

# An E Box Comprises a Positional Sensor for Regional Differences in Skeletal Muscle Gene Expression and Methylation

Elena Ceccarelli,<sup>1</sup> Michael J. McGrew,<sup>2</sup> Tom Nguyen,<sup>3</sup> Uta Grieshammer,<sup>4</sup> David Horgan, Stephen H. Hughes,\* and Nadia Rosenthal<sup>5</sup>

Cardiovascular Research Center, Massachusetts General Hospital—East, Charlestown, Massachusetts 02129; and \*National Cancer Institute—Frederick Cancer Research Program, ABL-Basic Research Program, Frederick, Maryland 21702-1013

To dissect the molecular mechanisms conferring positional information in skeletal muscles, we characterized the control elements responsible for the positionally restricted expression patterns of a muscle-specific transgene reporter, driven by regulatory sequences from the MLC1/3 locus. These sequences have previously been shown to generate graded transgene expression in the segmented axial muscles and their myotomal precursors, fortuitously marking their positional address. An evolutionarily conserved E box in the MLC enhancer core, not recognized by MyoD, is a target for a nuclear protein complex, present in a variety of tissues, which includes Hox proteins and Zbu1, a DNA-binding member of the SW12/SNF2 gene family. Mutation of this E box in the MLC enhancer has only a modest positive effect on linked CAT gene expression in transfected muscle cells, but when introduced into transgenic mice the same mutation elevates CAT transgene expression in skeletal muscles, specifically releasing the rostral restriction on MLC-CAT transgene expression in the segmented axial musculature. Increased transgene activity resulting from the E box mutation in the MLC enhancer correlates with reduced DNA methylation of the distal transgenic MLC1 promoter as well as in the enhancer itself. These results identify an E box and the proteins that bind to it as a positional sensor responsible for regional differences in axial skeletal muscle gene expression and accessibility. © 1999 Academic Press

## INTRODUCTION

Significant progress has been made over the past few years in defining the molecular circuitry that gives rise to skeletal muscle lineages in vertebrate embryos. Identification of transcription factors controlling the tissue- and stage-specific expression of the myogenic phenotype has led to the elucidation of regulatory pathways underlying the

generation of skeletal muscles from their mesodermal precursors during development.

Less is understood about the mechanisms responsible for conferring positional identity on individual muscle groups within the developing embryo and retaining these differences in the adult musculature. The complex and reproducible patterns of contractile tissues formed in the embryo, and maintained in the adult, are likely to be specified by common pathways patterning functionally related tissues, such as nerves, tendon, and bone.

We have previously described a transgenic model that responds to positional cues in developing skeletal muscles along the mouse anteroposterior axis and maintains positionally restricted patterns of expression in adult musculature (Donoghue *et al.*, 1991; Grieshammer *et al.*, 1992). Expression of a variety of transgene reporters, driven by a 1200-bp MLC1 promoter and a 920-bp downstream enhancer from the MLC1/3 locus, is graded in the developing somites and retains a rostrocaudal gradient of expression in

<sup>1</sup> Present address: SBGM Bat.142 CEA-Saclay, 91191 Gif-sur-Yvette France.

<sup>2</sup> Present address: Developmental Biology Institute of Marseille Campus de Luminy Case 907 13288 Marseille Cedex 9 France.

<sup>3</sup> Present address: Millenium Pharmaceuticals, Inc., Cambridge, MA 02139.

<sup>4</sup> Present address: Department of Anatomy, School of Medicine, University of California, San Francisco, CA 94143-0452.

<sup>5</sup> To whom correspondence should be addressed. Fax: (617) 724-9561.



enhancer seemed a particularly appealing candidate, since Hox proteins are involved in the establishment of positional identity in other somitically derived tissues, such as the axial skeleton, and have been implicated in the regulation of muscle patterning in both vertebrate and invertebrate embryos (reviewed in Olson and Rosenthal, 1994). Surprisingly, mutation of the Hox sites in the transgenic MLC enhancer did not lead to changes in the graded pattern of transgene expression (Rao *et al.*, 1996), suggesting that other elements in the MLC transgene must be responsive to regional differences in developing axial muscles.

In this report we demonstrate that the upstream E box (termed BMW) adjoining the Hox binding sites in the MLC enhancer acts as a weak negative regulatory element in muscle cell lines. It does not bind MyoD, but instead is recognized by a nuclear complex present in a variety of cell types. In transgenic mice, mutation of the BMW box alone is sufficient to increase transgene expression in a variety of muscle groups. The BMW mutation also releases the repression of an otherwise unmodified MLC transgene in anterior axial muscles, significantly diminishing the gradient both in activity and in methylation of the MLC transgenic sequences in axial musculature. These data implicate an E box in the MLC enhancer, and the protein complex that binds to it, as a positional sensor in the establishment of regionally restricted muscle gene expression and accessibility along the anteroposterior axis.

## MATERIAL AND METHODS

**Plasmids and site-directed mutagenesis.** The MSV-MyoD expression vector (Davis *et al.*, 1987) was a gift from A. Lassar and H. Weintraub. Plasmid MLC1CATBMW (abbreviated BMW in figures) was generated by polymerase chain reaction (PCR) mutagenesis (Libri *et al.*, 1995) of pMLC1CAT920 (Rosenthal *et al.*, 1989) (abbreviated WT in figures). The BMW site was mutated at the same residues shown by the methylation interference experiments to be important for the binding of the BMW protein complex (pBMW), using the following oligonucleotides:

BMWmut 5'-TGAGCGGTTACACTACCTAATTTAA-  
TAAGGTTAGGTCC-3'

5'MLC 5'-CGCGGATCCGCTATTAATCCC-3'

3'MLC 5'-CGCGGATCCGAGCTTTAAGATAC-3'

The mutagenic oligonucleotide (BMWmut), the 5'MLC oligonucleotide, and 100 ng of the wild-type template (pMLC1CAT920) were used to obtain a primary PCR product. A second PCR was performed using as primers the primary PCR product and the oligonucleotide 3'MLC. The resulting PCR product was then cloned in the *Bam*HI sites of the reporter plasmid pMLC1CAT. Vent polymerase was used to minimize PCR error and the entire mutant MLC enhancer was double-strand sequenced.

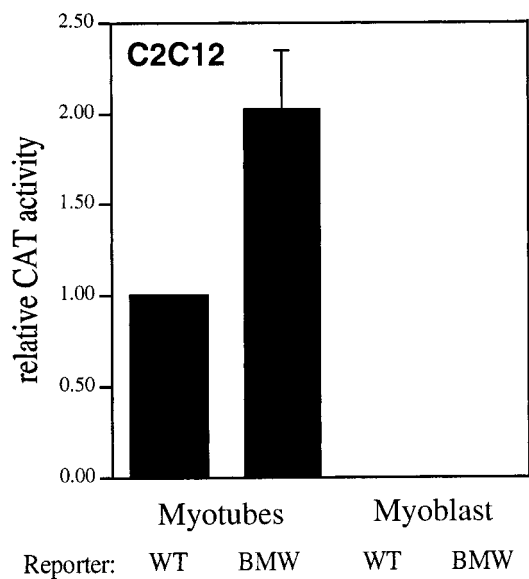
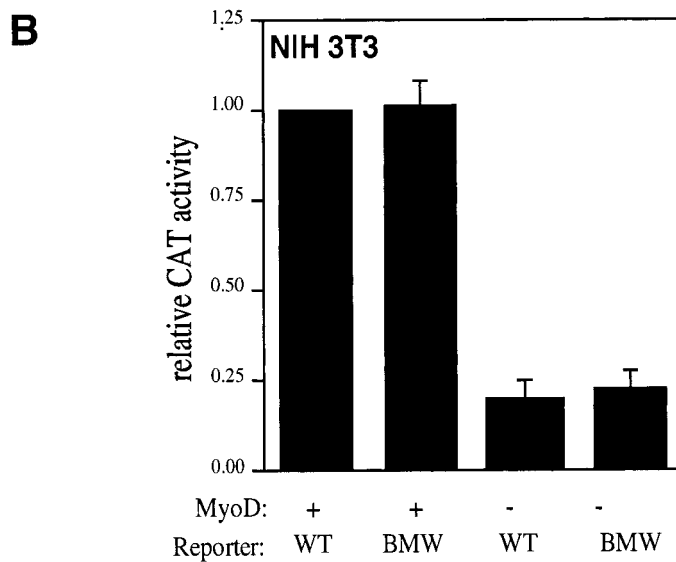
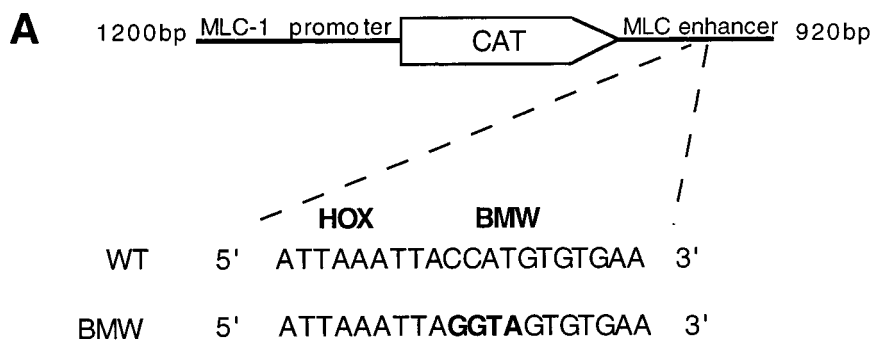
**Transgenic mice.** A 3.7-kb *Acc*I fragment was excised from pMLC1CATBMW expression cassettes (see Fig. 2) and injected into pronuclei of fertilized mouse eggs as described by Hogan *et al.*

