In Vivo Footprinting of a Muscle Specific Enhancer by Ligation Mediated PCR

PAUL R. MUELLER AND BARBARA WOLD

In vivo protein-DNA interactions at the developmentally regulated enhancer of the mouse muscle creatine kinase (MCK) gene were examined by a newly developed polymerase chain reaction (PCR) footprinting procedure. This ligation mediated, single-sided PCR technique permits the exponential amplification of an entire sequence ladder. Several footprints were detected in terminally differentiated muscle cells where the MCK gene is actively transcribed. None were observed in myogenic cells prior to differentiation or in nonmuscle cells. Two footprints appear to correspond to sites that can bind the myogenic regulator MyoD1 in vitro, whereas two others represent muscle specific use of apparently general factors. Because MyoD1 is synthesized by undifferentiated myoblasts, these data imply that additional regulatory mechanisms must restrict the interaction between this protein and its target site prior to differentiation.

THE DEVELOPMENTAL PROGRESSION FROM MESODERMAL precursor cell to determined, proliferating myoblast and from myoblast to postmitotic, differentiated muscle cell (myocyte) involves a cascade of regulatory changes. The determination step that produces myoblasts occurs when the developmental potential of a precursor is restricted to the myogenic lineage. A significant inroad to understanding this developmental decision comes from the recent identification of a family of genes whose ectopic expression in cultured cells can recruit otherwise nonmyogenic cells to function as myoblasts. Several of these genes have been cloned, including MyoD1 (*1*), myogenin (*2*), and Myf5 (*3*). Their products are nuclear proteins, and all share some sequence similarity with other important regulatory molecules. These include the *myc* family of onco-proteins (*4*) and the immunoglobulin enhancer binding proteins (E12 and E47) (*5*) in mammals, and the protein products of the *daughterless*, *twist*, and *achaete-scute* genes (*6*) in *Drosophila*. Myoblasts can, in turn, be triggered to undergo differentiation in cell culture by altering the signals provided by growth factors, hormones, and extracellular matrix components. Differentiation of skeletal muscle ultimately includes cell cycle withdrawal, transcriptional activation of muscle specific genes, assembly of muscle structures, and cell fusion to produce multinucleated myotubes.

P. R. Mueller and B. Wold are in the Division of Biology at the California Institute of Technology, 156-29, Pasadena, CA 91125.

Transcriptional regulation of myoblast and myocyte specific genes is central to the execution of this developmental pathway, but knowledge of how this regulation is achieved is limited. For example, recent studies suggest that MyoD1 can act as a positive transcriptional regulator by binding to sites in several myocyte specific genes (*7*). However, MyoD1 is also expressed in proliferating myoblasts where it can positively regulate expression of its own promoter, but where myocyte specific genes are transcriptionally silent (*8*). This raises the question of how MyoD1 and its relatives act differently in myoblasts versus myocytes. Moreover, genes expressed specifically in myocytes bind some factors in vitro that are widely distributed in nonmuscle cells and myoblasts as well as in myocytes (*9–11*). Whether these factors interact in the cell with myocyte specific genes when they are transcriptionally silent is not yet known.

In vivo footprinting can answer some of these questions by providing information on when and how proteins occupy a given regulatory region of DNA in the living cell. These experiments can be especially useful when taken together with genetic characterization of cis-acting elements and in vitro DNA binding studies of the relevant factors. Sequence inspection, reverse genetic analysis, and gel retardation experiments have identified a number of potentially important cis-acting sequence elements in the upstream enhancer of muscle creatine kinase (MCK) (*9, 12, 13*). Here, we have used genomic footprinting to examine protein-DNA interactions at this region in cells that express MCK (myocytes) and in cells that do not (nonmyogenic cells and myoblasts).

Despite the information provided by in vivo footprinting and the development of several genomic sequencing strategies for this purpose (*14, 15*), application of these techniques has been limited. In vivo footprinting of a single-copy regulatory region in large genomes (mammals) by established strategies is technically challenging. Large cell numbers are required, and experiments often have an unacceptable signal-to-noise ratio. Our genomic footprinting method largely eliminates these problems and enables the in vivo footprinting of relatively small numbers of cells (about 10^5 nuclei) or dissected tissues. It is based on exponential amplification of a genomic sequence ladder by the polymerase chain reaction (PCR) (*16*). Although presented as a footprinting technique, this method should be generally applicable to any PCR problem in which only one end of the region to be amplified is known. For example, Pfeifer *et al.* (*17*) have adapted it to the study of in vivo methylation patterns and genomic sequencing.

Footprinting with ligation mediated, single-sided PCR. In vivo footprints are visualized by comparing samples of DNA that have been exposed to nucleases or alkylating agents in the cell (in vivo) with samples exposed to these agents after the DNA has been

extracted from cells and deproteinized (in vitro or naked) (*18*). Dimethyl sulfate (DMS) is commonly used as the alkylating agent because cell membranes are freely and rapidly permeable to it (*18*). Proteins bound to DNA often alter the accessibility of DMS to guanines at or near the binding site (*19*). After purification of the DNA, both in vitro and in vivo DMS-treated samples are quantitatively cleaved at the methylated guanine residues with piperidine (*20*) and then compared to reveal the footprint.

The PCR consists of repeated cycles of template denaturation, primer annealing, and DNA polymerase extension to exponentially amplify a segment of DNA located between two primers. Each cycle doubles the number of templates, and after 25 to 30 cycles a single-copy gene can be amplified more than 10^6-fold (*16*). Conventional PCR is not immediately applicable to sequencing or footprinting because it requires two defined ends. A sequence or footprint ladder is composed of a population of related nucleic acid fragments. One end of each fragment is fixed by a primer or restriction cut and is therefore the same for all, whereas the other end is determined by variable chemical cleavage or chain termination and is therefore unique for each fragment. To apply PCR to a sequence ladder, we have introduced a simple ligation step that adds a common oligonucleotide sequence to the unique end of each member. A primer complementary to this new common sequence is then used, together with a primer complementary to the original fixed end, for simultaneous exponential amplification of all members of the sequence ladder. The procedure has high selectivity and specificity that are derived from the design of the ligation step and the choice of primers (Fig. 1). It also has high fidelity; a footprint consists of subtle differences in the starting concentrations of particular members of a sequence ladder, and these differences are reproducibly retained through the amplification.

