **4**:1961–1969.
20. **Mar, J. H., and C. P. Ordahi.** 1990. M-CAT binding factor, a novel *trans*-acting factor governing muscle-specific transcription. Mol. Cell. Biol. **10**:4271–4283.
21. **McGinnis, W., and R. Krumlauf.** 1992. Homeobox genes and axial patterning. Cell **68**:283–302.
22. **Miskimins, W. K., M. P. Roberts, A. McClelland, and F. H. Ruddle.** 1985. Use of a protein-blotting procedure and a specific DNA probe to identify nuclear proteins that recognize the promoter region of the transferrin receptor gene. Proc. Natl. Acad. Sci. USA **82**:6741–6744.
23. **Montarras, D., J. Chelly, E. Bober, H. Arnold, M. Ott, F. Gros, and C. Pinset.** 1991. Developmental patterns in the expression of myf5, myoD, myogenin, and MRF4 during myogenesis. New Biol. **3**:592–600.
24. **Navankasattusas, S., H. Zhu, A. V. Garcia, S. M. Evans, and K. R. Chien.** 1992. A ubiquitous factor (HF-1a) and a distinct muscle factor (HF-1b/MEF-2) form an E-box-independent pathway for cardiac muscle gene expression. Mol. Cell. Biol. **12**:1469–1479.
25. **Parmacek, M. S., A. J. Vora, T. Shen, E. Barr, F. Jung, and J. M. Leiden.** 1992. Identification and characterization of a cardiac-specific transcriptional regulatory element in the slow/cardiac troponin C gene. Mol. Cell. Biol. **12**:1967–1976.
26. **Parsons, W. J., J. A. Richardson, K. H. Graves, R. S. Williams, and R. W. Moreadith.** 1992. Chamber-specific and transmurally heterogeneous transcription directed by the human myoglobin promoter in hearts of transgenic mice. Submitted for publication.
27. **Piette, J., J. L. Bessereau, M. Huchet, and J. P. Changeux.** 1990. Two adjacent MyoD1-binding sites regulate expression of the acetylcholine receptor alpha-subunit gene. Nature (London) **345**:353–355.
28. **Sartorelli, V., K. A. Webster, and L. Kedes.** 1990. Muscle-specific expression of the cardiac alpha-actin gene requires myoD1, CArG-box binding factor, and Sp1. Genes Dev. **4**:1811–1822.
29. **Sassoon, D., G. Lyons, W. Wright, V. Lin, A. Lassar, H. Weintraub, and M. Buckingham.** 1989. Expression of two myogenic regulatory factors myogenin and MyoD1 during mouse embryogenesis. Nature (London) **341**:303–307.
30. **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.
31. **Thompson, W. R., B. Nadal-Ginard, and V. Mahdavi.** 1991. A MyoD1-independent muscle-specific enhancer controls the expression of the β-myosin heavy chain gene in skeletal and cardiac muscle cells. J. Biol. Chem. **266**:22678–22688.
32. **Underwood, L. E., and R. S. Williams.** 1987. Pretranslational regulation of myoglobin gene expression. Am. J. Physiol. **252**: C450–C453.
33. **Walsh, K.** 1989. Cross-binding of factors to functionally different promoter elements in c-*fos* and skeletal actin genes. Mol. Cell. Biol. **9**:2191–2201.
34. **Wefald, F. C., B. H. Devlin, and R. S. Williams.** 1990. Functional heterogeneity of mammalian TATA-box sequences revealed by interaction with a cell-specific enhancer. Nature (London) **344**:260–262.
35. **Weintraub, H., R. Davis, S. Tapscott, M. Thayer, M. Kravse, R. Benezra, T. K. Blackwell, D. Turner, R. Rupp, S. Hollenberg, Y. Zhvang, and A. Lassar.** 1991. The myoD gene family: nodal point during specification of the muscle cell lineage. Science **251**:761–766.
36. **Weller, P. A., A. J. Jeffreys, V. Wilson, and A. Blanchetot.** 1984. Organization of the human myoglobin gene. EMBO J. **3**:439–446.
37. **Weller, P. A., M. Price, H. Isenberg, Y. H. Edwards, and A. J. Jeffreys.** 1986. Myoglobin expression: early induction and subsequent modulation of myoglobin and myoglobin mRNA during myogenesis. Mol. Cell. Biol. **6**:4539–4547.
38. **Wentworth, B. M., M. Donoghue, J. C. Engert, E. B. Berglund, and N. Rosenthal.** 1991. Paired myoD-binding sites regulate myosin light chain gene expression. Proc. Natl. Acad. Sci. USA **88**:1242–1246.
39. **Yu, C. Y., K. Motamed, J. Chen, A. D. Bailey, and C. K. J. Shen.** 1991. The CACC box upstream of human embryonic ε globin gene binds Sp1 and is a functional promoter element in vitro and in vivo. J. Biol. Chem. **266**:8907–8915.