(1986). Potential founders were screened for transgene integration by tail blotting with <sup>32</sup>P-labeled CAT DNA probe (Rosenthal *et al.*, 1989). The three founder lines (BMW Nos. 3, 4, and 9) expressing the transgene were established by mating positive animals to (C57BL6 × CBA) F<sub>1</sub> mice. F<sub>1</sub> progenies were checked for the presence of the transgene by PCR amplification of tail genomic DNA with primers specific for the CAT gene. Transgene copy number was determined by Southern blot (30–50 copies in BMW 3 and 4, <5 copies in BMW 9) and by comparison to transgenic mice carrying 40 copies of the wild-type MLC genomic fragment (line 7; Grieshammer *et al.*). Line 7 was chosen as a representative wild-type MLC-CAT line because it had a transgene copy number comparable to that of the BMW mutant lines, and because it expressed the MLC-CAT transgene at levels similar to those measured in multiple independently derived MLC-CAT lines (see Rosenthal *et al.*, 1989).

**Cell culture, DNA transfection, and CAT assays.** NIH3T3 mouse fibroblasts were propagated in Dulbecco's modified Eagle medium (DMEM) containing 10% fetal calf serum (FCS), 2% glutamine, and 50 units/ml penicillin–streptomycin, and seeded for transfection at 1 × 10<sup>6</sup> cells/100-mm plate. The C2C12 mouse skeletal muscle cell line (Blau *et al.*, 1985) was cultured in DMEM supplemented with 20% FCS and seeded for transfections on 100-mm plates at 1 × 10<sup>5</sup> cells for myoblasts and 4 × 10<sup>5</sup> cells for myotubes. Plasmids were transfected into NIH3T3 and C2C12 cells by calcium phosphate coprecipitation (Graham *et al.*, 1975) using 10 μg of reporter construct, 5 μg of MSV-myoD expression vector, and 2 μg of a SV40-luciferase control plasmid. The medium was changed after 16 h (DMEM containing 10% FCS for NIH3T3 cells, 20% FCS for myoblasts, and 2% horse serum for myotubes). Forty-eight hours later, cells were harvested according to published protocols (Rosenthal, 1987). Luciferase enzymatic assays were carried out to normalize the amount of cell extract used for CAT assays. CAT enzymatic assays were performed according to Seed and Sheen (1988). For CAT assays on transgenic mice, mouse tissues were dissected and frozen in dry ice, and protein extract was prepared by successive cycles of freeze/thawing. CAT assays were determined on adult heterozygotes (5 weeks or older). Three mice from each founder line were analyzed. The amount of extract used for the CAT assays was normalized to total protein content. CAT assays were performed as described by Gorman *et al.* (1982). In Fig. 3, each experimental set was normalized to the EDL muscle (set to 100). MLC line 7 Absolute values for each muscle varied less than 10× between the four lines, with line 9 being the weakest expressor. Relative values of absolute CAT activity for the four lines (Nos. 3, 4, 7, 9) in masseter (a weakly expressing muscle) and EDL (a strongly expressing muscle), expressed as fold difference in absolute CAT activity, setting No. 7 masseter at 1, were as follows: 1,3,7,1 for masseter; 28, 45, 42, 11 for EDL.

**Detection of genomic methylation.** Preparation of genomic DNA and quantitative analysis of transgene methylation were performed essentially as described (Grieshammer *et al.*, 1995), using a ligation-mediated PCR procedure (McGrew and Rosenthal, 1993), with PCR programs shortened to 30 s for denaturation, 45 s for annealing, and 30 s for extension. Analyses were performed on multiple DNA samples prepared from at least three different animals in each line, and were repeated at least twice for each sample.

**Electrophoresis mobility shift assay (EMSA) and methylation interference footprinting.** Binding and reaction conditions used for EMSAs and methylation interference footprinting have been described in Wentworth *et al.* (1991). Approximately 400 ng of



purified fusion protein (GST-MyoD) or 10  $\mu$ g of nuclear extract was used for EMSAs in a reaction volume of 20  $\mu$ l containing 10 mM Hepes (pH 7.9), 50 mM NaCl, and 1  $\mu$ g of poly(dI-C). Nuclear extracts from HeLa, C2C12, and NIH3T3 cell lines were prepared by the method of Dignam *et al.* (1983). Jurkat, 70Z, and 70Z cells treated with lipopolisaccharide (LPS) nuclear extracts were a generous gift of J. Sharon. Total protein concentration was determined spectrophotometrically using a standard Bradford assay. To generate larger probes for EMSAs, fragments of the 173-bp MLC enhancer (see Fig. 5A, wt173 and mABC) were  $^{32}$ P end-labeled on the sense or antisense strand of the vector pCAT173 at the polylinker restriction sites (Wentworth *et al.*, 1991). The following complementary oligonucleotides were synthesized:

BMW 20-mer 5'-TCGACAAATTACCATGTGTG-3'

BMW 5'-GGACCTAACCTTATTAATTACCATG-TGTGAACCGCTCA-3'

mBMW 5'-GGACCTAACCTTATTAATTAGGTAG-TGTGAACCGCTCA-3'

A 5'-GTCGCAGCAGGTGCA-3'

C 5'-CGAAGCAACAGGTGCCT-3'

AT 5'-TTTTAAAAATAACTTTT-3'

For EMSAs, the oligonucleotides listed above were annealed and end-labeled with  $^{32}$ P using the Klenow fragment (Promega). In competition experiments, double-stranded annealed oligonucleotides were used at a 100-fold molar excess unless indicated otherwise in the figure legends. For the methylation interference footprinting, the antisense fragment of the 173-bp MLC enhancer (mABC) was  $^{32}$ P end-labeled as described in Wentworth *et al.* (1991).