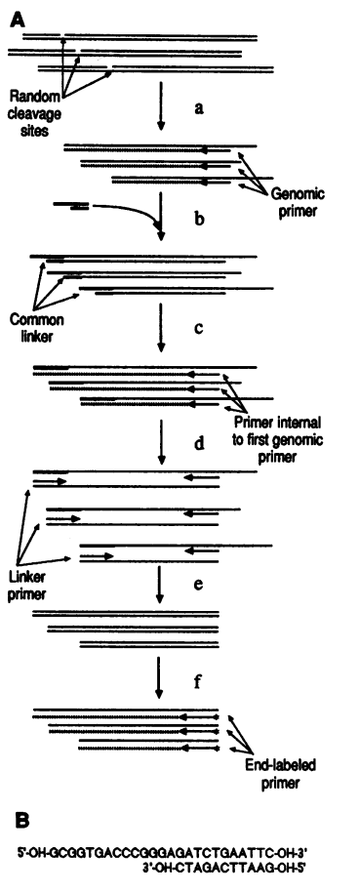

Fig. 1. Schematic of ligation mediated PCR footprinting or sequencing. (**A**) Starting material is genomic DNA that has been treated with Maxam and Gilbert sequencing chemistry (*20*) . This leaves 5′ and 3′ phosphates (*20*). The first step (a) defines the fixed end of the sequence ladder by denaturing the cleaved genomic DNA and annealing a gene specific primer. Extension (dotted line) of this primer to the variable cleavage site generates a family of blunt-ended duplex molecules, which are substrates for T4 DNA ligase–catalyzed addition (b) of a specially designed linker (heavy lines; see below) to each member of the sequence ladder, thereby providing each with a common, defined end. The genomic DNA provides the 5′ phosphate used in the ligation. Specificity is provided by the fact that the bulk of the genomic DNA does not serve as a substrate in the ligation reaction because it lacks a blunt double-stranded end. The DNA is then denatured (c), and a second gene specific primer is annealed and extended. To increase specificity, the second gene specific primer is positioned so that its extending end is 3′ to that of the first primer. The original genomic DNA is used again as a template in this reaction, only now it has the longer strand of the linker covalently attached to it, and the extension product reads through this added sequence. Each member of the sequence ladder now has two defined ends (the common linker primer and the second gene specific primer), and is suitable for PCR (*16*). After 16 rounds of PCR (d and e), the sequence ladder is amplified ~10^4-fold. It is visualized by primer extension of an end-labeled, third primer (f). Its appearance is that of the corresponding sequence ladder, except that it is uniformly longer by the additional length of the linker. The third primer should overlap the second primer, and also be positioned so that its extending end is 3′ to that of the second primer. The sequence ladder can also be visualized by filter blotting (*17*). (**B**) Structure of the common linker. The linker shown is but one example of the possible sequences that could be used. It is important that: (i) the linker contains no 5′ phosphates and is staggered to eliminate self-ligation and assure directionality in ligation, respectively; (ii) the duplex between the long and short oligomers is stable at ligation conditions, but not at PCR temperatures; and (iii) the longer oligomer should have a comparable T_m (melting temperature) to that of the second gene-specific primer (see above).

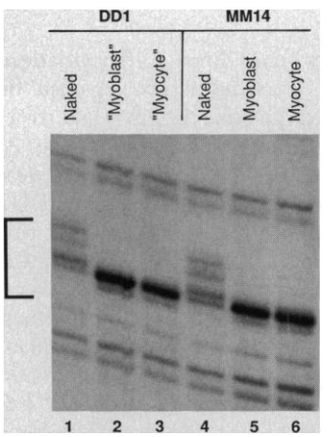

Fig. 2. In vivo footprint of metallothionein-I promoter visualized by ligation mediated, single-sided PCR. A footprint is ~185 bp upstream of transcription start on the coding strand of the MT-I promoter (*15*) in both DD1 (lanes 1 to 3) and MM14 (lanes 4 to 6) cells. The Sp1 consensus site is bracketed on left. Naked DNA's are genomic control samples from DD1 or MM14 cells that were treated with DMS in vitro. Myoblast and myocyte DNA's are from the same cells grown under proliferation or differentiation conditions, respectively, and treated with DMS in vivo. In vivo labels for DD1 cells are in quotations because these cells are differentiation defective and therefore do not form true myoblasts or myocytes. Cell culture and DNA preparations as in Figs. 3 and 4. All primers and oligomers were gel purified. For first strand synthesis, 3 μg of DMS-piperidine treated DNA (*15*) and 0.3 pmol of primer 1 (*44*) were suspended in 15 μl of 40 m*M* tris, *p*H 7.7, 50 m*M* NaCl. The sample was heated at 95°C for 2 minutes and then incubated at 60°C for 30 minutes. Hybridization was stopped by transferring to ice; a solution of 7.5 μl of 20 m*M* $MgCl_2$, 20 m*M* dithiothreitol (DTT), and 0.02 m*M* of each deoxynucleoside triphosphate (dNTP) was added, then 1.5 μl of a 1:4 dilution of Sequenase version 1.0 (USB) [diluted in 10 m*M* tris (*p*H 7.5), 1 m*M* EDTA] was added, and the sample was incubated at 47°C for 5 minutes. The reaction was stopped by heating at 60°C for 5 minutes, then adding 6 μl of 310 m*M* tris (*p*H 7.7), and then heating for 10 minutes at 67°C. For ligation of linker, the sample was transferred to ice, and a solution of 20 μl of 17.5 m*M* $MgCl_2$, 42.3 m*M* DTT, and BSA at 125 μg/ml was added, then 25 μl of ligation mixture [10 m*M* $MgCl_2$, 20 m*M* DDT, 3 m*M* ATP, BSA at 50 μg/ml, with 5 μl of PCR linker mix (20 pmol of linker per microliter in 250 m*M* tris, *p*H 7.7) and 3 Weiss units of T4 ligase per 25 μl] was added. The linker was prepared as described (*45*). After incubation overnight at 15°C, the reaction was stopped by heating to 70°C for 10 minutes. The sample was precipitated in the presence of 10 μg of yeast carrier tRNA. For the PCR reaction; the precipitated samples were washed once with 75 percent ethanol and then suspended in water. 20 μl of 5× Taq buffer (200 m*M* NaCl, 25 m*M* tris *p*H 8.9, 25 m*M* $MgCl_2$, 0.05 percent w/v gelatin) was added along with 20 nmol of each dNTP, 10 pmol of a primer 2 (*44*), 10 pmol of the longer oligomer of the linker (Fig. 1B), and 5 units of Taq polymerase (Cetus). The volume was adjusted to 100 μl with H_2O. Samples were covered with 90 μl of mineral oil, heated to 94°C for 1 minute, and then manually cycled (denatured for 1 minute at 94°C, hybridized for 2 minutes at 63°C, and extended for 3 minutes at 76°C) 16 times. Samples were placed on ice, 1 to 5 pmol of an end-labeled (*46*) primer 3 (*44*) was added, along with 2.5 units of Taq polymerase and 20 nmol of each dNTP. Samples were heated to 94°C for 2 minutes, hybridized at 66°C for 2 minutes, and extended at 76°C for 10 minutes. Polymerase activity was stopped by chilling on ice, adding 295 μl of 260 m*M* sodium acetate, 10 m*M* tris *p*H 7.5, and 4 m*M* EDTA, and, extracting with a mixture of 50 parts phenol, 50 parts chloroform, and 1 part isoamyl alcohol. The samples were precipitated, resuspended in loading dye, and half of each sample was placed on each lane of a standard sequencing gel (*20*).

