MOLECULAR AND CELLULAR BIOLOGY, July 1996, p. 3742–3755 Vol. 16, No. 7
0270-7306/96/$04.00+0
Copyright © 1996, American Society for Microbiology

Flanking Sequences Modulate the Cell Specificity of M-CAT Elements

SARAH B. LARKIN, IAIN K. G. FARRANCE,† AND CHARLES P. ORDAHL*

Department of Anatomy and Cardiovascular Research Institute, University of California San Francisco, San Francisco, California 94143

Received 21 November 1995/Returned for modification 5 January 1996/Accepted 2 April 1996

M-CAT elements mediate both muscle-specific and non-muscle-specific transcription. We used artificial promoters to dissect M-CAT elements derived from the cardiac troponin T promoter, whose regulation is highly striated muscle specific. We show that muscle-specific M-CAT-dependent expression requires two distinct components: the core heptameric M-CAT motif (5′-CATTCCT-3′), which constitutes the canonical binding site for TEF-1-related proteins, and specific sequences immediately flanking the core motif that bind an additional factor(s). These factors are found in higher-order M-CAT DNA-protein complexes with TEF-1 proteins. Non-muscle-specific promoters are produced when the sequences flanking the M-CAT motif are removed or modified to match those of non-muscle-specific promoters such as the simian virus 40 promoter. Moreover, a mutation of the 5′-flanking region of the cardiac troponin T M-CAT-1 element upregulated expression in nonmuscle cells. That mutation also disrupts a potential E box that apparently does not bind myogenic basic helix-loop-helix proteins. We propose a model in which M-CAT motifs are potentially active in many cell types but are modulated through protein binding to specific flanking sequences. In nonmuscle cells, these flanking sequences bind a factor(s) that represses M-CAT-dependent activity. In muscle cells, on the other hand, the factor(s) binding to these flanking sequences contributes to both the cell specificity and the overall transcriptional strength of M-CAT-dependent promoters.

M-CAT elements are transcriptional regulatory motifs with the canonical sequence 5′-CATTCCT-3′ (8, 14, 37) that are required for the cell-specific transcription of the cardiac troponin T (cTNT) gene in both embryonic skeletal and cardiac muscle (21, 36–38). Recently, M-CAT elements have been implicated in the cell-specific regulation of many cardiac promoters (15, 20, 27, 29, 33, 35, 40, 41, 45, 48, 55). Despite the clear role for M-CAT elements in striated muscle-specific transcription, several nonmuscle promoters have also been shown to be governed by M-CAT elements. The simian virus 40 (SV40) enhancer is regulated in part by three binding motifs, dubbed GTIIC (5′-CATTCCA-3′), SphI (5′-CATGCTT-3′), and SphII (5′-CATACTT-3′) (11, 12, 60), that differ from the canonical M-CAT motif at nucleotide positions that were shown to result in functional M-CAT motifs in the context of the cTNT promoter (14). The SV40 enhancer and the closely related human papillomavirus type 16 E6 and E7 enhancer are active in a variety of nonmuscle cell lines such as HeLa cells, P19 and F9 teratocarcinoma cells, and keratinocytes (11, 16, 22, 42). M-CAT-binding sites are active in two-cell mouse embryos and ES cell lines (39). In addition, the M-CAT-dependent somatomammotropin promoter directs cell-specific expression in placental cells (25). Until now, it was not known how M-CAT elements might control muscle-specific expression on the one hand and nonspecific expression on the other.

It is not uncommon for similar binding motifs to mediate different effects in different cell types. For example, E boxes (5′-CANNTG-3′) govern muscle-specific transcription through binding of members of the myoD family of basic helix-loop-helix (bHLH) transcription factors (collectively termed MDFs) that are themselves muscle specific (13, 43, 56). Likewise, in *Drosophila* development, very early neuronal specification is achieved through the action of bHLH proteins of the achaete-scute complex that, at least at this stage of development, are found only in these proneuronal cells (reviewed in reference 23). In B lymphocytes, however, E-box sites activate lymphoid cell-specific expression through the binding of ubiquitous bHLH factors such as the E2A gene products (2, 61). The specificity in this case may be generated by the combined effects of slightly increased E2A gene product concentration and a developmentally programmed decrease in the concentration of the negative-regulator HLH protein, Id (reviewed in references 3 and 26). Lymphoid cell-specific displacement of the ubiquitous repressor protein ZEB may also play a part (17).

Similarly, octamer sites can direct either cell-specific or ubiquitous expression. The ubiquitous POU-domain protein Oct-1 activates ubiquitous small nuclear RNA and histone H2B gene transcription via octamer sites. Octamer-dependent lymphoid cell-specific immunoglobulin gene activation is likely to involve both Oct-1 and the closely related but lymphoid cell-specific Oct-2 protein, probably in collaboration with a lymphoid cell-specific coactivator, OCA-B (10, 34). However, the transcriptional activating properties of Oct-1 and Oct-2 are distinct, and Oct-2 is absolutely required for later stages of B-cell differentiation (9, 10, 30, 53, 54). Thus, cell-specific gene regulation in muscle, neuronal progenitors, and B lymphocytes can be mediated by common DNA-binding motifs and is achieved by the binding of different factors or combinations of factors and/or cofactors in the different tissues.

M-CAT elements bind TEF-1 or TEF-1-related proteins both in muscle extracts and in nonmuscle extracts (14, 20, 22, 27, 35, 40, 48, 49, 58), with the exception of the placenta-specific somatomammotropin M-CAT site (24). Recent experiments show that there are multiple TEF-1-related genes and that mRNAs from two of these TEF-1-related genes show distinct tissue distributions, one enriched in skeletal muscle and the other enriched in cardiac muscle (1a, 51). Moreover, we have demonstrated that a muscle-specific complex is

* Corresponding author. Mailing address: Department of Anatomy, S1334, University of California San Francisco, San Francisco, CA 94143-0452. Phone: (415) 476-4051. Fax: (415) 476-4845.

† Present address: Cardiovascular Research Institute and Division of Cardiology, VA Medical Center, San Francisco, CA 94121.

formed on M-CAT elements. This complex contains a TEF-1 protein that is muscle enriched but is also found in nonmuscle tissues (14a). Thus, it is probable that TEF-1 proteins will provide a measure of tissue specificity by virtue of their muscle enrichment and participation by as yet undefined interactions in muscle-specific complexes on the core M-CAT motif.

Here, we report experiments that address the mechanism(s) controlling the tissue-specific activity of M-CAT elements. M-CAT elements were studied independently, in artificial promoters, to test their capacity to mediate tissue specificity in the absence of other promoter elements. We show that the canonical 7-nucleotide M-CAT core motif that constitutes the TEF-1-binding site is insufficient alone to mediate muscle specificity. However, the sequence flanking the cTNT promoter M-CAT core motifs strongly contributes to tissue specificity by repressing M-CAT-dependent promoter activity in nonmuscle tissues, coincident with the appearance of a second binding activity. The sequence-specific binding of nuclear factors to these flanking sequences implicates additional factors that, in collaboration with TEF-1 and TEF-1-related proteins, form a muscle-specific transcription complex.

Note on nomenclature. For the purpose of this article, the heptameric M-CAT sequence itself (CATTCCT) will be referred to as the core motif, while "M-CAT element" refers to this core motif surrounded by native flanking sequence as defined by footprint (38), or variants of this sequence. Furthermore, the M-CAT-binding factors in muscle nuclear extracts are composed of the products of several different TEF-1-related genes (1a, 14a). In the interest of brevity, TEF-1 and TEF-1-related proteins, recognized by a polyclonal antibody raised against a common epitope, are hereafter collectively referred to as TEF-1 proteins.

MATERIALS AND METHODS

Tissue culture and CAT assay. All tissues were harvested from embryonic day 12 chicks. Fibroblasts were isolated from skin that was peeled off the breast, flanks, and leg. Breast muscle and livers were dissected free of the connective tissue and gallbladder, respectively. After being minced, the tissues were washed in serum-free medium and disrupted in 0.25% trypsin with periodic pipetting. When discrete pieces of tissue were no longer visible (3 to 10 min, depending on the tissue), the cells were pelleted, resuspended in medium, filtered through 20-μm-pore-size nylon mesh, and counted. Muscle cells were plated at 1.25×10^6 cells per 60-mm-diameter collagen-coated plate. Liver cells (consisting of a mixed population of fibroblasts and hepatocytes) were plated at 1.5×10^6 cells per 60-mm collagen-coated plate. Skin-derived fibroblast cells were plated at 0.75×10^6 cells per uncoated plate. The cells were allowed to attach for 24 h and then transfected by standard calcium phosphate methods (19, 47). The cells were exposed to the transfection cocktail for 16 h before the medium was changed. At 48 h later, cells were harvested and lysed by three freeze-thaw cycles. Chloramphenicol acetyltransferase (CAT) activity was assayed by standard techniques (47).

Plasmid construction. All oligonucleotides were obtained from Operon Technologies, Alameda, Calif., and designed with 5′ cohesive termini to enable them to multimerize only head to tail (see Table 1). Oligonucleotides were treated with kinase, mixed with various proportions of non-kinase-treated oligonucleotide (to limit the multimerization reaction), and ligated. After ligation overnight, overhanging ends were filled with Klenow polymerase. Multimers in the appropriate size range were isolated from a 6% nondenaturing polyacrylamide gel, purified, and ligated into −49cTNT.CAT, into the *Cla*I site in the polylinker just upstream of the cTNT TATA (38) (see Fig. 1A, 2A, etc., for diagrammatic representations of the resultant multimers). All constructs were sequenced to verify the multimer copy number, sequence and orientation. All multimers presented here contain either five or six element copies. Variation of the copy number from 4 to 6 did not affect relative efficiency (31a). However, it was noted that multimers in the sense orientation relative to the TATA box produced less efficient promoters in muscle cells than did those multimerized in the antisense orientation, although orientation did not affect the activity of derived promoters in the nonmuscle tissues tested (31a). M-CAT-1/Emtcat is the only promoter presented in this study in which multimers are in the sense orientation, since they were never derived in the antisense orientation despite multiple attempts.

Nuclear extract preparation, gel retardation assay, and Western analysis of retarded complexes. Nuclear extracts were prepared from skeletal muscle (breast and leg), heart, lung, and brain tissue from embryonic day 12 chickens as described previously (38). High-resolution gel retardation assays were performed as described previously (14a) with 0.3 ng of radiolabelled probe, 4 to 6 μg of nuclear protein extract, 1 μg of poly(dI-dC), and a 250-fold excess of competitor DNA, where appropriate (unless otherwise noted in the figure legends). Anti-mouse myoD antibodies were purchased from Santa Cruz Biotechnology, Santa Cruz, Calif., and gel retardation supershift analysis was performed by preincubating the binding-reaction mixtures for 15 min at room temperature with 6 μg of affinity-purified antiserum prior to addition of radiolabelled probe. After addition of the probe, the binding reaction was continued as usual with a 20-min room temperature incubation. Western blot (immunoblot) analysis of DNA-protein complexes was performed as described previously (14a).

Methylation interference assay. Methylation interference was performed with dimethyl sulfate (Fluka) essentially as described previously (1) with the following modification. Following a fivefold scale-up of the gel retardation incubation mixture, the entire gel was transferred to a DEAE membrane with a semidry blotter for 1 to 2 hours in 2.5× Tris-borate-EDTA (TBE) and the membrane was exposed to film for 3 to 12 h. Following visualization and excision of the required bands, DNA was eluted by incubating the DEAE strips at 52°C for 1 h in high-salt elution buffer (1 M NaCl, 50 mM Tris · HCl [pH 8], 10 mM EDTA), extracting the DNA with phenol-chloroform, and precipitating it with ethanol. The DNA was cleaved in 1 M piperidine (Fluka) and purified by multiple rounds of lyophilization as described previously (1). Cleaved ladders were resolved on a 16% denaturing polyacrylamide gel.