## RESULTS

### An E Box in the MLC Enhancer Acts as a Negative Element

The MLC enhancer core presents a high degree of conservation between mouse, rat, and human over a region that has been previously shown to be important for muscle-specific expression both in tissue culture (Rosenthal, 1989; Rosenthal *et al.*, 1990; Wentworth *et al.*, 1991) and in transgenic mice (Rosenthal *et al.*, 1989). Multiple conserved regions in the MLC enhancer core constitute targets for specific myogenic factors (Fig. 1A) including members of the MyoD (sites A, B, and C) and MEF2 (A/T rich site)

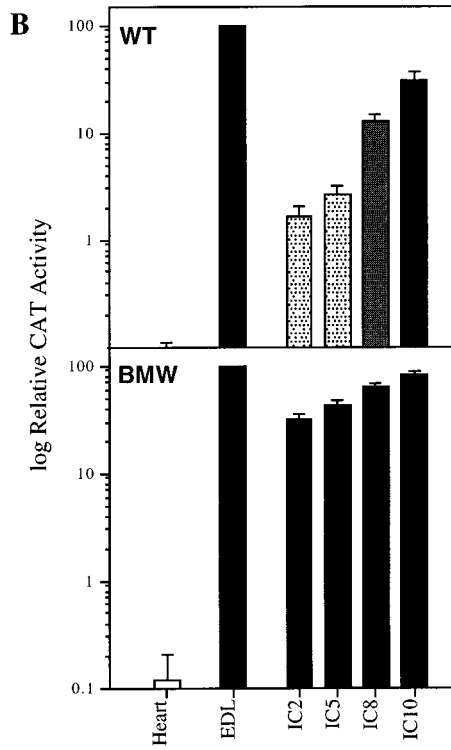
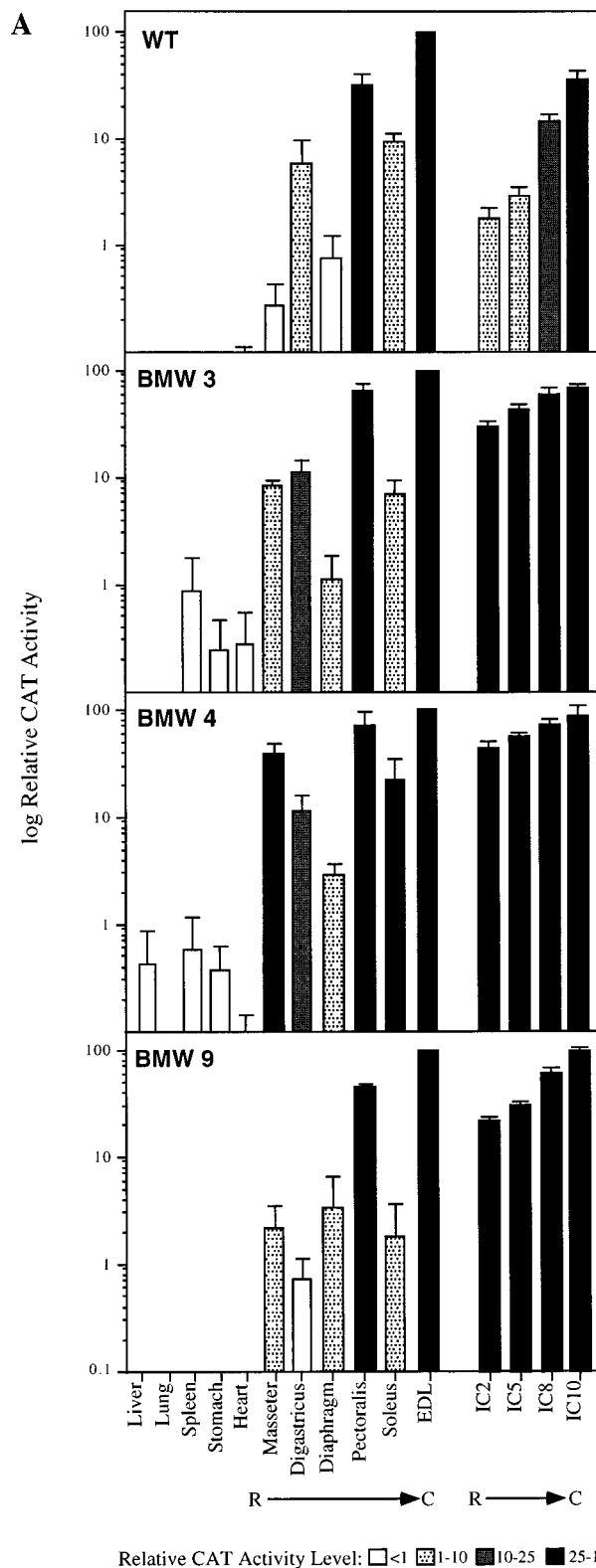
families. In addition, a pair of consensus Hox protein binding sites (CTTATTAAATTA) (Rao *et al.*, 1996) and an adjacent E box (CATGTG), termed BMW, also show a high degree of conservation (Fig. 1A), suggesting a possible function in MLC enhancer activity.

To investigate whether the BMW box was involved in enhancer-mediated transcriptional activation, an MLC enhancer variant in which the BMW motif was altered by oligonucleotide-directed mutagenesis (Fig. 2A) was tested for its ability to drive transcription from an otherwise unmodified MLC1 promoter-CAT expression cassette in transient transfections of C2C12 muscle and NIH3T3 fibroblast cell lines. In NIH3T3 cells, muscle-specific activity of both wild-type (WT) and mutant (BMW) MLC enhancers was tested by transactivation with a cotransfected MyoD expression vector. As previously shown (Wentworth *et al.*, 1991), MyoD efficiently transactivated wild-type MLC enhancer-driven CAT gene expression. The BMW-mutated enhancer responded to MyoD transactivation as strongly as did the wild-type enhancer in NIH3T3 fibroblasts (Fig. 2B, top). In C2C12 differentiated myotubes, however, the BMW mutation resulted in a twofold increase in CAT activity over that driven by a WT enhancer (Fig. 2B, bottom). Thus, unlike other E boxes in the MLC enhancer, the BMW box is not a functional site for transactivation by MyoD and appears to have a repressive role on the muscle-specific action of the MLC enhancer.

### Mutation of the BMW Box Releases the Gradient of MLC-CAT Transgene Expression in Axial Muscles

To assess the function of the BMW box *in vivo*, where it might affect temporal or spatial activation of linked reporter gene activity, transgenic mice were generated with the MLC-CAT construct carrying the BMW mutation in the downstream enhancer. Five germline-integrated founder lines were established, and CAT transgene activity was determined in protein extracts of muscle and liver tissue from representative animals of the five lines. Two lines did not express the transgene in either tissue. The other three lines (BMW 3, 4, and 9) all expressed high levels of transgene expression and were analyzed for CAT activity in muscle and nonmuscle tissue. As shown in Fig. 3A, mice carrying an MLC-CAT cassette driven by the wild-type MLC enhancer expressed the transgene in a complex pattern. As previously described (Donoghue *et al.*, 1991; Grieshammer *et al.*, 1992), transgene expression was low in the masseter and diaphragm and graded in the intercostal muscles, with very low expression levels in liver, lung, spleen, stomach, and heart. In comparison, mice carrying an

**FIG. 2.** (A) Structure of the WT and BMW MLC-CAT expression cassette used in transfection experiments and to generate transgenic mice. (B) MLC-CAT activity in nonmuscle (top) and muscle (bottom) cells transfected with wild-type and BMW-mutated MLC-CAT transgenes. In NIH3T3 cells, MLC enhancer activity was tested by transactivation with a cotransfected MyoD expression vector. Value for each bar is the average of at least three experiments (with standard error).





MLC-CAT cassette driven by the BMW mutant enhancer retained muscle-specific transgene expression, but at a consistently higher level, specifically in the masseter and diaphragm. Moreover, the BMW mutation largely alleviated CAT transgene repression in the rostral intercostal muscles, which expressed 10- to 15-fold higher levels of transgene activity compared with the wild-type transgene (average values from all BMW lines is shown in Fig. 3B). Thus the BMW box acts as a repressor of MLC transgene expression, most notably in anterior axial muscle groups.

### **Positionally Restricted Methylation of Transgenic MLC Regulatory Sequences Is Relieved by Mutation of the BMW E Box**

Multiple lines of evidence from this and other laboratories (Donoghue *et al.*, 1992b; Grieshammer *et al.*, 1995) have shown that positional differences in MLC-CAT transgene expression, both in intact muscles and in primary transgenic muscle cell cultures, cannot be attributed to differences in transcriptional competence, but rather reflect the extent of CpG methylation in both MLC1 promoter and MLC enhancer sequences. Taken together, these studies implicate selective methylation in the maintenance of different gene expression patterns in skeletal muscles, and establish the downstream enhancer of the MLC1/3 locus as a target for epigenetic modification.