We tested this PCR footprinting technique on the mouse metallothionein I (MT-I) promoter, which has been well characterized both in vivo and in vitro. Previously, we had used cells containing more than 100 copies of the MT-I promoter to observe in vivo interactions and found a prominent footprint at the upstream Sp1 binding site (Sp1-A) (*15*). Using PCR footprinting, we reexamined this region in MM14 and DD1 cells that contain only a single copy of the MT-I gene per haploid genome. The expected footprint is apparent in both MM14 and DD1 lines, as indicated by comparison of the naked DNA control sample (Fig. 2, lanes 1 and 4) with the in vivo DNA sample from cells grown under either proliferation (lanes 2 and 5) or differentiation (lanes 3 and 6) conditions. This result illustrates the sensitivity of the technique; the data shown are from a 9-hour, screened exposure on Kodak XAR-P film. These footprints correlate with the observed basal expression of MT-I in these cells (Fig. 3), and establish that the data obtained from ligation mediated PCR footprinting are consistent with data from more conventional methods.

Expression of muscle regulatory genes during differentiation. Differentiation of myoblasts into myocytes is accompanied by complex changes in the expression of muscle regulatory and structural gene products as well as some housekeeping genes (Fig. 3) (*21*). Levels of RNA's relevant to this footprinting study were examined in MM14 cells, a permanent myogenic cell line (*22*), DD1 cells, a differentiation defective derivative of these (*23*), and two other myogenic lines, aza-myoblasts and BC_3H1 cells. Under differentiation conditions, there was a decrease in expression of proliferation related genes, such as *c-myc*, and general housekeeping genes, such as MT-I, in both myogenic and differentiation defective lines. Myocyte specific gene products such as MCK (Fig. 3) and myosin heavy chain (*24*) increased during differentiation in MM14 but remained absent from DD1 cells. Two of the three known myogenic regulators are expressed in the MM14 cells; MyoD1 remains constant before and after differentiation, whereas myogenin is activated upon differentiation. The third regulator, Myf5, could not be detected in MM14 cells, and all three are absent in DD1 cells.

In vivo footprinting of the muscle creatine kinase enhancer. MCK expression is activated during muscle differentiation (Fig. 3) (*25*). The MCK transcriptional control elements that have been identified include an intronic enhancer region, a proximal promoter region, and an upstream enhancer region (*12, 13, 26*). We have focused on the upstream enhancer region because it confers high-level stage and tissue specific expression on a reporter gene in both cell culture (*12, 13, 27*) and transgenic mice (*28*). This enhancer is located about 1 kb upstream of the transcription start and contains sequence motifs similar to recognition sites of several putative general and muscle specific transcription factors (see below). Four of these potential binding sites are occupied in MM14 myocytes (Fig. 4), though none are detectably occupied in MM14 myoblasts or DD1 cells. Multiple, independent DNA preparations from each cell type were tested, and all interactions were highly reproducible (summarizd in Fig. 5).

A myocyte specific in vivo footprint is near the upstream end of an adenine-rich sequence (CTAAAAATAACCC) located at −1077 (Fig. 4, lane 3). The A-T–rich character of this sequence prevents us from observing additional interactions in this region employing the DMS-guanine reaction. Factors from extracts of both muscle and nonmuscle cells bind to this sequence in vitro (*9*), suggesting that the in vivo interaction observed here results from myocyte specific use of factors present in many other cell types, including myoblasts. Evidence that this A-rich sequence may be acting as a positive, myocyte specific regulatory element comes from studies of the chicken myosin light chain 2-A (MLC2-A) promoter (*10*) and the rat MCK enhancer (*11*). In both cases, genetic and biochemical studies show that it contributes to muscle specific expression and binds one or more proteins present in extracts both from muscle and nonmuscle cells.

The enhancer contains an exact match to the in vitro binding site of transcription factor AP-2 (*29, 30*). Although AP-2 is not a myocyte specific factor (*29*), the in vivo footprint is restricted to differentiated MM14 cells (Fig. 4, lane 9, and Fig. 5). The protections observed are partial, an indication that the site may be occupied only part of the time, on average, or that the nature of the physical interaction between protein and DNA only partially occludes DMS accessibility. In either case, the reproducibility of this observation in multiple experiments indicates that it is a genuine footprint. There were no footprints at this site in proliferating MM14 myoblasts, differentiation defective DD1 cells (Fig. 4, lanes 8 and 11 and 12), Balb/c 3T3 fibroblasts, or L cells (*24*). This myocyte specific pattern contrasts with the in vitro interactions detected with gel-mobility shift assays, in which factors from both muscle and nonmuscle cell extracts bind in this region (*9*).