RESULTS

Multimerized M-CAT elements direct striated-muscle-specific reporter gene expression. Striated-muscle-specific expression of the cTNT promoter requires two M-CAT elements that are found within 99 bases upstream of the transcription initiation site (36, 37). The distal M-CAT element (M-CAT-1) and the proximal M-CAT element (M-CAT-2) are separated in the native promoter by a center-to-center spacing of 23 nucleotides, and perturbation of this spacing reduced promoter activity (37). Mutation of either M-CAT element abolishes promoter activity, indicating that a single M-CAT is insufficient to drive expression in the context of the native promoter. We created artificial promoters containing multiple copies of either M-CAT-1 or M-CAT-2 to study the contribution of each site, in isolation, to tissue specificity and efficiency of expression (Fig. 1). In these artificial promoters, the flanking sequences surrounding each M-CAT element included those previously shown to be protected during DNase I footprinting (38). The constructs also preserved the native 23-bp center-to-center spacing between M-CAT elements (38). The tissue specificity of these artificial promoters was assessed by transfection into primary cultures of skeletal muscle, skin fibroblasts, and hepatocytes.

In this multimeric form, the M-CAT-1 element directs very high-level striated-muscle-specific expression (Fig. 1B). M-CAT-1-dependent reporter activity is 183-fold higher in muscle than liver cells and 284-fold higher in muscle than in skin cells. Multimerization of the M-CAT-2 element results in a much weaker promoter (note the scale in Fig. 1C) but one that retains the ability to direct muscle-specific reporter activity (9.6-fold higher than in liver, undetectable in skin). The activity of M-CAT-1 or M-CAT-2 multimers is eliminated by mutation (Table 1) of the core M-CAT motif (data not shown). Thus, in multimeric form, both promoter elements carry inherent muscle specificity but show substantial quantitative differences in the levels of promoter activity they support.

Tissue specificity is dependent on the configuration of the multimerized M-CATs. Earlier studies showed that the multimerized SV40 GTIIC element (CATTCCA) supported expression in a range of cultured cell lines of nonmuscle origin (16). This multimerization differed from the M-CAT multimers described above not only in one base change at position 7 in the core motif but also in the configuration of the multimers. The GTIIC elements were multimerized in tandem pairs separated by 2 bp (center-to-center spacing, 9 bp) and

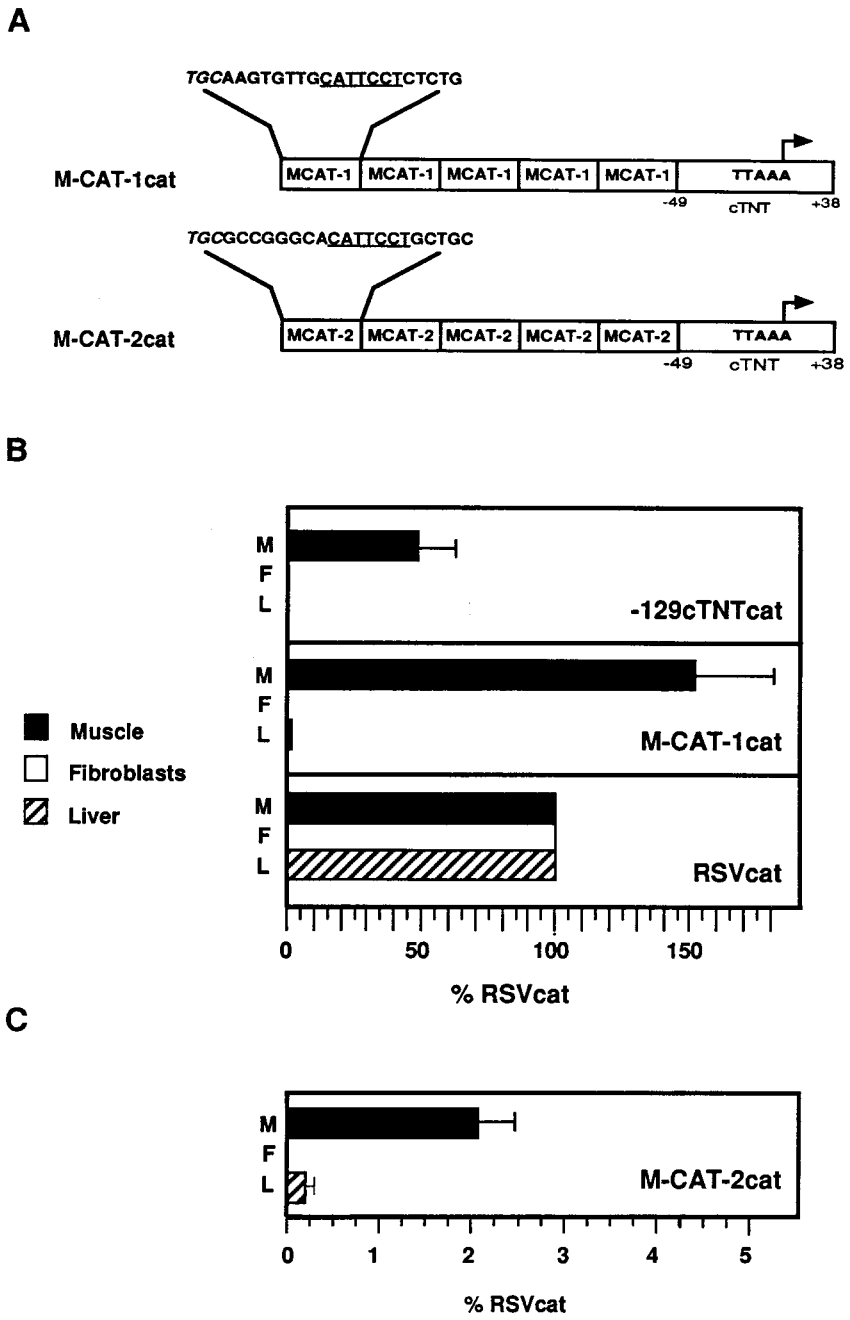


FIG. 1. cTNT promoter M-CAT elements, when multimerized, can confer muscle specificity upon a basal promoter. (A) The sequences of the 23-bp oligonucleotides covering the consensus M-CAT elements, M-CAT-1 and M-CAT-2, and their flanking sequences derived from the cTNT promoter are shown. The oligonucleotides were multimerized and inserted, in the antisense orientation, directly upstream of the cTNT TATA box linked to the CAT reporter gene, to yield M-CAT-1cat and M-CAT-2cat, respectively. M-CAT elements are shown here in the sense orientation for legibility. (B) M-CAT-1cat directs muscle-specific reporter gene expression in primary cultures of skeletal muscle (M), embryonic skin fibroblasts (F), and liver cells (L). Reporter activity was normalized to RSVcat activity in each cell type. −129cTNTcat is shown for comparison in the top panel. Error bars correspond to the standard error of the mean of six or more independent transfections, in all figures. (C) M-CAT-2cat expression is very much weaker than M-CAT-1cat expression (note the scales in panels B and C) but is also muscle specific.

with each pair then separated by 29 bp (16). This tandem arrangement resembles the spacing of the diverged M-CAT variants SphI and SphII found in the SV40 enhancer.

We created multimers with the core muscle M-CAT motif (CATTCCT) but configured as above, with a 9-bp center-to-center spacing (see Table 2), and inserted them into artificial promoters as described previously (M-CAT×2cat [Fig. 2A]). This artificial promoter directs moderate to low-level expression in all cell types tested (Fig. 2B, top panel). The number of copies, from six to eight, did not significantly affect the gross activity or tissue specificity of the resultant chimeric promoter (data not shown). Next, we investigated the contribution of correct flanking sequence to the muscle specificity of multimerized M-CATs. We multimerized a consensus M-CAT motif (CATTCCT) to give 23-bp center-to-center spacing but with the flanking sequence derived from the sequence surrounding

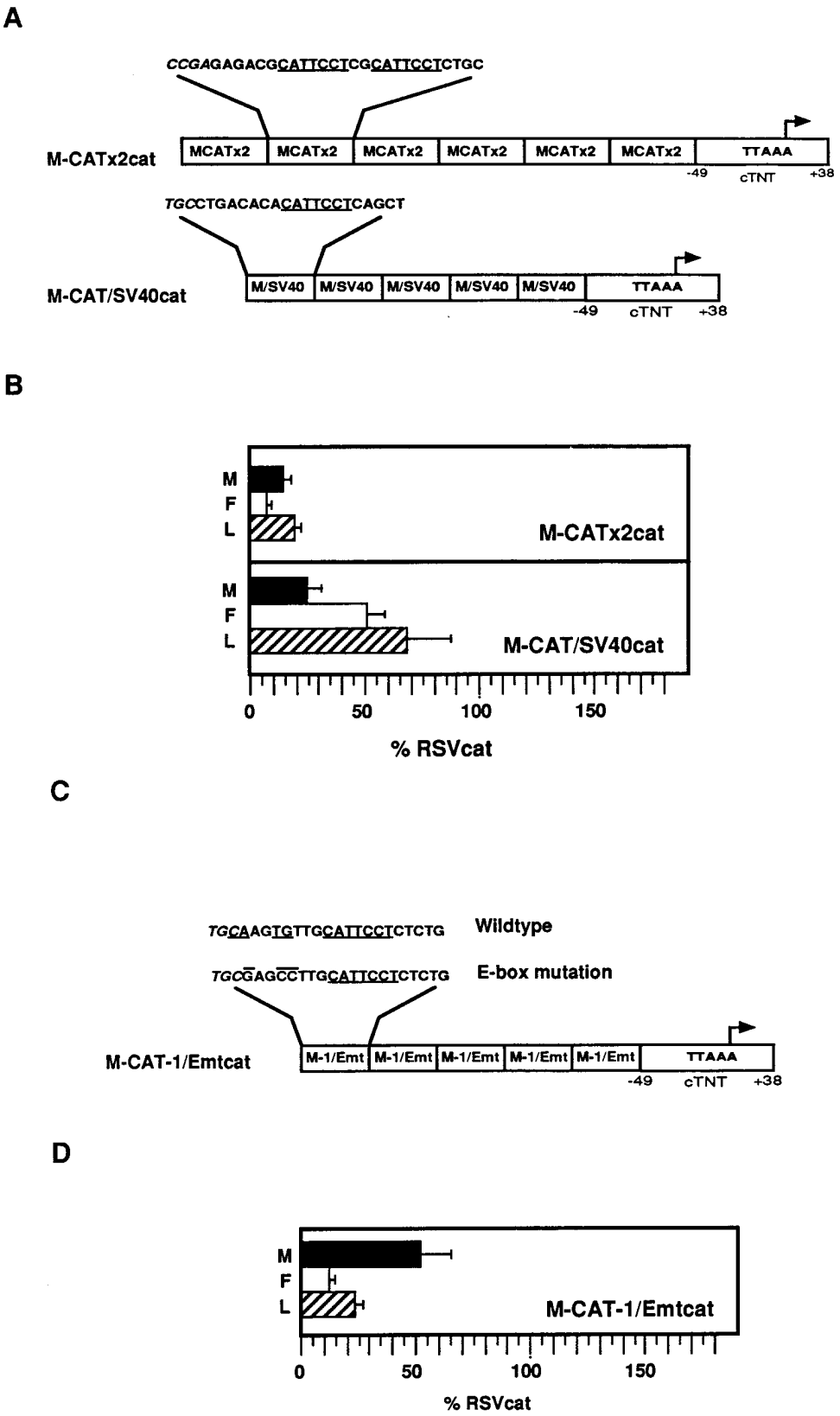


FIG. 2. Muscle specificity requires the correct flanking DNA with respect to both spacing and sequence. (A) Sequences are shown of M-CAT variant elements in which the native flanking sequence of the cTNT M-CAT elements is either removed, generating the tandem element M-CAT×2, or substituted by SV40 enhancer-derived sequence, to yield the M-CAT/SV40 element. These elements were multimerized and fused to the cTNT TATA in the antisense orientation to construct the M-CAT×2cat and M-CAT/SV40cat reporters, respectively. M-CAT elements are shown here in the sense orientation for legibility. (B) M-CAT×2cat directs comparable expression levels in skeletal muscle (M), embryonic skin fibroblasts (F), and liver (L). Similarly, MCAT/SV40cat directs reporter expression in all three cell types. Reporter activity is normalized to RSVcat activity in each cell type. (C) The E-box consensus (CANNTG) located in the 5′ flank of the M-CAT-1 element was mutated (CAAGTG to CGAGCC) to generate the element MCAT-1/Emt, and this element was multimerized and fused to the cTNT TATA in the sense orientation to construct M-CAT-1/Emtcat. (D) M-CAT-1/Emtcat is capable of directing expression in both skin fibroblasts (F) and liver (L). Higher-level expression of this construct was maintained in muscle (M): following normalization to RSVcat, CAT activity in muscle cells was twice that in liver cells and four times that in skin fibroblasts.