To determine whether increases in MLC-CAT transgene expression elicited by mutation of the BMW E box correlated with changes in the methylation status of transgene sequences, we examined methylation of selected CpG residues in the transgenic MLC1 promoter and MLC enhancer, by a quantitative PCR-based analysis (Grieshammer *et al.*, 1995; McGrew and Rosenthal, 1993) of DNA from the muscles of wild-type and BMW mutant lines. As seen in Fig. 4, mutation of the BMW E box released the rostrocaudal gradient of MLC enhancer methylation in intercostal muscles, and resulted in demethylation of the enhancer in selected muscles such as the diaphragm. Unexpectedly, the effect of the BMW mutation was more pronounced on the transgenic MLC1 promoter, which was significantly less methylated in rostral intercostal muscles, as well as in the diaphragm and masseter. The changes in methylation of CpGs in the proximal enhancer and distal promoter resulting from the BMW mutation correlated well with corresponding changes in expression of the MLC-CAT transgene (Fig. 3), supporting a role for methylation in the maintenance

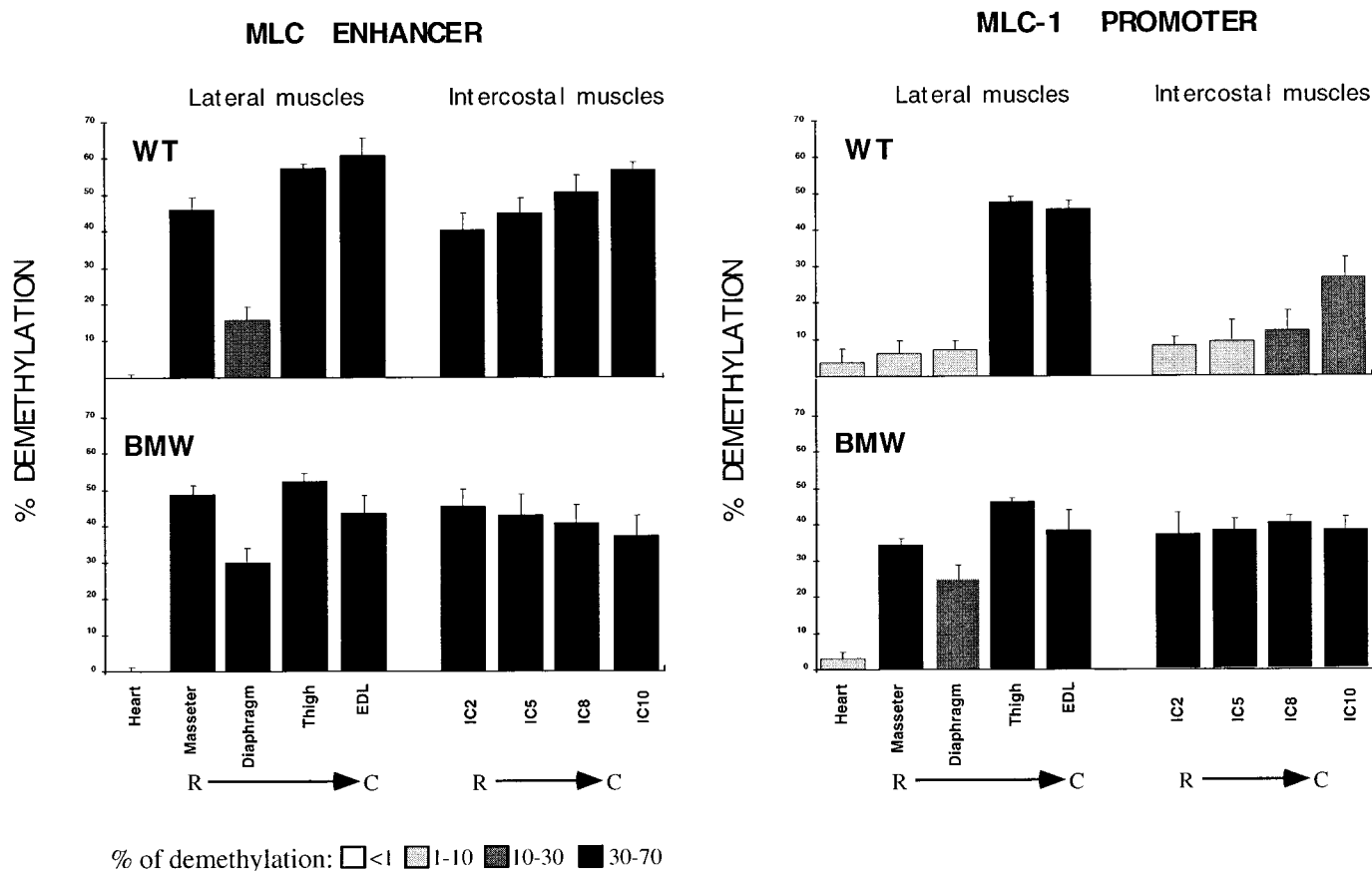
of positional differences in MLC-CAT transgene transcription.

### **The BMW Box Is Recognized by a Nuclear Protein Complex but Not by MyoD**

The results described above implicate the BMW E box, and the proteins that bind to it, in the rostrocaudal gradient of MLC-CAT transgene expression. Bacterially expressed GST-MyoD fusion protein binds specifically to other E boxes (A, B, and C) in the enhancer (Wentworth *et al.*, 1991), although MyoD-containing complexes derived from cell extracts do not recognize the B box (Rosenthal *et al.*, 1990), presumably because it does not represent an optimal binding site for MyoD-E protein heterodimers (Huang *et al.*, 1996). The A, B, and C motifs appear to be the only bona fide MyoD binding sites in the enhancer, since a 173-bp enhancer fragment in which all three sites were mutagenized did not form a complex with GST-MyoD (Fig. 5A, lane 2). As seen in Fig. 5B (lane 1), the GST-MyoD fusion protein did not bind a 20-bp oligonucleotide probe containing the BMW site. Thus, the upstream BMW box is not a target for MyoD *in vitro*.

Since previous DNase protection analysis of the MLC enhancer with a HeLa cell extract revealed a footprint over the BMW site (B. M. Wentworth and N.R., unpublished observations) gel shift assays with nuclear extracts from HeLa cells were performed to identify nuclear protein complexes that interact with this sequence. As seen in Fig. 5B, a protein-DNA complex was observed with HeLa nuclear extracts using the BMW site as probe (lane 4). The specificity of this interaction was confirmed by competition experiments with a 100-fold molar excess of unlabeled BMW oligonucleotide or an unrelated AT-rich oligonucleotide constituting the MEF2 site in the MLC enhancer (Fig. 5B, lanes 5 and 6). To further demonstrate the binding specificity of the HeLa protein complex with the BMW site, competition experiments were carried out using an excess of unlabeled oligonucleotides constituting the E boxes in BMW, A or C sites (Fig. 5C, lanes 2-4). Only the BMW E box was able to compete the binding of the HeLa nuclear extract, while the E boxes in sites A and C could not. The same sequence specificity of complex binding was observed in gel shift assays using C2C12 nuclear extracts (data not shown and see below). These results demonstrate that the BMW E box in the MLC enhancer is recognized specifically by a ubiquitous nuclear protein complex, termed pBMW, that is different from that formed over sites A and C.