The AP-2 site and the adenine-rich element flank a 110–base pair (bp) central core that, by itself, retains most of the activity of the upstream enhancer (*9*). Two sequences similar to elements found in the immunoglobulin kappa (κ) (*31*) and heavy (H) chain (*32*) cellular enhancers are present in the core, and both are critical for the activity of the murine MCK enhancer. Deletion of the κ chain enhancer–like sequence results in an ~10-fold decrease in enhancer activity (*33*), and mutation of the H chain enhancer–like sequence results in decrease of about 25-fold (*9*). Buskin and Hauschka have reported a myocyte specific binding activity, MEF-I (myocyte enhancing factor one) (*9*), that interacts with the H chain enhancer–like sequence in vitro. We have therefore identified the H chain enhancer–like site as MEF-I in Figs. 4 and 5. In vivo footprints where found at both the κ chain enhancer–like and MEF-I (H chain enhancer–like) sites in MM14 myocytes, but not in the other cell types tested (Fig. 4, compare lanes 3 and 9 to lanes 2, 5, 6, 8, 11, and 12).

Based on similarity to cis-acting sequence motifs, it has been suggested that the CarG and *sphI* elements present in the MCK

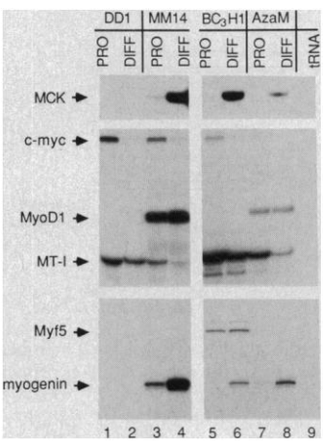

Fig. 3. (lower right) RNA analysis of muscle and nonmuscle specific genes. Total RNA's from MM14 (*22*) (lanes 3 and 4), BC_3H1 (*47*) (lanes 5 and 6), aza-myoblasts (*48*) (AzaM, lanes 7 and 8), and DD1 (*23*) (lanes 1 and 2), were examined. Lane 9 contains yeast tRNA (30 μg). Cells were grown under proliferation (PRO) or differentiation (DIFF) conditions. Each left (DD1 and MM14) and right (BC_3H1 and AzaM) set of panels are from the same gel. The weak band above myogenin in lane 4 is independent of Myf5. The top set of panels shows the extension of 2 μg of RNA with an MCK primer to create a 72-nt fragment. The middle set of panels shows RNase protection of 10 μg of RNA with *c-myc*, MyoD1, and MT-I probes simultaneously to create 159-, 92-, and 67-nt fragments, respectively. The bottom set of panels shows RNase protection of 6 μg of RNA with Myf5 and myogenin probes simultaneously to create 197- and 175-nt fragments, respectively. The DD1 and MM14 RNA's were purified by the guanidinium-CsCl method (*15*). Cells were immunostained with myosin heavy chain antibodies (*24*) to determine the percentage of cells that had differentiated (*23*). There were no detectable myocytes in the DD1 cultures in either type of media. MM14 cells in proliferation and differentiation media were 6 and 86 percent myosin positive, respectively. The procedure for the MCK primer extension was as described (*28*) and for RNase protection was as described (*15*).

enhancer may be important for its function (*12, 13*). In the context of other muscle specific genes, the CArG element has been shown to be important for expression (*34*). The *sphI* element has been shown to be important for the activity of the SV40 enhancer (*35*). However, neither of these sites in the MCK enhancer were detectably occupied in any of the cell lines we tested (Fig. 5) (*24*).

With the exception of MM14 myocytes, the cell types tested had no convincing protein-DNA interactions at the MCK enhancer. There was, however, a reproducible difference between in vivo and in vitro DMS treated DNA samples. It consists of two in vivo hypersensitivities at adenines −1154 and −1152, just upstream of the MEF-I site (Fig. 4, lanes 8 and 9 and 11 and 12; Fig. 5) (*24*). Although the piperidine cleavage reaction used favors strand scission at alkylated guanines, it may reveal, with reduced sensitivity, adenine residues that are particularly reactive with DMS. Hypersensitivity to DMS alkylation may be caused by torsional strain on the DNA or by proteins closely interacting with DNA to create local hydrophobic pockets (*19*). We now favor the former possibility, in part because there are no associated protected residues that, in our experience, tend to be a better general indicator of protein-DNA interactions. If this does represent protein binding, it differs from the other footprints observed because it occurs in all cell types and does not change when MCK is expressed.

In vivo footprinting of DDI, the differentiation defective derivative of MM14, permitted us to ask whether any interactions observed in fully differentiated myocytes are actually dependent on the switch from growth factor–rich proliferation medium to growth factor–poor differentiation medium, irrespective of differentiation itself. In particular, it seemed possible that in vivo binding by putative general factors like AP-2 might be regulated in response to growth signals without requiring overt muscle differentiation. Although this would be an elegant mechanism for linking withdrawal from the cell cycle with expression of differentiation specific genes, we found no support for this possibility. DD1 cells displayed no in vivo footprints under either culture condition (Fig. 4, lanes 4 to 6, and 10 to 12).

Regulation of DNA binding activities during myogenesis. All in vivo protein-DNA interactions detected at the MCK enhancer were confined to differentiated myocytes in which MCK is actively transcribed. Although the correlation of in vivo footprints with gene