the GTIIC element in the SV40 early enhancer. As before, this multimer was cloned upstream of the cTNT TATA (M-CAT/SV40cat [Fig. 2A]). Figure 2B (bottom panel) shows that this chimeric promoter directs efficient reporter gene expression in all three cell types. Thus, muscle-specific expression is dependent upon the presence and sequence of the DNA flanking the core M-CAT motifs.

The 5′-flanking region of the M-CAT-1 element mediates tissue-specific repression. The 5′-flanking region of the M-CAT-1 element contains a potential E box (38), which is predicted to be a low-affinity MDF-binding site by comparison with the consensus derived from authentic MDF-dependent promoter comparison and various PCR techniques (5, 57, 59). Deletion of this E box decreased promoter activity modestly but did not affect tissue specificity (38). To examine the role of this E box in M-CAT-1 multimers, an M-CAT-1 element with an E-box mutation was used to construct an artificial promoter as described above (M-CAT-1/Emtcat [Fig. 2C]). Mutation of three bases within the E box in the M-CAT-1 oligomer greatly increased expression in skin fibroblasts and liver (Fig. 2D). Thus, mutation of the E box or some overlapping site appears to relieve a repression mechanism that blocks expression in nonmuscle cells. Since MyoD family members are not found in these nonmuscle tissues, the factor(s) responsible for this repression must be either a non-MDF E-box-binding factor or a factor which recognizes an overlapping site.

Multimerization of M-CAT-1/Emt produced a promoter substantially weaker in muscle (Fig. 2D) than its wild-type equivalent (M-CAT-1cat; Fig. 1B, middle panel). However, we believe that the relatively low activity of the M-CAT-1/Emt artificial promoter in muscle reflects at least in part the sense orientation of the multimers in this construct (see Materials and Methods). Thus, the mutation of three bases in the region of the E box relieves repression in nonmuscle tissues and may affect activity in muscle cells.

Gel retardation analysis of the M-CAT elements. Gel retardation analysis was performed to examine the binding of proteins to the flanking regions of each M-CAT element. The complexes resulting from the incubation of the 23-bp M-CAT-1 element with embryonic skeletal muscle extract are shown in Fig. 3A. The three high-mobility complexes labeled C1, C2, and C3 are described in detail elsewhere (14a) and are generated by the interactions of three distinct TEF-1 polypeptides with the core M-CAT motif (Fig. 3A, lane 1). These complexes are competed for by a 250-fold excess of any unlabeled oligonucleotide containing an intact M-CAT core motif, although the affinities vary (Fig. 3A, lanes 2, 4, 6, 7, 9, and 10). M-CAT mutant DNA and MCK E-box DNA fail to compete for these major complexes (Fig. 3A, lanes 3, 5, and 8) indicating sequence-specific DNA-protein binding.

In addition to these major DNA-protein complexes, two lower-mobility complexes (LMC 1A and LMC 1B) are generated with the M-CAT-1 probe (Fig. 3A, lane 1). These complexes are self-competed for (Fig. 3A, lane 2). LMC 1B (and to a lesser extent, LMC 1A) is also competed for by the M-CAT-1m element, in which the core M-CAT motif is mutated, leaving only the flanking sequence intact (lane 3). However, neither LMC 1A nor LMC 1B can be competed for by M-CAT-2 elements whether the core motif is intact or mutated (lanes 4 and 5). Similarly, the oligonucleotide containing the core M-CAT motif surrounded by SV40 enhancer flanking sequence is unable to compete for either LMC 1A and LMC 1B (lane 6). A chimeric M-CAT element, containing the 5′-flanking sequence of the M-CAT-1 element and the 3′-flanking sequence of the M-CAT-2 element, competes for both LMC 1A and LMC 1B (M-CAT-1/2; lane 9), but the reverse chimera cannot (M-CAT-2/1; lane 10). We conclude that it is the 5′-flanking sequence of M-CAT-1 that is most important in the formation of LMC 1A and LMC 1B.

Gel retardation analysis of the M-CAT-2 probe in the presence of embryonic skeletal muscle nuclear extract reveals a very similar situation (Fig. 4). There are, as before, three major retarded complexes of relatively high mobility that are generated by TEF-1 polypeptides bound to the core M-CAT motif (C1, C2, and C3 in Fig. 4). These appear to be identical in mobility to those formed on the M-CAT-1 probe, and, indeed, these probes cross-compete for C1, C2, and C3 (Fig. 3A, lane 4; Fig. 4: lane 2). As judged by such competition experiments, the M-CAT-1 element constitutes a higher-affinity binding site than does the M-CAT-2 element. These major complexes are also competed for by a 250-fold excess of any competitor DNA containing an intact M-CAT core motif (Fig. 4, lanes 2, 4, 6, 7, 9, and 10). Competitor DNA containing a mutation within the core M-CAT motif fails to compete for any of C1, C2, or C3 (lanes 3 and 5), as does an unrelated DNA sequence (the MCK E-box [lane 8]).

M-CAT-2 interaction with embryonic skeletal muscle nuclear extract also formed sequence-specific low-mobility complexes (Fig. 4, LMC 2A and LMC 2B). LMC 2A is a doublet whose component bands apparently have identical sequence specificity and tissue distribution. LMC 2A can be competed for only by wild-type M-CAT-2 oligonucleotide with its intact flanking sequences (Fig. 4, lane 4). However, because LMC 2A is formed only when the non-cTNT-derived cohesive termini (Table 1) of the M-CAT-2 23-bp oligonucleotide are filled with Klenow polymerase (31a), the significance of LMC 2A is uncertain and we do not consider it further here.

LMC 2B, by contrast, appears to be an authentic equivalent of the low-mobility complexes formed with the M-CAT-1 element. LMC 2B is competed for by competitor DNAs that include the M-CAT-2-flanking sequence, irrespective of whether the M-CAT motif itself is intact or mutated (Fig. 4, lanes 4 and 5). LMC 2B is not competed for by M-CAT-1 elements, whether their core motif is intact or mutated (Fig. 4, lanes 2 and 3), indicating that different proteins probably interact with the flanking regions of the M-CAT-1 and M-CAT-2 elements. As is the case for LMC 1A and LMC 1B, an oligonucleotide containing the core M-CAT motif surrounded by SV40 enhancer flanking sequence is unable to compete for LMC 2B (lane 6). A chimeric M-CAT element containing the M-CAT-2 5′-flanking region and the M-CAT-1 3′-flanking region can compete (M-CAT-2/1 [lane 10]), but the reverse chimera cannot (M-CAT-1/2 [lane 9]). Thus, LMC 2B depends upon the 5′-flanking sequence of M-CAT-2 for its formation.

The M-CAT variants M-CAT/SV40 and M-CAT-1/Emt were radiolabelled and subjected to gel retardation analysis. Neither formed detectable low-mobility complexes (data not shown). This result, together with the competition experiments (above), indicates that the flanking regions surrounding M-CAT core motifs are important for the formation of low-mobility protein–DNA complexes. We tentatively conclude that the flanking sequences of each M-CAT motif of the cTNT promoter interact with a distinct protein(s) that is likely to be involved in the muscle-specific activity of these elements.

Contribution of the E box to the formation of LMC 1A and 1B. Since a mutation within the nominal E box had a profound effect on the tissue specificity of the M-CAT-1 element, we tried to compete for LMC 1A and LMC 1B with excess oligonucleotide containing the E-box mutation and found that it was unable to compete well for either complex (Fig. 3A, lane 7). We next investigated the competition ability of the well-characterized, muscle-specific MCK E box (defined as a high-