**FIG. 3.** Wild-type and BMW MLC-CAT transgene expression in adult skeletal muscle. (A) CAT assays were performed on protein extracts from various adult tissues (IC, intercostal muscle; EDL, extensor digitalis longus) isolated from three independently generated BMW MLC-CAT transgenic lines (BMW Nos. 3, 4, 9) and a wild-type MLC-CAT transgenic line (No. 7) (Grieshammer *et al.*, 1992). (B) Expression in the rostral intercostal muscle of three independent BMW-mutated MLC-CAT transgenic lines compared with wild type (WT, No. 7). Values (with standard error) from three to six independent experiments were averaged and normalized to the expression of CAT in the EDL (set to 100). Note the log scale of relative CAT activity. The shading density reflects the relative difference in CAT activity (see bottom of the figure). The relative position of the muscle is indicated at the bottom: R, rostral; C, caudal.



**FIG. 4.** Methylation of promoter and enhancer sequences in wild-type (WT) and BMW-mutated MLC-CAT transgenes. Demethylation in the MLC enhancer (*HhaI* site) and in the MLC1 promoter (*ThaI* site) analyzed by LMPCR. Genomic DNA was isolated from adult tissues (IC, intercostal muscle; EDL, extensor digitalis longus) in wild-type MLC-CAT and multiple BMW MLC-CAT lines described in Fig. 3. Percentage demethylation was calculated as previously described (Grieshammer *et al.*, 1995) from densitometric scans of labeled LMPCR products analyzed by gel electrophoresis and autoradiography. Values (with standard error) from at least three independent experiments were averaged. The gray shading reflects the relative levels of demethylation in each muscle group. The relative position of the muscles is indicated at the bottom: R, rostral; C, caudal.

### Mapping pBMW Binding Sites within the BMW Box

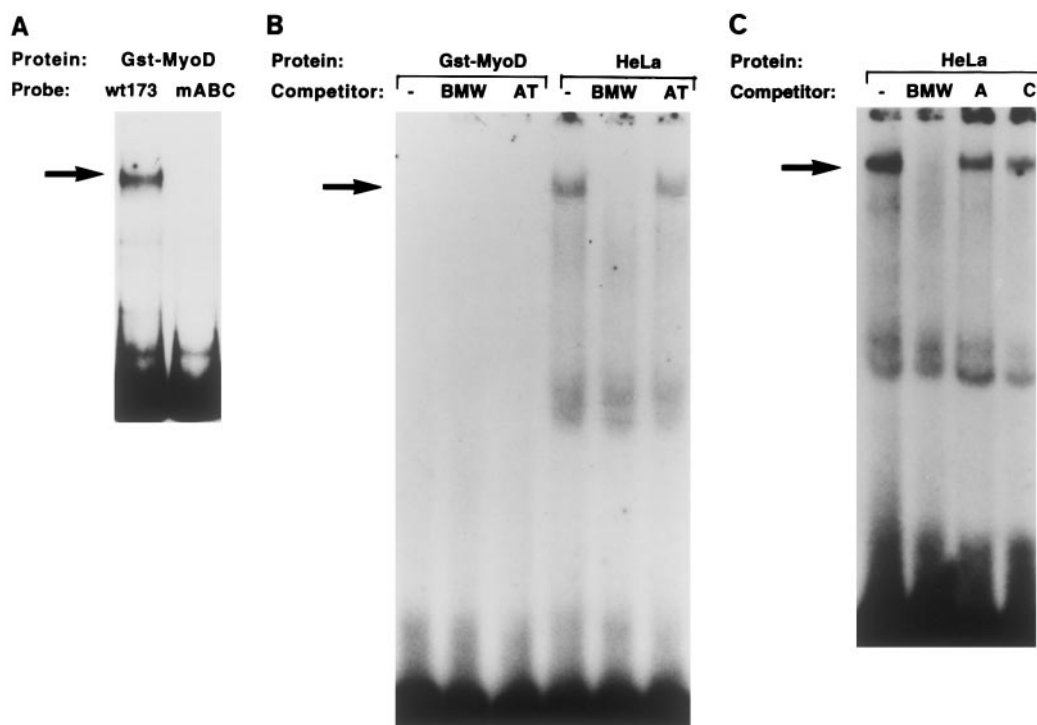
The specific nucleotides in the MLC enhancer to which pBMW binds were mapped by methylation interference assays on an MLC enhancer fragment carrying mutations in sites A, B, and C. Comparison of the methylation pattern seen on the protein-bound DNA sample (Fig. 6A, lane B) with that on free DNA (Fig. 6A, lane F) established positions of protein-DNA interaction. As shown in Fig. 6A, pBMW contacted two G residues on the antisense strand at the BMW box. To further define the bases necessary for pBMW binding, we generated a mutated oligonucleotide of 40 bp (mBMW) containing the adjoining Hox and BMW sites in the MLC enhancer. Mutations were introduced in the two G residues shown to be important for the binding of pBMW by methylation interference, and in two adjacent bases of

the E box element. EMSAs were carried out using a 40-bp wild-type BMW probe and either wild type or the mBMW oligonucleotide. As shown in Fig. 6B, the BMW complex in C2C12 nuclear extracts was not competed by the unlabeled mBMW oligonucleotide, even at 300-fold molar excess (Fig. 6B, lane 7), while a 50-fold molar excess of unlabeled BMW oligonucleotide was sufficient to compete for the binding of this complex (Fig. 6B, lane 2). This analysis confirms that contact nucleotides identified by footprinting analysis of the BMW box are essential for pBMW binding.

### The BMW Protein Complex Is Abundant in Undifferentiated Cell Types

A variety of cell lines were screened for the presence of the pBMW complex by gel shift assays, using the BMW box as probe. As shown in Fig. 7A, nuclear extracts from HeLa cells,





**FIG. 5.** (A) MLC enhancer fragments  $^{32}\text{P}$ -labeled wild type (wt173, lane 1) or mutated at the three MyoD binding sites (mABC, lane 2) were combined with GST-MyoD bacterial fusion protein and analyzed by electrophoretic mobility shift assay. (B) GST-MyoD fusion protein (lanes 1–3) or nuclear extracts from HeLa cells (lanes 4–6) were incubated with a  $^{32}\text{P}$ -labeled 20-bp oligonucleotide containing the BMW site. Competition experiments were performed with 100-fold molar excess of unlabeled BMW oligonucleotide (lanes 2 and 5) or A/T-rich oligonucleotide (lanes 3 and 6). (C) Nuclear extracts from HeLa cells were incubated with the 20-bp BMW probe and competition experiments were carried out using 100-fold molar excess of unlabeled BMW oligonucleotide (lane 2) or oligonucleotides containing either the A box (lane 3) or the C box (lane 4).

NIH 3T3 fibroblasts, C2C12 myoblasts, and myotubes and Jurkat cells (a T-lymphocyte line) formed a protein complex of the same mobility as pBMW, whereas L6 cells, 10T1/2, and foreskin fibroblasts (FS4) did not (data not shown). The relatively low level of protein binding in extracts from C2C12 myotubes relative to C2C12 myoblasts (compare lanes 4 and 3 in Fig. 7A) suggests that pBMW might be more abundant in undifferentiated cells, consistent with the negative regulatory activity of the BMW box in differentiated myocytes. Similarly, extracts from undifferentiated 70Z pre-B cells formed a sequence-specific complex with the BMW probe (Fig. 7B, lanes 1–4), while the same complex was absent in nuclear extract from 70Z cells differentiated into mature B cells by LPS treatment (Fig. 7B, lane 5). Taken together, these results suggest a general inhibitory function for the pBMW complex in regulating differentiation-specific gene expression.