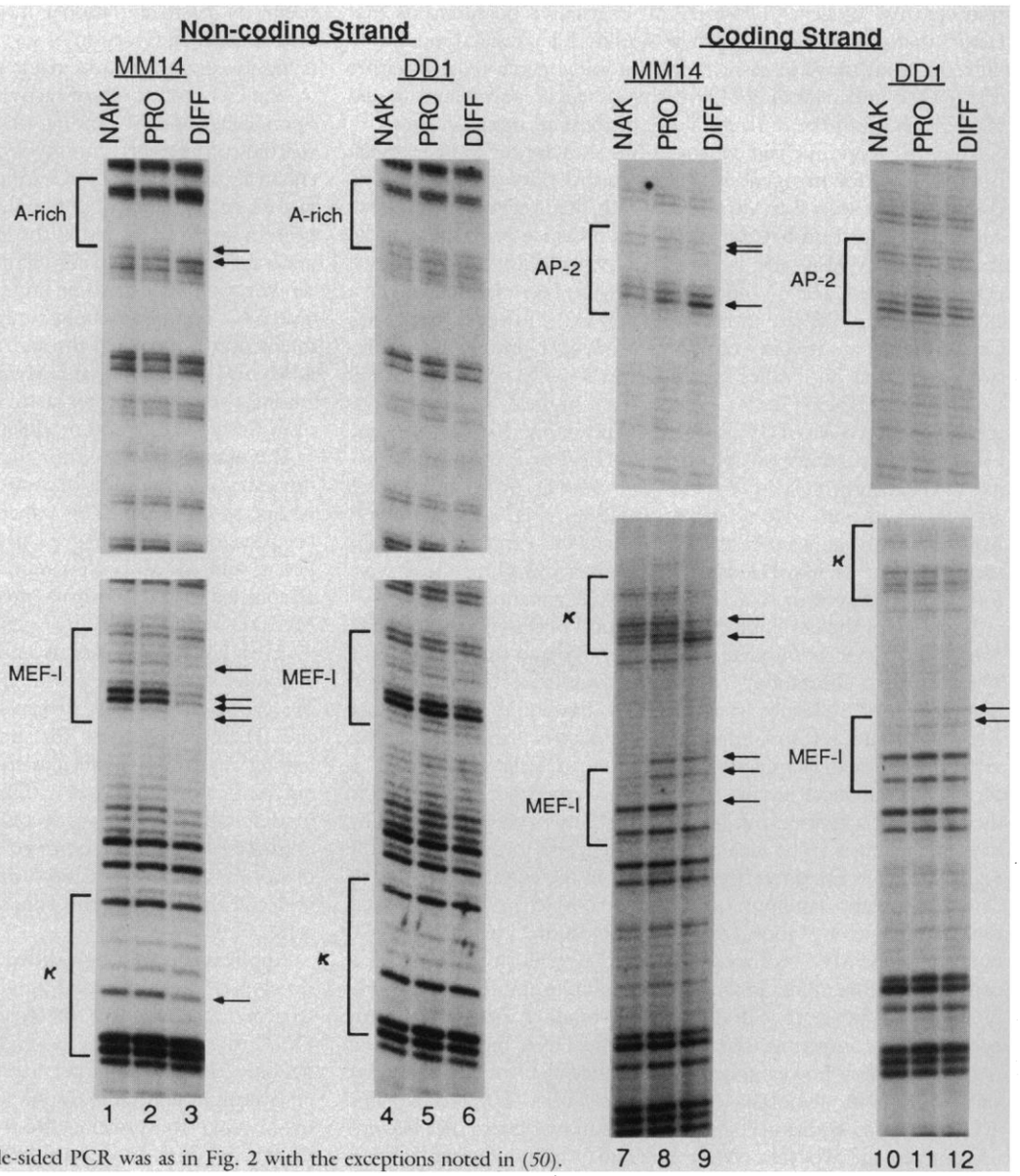

Fig. 4. In vivo footprinting of MCK enhancer in MM14 muscle cells and a differentiation defective derivative, DD1 cells. Noncoding and coding strands were visualized by ligation mediated PCR footprinting (Figs. 1 and 2). Cell lines are labeled at the top of each set of footprint ladders. In vitro DMS treated "naked" DNA (NAK, lanes 1, 4, 7, or 10) are compared to in vivo DMS treated DNA from cells grown in proliferation (PRO, lanes 2, 5, 8, or 11) or differentiation (DIFF, lanes 3, 6, 9, or 12) media. To conserve space, irrelevant parts of the footprint ladder are not shown; they contained no footprints (*24*). Brackets on the left of each footprint ladder identify the location of consensus sequences (see text for identity). Arrows on the right of each ladder mark bases that were consistently protected or hypersensitive in multiple experiments on independent preparations of DNA. These data are summarized in Fig. 5. Occasionally, we have observed spurious fluctuation in the intensity of an individual band. For example, in lane 3 between the MEF-I and κ chain enhancer–like sites where there is an apparent protection, this was seen only in the experiment shown, but not in several others, and is therefore not considered a legitimate footprint. Therefore multiple experiments are necessary; in all cases shown here, multiple experiments confirmed the authenticity of the indicated footprints. MM14 and DD1 cells were grown (*27*) and treated with DMS in vivo (*15*) as described. DNA for in vitro and in vivo DMS treatment was harvested from cells as per (*15*) or by adaptation (*49*) of the RNA isolation procedure used in Fig. 3. In vitro DMS treatment (naked DNA) was as described (*15*) except that 2.5 μl of DMS per 1 ml of DNA solution was used. Piperidine treatment of DNA was as in (*15*). Ligation mediated single-sided PCR was as in Fig. 2 with the exceptions noted in (*50*).

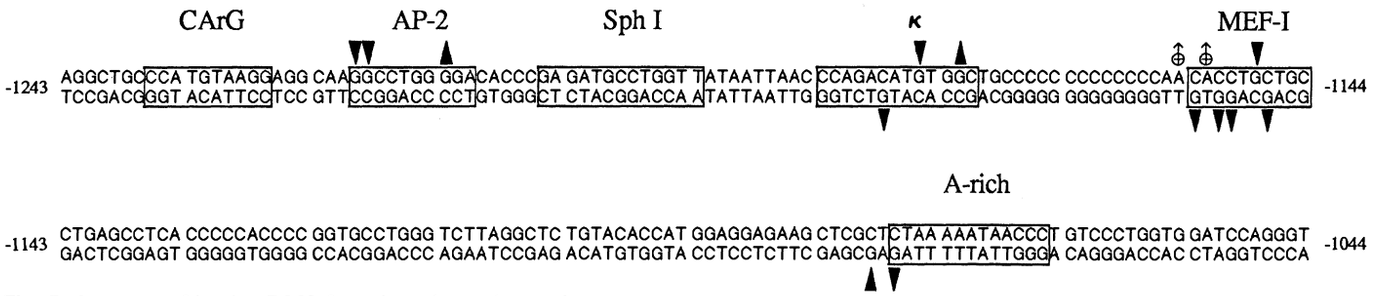

Fig. 5. Summary of in vivo DMS footprints observed over the upstream MCK enhancer. Upstream sequence (*12*) of MCK from −1243 to −1044 is shown. The enhancer as defined by (*12*) is from −1256 to −1050, and the core enhancer as defined by (*9*) is from −1207 to −1097. Sequence similarities to known factor binding sites are boxed. Changes in sensitivity to DMS (Fig. 4) are indicated; all interactions were reproducible. Identification of protected and hypersensitive bases was made by counting bases on overexposed gels and alignment with marker DNA ladders. Hypersensitivities (▲) and protections (▼) were observed in MM14 myocytes only, with the exception of the two adenine hypersensitivities (δ) at −1153 and −1155, which were observed in all cell types.

activity is direct, their relation to specific in vitro binding activities from various cell types is not so straightforward (Fig. 6). For example, even though the A-rich and AP-2–like sites are occupied only in myocytes in vivo, they can be occupied in vitro by factors present in many cell types (*9*). This suggests developmentally restricted use of general factors. An alternative possibility is that factors that ultimately occupy these sites in differentiated muscle are different from those in extracts from nonmyogenic cells. In either case, a clear understanding of how this enhancer is regulated should take into account the differential use of these recognition sites.