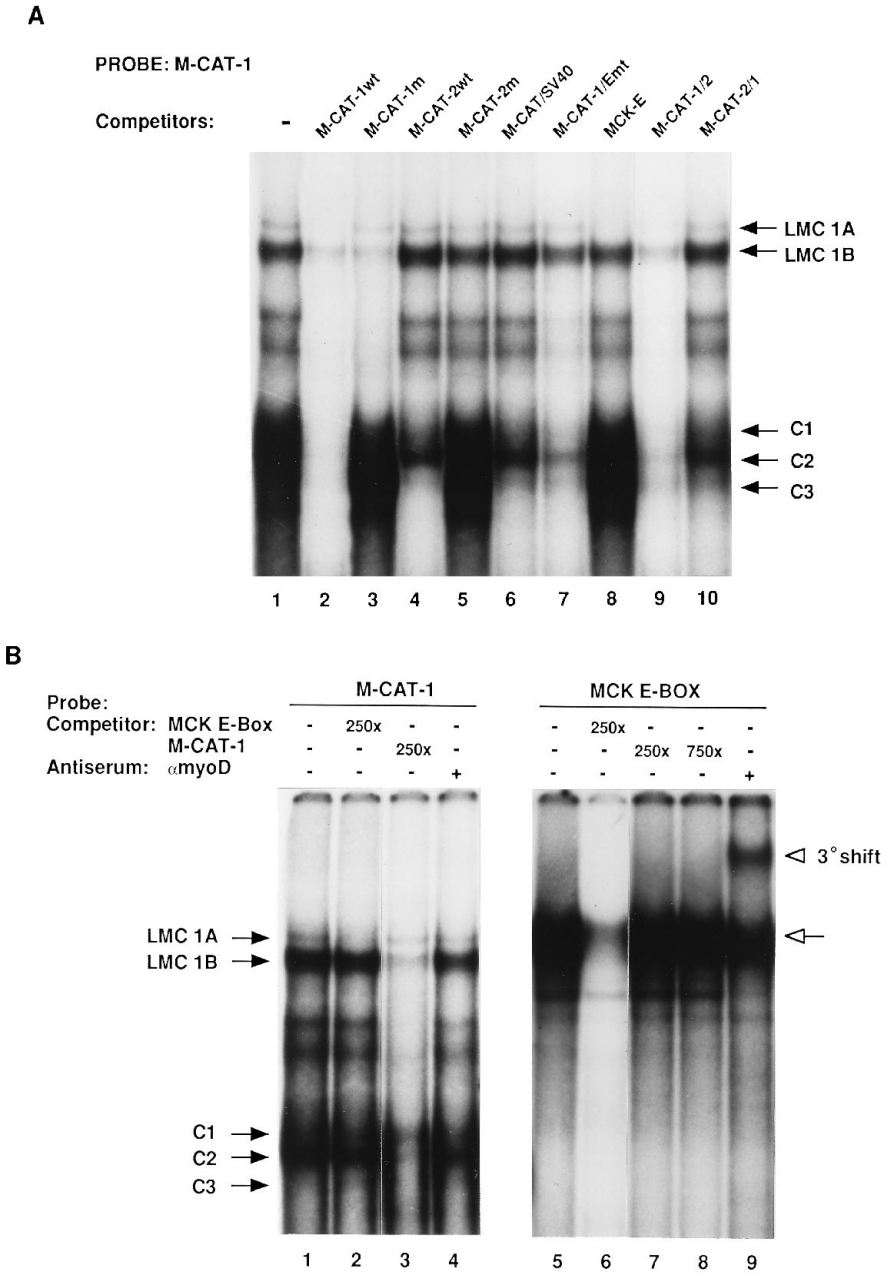


FIG. 3. Protein complexes bind in a sequence-specific manner to the 5′-flanking sequence of the M-CAT-1 element. (A) The 23-bp M-CAT-1 element was radiolabelled and subjected to gel retardation analysis in the presence of skeletal muscle nuclear extract. Protein-DNA complexes were challenged with a 250-fold excess of unlabelled competitor oligonucleotides as labelled above the lanes. The TEF-1/M-CAT complexes, C1, C2, and C3, are labelled on the right, as are LMC 1A and LMC 1B. Intermediate-mobility, lower-intensity complexes that appear to show similar sequence specificity to LMC 1B upon competition were also formed. Their significance is not clear. (B) Comparative gel retardation analysis of the M-CAT-1 element and the MCK E box in the presence of skeletal muscle nuclear extract. Gel retardation analysis of M-CAT-1 is shown in lanes 1 to 4. Analysis of the MCK E-box is shown in lanes 5 to 9. Labelled probes were adjusted to equal specific activity, and lanes shown are taken from a single exposure of the same experiment to demonstrate comparative avidities of complex formation. Protein-DNA complexes were challenged with unlabelled competitor oligonucleotides: lanes 1, 4, 5, and 9, none; lanes 2 and 6, 250-fold molar excess of MCK E-box; lanes 3 and 7, 250-fold molar excess of M-CAT-1 element; lane 8, 750-fold molar excess of M-CAT-1 element. Chicken skeletal muscle nuclear extracts in lanes 4 and 9 were preincubated with 6 μg of anti-mouse myoD antiserum. TEF-1/M-CAT complexes, C1, C2, and C3, are labelled to the left of the figure, as are LMC 1A and LMC 1B. The presumptive E-box–MDF complex is labelled with an open arrow to the right of the figure, and the tertiary chicken myoD/αmyoD antibody supershift is labelled with an open arrowhead to the right of the figure.

affinity myoD:E12-binding site [5, 6, 59]). LMC 1A was competed for, whereas LMC 1B was unaffected (Fig. 3A, lane 8, and Fig. 3B, lane 2). Thus, LMC 1A and LMC 1B result from the binding of proteins with distinct sequence specificities.

Since LMC 1A was competed for by a canonical MDF E box, we investigated whether it might contain an MDF component. DNA-protein complexes formed between the MCK E box and muscle nuclear extracts have similar mobility to LMC 1A (Fig. 3B, compare lanes 1 and 6). If these complexes are formed by interaction with the same protein(s), the affinity of the M-CAT-1 E box is clearly lower than that of the MCK E box, as judged by the relative intensity of the relevant bands (Fig. 3B,

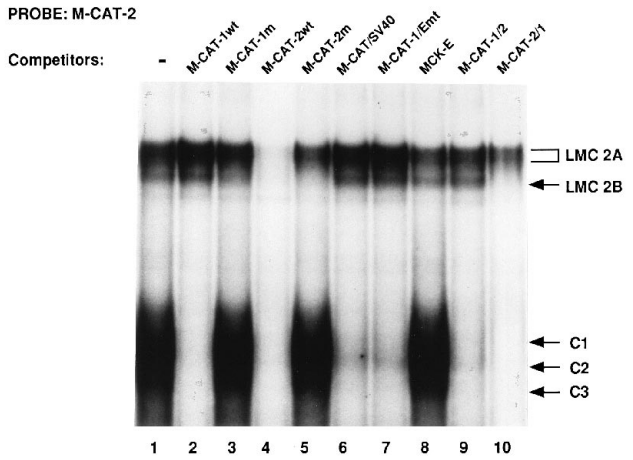


FIG. 4. Protein complexes bind in a sequence-specific manner to the 5′-flanking sequence of the M-CAT-2 element. The 23-bp M-CAT-2 element (sequence shown in Fig. 1A and Table 1) was radiolabelled and subjected to gel retardation analysis in the presence of skeletal muscle nuclear extract. Protein-DNA complexes were challenged with a 250-fold excess of unlabelled competitor oligonucleotides as indicated above the lanes. The high-mobility TEF-1/M-CAT complexes, C1, C2 and C3, are labelled to the right of the figure, as are LMC 2A and LMC 2B.

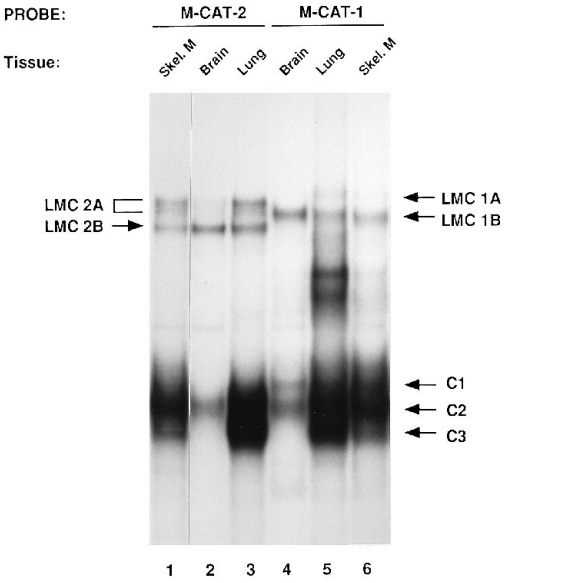


FIG. 5. Tissue distribution of the nuclear factors binding the M-CAT-1 and M-CAT-2 elements. Gel retardation analysis of DNA-protein complex formation on the M-CAT-2 probe (lanes 1 to 3) and on the M-CAT-1 probe (lanes 4 to 6) in the presence of nuclear extracts from skeletal muscle (Skel. M.), brain, and lung tissue. The high-mobility TEF-1/M-CAT complexes, C1, C2, and C3, are labelled to the right of the figure, as are LMC 1A and LMC 1B, formed on the M-CAT-1 probe. LMC 2A and LMC 2B, formed on the M-CAT-2 probe, are labelled to the left of the figure.

compare lanes 1 and 5). However, whereas the MCK E-box–chicken muscle extract complexes were supershifted in part with anti-murine myoD antibody (lane 9), LMC 1A was unaffected (lane 4), suggesting that chicken myoD may not participate in LMC 1A formation. Moreover, LMC 1A is formed in several nonmuscle tissues (Fig. 5, lane 5, and data not shown). Thus, while LMC 1A shares the DNA sequence-binding specificity of bHLH proteins, it is unlikely to belong to the muscle-specific myoD family of bHLH proteins.

LMC 1A is a low-intensity complex compared with LMC 1B, and hence its functional contribution may be similarly small. Moreover, LMC 1A is not found in several nonmuscle tissues tested (Fig. 5 and data not shown), inconsistent with a key role for this activity in repression of M-CAT elements in such tissues. We tentatively conclude that LMC 1A is not the activity responsible for the repression in nonmuscle cells.

LMC 1B, by contrast, is a reasonable candidate for the factor(s) that interacts with the 5′-flanking sequence of the M-CAT-1 element to mediate the repression of its activity in nonmuscle tissues. LMC 1B can be detected in nuclear extracts from all nonmuscle tissues tested (Fig. 5 and data not shown). LMC 1B cannot be competed for by a consensus E box (Fig. 3A, lane 8; Fig. 3B, lane 2), nor is it recognized by the mouse anti-myoD antiserum (Fig. 3B, lane 4); therefore, we conclude that this complex is generated by a non-MDF factor with sequence specificity of binding distinct from that of MDFs as well as of LMC 1A. Interestingly, LMC 1B derived from muscle appears to consistently migrate fractionally faster than LMC 1B derived from nonmuscle tissue (Fig. 5, compare lane 6 with lanes 4 and 5).

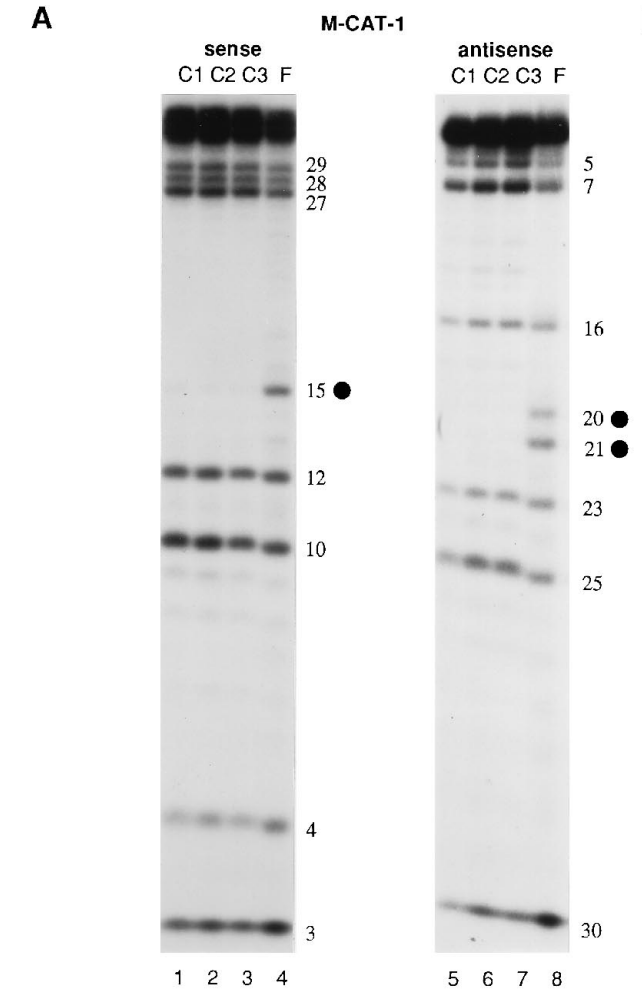
TABLE 1. Oligonucleotides used in the construction of M-CAT multimer-based synthetic promoters and as competitor oligonucleotides

Name	Sense strand sequence[a]	Length (nt)
M-CAT-1	**TGC**AAGTGTTGCATTCCTCTCTG	23
M-CAT-1mt	**TGC**AAGTGTTGCcccaCTCTCTG	23
M-CAT-2	**TGC**GCCGGGCACATTCCTGCTGC	23
M-CAT-2mt	**TGC**GCCGGGCACcccaCTGCTGC	23
M-CAT×2	**CCGA**GAGACGCATTCCTCGCATTCCTCTGC	30
M-CAT/SV40	**TGC**CTGACACACATTCCTCAGCT	23
M-CAT-1Emt	**TGC**gACccTTGCATTCCTCTCTG	23
M-CAT-1/2	**TGC**AAGTGTTGCATTCCTGCTGC	23
M-CAT-2/1	**TGC**GCCGGGCACATTCCTCTCzTG	23
MCK E box	GATCCGGCTCAGGCAGCAGGTGTTGG	26

[a] Cohesive termini are denoted by boldface type, the core M-CAT motif(s) is indicated by underlining, and mutated bases are denoted by lowercase type. The MCK E box (MEF-1 site) sequence is derived from the right-hand site of the muscle creatine kinase (MCK) enhancer (6).