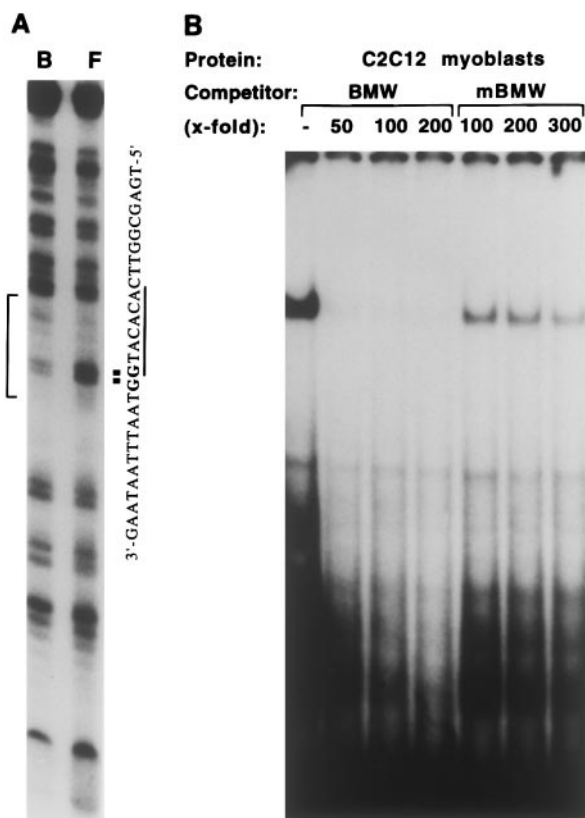
## DISCUSSION

### *The BMW E Box Is a Positional Sensor*

The discovery that transgenic reporters driven by regulatory elements from the MLC1/3 locus faithfully mark

the positional address in segmented axial muscles by graded expression along the AP axis (Donoghue *et al.*, 1991; Grieshammer *et al.*, 1992) provided an experimental model to investigate the molecular basis of regional identity in these tissues. This avenue of investigation has now led to the identification of a single *cis*-acting element in the MLC enhancer that responds to regional regulators in the segmented axial muscles. Mutation of the BMW E box in this element releases the rostral restriction of linked transgene expression and alters the methylation of both proximal and distal regulatory sequences in the transgene.

Molecular studies of axial specification in invertebrates have indicated that skeletal muscles within individual body segments are specified by common pathways involved in establishing segmental identity in other cell types. The establishment of the embryonic body plan is controlled at least in part by the homeodomain gene family, which regulates downstream gene expression programs in spatially and temporally restricted patterns to specify the body axes. Although a cell-autonomous role for homeobox genes in the positional identity of *Drosophila* muscles has been demonstrated (Michelson,



**FIG. 6.** (A) Methylation interference footprinting of HeLa nuclear extract bound to the MLC enhancer mutated at sites A, B, and C. The mABC enhancer fragment was labeled on the anti-sense strand and binding to nuclear extracts was assayed by methylation interference on bound (lane B) and free (lane F) DNA. The footprint pattern is shown as well as the BMW site (underline) near the F lane. (B) Nuclear extracts from C2C12 cells were incubated with a  $^{32}\text{P}$ -labeled 40-bp oligonucleotide containing the HOX and BMW conserved regions of the MLC enhancer core. Competition experiments were performed with 50 (lane 2)-, 100 (lane 3)-, and 200 (lane 4)-fold molar excess of unlabeled 40-mer BMW oligonucleotide or with 100 (lane 5)-, 200 (lane 6)-, and 300 (lane 7)-fold molar excess of unlabeled mBMW mutant oligonucleotide.

1994), identification of similar mechanisms conferring positional specificity on segmented muscles in vertebrates has been elusive, partly because it is more difficult to score morphological differences in muscles associated with different vertebral segments than in the segmented musculature of flies (reviewed in Olson and Rosenthal, 1994). By identifying a positional sensor in vertebrate muscle capable of modulating the accessibility of both proximal and distal regulatory DNA sequences to the transcriptional apparatus, this study provides a molecular model of regional regulation in skeletal muscle whose mechanistic details can now be further dissected.

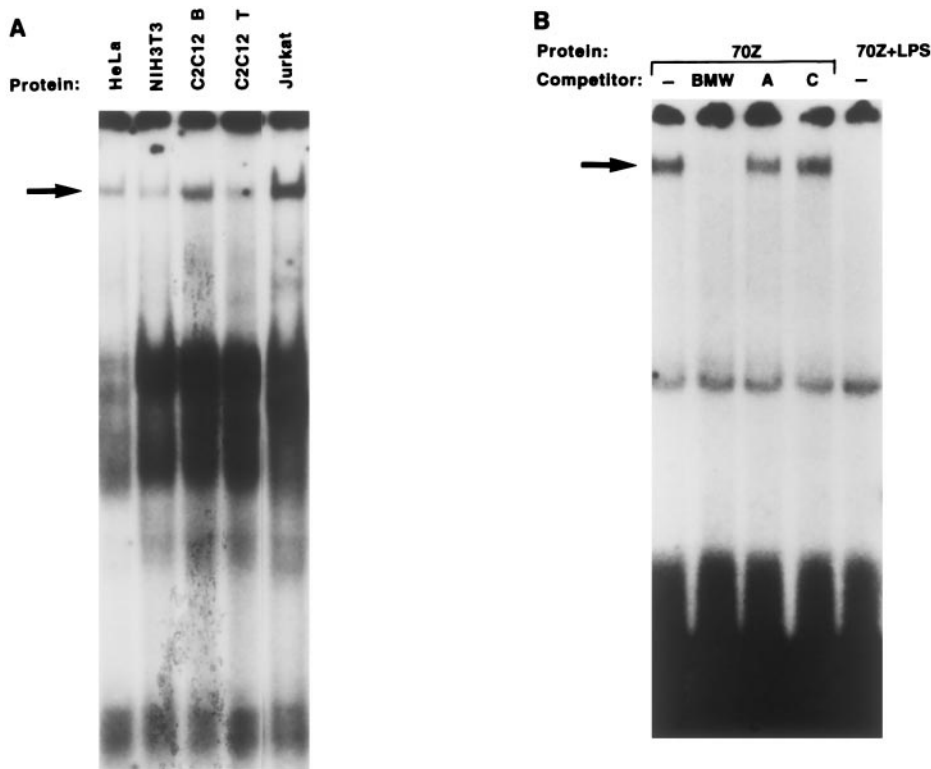
### Role of Myogenic Factors in MLC-CAT Transgene Expression

The characterization of proteins binding to the BMW box suggests that it is not responsive to the MyoD family of bHLH myogenic factors. Pathways distinct from those responsible for the myogenic phenotype may therefore be responsible for conferring positional identity on the mammalian axial musculature. Gene knockout phenotypes indicate that at least some of these factors, such as myogenin, are critical for establishment of differentiation-specific gene expression programs, while others, such as myf5, specify skeletal muscle lineages in the early embryo (Arnold and Braun, 1996). None of the myogenic bHLH factors characterized to date are likely to confer positional information in the embryo, although they are clearly involved in establishing muscle cell-specific patterns of gene expression, operating through other E boxes in the MLC enhancer. It is therefore not surprising that a DNA regulatory element that participates in positional sensing during embryonic development should interact with factors other than those belonging to the MyoD family.

### The pBMW Complex and Repression of Gene Activity

The pBMW nuclear protein complex defined in this study is potentially responsible for establishing positional information through its interaction with an E box, a sequence motif more commonly present in muscle control elements as a target for myogenic bHLH factors. The possibility remains that one of the unknown proteins in the pBMW complex is actually a bHLH factor acting in concert with other factors, such as Zbu1 and Hox proteins, which are known to bind the BMW element (Gong *et al.*, 1997; Rao *et al.*, 1996). The BMW sequence (CATGTG) diverges from the optimal sequence for myogenic factors, binding as heterodimers with E proteins (CAGGTG), although the CANNTG core of the BMW box is identical to that of the MCK enhancer Left-E box, which is similarly adjacent to an A/T-rich site. In contrast to the BMW box, mutations of the Left-E box lead to decreased expression in skeletal muscle cells (Amacher *et al.*, 1993). Notably, multimerized MCK Left-E boxes are more active in NIH3T3 cells and in myoblasts than multimerized MCK Right-E boxes which contain the higher-affinity MyoD binding site (Apone and Hauschka, 1995).