The MEF-I and κ chain enhancer–like sites appear to be recognition elements for myogenesis-specific factors. However, a detailed comparison of several in vitro assays with the in vivo data suggests additional regulation beyond the simple presence or absence of the factors. The MEF-I site is bound in vitro only when myocyte extracts are used, and no binding activity has been detected at the κ chain enhancer–like site in these extracts (*9*). However, both sites can be bound in vitro by recombinant MyoD1 protein (*7*). Moreover, polyclonal antibodies raised against MyoD1 recognize MEF-I–DNA complexes (*7, 36*). These data suggest a similarity or identity between MyoD1 and MEF-I, and imply that MyoD1 may be, at least in part, responsible for the MEF-I or κ chain enhancer–like in vivo footprints (Fig. 6). However, MyoD1 RNA and protein are present in both myoblasts and myocytes (Fig. 3) (*8*), whereas MEF-I activity appears restricted to myocyte extracts. Thus, the mere presence of MyoD1 is not sufficient to produce the in vivo footprints observed or to activate the MCK enhancer.

A model for the developmental regulation of this myocyte specific enhancer must accommodate both the in vivo and the in vitro data. There are two substantially different possibilities. One of these is that access of sequence specific DNA binding proteins to the recognition sites is restricted until differentiation is triggered. In this view, the failure of factors present in nonmuscle cells or myoblasts to act on this enhancer in vivo would be governed not by changes in their intrinsic activities, but by the availability of the MCK binding sites in chromatin. The idea of restricted access to developmentally regulated genes has often been discussed in the context of open and closed chromatin configurations (*37*) or covalent modifications of DNA such as methylation (*38*). Our data on the absence of in vivo binding to the MCK enhancer by factors present in all cell types is consistent with a model in which accessibility is limited. Accessibility may also be restricted if potential binding sites are already occupied by competing, sequence specific DNA binding proteins (*39*). We did not, however, detect occupancy of any myocyte specific sites in vivo in myoblasts or nonmuscle cells, nor did we find evidence of any additional sites occupied in myoblasts that become unoccupied in myocytes. While there may be undetected interactions, the complete absence of footprints in myoblasts and nonmuscle cells is striking.

An alternative possibility is that the intrinsic activity of the enhancer binding factors themselves is regulated, rather than the accessibility of their binding sites. This regulation of activity could occur by post-translational modification or via interplay among factors. If the interaction of several factors with the MCK enhancer is highly cooperative, the triggering event may be a modest change in the concentration or activity of only one of these factors. Specifically, MEF-I binding activity appears to be completely restricted to differentiating myocytes (*9*), and may therefore regulate enhancer activity by nucleating cooperative binding of general factors. MEF-I itself might be subject to further regulation. Protein-protein interactions within the *myc*-MyoD1 regulatory family could generate different binding affinities and specificities. For example, dimerization appears to be important for activity of the related E12 and E47 immunoglobulin enhancer binding proteins (*5*). This interaction is mediated through the structural similarity also found in MyoD1, myogenin, and Myf5, and by analogy, such interactions among the myogenic regulators could provide an elegant mechanism for activating and modulating muscle specific enhancers.

The picture presented for the MCK enhancer in MM14 skeletal myocytes cannot fully account for its known activity in other myogenic cell types. This enhancer can drive muscle specific transcription of a reporter gene in cardiac muscle where Myf5, myogenin, and MyoD1 expression is absent (*28*). In addition, some established MCK positive myogenic cell lines do not express MyoD1, but do express myogenin or Myf5 (Fig. 3, lanes 5 to 8) (*2, 3*). How MCK expression is activated in these MyoD1 negative cells is not yet known. One possibility is that more than one member of the *myc*-MyoD1 regulatory family could interact at the same binding site. There is precedent for such overlapping binding specificities among regulatory molecules in other systems (*40*). Alternatively, different recognition sites within the enhancer segment may be used in different types of myocytes. Comparison of the MM14 case presented here with studies of similar design in MyoD1 negative myocytes should reveal how different combinations of regulatory molecules can activate this enhancer in related but distinct myogenic cells.

Applications of single-sided ligation mediated PCR. We have developed a ligation-based, single-sided PCR strategy and applied it to genomic sequencing. Its sensitivity is high. From 1 μg of cellular DNA, the sequence of a single copy segment of a mammalian genome (about 3.3 pg per haploid genome) can be obtained from an overnight autoradiographic exposure. This opens the possibility of in vivo footprinting from small cell numbers and specific dissected tissues. In view of the high sensitivity and specificity of the

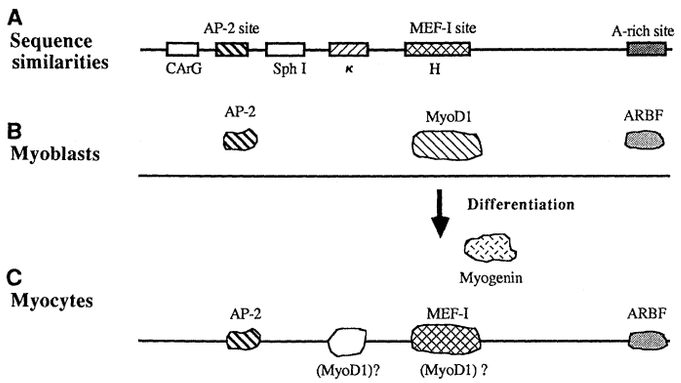

Fig. 6. Schematic representation of interactions between protein and DNA at the MCK enhancer. See text for discussion. (**A**) Location of sequence motifs similar to those of known binding factors. (**B**) Myoblasts contain factors that can bind the MCK enhancer in vitro, but no interactions are observed in vivo. (**C**) On differentiation, the MCK enhancer is occupied at four of these sequence similarities. In addition, myogenin (*2*), a factor with substantial protein similarities to MyoD1 is expressed. ARBF; A rich binding factor.

technique, we expect that the lower limit of the number of nuclei per sample will now be governed by statistical considerations, which are quite complex for this procedure (*17*). In principle, if the number of founder molecules representing a given member of the sequence ladder before amplification is too low, fluctuation among samples could give artifactual variation in the intensity of the corresponding band at the end of the amplification procedure. This might be mistaken for a legitimate footprint. In practice, we have not worked with cell numbers below 3×10^5 per reaction. We have empirically determined that under these conditions multiple analyses of the same sequence ladder are generally free of detectable under- or over-representation of individual bands. For example, the similar intensity of bands in all DNA samples from DD1 cells illustrates typical reproducibility (Fig. 4, lanes 4 to 6 and 10 to 12). The few spurious variations observed were resolved by comparing several independent experiments, and only protections or hypersensitivities that appear in each experiment are identified as footprints in Figs. 4 and 5.