Methylation interference footprints of low-mobility complexes. We used methylation interference footprinting to compare the DNA-protein interactions in low-mobility complexes with those in the high-mobility complexes, C1, C2 and C3, that contain TEF-1 proteins bound to the core M-CAT motif (14a). Methylation of G^{20} and G^{21} (antisense strand), at the center of the core M-CAT motif, abolished formation of C1, C2, and C3 on both M-CAT-1 and M-CAT-2 elements (Fig. 6A, compare lanes 5 to 7 with free M-CAT-1 probe in lane 8; Fig. 6B, compare lanes 13 to 15 with free M-CAT-2 probe in lane 16). In the case of M-CAT-1, methylation of residue G^{15} (sense strand) immediately 5′ of the M-CAT motif is also incompatible with formation of complexes C1, C2, and C3 (Fig. 6A, compare lanes 1 to 3 with free probe in lane 4). Thus, interaction of a TEF-1 protein(s) with the M-CAT element is sensitive to G methylation over a narrow area centered on the core of the M-CAT motif.

Analysis of the low-mobility complexes indicates a broader, more complex footprint (Fig. 7). LMC 1A formation is abolished by methylation at residue G^{10} and reduced by methylation at residues G^4 and G^{12} on the sense strand (Fig. 7A, compare lane 3 with free probe in lanes 1 and 4). LMC 1A is abolished by methylation at residues G^5 and G^7 on the antisense strand and slightly reduced by methylation at residues



C

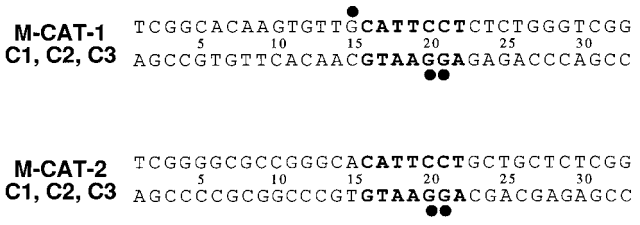
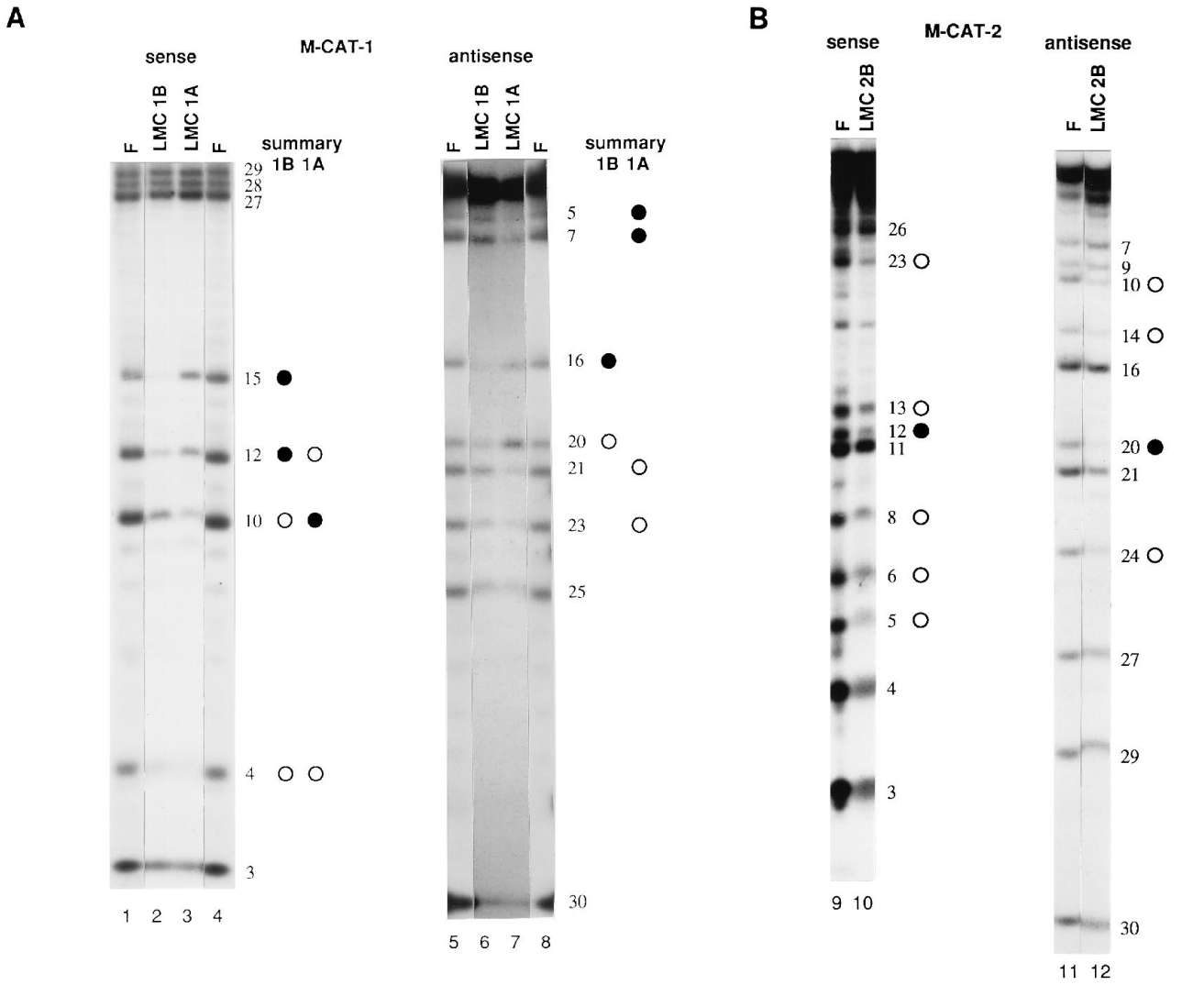


FIG. 6. Methylation interference analysis of the TEF-1/M-CAT complexes, C1, C2, and C3. (A) A 33-mer M-CAT-1 probe was radiolabelled on the sense or antisense strand, and muscle nuclear protein-DNA interactions within the TEF-1/M-CAT complexes, C1, C2 and C3, were examined by methylation interference analysis. Cleavage patterns of C1-, C2-, and C3-bound probe are shown in lanes 1 to 3 (sense strand) and lanes 5 to 7 (antisense strand), respectively. The cleavage patterns for unbound probe are shown in lanes 4 (sense strand) and 8 (antisense strand). A summary of the methylated G residues which abolish complex formation is shown to the right of each panel (solid circles). Residue numbering is from the 5′ end of the duplex 33-mer probes, as shown in panel C. (B) A 33-mer M-CAT-2 probe was radiolabelled on the sense or antisense strand, and muscle nuclear protein-DNA interactions within the TEF-1/M-CAT complexes, C1, C2, and C3, were examined by methylation interference analysis. Cleavage patterns of C1-, C2- and C3-bound probe are shown in lanes 9 to 11 (sense strand) and lanes 13 to 15 (antisense strand), respectively. The cleavage pattern for unbound probe is shown in lanes 12 (sense strand) and 16 (antisense strand). A summary of the methylated G residues which abolish complex formation is shown to the right of each panel (solid circles). (C) Summary of the methylation interference analysis of TEF-1/M-CAT complexes, C1, C2, and C3, formed on the M-CAT-1 and M-CAT-2 elements of the cTNT promoter. G residues whose methylation abolished complex formation are denoted by solid circles. Numbering is from the 5′ end of the double-stranded oligonucleotides. The M-CAT-1 oligonucleotide spans bp −109 to −83 of the cTNT promoter, and the M-CAT-2 oligonucleotide spans bp −85 to −59.

G^{21} and G^{23} (Fig. 7A, compare lane 7 with free probe in lanes 5 and 8). Thus, the LMC 1A methylation interference pattern is consistent with a protein(s) binding to the E-box region (Fig. 7C).

The LMC 1B pattern of sensitivity to G methylation is distinct from that seen with either C1, C2, and C3 or LMC 1A. LMC 1B formation is abolished by methylation of G^{15} and G^{12} and reduced by methylation at G^{10} or G^{4} on the sense strand (Fig. 7A, compare lane 2 with free probe in lanes 1 and 4). Methylation of G^{16} on the antisense strand abolished LMC 1B formation, and methylation of G^{20} reduced its formation (Fig. 7A, compare lane 6 with free probe in lanes 5 and 8). Thus, the LMC 1B footprint encompasses the 5′ end of the M-CAT motif as well as the flanking sequence immediately upstream, including part of the E-box consensus (Fig. 7C).

Formation of LMC 2B is incompatible with methylation of G^{12} and is reduced by methylation at G^{5}, G^{6}, G^{8}, G^{13}, or G^{23} on the sense strand (Fig. 7B, compare lane 10 with free probe in lane 9). LMC 2B formation is abolished by methylation of G^{20} and reduced by methylation of G^{24}, G^{14}, or G^{10} on the antisense strand (Fig. 7B, compare lane 12 with free probe in lane 11). Thus, the LMC 2B methylation interference pattern



C

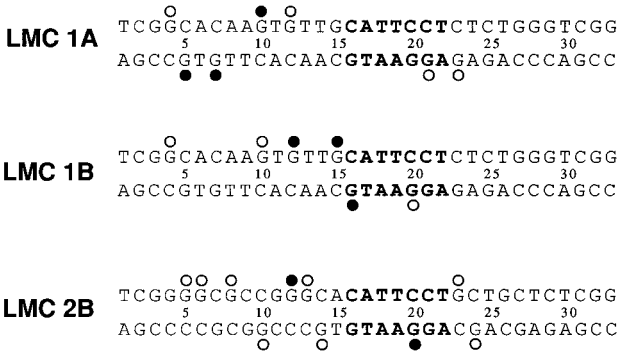


FIG. 7. Methylation interference analysis of LMC 1A, LMC 1B, and LMC 2B. (A) A 33-mer M-CAT-1 probe was radiolabelled on the sense or antisense strand, and muscle nuclear protein-DNA interactions within LMC 1A and LMC 1B were examined by methylation interference analysis. Cleavage patterns of LMC 1B-bound probe are shown in lanes 2 (sense strand) and 6 (antisense strand); cleavage patterns of LMC 1A-bound probe are shown in lanes 3 (sense strand) and 7 (antisense strand). The cleavage pattern for unbound probe is shown in lanes 1 and 4 (sense strand) and lanes 5 and 8 (antisense strand). A summary of the methylated G residues which interfere completely (solid circles) and partially (open circles) with LMC 1B formation is shown to the right of each panel, and a similar summary for LMC 1A is shown to the far right of each panel. The complete and partial scores result from comparison of three independent experiments for LMC 1B and two experiments for LMC 1A. (B) A 33-mer M-CAT-2 probe was radiolabelled on the sense or antisense strand, and muscle nuclear protein-DNA interactions within LMC 2B were examined by methylation interference analysis. Cleavage patterns of LMC 2B-bound probe are shown in lanes 10 (sense strand) and 12 (antisense strand). The cleavage pattern for unbound probe is shown in lanes 9 (sense strand) and 11 (antisense strand). A summary of the methylated G residues which interfere completely (solid circles) and partially (open circles) with LMC 2B formation is shown to the right of each panel. Complete and partial scores result from comparison of three independent experiments. (C) Summary of the methylation interference analysis of LMC 1A, LMC 1B, and LMC 2B formed on the M-CAT-1 and M-CAT-2 elements of the cTNT promoter. G residues whose methylation abolished or partially interfered with complex formation are denoted by solid and open circles, respectively. Numbering is from the 5′ end of the double-stranded oligonucleotides.