Other sites to which the BMW box bears resemblance lie in the regulatory elements of B-lymphocyte-specific genes. A search of the kappa light chain locus for sequences sharing homology to the BMW box revealed conserved BMW-like motifs present in the kappa light chain promoter (Boyd *et al.*, 1986) and enhancer elements (Picard and Schaffner, 1984) (Fig. 1B). The pBMW complex is present in the pre-B cell line 70Z, but absent from 70Z cells treated with LPS (see Fig. 7), which elicits cellular responses leading to the expression of B-cell-specific genes such as the kappa light chain immunoglobulin. It is therefore possible



**FIG. 7.** (A) Mobility shift assays were performed with nuclear extracts from HeLa cells (lane 1), NIH3T3 cells (lane 2), C2C12 myoblasts (lane 3), myotubes (lane 4), and Jurkat cells (lane 5).  $^{32}$ P-labeled 20-mer BMW oligonucleotide was combined with the extracts and analyzed by gel electrophoresis. (B) Nuclear extracts from the pre-B 70Z cells and mature B 70Z cells [after treatment with lipopolysaccharide (LPS)] were incubated with the 20-bp BMW probe (lanes 1 and 5, respectively), and to assess the specificity of the complex formed by the 70Z nuclear extracts, competition experiments were carried out using 100-fold molar excess of unlabeled BMW oligonucleotide (lane 2) or oligonucleotides containing either the A box (lane 3) or the C box (lane 4).

that a variety of differentiation-specific genes might be targets for the pBMW complex, which may participate in their repression, in either undifferentiated cells or, postmitotically, in nonpermissive tissue types.

Although the identity of each protein in the pBMW complex has yet to be determined, previous studies indicate that proteins from the Hox class of regulatory factors bind directly to the BMW site and/or the adjoining AT-rich motifs (Rao *et al.*, 1996; Gong *et al.*, 1997). These proteins might be acting directly to specify positionally restricted expression patterns on structural genes, as suggested by the requirement for one of these proteins, Hoxc-8, in maximal MLC-CAT transgene expression in selected muscles (Houghton and Rosenthal, 1999). This would be an unorthodox role for clustered Hox genes, which have been considered master regulators of the vertebrate body plan. Hox gene products are unexpectedly abundant in neonatal and adult muscle, however (Houghton and Rosenthal, 1999), and it remains to be seen whether they play a secondary role in regulating structural genes in these tissues during development or later in regeneration, perhaps through interaction with intermediary factors that control

gene accessibility. Another protein previously characterized by its affinity for the BMW site is Zbu1 (Gong *et al.*, 1997), a novel DNA-binding member of the SWI2/SNF2 family of transcription cofactors which have been implicated in a number of DNA transactions such as chromatin remodeling and transcription (Tsukiyama and Wu, 1997). Originally isolated using the BMW box as a DNA target, Zbu1 may have a more restricted DNA specificity in the context of the pBMW complex. The uniformly high accumulation of Zbu1 protein in postmitotic tissues such as skeletal muscle, heart, and brain (S. Kaushal and N. R. unpublished results) and the increased MLC-CAT transgene expression resulting from mutation of its target BMW box in the MLC enhancer (Fig. 3) suggest that Zbu1 may be collaborating with other factors in the pBMW complex to specify regional differences in gene accessibility in differentiated tissues. Since Zbu1 itself is expressed relatively late in embryonic development, and its transcripts are not graded in axial muscles, it might play an ancillary role, possibly through chromatin remodeling, to establish positional information in concert with other proteins that are themselves differentially distributed.

### **Control of MLC1 Promoter Methylation by a Distal Regulatory Element**

The most dramatic effect of the BMW mutation on MLC-CAT transgene methylation is on the MLC1 promoter, which would be situated several thousand bases away from the MLC enhancer in either direction among tandem arrays of MLC-CAT transcription units. The ability of the BMW box to effect such distal changes in epigenetic DNA modification is similar to the action of other known methylation regulators, specifically the one between the H19 and IGF2 genes (Bartolomei and Tilghman, 1997; Surani, 1998; Webber *et al.*, 1998), where a short sequence element has been identified as a regulator of parental imprinting through selective methylation (Birger *et al.*, 1999). In the endogenous MLC1/3 locus, control of MLC1 promoter activity by methylation appears to be site specific, since the internal MLC3 promoter is ubiquitously demethylated (McGrew *et al.*, 1996). In contrast, control of MLC1 promoter activity by methylation is tissue specific, since nonmuscle tissues, such as the heart, do not exhibit any changes in the highly methylated state of MLC-CAT transgene sequences carrying the BMW mutation (Fig. 4).

Although the relative chromatin configuration of transgenic wild-type and mutant MLC sequences has yet to be analyzed, a general cause-and-effect relationship between DNA methylation and the regulation of gene accessibility has been suggested by the interaction of the methyl CpG-binding protein MeCP2 and histone deacetylase activity, leading to acetylation of histones and transcriptionally repressive chromatin (Eden *et al.*, 1998; Jones *et al.*, 1998; Nan *et al.*, 1998). Recent demonstration of the synergy between demethylation and histone deacetylase inhibition in the reactivation of gene expression in cancer (Cameron *et al.*, 1999) establishes a dynamic interaction between the two processes, and further demonstrates the dominance of DNA methylation in the stable maintenance of gene repression. Extrapolating these concepts to differentiation-specific gene regulation would suggest that the release of MLC-CAT transgene repression in selected skeletal muscles by mutation of the BMW box is causally related to the demethylation of promoter and enhancer sequences seen in the transgene.

### **Specification of Muscle Positional Identity by Gene Accessibility**

What is the function of the BMW box in the endogenous MLC1/3 locus, where other regulatory sites with similar protein binding capacities may exist? Although the steady-state levels of transcripts from the endogenous MLC1/3 locus are not graded, increased activity of the BMW-mutated MLC enhancer in muscle cell transfections suggests a repressive role. The variable effect of the BMW mutation on MLC-CAT transgene activity in different muscle groups *in vivo* argues for a more complex role in maintaining regional distinctions of structural gene expression. Variations in gene accessibility could be achieved

through the recognition of regulatory elements such as the BMW box by specific DNA-binding protein complexes, followed by local DNA methylation which may lock genes into a silenced chromatin configuration (Cameron *et al.*, 1999). Thus the fortuitous phenomenon of graded MLC-CAT transgene expression provides a model for the molecular analysis of regional gene regulation. Although endogenous muscle-specific genes regulated in this pattern have yet to be identified, it follows that the initial abundance or composition of the pBMW complex in different tissues may determine the relative accessibility of specific genes to transcriptional activation, perhaps through controlling methylation patterns. From a survey of cell types where the pBMW complex is present it appears that the complex may be active while cells are still in a proliferative state. The identification of a molecular target for positional restriction of transcriptional activity through epigenetic mechanisms should hasten progress in elucidating general processes underlying regional regulation of tissue-specific genes.

### **ACKNOWLEDGMENTS**

We are grateful to Bruce M. Wentworth for contributing unpublished footprinting data on the eponymous BMW box; to Natalia Bogdanova, Serafima Zaltsman, and Esfir Slonimsky for invaluable assistance with CAT assays and genetic analysis of transgenic mice; and to members of the Rosenthal laboratory for comments and discussion. E.C. dedicates this paper entirely to Billie. E.C. was supported by a grant from Telethon-Italia Organization. S.H.H. was supported by the National Cancer Institute, DHHS, under Contract N01-CO-74101 with ABL. This work was supported by grants to N.R. from the National Institutes of Health (R01AR41926) and by a sponsored research agreement from Bristol Myers-Squibb.