A second limitation on in vivo footprinting is heterogeneity of the starting cell population. If the cells are not uniform with respect to expression of the gene of interest, footprints may be obscured by background from physiologically distinct nuclei. In our experiments, we have studied clonally derived cell lines, and care was taken to achieve physiologic uniformity: cell populations of myoblasts were as free of prematurely differentiated myocytes as possible (less than 6 to 8 percent myocytes), and differentiated myocyte preparations were harvested when very few undifferentiated precursors remained (less than 12 to 14 percent myoblasts), as determined by immunostaining for myosin heavy chain (*24*).

Ligation mediated PCR can be adapted for uses other than in vivo footprinting. For example, it has been used to determine the in vivo methylation pattern for genes subject to differential methylation during development (*17*). It can also be used to clone a new segment of genomic DNA beginning, for instance, with a primer positioned near the 5′ end of a known mRNA sequence to produce a nested series extending into an unknown promoter (*41*). The procedure has been adapted for sequencing all four bases and has been used with multiplexing (*17*), which permits several different sequences to be determined simultaneously (*42*). It may therefore be possible to conduct a genomic sequence walk through a single copy gene, beginning from a region of known sequence and proceeding through new sequence by successive steps of one-sided PCR, thereby entirely bypassing cloning steps.

For applications that require single-base resolution, this method differs significantly from another one-sided PCR strategy reported recently in which terminal transferase is used to add an oligo-dG (deoxyguanylate) or -dA (deoxyadenylate) tail to one end of each substrate molecule (*43*). This homopolymeric tail is expected to be somewhat variable in length, and terminal transferase will make additions to random single-stranded ends as well as to blunt-end duplexes. By contrast, the ligation method adds a uniform, defined sequence to the end of each molecule, and takes advantage of the high specificity of DNA ligase for a blunt-end duplexed substrate. These design features, together with other details of the ligation based procedure, reduce nonspecific background and provide the resolution required for genomic footprinting, sequencing, and methylation studies.

REFERENCES AND NOTES

1. R. L. Davis, H. Weintraub, A. B. Lassar, *Cell* **51**, 987 (1987).
2. W. E. Wright, D. A. Sassoon, V. K. Lin, *ibid.* **56**, 607 (1989); D. G. Edmondson and E. N. Olsen, *Genes Dev.* **3**, 628 (1989).
3. T. Braun, G. Buschhausen-Denker, E. Bober, E. Tannich, H. H. Arnold, *EMBO J.* **8**, 701 (1989).
4. R. A. DePinho, K. S. Hatton A. Tesfaye, G. D. Yancopoulos, F. W. Alt, *Genes Dev.* **1**, 1311 (1987).
5. C. Murre, P. S. McCaw, D. Baltimore, *Cell* **56**, 777 (1989).
6. M. Caudy *et al.*, *ibid.* **55**, 1061 (1988); C. Cronmiller, P. Schedl, T. Y. Cline, *Genes Dev.* **2**, 1666 (1988); B. Thisse, C. Stoetzel, C. Gorostiza-Thisse, F. Perrin-Schmitt, *EMBO J.* **7**, 2175 (1988); R. Villares and C. V. Cabrera, *Cell* **50**, 415 (1987).
7. A. B. Lassar *et al.*, *Cell* **58**, 823 (1989).
8. S. T. Tapscott *et al.*, *Science* **242**, 405 (1988).
9. J. N. Buskin and S. D. Hauschka, *Mol. Cell. Biol.* **9**, 2627 (1989).
10. T. Braun, E. Tannich, G. Buschhausen-Denker, H. H. Arnold, *ibid.*, p. 2513.
11. R. A. Horlick and P. A. Benfield, *ibid.*, p. 2396.
12. J. Jaynes, J. E. Johnson, J. N. Buskin, C. L. Gartside, S. D. Hauschka, *ibid.* **8**, 62 (1988).
13. E. A. Sternberg *et al.*, *ibid.*, p. 2896 (1988).
14. G. M. Church and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1991 (1984); J. M. Huibregtse and D. R. Engelke, *Gene* **44**, 151 (1986); P. D. Jackson and G. Felsenfeld, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2296 (1985).
15. P. R. Mueller, S. J. Salser, B. Wold, *Genes Dev.* **2**, 412 (1988).
16. For reviews, see: R. K. Saiki *et al.*, *Science* **239**, 487 (1988); T. J. White, N. Arnheim, H. A. Erlich, *Trends Gen.* **5**, 185 (1989).
17. G. P. Pfeifer, S. D. Steigerwald, P. R. Mueller, B. Wold, A. D. Riggs, *Science* **246**, 810 (1989).
18. E. Giniger, S. M. Varnum, M. Ptashne, *Cell* **40**, 767 (1985); A. Ephrussi, G. M. Church, S. Tonegawa, W. Gilbert, *Science* **227**, 134 (1985).
19. W. Gilbert, A. Maxam, A. Mirzabekov, in *Control of Ribosome Synthesis*, Alfred Benzon Symposium IX, N. O. Kjeldgaard and O. Maaloe, Eds. (Academic Press, New York, 1976), pp. 139–148; L. Johnsrud, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5314 (1978).
20. A. M. Maxam and W. Gilbert, *Methods Enzymol.* **65**, 499 (1980).
21. For example, see: P. Gunning *et al.*, *Mol. Cell. Biol.* **7**, 4100 (1987).
22. T. A. Lindhart, C. H. Clegg, S. D. Hauschka, *J. Supramol. Struct.* **14**, 483 (1980).
23. R. W. Lim and S. D. Hauschka, *Dev. Biol.* **105**, 48 (1984).
24. P. R. Mueller and B. Wold, unpublished results.
25. J. S. Chamberlain, J. B. Jaynes, S. D. Hauschka, *Mol. Cell. Biol.* **5**, 484 (1985).
26. J. B. Jaynes, J. S. Chamberlain, J. N. Buskin, J. E. Johnson, S. D. Hauschka, *ibid.* **6**, 2855 (1986).
27. J. E. Johnson, C. L. Gartside, J. B. Jaynes, S. D. Hauschka, *Dev. Biol.* **134**, 258 (1989).
28. J. E. Johnson, B. Wold, S. D. Hauschka, *Mol. Cell. Biol.* **9**, 3393 (1989).
29. P. J. Mitchell, C. Wang, R. Tjian, *Cell* **50**, 847 (1987).
30. M. Imagawa, R. Chiu, M. Karin, *ibid.* **51**, 251 (1987).
31. D. Picard and W. Schaffner, *Nature* **307**, 80 (1984).
32. J. Banerji, L. Olson, W. Schaffner, *Cell* **33**, 729 (1983).
33. J. E. Johnson, thesis, University of Washington, Seattle (1989).
34. For review, see A. Taylor, H. P. Erba, G. O. Muscat, L. Kedes, *Genomics* **3**, 323 (1988).
35. I. Davidson *et al.*, *Nature* **323**, 544 (1986).
36. J. Buskin, A. B. Lassar, R. L. Davis, H. Weintraub, S. D. Hauschka, *J. Cell Biol.* **107**, 98a (1988).
37. For review, see H. Weintraub, *Cell* **42**, 705 (1985).
38. For reviews, see A. D. Riggs and P. A. Jones, *Adv. Cancer Res.* **40**, 1 (1983); R. Holliday, *Science* **238**, 163 (1987).
39. K. Zinn and T. Maniatis, *Cell* **45**, 611 (1986).
40. C. Desplan, J. Theis, P. H. O'Farrell, *ibid.* **54**, 1081 (1988); T. Hoey and M. Levine, *Nature* **332**, 858 (1988); J. B. Jaynes and P. H. O'Farrell, *ibid.* **336**, 744 (1988).
41. L. Fors, R. Saavedra, L. Hood, unpublished results.
42. G. M. Church and S. Kieffer-Higgens, *Science* **240**, 185 (1988).
43. M. A. Frohman, M. K. Dush, G. R. Martin, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8998