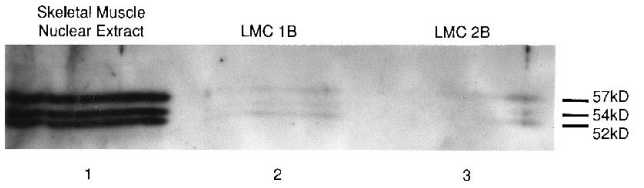


FIG. 8. Western blot analysis of proteins eluted from LMC 1B and LMC 2B. Lanes: 1, 5 μg of unfractionated skeletal muscle nuclear proteins run on the same gel for comparison; 2, polypeptides eluted from the LMC 1B protein/M-CAT-1 complex; 3, polypeptides eluted from the LMC 2B protein/M-CAT-2 complex. Detection of TEF-1 in these complexes is specific, since an intermediate-mobility complex formed in brain nuclear extract was tested and contained no TEF-1 proteins by this assay. The estimated sizes in kilodaltons of the TEF-1 polypeptide bands are indicated to the right of the panel.

encompasses the entire M-CAT motif as well as the proximal part of the 5′-flanking sequence (Fig. 7C). The experiments shown were performed with skeletal muscle extract; however, essentially identical results were obtained when a brain nuclear extract was used (data not shown).

The LMC 1B and LMC 2B footprints partially overlap the M-CAT core motif, although the sensitivity to methylation of antisense residue G^{21} (i.e., the 3′ end of the motif) is less pronounced in the low-mobility complexes than in the high-mobility complexes, C1, C2, and C3. This could indicate that the protein(s) in LMC 1B and 2B displaces the TEF-1 proteins bound to the M-CAT core motif by virtue of an overlapping recognition site or, alternatively, that TEF-1 proteins participate in the formation of the low-mobility complexes with a coincident alteration in their interaction with the core M-CAT motif. This led us to ask if the TEF-1 proteins present in the major high-mobility complexes C1, C2, and C3 are also present in LMC 1B and LMC 2B.

Higher-order complexes contain TEF-1 polypeptides. Immunoblotting with a polyclonal antibody broadly reactive against TEF-1 proteins revealed three TEF-1-related polypeptides in extracts of striated muscle (Fig. 8, lane 1). These three polypeptides are also isolated from gel shift high-mobility complexes C1, C2, and C3 (14a). To determine if TEF-1 proteins are present in the low-mobility complexes, LMC 1B and LMC 2B were gel isolated and subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis and immunoblotting with the anti-TEF-1 antibody. LMC 1A was insufficiently abundant to be analyzed in this way. However, both LMC 1B and LMC 2B contained the same three TEF-1 polypeptides that are found in unfractionated muscle nuclear extracts (Fig. 8, compare lanes 2 and 3 with lane 1). We conclude from this experiment that TEF-1 proteins are components of the low-mobility complexes and therefore that TEF-1 proteins interact with other proteins to form higher-order complexes on the M-CAT-1 and M-CAT-2 elements.

DISCUSSION

The present work shows that muscle-specific M-CAT elements are complex protein-binding sites that depend upon both a core M-CAT motif (CATTCCT) and the flanking sequences surrounding that motif. The core motif is a binding site for TEF-1 proteins, and this binding interaction is essential for the function of M-CAT elements. Flanking sequences modulate expression directed by M-CAT motifs. Non-muscle-specific artificial promoters were obtained when flanking sequences from the SV40 enhancer were used to surround a canonical core M-CAT motif or when intervening flanking sequence was removed (Fig. 2B). This explains why M-CAT elements from the SV40 genome are active in HeLa, P19, and F9 cells and in keratinocytes (11, 16, 22, 42, 60), as well as the nonmuscle cell types used here. Strong non-muscle-specific promoters were produced when the spacing, center-to-center, between M-CATs was 23 bp or greater (Fig. 2B, bottom panel, and data not shown). The promoter produced was weaker when the M-CAT elements were closely apposed (Fig. 2B, top panel).

Muscle-specific artificial promoters were obtained, on the other hand, when flanking sequences from the cTNT promoter were used to surround canonical core M-CAT motifs (Fig. 1B and C). Flanking sequence appears to be responsible for mediating repression of the cTNT-derived M-CAT elements in nonmuscle cells. For example, the M-CAT-1/Emt variant contains a mutation within the E-box site found in the 5′-flanking sequence of the cTNT M-CAT-1 element. The promoter derived by multimerization of this element was efficiently expressed in liver and fibroblast cells (Fig. 2D), unlike the native cTNT promoter or artificial promoters containing native flanking sequence, which were inactive in these cell types.

M-CAT element-dependent promoters are also absolutely dependent upon an intact core M-CAT motif for activity, indicating that protein interaction with the core M-CAT motif is a minimum requisite for function of the compound element (16, 37, 38; also see above). M-CAT core motifs are binding targets for TEF-1 proteins (14, 20, 22, 27, 35, 49, 51, 58). The dependence of tissue specificity upon appropriate flanking sequences might result from the binding of additional proteins, which modulate the activity of TEF-1 proteins bound to the M-CAT core motif. Alternatively, the modulation might be mediated by a change in interaction between bound TEF-1 proteins and the different flanking sequences.

To distinguish between the above possibilities, we looked for specific interactions between flanking sequences and proteins present in muscle and nonmuscle extracts. We found that any M-CAT element variant that contained an intact core M-CAT motif, regardless of flanking sequence, could compete for the three high-mobility complexes, C1, C2, and C3, although the tissue-specific activity of the elements varied widely (Fig. 3 and 4). Therefore, we conclude that this binding to the core motif alone is probably insufficient to confer cell-specific regulation. However, gel retardation assays revealed that higher-order complexes also form on M-CAT-1 and M-CAT-2 elements. The formation of higher-order complexes depends on specific cTNT flanking DNA sequences because M-CAT elements with non-cTNT-derived flanking sequence neither form such higher-order complexes (data not shown) nor compete for their formation (Fig. 3 and 4). Moreover, the ability of specific flanking sequences to form and compete for higher-order complex formation (Fig. 3 and 4 and data not shown) directly correlates with their ability to support muscle-specific transcription in artificial promoters (Fig. 1 and 2).

Different higher-order complexes form on M-CAT-1 and M-CAT-2 elements, as judged by their distinct mobilities (Fig. 5) and the inability of the probes to cross-compete (Fig. 3 and 4). Competition experiments with chimeric M-CAT elements indicate that low-mobility complex formation depends on the 5′ sequence of cTNT M-CAT elements (Fig. 3 and 4). Moreover, competition experiments with the M-CAT-1/Emt variant indicate that a localized 3-nucleotide mutation in the 5′-flanking sequence of M-CAT-1 disrupts binding of LMC 1A and LMC 1B (Fig. 3). These data are consistent with the notion that additional proteins bind in a sequence-specific manner to the flanking regions of cTNT-derived M-CAT elements to form LMC 1A, LMC 1B, and LMC 2B.

LMC 1A and LMC 1B both appear to result from the bind-

ing of a protein at, or very near, the E box, because a 3-nucleotide mutation in this site in M-CAT-1/Emt greatly reduces the ability of that element to compete for either complex (Fig. 3A, lane 7). LMC 1A formation appears to result from binding of a factor to the E box itself, since the methylation interference footprint derived for LMC 1A is strongest over the E box (Fig. 7A and C) and the MCK E box can compete for this complex (Fig. 3). However, the limited tissue distribution and low relative intensity of LMC 1A lead us to tentatively conclude that its functional significance is likely to be minor.

LMC 1B has a binding specificity distinct from LMC 1A and MDFs, since it is not competed for by the MCK E box (Fig. 3). This complex is formed in all tissues examined (liver, kidney, gizzard, lung, brain, and skeletal and cardiac muscle [Fig. 5 and data not shown]). LMC 1B is therefore also unlikely to contain MDF family members. Its methylation interference footprint encompasses the 5′ end of the core M-CAT motif and the 3′ end of the E box (Fig. 7A and C). LMC 1B competition experiments with mutations through the 5′-flanking DNA of the M-CAT-1 element support the notion that LMC 1B recognizes a site overlapping but distinct from the E box, in agreement with the methylation interference data presented here (31a). The LMC 1B complex may mediate repression of the M-CAT-1 element in nonmuscle cells because a mutation that disrupts its formation correlates with increased M-CAT-dependent promoter activity in nonmuscle cells (Fig. 2D). It has been noted that LMC 1B derived from skeletal muscle tissues appears to migrate fractionally faster than does nonmuscle-derived LMC 1B (Fig. 5). We hypothesize that this slight difference in electrophoretic mobility is indicative of a posttranslational modification, such as phosphorylation, of a component of the complex that permits a muscle-specific release of the repression.

The M-CAT-2 low-mobility complex, LMC 2B, has characteristics similar to those of LMC 1B. Besides having similar electrophoretic mobility and broad methylation interference footprints, both contain TEF-1 proteins (see below) and neither is competed for by M-CAT elements containing non-cTNT-flanking regions. Like LMC 1B, LMC 2B is likely to be involved in mediating the cell specificity of the M-CAT-2 element, because LMC 2B is generated by sequence-specific interaction of a nuclear protein(s) with the M-CAT-2 flanking region and this flanking region is required for the tissue specificity of M-CAT-2. Nevertheless, the flanking sequences recognized in each M-CAT element are quite different, and M-CAT-1 and M-CAT-2 do not cross-compete for low-mobility complex formation. Thus, either LMC 1B and LMC 2B contain different additional proteins or, if the same protein(s) is involved in the formation of both complexes, it must interact with the flanking sequences in a substantially different manner.

Methylation interference footprinting of LMC 1B or LMC 2B gave a broad footprint overlapping both the M-CAT site and the immediate 5′-flanking sequence of both M-CAT-1 and M-CAT-2 elements (Fig. 7). Moreover, each of the three immunologically related forms of TEF-1 protein found in muscle nuclear extracts was also found to be present in LMC 1B and LMC 2B (Fig. 8). However, the interaction of TEF-1 proteins with the cognate M-CAT-binding motif was subtly altered in the low-mobility complexes in that the methylation footprints were shifted toward the 5′ region of the element, compared with the TEF-1 footprint of the high-mobility complexes, C1, C2, and C3 (compare Fig. 6C and 7C). We conclude that LMC 1B and LMC 2B consist of higher-order complexes of TEF-1 proteins and other auxiliary proteins. Low-mobility complexes may consist of multiple species with distinct TEF-1 protein constituents (as do the high-mobility bands, C1, C2, and C3 [14a]), but size differences in this range may not be resolved under these gel shift conditions. The auxiliary proteins may differ between LMC 1B and LMC 2B, but both alter the binding of TEF-1 proteins to their cognate motif in a similar manner. The displacement of the TEF-1 footprint in these higher-order complexes may be a measure of the tightness of the interaction between TEF-1 proteins and the auxiliary protein(s), such that TEF-1 proteins are tilted off their cognate site toward the 5′ binding site of the second protein (Fig. 7C). Alternatively, the altered footprint may represent changes in the combination of TEF-1 proteins bound on these elements.