### **REFERENCES**

- Amacher, S. L., Buskin, J. N., and Hauschka, S. D. (1993). Multiple regulatory elements contribute differentially to muscle creatine kinase enhancer activity in skeletal and cardiac muscle. *Mol. Cell. Biol.* **13**, 2753-2764.
- Apone, S., and Hauschka, S. D. (1995). Muscle gene E-box control elements. *J. Biol. Chem.* **270**, 21420-21427.
- Arnold, H. H., and Braun, T. (1996). Targeted inactivation of myogenic factor genes reveals their role during mouse myogenesis: A review. *Int. J. Dev. Biol.* **40**, 345-353.
- Bartolomei, M. S., and Tilghman, S. M. (1997). Genomic imprinting in mammals. *Annu. Rev. Genet.* **31**, 493-525.
- Birger, Y., Shemer, R., Perk, J., and Razin, A. (1999). The imprinting box of the mouse Igf2r gene [In Process Citation]. *Nature* **397**, 84-88.
- Blau, H. M., Pavlath, G. K., Hardeman, E. C., Chiu, C. P., Silberstein, L., Webster, S. G., Miller, S. C., and Webster, C. (1985). Plasticity of the differentiated state. *Science* **230**, 758-766.
- Boyd, R. T., Goldrick, M. M., and Gottlieb, P. D. (1986). Structural differences in a single gene encoding the V kappa Ser group of light chains explain the existence of two mouse light-chain genetic markers. *Proc. Natl. Acad. Sci. USA* **83**, 9134-9138.



- Cameron, E. E., Bachman, K. E., Myohanen, S., Herman, J. G., and Baylin, S. B. (1999). Synergy of demethylation and histone deacetylase inhibition in the reexpression of genes silenced in cancer [In Process Citation]. *Nature Genet.* **21**, 103–107.
- Davis, R. L., Weintraub, H., and Lassar, A. B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987–1000.
- Dignam, J. D., Martin, P. L., Shastry, B. S., and Roeder, R. G. (1983). Eukaryotic gene transcription with purified components. *Methods Enzymol.* **101**, 582–598.
- Donoghue, M. J., Merlie, J. P., Rosenthal, N., and Sanes, J. R. (1991). Rostrocaudal gradient of transgene expression in adult skeletal muscle. *Proc. Natl. Acad. Sci. USA* **88**, 5847–5851.
- Donoghue, M. J., Morris-Valero, R., Johnson, Y. R., Merlie, J. P., and Sanes, J. R. (1992a). Mammalian muscle cells bear a cell-autonomous, heritable memory of their rostrocaudal position [see Comments]. *Cell* **69**, 67–77.
- Donoghue, M. J., Patton, B. L., Sanes, J. R., and Merlie, J. P. (1992b). An axial gradient of transgene methylation in murine skeletal muscle: Genomic imprint of rostrocaudal position. *Development* **116**, 1101–1112.
- Eden, S., Hashimshony, T., Keshet, I., Cedar, H., and Thorne, A. W. (1998). DNA methylation models histone acetylation [Letter] [In Process Citation]. *Nature* **394**, 842.
- Gong, X., Kaushal, S., Ceccarelli, E., Bogdanova, N., Neville, C., Nguyen, T., Clark, H., Khatib, Z. A., Valentine, M., Look, A. T., and Rosenthal, N. (1997). Developmental regulation of Zbu1, a DNA-binding member of the SWI2/SNF2 family. *Dev. Biol.* **183**, 166–182.
- Gorman, C. M., Moffat, L. F., and Howard, B. H. (1982). Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* **2**, 1044–1051.
- Graham, F. L., Abrahams, P. J., Mulder, C., Heijneker, H. L., Warnaar, S. O., De Vries, F. A., Fiers, W., and Van Der Eb, A. J. (1975). Studies on in vitro transformation by DNA and DNA fragments of human adenoviruses and simian virus 40. *Cold Spring Harbor Symp. Quant. Biol.* **39**(Pt. 1), 637–650.
- Grienshammer, U., Sassoon, D., and Rosenthal, N. (1992). A transgene target for positional regulators marks early rostrocaudal specification of myogenic lineages. *Cell* **69**, 79–93.
- Grienshammer, U., Sassoon, D., and Rosenthal, N. (1993). Positional specification during muscle development. In "Molecular Basis of Morphogenesis," pp. 177–187.
- Grienshammer, U., McGrew, M. J., and Rosenthal, N. (1995). Role of methylation in maintenance of positionally restricted transgene expression in developing muscle. *Development* **121**, 2245–2253.
- Hogan, B., Constantini, F., and Lacy, E. (1986). "Manipulating the Mouse Embryo: A Laboratory Manual." Cold Spring Harbor Laboratory, New York.
- Houghton, L., and Rosenthal, N. (1999). Regulation of a muscle-specific transgene by persistent expression of Hox genes in post-natal murine muscle. *Dev. Dynamics*, in press.
- Huang, J., Blackwell, T. K., Kedes, L., and Weintraub, H. (1996). Differences between MyoD DNA binding and activation site requirements revealed by functional random sequence selection. *Mol. Cell. Biol.* **16**, 3893–3900.
- Jones, P. L., Veenstra, G. J., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J., and Wolffe, A. P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genet.* **19**, 187–191.
- Libri, D., Stutz, F., McCarthy, T., and Rosbash, M. (1995). RNA structural patterns and splicing: Molecular basis for an RNA-based enhancer. *RNA* **1**, 425–436.
- McGrew, M. J., and Rosenthal, N. (1993). Quantitation of genomic methylation using ligation-mediated PCR. *Biotechniques* **15**, 722–729.
- McGrew, M. J., Bogdanova, N., Hasegawa, K., Hughes, S. H., Kitsis, R. N., and Rosenthal, N. (1996). Distinct gene expression patterns in skeletal and cardiac muscle are dependent on common regulatory sequences in the MLC1/3 locus. *Mol. Cell. Biol.* **16**, 4524–4534.
- Michelson, A. M. (1994). Muscle pattern diversification in *Drosophila* is determined by the autonomous function of homeotic genes in the embryonic mesoderm. *Development* **120**, 755–768.
- Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex [see Comments]. *Nature* **393**, 386–389.
- Olson, E. N., and Rosenthal, N. (1994). Homeobox genes and muscle patterning. *Cell* **79**, 9–12.
- Picard, D., and Schaffner, W. (1984). A lymphocyte-specific enhancer in the mouse immunoglobulin kappa gene. *Nature* **307**, 80–82.
- Rao, M. V., Donoghue, M. J., Merlie, J. P., and Sanes, J. R. (1996). Distinct regulatory elements control muscle-specific, fiber-type-selective, and axially graded expression of a myosin light-chain gene in transgenic mice. *Mol. Cell. Biol.* **16**, 3909–3922.
- Rosenthal, N. (1987). In "Methods in Enzymology" (S. L. Berger and A. R. Kimmel, Eds.), pp. 704–720. Academic Press, Boston.
- Rosenthal, N. (1989). Muscle cell differentiation. *Curr. Opin. Cell Biol.* **1**, 1094–1101.
- Rosenthal, N., Kornhauser, J. M., Donoghue, M., Rosen, K. M., and Merlie, J. P. (1989). Myosin light chain enhancer activates muscle-specific, developmentally regulated gene expression in transgenic mice. *Proc. Natl. Acad. Sci. USA* **86**, 7780–7784.
- Rosenthal, N., Berglund, E. B., Wentworth, B. M., Donoghue, M., Winter, B., Bober, E., Braun, T., and Arnold, H. H. (1990). A highly conserved enhancer downstream of the human MLC1/3 locus is a target for multiple myogenic determination factors. *Nucleic Acids Res.* **18**, 6239–6246.
- Seed, B., and Sheen, H. H. (1988). A simple phase extraction assay for chloramphenicol acetyltransferase activity. *Gene* **67**, 271–277.
- Surani, M. A. (1998). Imprinting and the initiation of gene silencing in the germ line. *Cell* **93**, 309–312.
- Tsukiyama, T., and Wu, C. (1997). Chromatin remodeling and transcription. *Curr. Opin. Genet. Dev.* **7**, 182–191.
- Webber, A. L., Ingram, R. S., Levorse, J. M., and Tilghman, S. M. (1998). Location of enhancers is essential for the imprinting of H19 and Igf2 genes. *Nature* **391**, 711–715.
- Wentworth, B. M., Donoghue, M., Engert, J. C., Berglund, E. B., and Rosenthal, N. (1991). Paired MyoD-binding sites regulate myosin light chain gene expression. *Proc. Natl. Acad. Sci. USA* **88**, 1242–1246.

Received March 10, 1999

Revised April 21, 1999

Accepted April 21, 1999