(1988); E. L. Loh, J. F. Elliott, S. Cwirla, L. L. Lanier, M. M. Davis, *Science* **243**, 217 (1989).
44. MT-I primers: 1, CGGAGTAAGTGAGGAGAAGGTACTC; 2, GGAGAAGGTACTCAGGACGTTGAAG; 3, GAAGTACTCAGGACGTTGAAGTCGTGG.
45. For hybridization, the oligomers of the linker (see Fig. 1B for sequence) were brought to a final concentration of 20 pmol/μl in 250 m*M* tris, *p*H 7.7, heated to 95°C for 5 minutes, transferred to 70°C, and then slowly cooled (~3 hours) to 4°C. The hybridized linker was stored at −20°C and thawed on ice as needed.
46. The labeled oligomer was phosphorylated with crude [γ-^{32}P]ATP (DuPont) and T4 polynucleotide kinase (NEB), and unincorporated ^{32}P was removed by Nensorb-20 columns (DuPont). Specific activity of the primer was 4×10^6 to 9×10^6 cpm/pmol.
47. D. Schubert *et al.*, *J. Cell. Biol.* **61**, 398 (1974).
48. A. B. Lassar, B. M. Paterson, H. Weintraub, *Cell* **47**, 649 (1986).
49. E. Meese and N. Blin, *Gene Anal. Tech.* **4**, 45 (1987).
50. For footprinting the MCK enhancer, we used 2 μg of DNA per reaction. The Sequenase reaction was performed at 45°C with 0.1 m*M* of each dNTP at and 0.1 pmol of primer. The final concentration of $MgCl_2$ in the Taq polymerase reactions was 10 m*M*. Although the conditions we used in this figure are acceptable, those in Fig. 2 are recommended; also see (*17*) for additional improvements. In lanes 1 to 3, ligation of the linker was allowed to proceed for only 2 hours; for all other lanes, ligation was for the usual 10 to 12 hours. This shortened ligation time did not change the pattern of footprints (*24*), but did result in a reduced intensity of particular regions of the sequence such as the stretch of guanines between the κ chain enhancer–like and MEF-sites [for example, compare lane 1 with lane 4, also see (*17*)]. Experiments with longer ligation times confirmed that there was no observable footprint in this region (*24*). MCK coding strand primers used were: 1, CAAACCTGCGGCTGAGGGGAAGTGG; 2, CTGCCCCTCACCCTGGATCCACCAG; 3, CTGCCCCTCACCCTGGATCCACCAGGG. MCK noncoding strand primers used were: 1, GCTCTGGTCTGCCTTCCACAGCTTG; 2, CATATTGTGTCCTGCTCTGGTCTGC; 3, GCTCTGGTCTGCCTTCCACAGCTTGGG. Because of the different length and G·C content of primers, different temperatures were used during the Taq polymerase hybridizations: coding strand primer 2, 66°C; coding strand primer 3, 72°C; noncoding strand primer 2, 66°C; noncoding strand primer 3, 69°C.
51. We thank J. Miner for providing template plasmids used in the RNase protection assays and aza-myoblast RNA's; S. Hauschka for the MM14 and DD1 cell lines; S. Sharp for BC_3H1 RNA's; P. Garrity, J. Johnson, U. Landegren, H. Weinhard, and P. Mathers for helpful discussions; S. Hauschka, J. Buskin, H. Arnold, E. Olsen, A. Lassar, and H. Weintraub for supplying information before publication; and N. Davidson, D. Anderson, K. Zinn, A. Riggs, G. Pfeifer, S. Steigerwald, and members of the Wold group for critical reading of this manuscript. Supported by NIH (General Medicine) grants RR07003 and GM35526 (B.W.). B.W. would like to dedicate this work to D.E.W.

9 July 1989; accepted 20 September 1989

"Tit for tat, I suppose. He was a vegetarian."