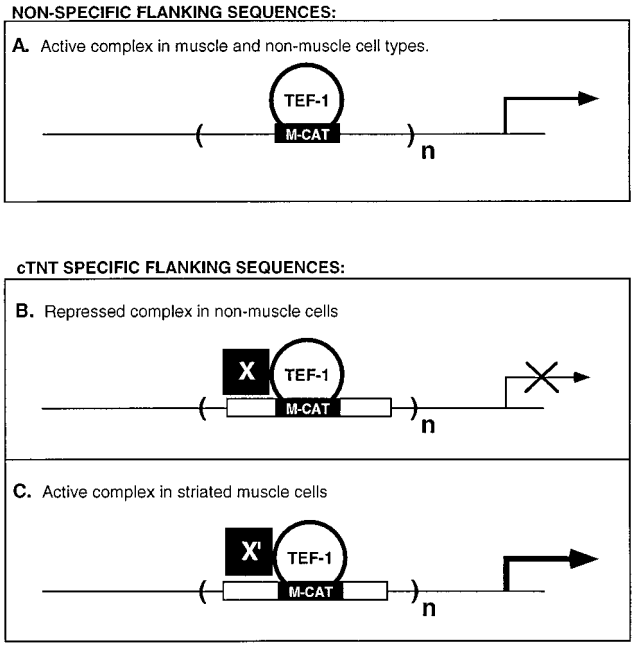


FIG. 9. Diagrammatic representation outlining the mechanism proposed for the tissue specificity of M-CAT elements containing cTNT-derived flanks. (A) Simple M-CAT elements bind TEF-1 proteins but have flanking sequences that do not bind additional factors. Such elements direct non-cell-specific expression. (B and C) Complex M-CAT elements bind TEF-1 proteins but have flanking sequence that binds an additional protein(s) (X/X′). The complex of TEF-1 protein(s) and X in nonmuscle cells is inactive (B), whereas the complex of TEF-1 protein(s) and X′ in nonmuscle cells is active (C). The difference between the complexes in panels B and C may lie in modifications of TEF-1 protein(s) or the X component (shown here as X versus X′) or differences in the primary components or stoichiometric balance within them.

In summary, muscle-specific M-CAT elements are a distinct subset of active M-CAT elements. Simple M-CAT elements, such as those found in the SV40 promoter, direct non-cell-specific gene expression. These consist of M-CAT core motifs that bind TEF-1 proteins and flanking sequences that do not bind protein (Fig. 9A). Muscle-specific M-CAT elements, on the other hand, consist of a core M-CAT motif that binds TEF-1 proteins plus flanking DNA that binds an additional protein(s) in a sequence-specific fashion. We hypothesize that the second binding determines whether the element is negatively regulated (in nonmuscle cells [Fig. 9B]) or positively regulated (in muscle cells [Fig. 9C]). Neither the identity nor the diversity of potential secondary binding proteins is known at this time (see below).

The model of M-CAT element-dependent muscle-specific transcription outlined above differs markedly from the mechanism by which E boxes are believed to activate muscle-specific transcription. E-box-binding factors (MDFs) are themselves muscle specific and, once expressed, are potentially capable of

TABLE 2. M-CAT elements and their flanking DNA in other muscle-specific promoters

Promoter and position	M-CAT and flanking sequence[a]	Tissue[b]	Reference(s)
cTNT			
Chicken (−66S)	GCGCCGGGCA**CATTCCT**GCTGCTCTGC	Sk, C	36
Chicken (−89S)	ACAAGTGTTG**CATTCCT**CTCTGGGCGC	Sk, C	37
βMHC			
Human (−284AS)	CCAGGCCTCA**CATTCCA**CAGCTGGCAG	C	15
Rabbit (−264AS)	TATCTCCTCG**CATTCCA**CTGCCTGTGG	Sk	48
Rat (−267AS)	ATATCCCTTA**CATTCCA**CAGCTCAC	Sk, C	55
Rat (−196S)	CTGAACATGC**CATACCA**CAACAATGAC	C, ↑ PKC	28
αMHC			
Rat (−236AS)	TCCTTGGGCA**CATTCCT**CCTCCCCAAA	Sk, C	40
Rat (−42AS)	CTTTATAGCT**CATTCCA**CGTGCCTGCT	C, ↑ cAMP	20
Myosin light chain 2			
Chicken (−35AS)	AAATACACCC**CATTCCA**GGCTAAAAAT	C[c]	46
s/cTNC			
Mouse (−50AS)	CCTCCTGCTA**CATTCCC**AGCCCAGCCC	C[c]	44
VSM α-actin			
Mouse (−175AS)	TCTTCCACTG**CATTCCT**CTGTTCTGCT	VSM, ↑ serum	7, 52
Skin α-actin			
Chicken (−60S)	AGCTTGCCG**CATTCCT**GGGGGCCGGG	C, ↑ TGF-β	35
Mouse (−63S)	AGGGCAGCAA**CATTCTT**CGGGGCGGT	↑ PKC	29
AChRβ			
Rat (−43S)	ACAGGTGCA**CATTCCT**GGGCGCCTCG	Sk	4
PGAM-M			
Human (−72AS)	TGCCAATCAG**CATTCCA**GGCGGTGGCA	Sk, C[c]	41
c-*mos*			
Rat (−1561S)	AGGCTTTATC**CATTTCT**GAGATAAAGA	Sk	32
Rat (−1541S)	ATAAAGATTT**CATTTCT**AATCTCAGTA	Sk	32
AE3 Cl^-/HCO_3^- channel			
Rodent/human (−190S)	GATCTCGGCA**CATTCCT**TCCATCTTAT	C[c]	33
Brain natriuretic peptide			
Rat (−101AS)	TTATCAGACA**CATTCCT**GCCTGCTGAG	C	18
Human (−117S)	CGGAGGGGCT**CATTCCC**GGGCCCTGAT	C[c]	31

[a] M-CAT motifs are shown in boldface text. Ten nucleotides of sequence flanking the M-CAT on each side is shown in every case for which this information is published. Nominal E-box sequences are underlined.

[b] Sk, C, or VSM indicates the tissue type in which the M-CAT-containing promoter was tested: skeletal myocytes or derived cell lines, cardiomyocytes, and vascular smooth muscle cells, respectively. Other notations refer to M-CATs which have been demonstrated to be the target response elements for the stimuli noted: cyclic AMP (cAMP), protein kinase C (PKC), or transforming growth factor beta (TGF-β).

[c] M-CAT element whose function has not yet been shown.

activating all appropriate E-box targets de novo. M-CAT elements, by contrast, bind TEF-1 proteins that are widely expressed in many cell types but whose activity is modified by the interaction of a cofactor(s) that binds to specific flanking regions. This added level of complexity may provide greater flexibility in the regulation of M-CAT-dependent promoters, by virtue of the conditional and modulatable activity of these compound elements.

The cTNT-derived M-CAT elements studied here were compared with other M-CAT elements implicated in the function of muscle-specific promoters (Table 2). None of these M-CAT elements are precisely conserved with respect to the immediate flanking sequence; however, it is noteworthy that nominal E boxes appear close to M-CAT elements in the cTNT, β myosin heavy chain (βMHC), αMHC, and acetylcholine receptor β (AChRβ) subunit promoters (4, 15, 20, 38). Mutation of the E box flanking the distal M-CAT in the human βMHC promoter reduces cardiac activity to one-fifth (15). The individual requirement for the E box which flanks the M-CAT in the AChRβ promoter has not been reported.

The M-CAT element and E box in the αMHC proximal promoter are particularly interesting. In this promoter, the M-CAT element overlaps the E box, and this complex element mediates cyclic AMP inducibility and is required for basal cardiac muscle-specific promoter activity (20). On gel retardation analysis, this element generates only a low-mobility complex, which is competed for by M-CAT elements and contains a TEF-1 protein(s) (20). The methylation interference footprint of that complex is similar to the footprint derived here for LMC 1B and LMC 2B (20). Thus, the auxiliary protein in LMC 1B may be related or identical to that binding the E-box/M-CAT element of the αMHC promoter.

Another E box, found in the AChRδ promoter, is capable of mediating both activation in myotubes and repression in myoblasts and fibroblasts (50). Unlike the M-CAT-1 E box, the AChRδ E box binds myoD and myogenin as well as a third, non-muscle-specific, unidentifiable factor. Furthermore, the AChRδ E-box is not found in association with another *cis*-acting sequence upon whose activity it depends (50); in contrast, mutation of the core M-CAT motif in M-CAT-1, which leaves the E box intact, abolishes all promoter activity of the mutant element in muscle (data not shown). This difference may reflect the differential ability of these E boxes to bind MDFs as well as the presumptive repressor protein. Thus, the third non-MDF factor bound to the AChRδ E box may be related to the M-CAT-1 auxiliary factor found in LMC 1B.

The nominal E-box-binding site may constitute one conserved binding site for TEF-1-associated cell-specific repres-

sors. The M-CAT elements found in other cardiac promoters which do not contain E boxes are nevertheless probably subject to similar repression mechanisms. The lack of overt conservation in these other M-CAT element-flanking sequences and indeed between those of M-CAT-1 and M-CAT-2 may point to the existence of a number of such auxiliary proteins, perhaps each facilitating the targeting of different second-messenger cascades to TEF-1 proteins. On the other hand, the diversity may reflect the ability of a single factor to bind a very loosely defined binding site once closely associated with a TEF-1-binding motif.

ACKNOWLEDGMENTS

We thank Nina Kostanian and Monique Benoualid for their expert technical assistance. We also thank Paul Simpson and David Gardner for critical reading of the manuscript and members of our laboratory for helpful discussion and comments.

This work was supported by NIH grants HL-35561 and HL-43821 and a postdoctoral fellowship from the NIH Fogarty International Center (F05 TWO4892-02) to S.B.L.

REFERENCES

1. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. A. Smith, J. G. Seidman, and K. Struhl (ed.).** 1987. Current protocols in molecular biology. Greene Publishing Associates and Wiley-Interscience, New York.

1a. **Azakie, A., S. B. Larkin, I. K. Farrance, G. Grenningloh, and C. P. Ordahl.** 1996. DTEF-1, a novel member of the transcription enhancer factor-1 (TEF-1) multigene family. J. Biol. Chem. **271:**8260–8265.

2. **Bain, G., E. C. Maandag, D. J. Izon, D. Amsen, A. M. Kruisbeek, B. C. Weintraub, I. Krop, M. S. Schlissel, A. J. Feeney, M. van Roon, et al.** 1994. E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. Cell **79:**885–892.

3. **Baker, S. J., and E. P. Reddy.** 1995. B cell differentiation: role of E2A and Pax5/BSAP transcription factors. Oncogene **11:**413–426.

4. **Berberich, C., I. Durr, M. Koenen, and V. Witzemann.** 1993. Two adjacent E box elements and a M-CAT box are involved in the muscle-specific regulation of the rat acetylcholine receptor beta subunit gene. Eur. J. Biochem. **216:**395–404.

5. **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.

6. **Buskin, J. N., and S. D. Hauschka.** 1989. Identification of a myocyte nuclear factor that binds to the muscle-specific enhancer of the mouse muscle creatine kinase gene. Mol. Cell. Biol. **9:**2627–2640.

7. **Cogan, J. G., S. Sun, E. S. Stoflet, L. J. Schmidt, M. J. Getz, and A. R. Strauch.** 1995. Plasticity of vascular smooth muscle alpha-actin gene transcription. Characterization of multiple, single-, and double-strand specific DNA-binding proteins in myoblasts and fibroblasts. J. Biol. Chem. **270:** 11310–11321.

8. **Cooper, T. A., and C. P. Ordahl.** 1985. A single cardiac troponin T gene generates embryonic and adult isoforms via developmentally regulated alternate splicing. J. Biol. Chem. **260:**11140–11148.

9. **Corcoran, L. M., and M. Karvelas.** 1994. Oct-2 is required early in T cell-independent B cell activation for G1 progression and for proliferation. Immunity **1:**635–645.

10. **Corcoran, L. M., M. Karvelas, G. J. Nossal, Z. S. Ye, T. Jacks, and D. Baltimore.** 1993. Oct-2, although not required for early B-cell development, is critical for later B-cell maturation and for postnatal survival. Genes Dev. **7:**570–582.

11. **Davidson, I., C. Fromental, P. Augereau, A. Wildeman, M. Zenke, and P. Chambon.** 1986. Cell-type specific protein binding to the enhancer of simian virus 40 in nuclear extracts. Nature (London) **323:**544–548.

12. **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.

13. **Edmondson, D. G., and E. N. Olson.** 1993. Helix-loop-helix proteins as regulators of muscle-specific transcription. J. Biol. Chem. **268:**755–758.

14. **Farrance, I. K., J. H. Mar, and C. P. Ordahl.** 1992. M-CAT binding factor is related to the SV40 enhancer binding factor, TEF-1. J. Biol. Chem. **267:** 17234–17240.

14a. **Farrance, I. K. G., and C. P. Ordahl.** 1996. The role of transcription enhancer factor-1 (TEF-1) related proteins in the formation of M-CAT binding complexes in muscle and nonmuscle tissues. J. Biol. Chem. **271:**8266–8274.

15. **Flink, I. L., J. G. Edwards, J. J. Bahl, C. C. Liew, M. Sole, and E. Morkin.** 1992. Characterization of a strong positive cis-acting element of the human beta-myosin heavy chain gene in fetal rat heart cells. J. Biol. Chem. **267:** 9917–9924.

16. **Fromental, C., M. Kanno, H. Nomiyama, and P. Chambon.** 1988. Cooperativity and hierarchical levels of functional organization in the SV40 enhancer. Cell **54:**943–953.

17. **Genetta, T., D. Ruezinsky, and T. Kadesch.** 1994. Displacement of an E-box-binding repressor by basic helix-loop-helix proteins: implications for B-cell specificity of the immunoglobulin heavy-chain enhancer. Mol. Cell. Biol. **14:**6153–6163.

18. **Glembotski, C.** 1995. Personal communication.

19. **Gorman, C. M.** 1985. High efficiency gene transfer into mammalian cells, p. 143–109. *In* D. M. Glover (ed.), DNA cloning. IRL Press, Oxford.

20. **Gupta, M. P., M. Gupta, and R. Zak.** 1994. An E-box/M-CAT hybrid motif and cognate binding protein(s) regulate the basal muscle-specific and cAMP-inducible expression of the rat cardiac alpha-myosin heavy chain gene. J. Biol. Chem. **269:**29677–29687.

21. **Iannello, R., J. H. Mar, and C. P. Ordahl.** 1991. Characterization of a promoter element required for transcription in myocardial cells. J. Biol. Chem. **266:**3309–3316.

22. **Ishiji, T., M. J. Lace, S. Parkkinen, R. D. Anderson, T. H. Haugen, T. P. Cripe, J. H. Xiao, I. Davidson, P. Chambon, and L. P. Turek.** 1992. Transcriptional enhancer factor (TEF)-1 and its cell-specific co-activator activate human papillomavirus-16 E6 and E7 oncogene transcription in keratinocytes and cervical carcinoma cells. EMBO J. **11:**2271–2281.

23. **Jan, Y. N., and L. Y. Jan.** 1993. HLH proteins, fly neurogenesis, and vertebrate myogenesis. Cell **75:**827–830.

24. **Jiang, S. W., and N. L. Eberhardt.** 1995. Involvement of a protein distinct from transcription enhancer factor-1 (TEF-1) in mediating human chorionic somatomammotropin gene enhancer function through the GT-IIC enhanson in choriocarcinoma and COS cells. J. Biol. Chem. **270:**13906–13915.

25. **Jiang, S. W., N. L. Eberhardt, T. P. Cripe, J. H. Xiao, I. Davidson, P. Chambon, and L. P. Turek.** 1994. The human chorionic somatomammotropin gene enhancer is composed of multiple DNA elements that are homologous to several SV40 enhansons. J. Biol. Chem. **269:**10384–10392.

26. **Kadesch, T.** 1992. Helix-loop-helix proteins in the regulation of immunoglobulin gene transcription. Immunol. Today **13:**31–36.

27. **Kariya, K., I. K. Farrance, and P. C. Simpson.** 1993. Transcriptional enhancer factor-1 in cardiac myocytes interacts with an alpha 1-adrenergic- and beta-protein kinase C-inducible element in the rat beta-myosin heavy chain promoter. J. Biol. Chem. **268:**26658–26662.

28. **Kariya, K., L. R. Karns, and P. C. Simpson.** 1994. An enhancer core element mediates stimulation of the rat beta-myosin heavy chain promoter by an alpha 1-adrenergic agonist and activated beta-protein kinase C in hypertrophy of cardiac myocytes. J. Biol. Chem. **269:**3775–3782.

29. **Karns, L. R., K. Kariya, and P. C. Simpson.** 1995. M-CAT, CArG, and Sp1 elements are required for alpha 1-adrenergic induction of the skeletal alpha-actin promoter during cardiac myocyte hypertrophy. Transcriptional enhancer factor-1 and protein kinase C as conserved transducers of the fetal program in cardiac growth. J. Biol. Chem. **270:**410–417.

30. **Konig, H., P. Pfisterer, L. M. Corcoran, and T. Wirth.** 1995. Identification of CD36 as the first gene dependent on the B-cell differentiation factor Oct-2. Genes Dev. **9:**1598–1607.

31. **LaPointe, M.** 1995. Personal communication.

31a. **Larkin, S. B.** Unpublished observations.

32. **Lenormand, J. L., and S. A. Leibovitch.** 1995. Identification of a novel regulatory element in the c-mos locus that activates transcription in somatic cells. Biochem. Biophys. Res. Commun. **210:**181–188.

33. **Linn, S. C., G. R. Askew, A. G. Menon, and G. E. Shull.** 1995. Conservation of an AE3 Cl-/HCO3- exchanger cardiac-specific exon and promoter region and AE3 mRNA expression patterns in murine and human hearts. Circ. Res. **76:**584–591.

34. **Luo, Y., and R. G. Roeder.** 1995. Cloning, functional characterization, and mechanism of action of the B-cell-specific transcriptional coactivator OCA-B. Mol. Cell. Biol. **15:**4115–4124.

35. **MacLellan, W. R., T. C. Lee, R. J. Schwartz, and M. D. Schneider.** 1994. Transforming growth factor-beta response elements of the skeletal alpha-actin gene. Combinatorial action of serum response factor, YY1, and the SV40 enhancer-binding protein, TEF-1. J. Biol. Chem. **269:**16754–16760.

36. **Mar, J. H., P. B. Antin, T. A. Cooper, and C. P. Ordahl.** 1988. Analysis of the upstream regions governing expression of the chicken cardiac troponin T gene in embryonic cardiac and skeletal muscle cells. J. Cell Biol. **107:**573–585.

37. **Mar, J. H., and C. P. Ordahl.** 1988. A conserved CATTCCT motif is required for skeletal muscle-specific activity of the cardiac troponin T gene promoter. Proc. Natl. Acad. Sci. USA **85:**6404–6408.

38. **Mar, J. H., and C. P. Ordahl.** 1990. M-CAT binding factor, a novel *trans*-acting factor governing muscle-specific transcription. Mol. Cell. Biol. **10:** 4271–4283.

39. **Melin, F., M. Miranda, N. Montreau, M. L. DePamphilis, and D. Blangy.** 1993. Transcription enhancer factor-1 (TEF-1) DNA binding sites can specifically enhance gene expression at the beginning of mouse development. EMBO J. **12:**4657–4666.

40. **Molkentin, J. D., and B. E. Markham.** 1994. An M-CAT binding factor and

an RSRF-related A-rich binding factor positively regulate expression of the alpha-cardiac myosin heavy-chain gene in vivo. Mol. Cell. Biol. **14:**5056–5065.

41. **Nakatsuji, Y., K. Hidaka, S. Tsujino, Y. Yamamoto, T. Mukai, T. Yanagihara, T. Kishimoto, and S. Sakoda.** 1992. A single MEF-2 site is a major positive regulatory element required for transcription of the muscle-specific subunit of the human phosphoglycerate mutase gene in skeletal and cardiac muscle cells. Mol. Cell. Biol. **12:**4384–4390.
42. **Ondek, B., A. Shepard, and W. Herr.** 1987. Discrete elements within the SV40 enhancer region display different cell-specific enhancer activities. EMBO J. **6:**1017–1025.
43. **Ordahl, C. P.** 1992. Developmental regulation of sarcomeric gene expression. Curr. Top. Dev. Biol. **26:**145–168.
44. **Parmacek, M. S., and J. M. Leiden.** 1989. Structure and expression of the murine slow/cardiac troponin C gene. J. Biol. Chem. **264:**13217–13225.
45. **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.
46. **Qasba, P., E. Lin, M. D. Zhou, A. Kumar, and M. A. Siddiqui.** 1992. A single transcription factor binds to two divergent sequence elements with a common function in cardiac myosin light chain-2 promoter. Mol. Cell. Biol. **12:**1107–1116.
47. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
48. **Shimizu, N., E. Dizon, and R. Zak.** 1992. Both muscle-specific and ubiquitous nuclear factors are required for muscle-specific expression of the myosin heavy-chain β gene in cultured cells. Mol. Cell. Biol. **12:**619–630.
49. **Shimizu, N., G. Smith, and S. Izumo.** 1993. Both a ubiquitous factor mTEF-1 and a distinct muscle-specific factor bind to the M-CAT motif of the myosin heavy chain beta gene. Nucleic Acids Res. **21:**4103–4110.
50. **Simon, A. M., and S. J. Burden.** 1993. An E box mediates activation and repression of the acetylcholine receptor delta-subunit gene during myogenesis. Mol. Cell. Biol. **13:**5133–40.
51. **Stewart, A. F., S. B. Larkin, I. K. Farrance, J. H. Mar, D. E. Hall, and C. P. Ordahl.** 1994. Muscle-enriched TEF-1 isoforms bind M-CAT elements from muscle-specific promoters and differentially activate transcription. J. Biol. Chem. **269:**3147–3150.
52. **Sun, S., E. S. Stoflet, J. G. Cogan, A. R. Strauch, and M. J. Getz.** 1995. Negative regulation of the vascular smooth muscle alpha-actin gene in fibroblasts and myoblasts: disruption of enhancer function by sequence-specific single-stranded-DNA-binding proteins. Mol. Cell. Biol. **15:**2429–2436.
53. **Tanaka, M., and W. Herr.** 1990. Differential transcriptional activation by Oct-1 and Oct-2: interdependent activation domains induce Oct-2 phosphorylation. Cell **60:**375–386.
54. **Tanaka, M., J. S. Lai, and W. Herr.** 1992. Promoter-selective activation domains in Oct-1 and Oct-2 direct differential activation of an snRNA and mRNA promoter. Cell **68:**755–767.
55. **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.
56. **Weintraub, H.** 1993. The MyoD family and myogenesis: redundancy, networks, and thresholds. Cell **75:**1241–1244.
57. **Wright, W. E., M. Binder, and W. Funk.** 1991. Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site. Mol. Cell. Biol. **11:**4104–4110.
58. **Xiao, J. H., I. Davidson, H. Matthes, J.-M. Garnier, and P. Chambon.** 1991. Cloning, expression, and transcriptional properties of the human enhancer factor TEF-1. Cell **65:**551–568.
59. **Yutzey, K. E., and S. F. Konieczny.** 1992. Different E-box regulatory sequences are functionally distinct when placed within the context of the troponin I enhancer. Nucleic Acids Res. **20:**5105–5113.
60. **Zenke, M., T. Grundstrom, H. Matthes, M. Wintzerith, C. Schatz, A. Wildeman, and P. Chambon.** 1986. Multiple sequence motifs are involved in SV40 enhancer function. EMBO J. **5:**387–397.
61. **Zhuang, Y., P. Soriano, and H. Weintraub.** 1994. The helix-loop-helix gene E2A is required for B cell formation. Cell **79:**875–884.