MOLECULAR AND CELLULAR BIOLOGY, July 1996, p. 3893–3900 Vol. 16, No. 7
0270-7306/96/$04.00+0

Differences between MyoD DNA Binding and Activation Site Requirements Revealed by Functional Random Sequence Selection

JING HUANG,[1,2] T. KEITH BLACKWELL,[3] LARRY KEDES,[2]* AND HAROLD WEINTRAUB[1]

Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, Washington 98104[1]; Institute for Genetic Medicine and Department of Biochemistry and Molecular Biology, USC School of Medicine, Los Angeles, California 90033[2]; and Center for Blood Research and Harvard Medical School, Boston, Massachusetts 02115[3]

Received 14 December 1995/Returned for modification 6 February 1996/Accepted 26 March 1996

A method has been developed for selecting functional enhancer/promoter sites from random DNA sequences in higher eukaryotic cells. Of sequences that were thus selected for transcriptional activation by the muscle-specific basic helix-loop-helix protein MyoD, only a subset are similar to the preferred in vitro binding consensus, and in the same promoter context an optimal in vitro binding site was inactive. Other sequences with full transcriptional activity instead exhibit sequence preferences that, remarkably, are generally either identical or very similar to those found in naturally occurring muscle-specific promoters. This first systematic examination of the relation between DNA binding and transcriptional activation by basic helix-loop-helix proteins indicates that binding per se is necessary but not sufficient for transcriptional activation by MyoD and implies a requirement for other DNA sequence-dependent interactions or conformations at its binding site.

The basic helix-loop-helix (bHLH) family of proteins (48) includes many that are pivotal in cell type determination, such as the MyoD group (including MyoD, myogenin, myf-5, and MRF4) for skeletal myogenesis (10, 20, 55, 73, 75). The bHLH proteins form homo- or heterodimers through the HLH domain and bind to DNA through the adjacent basic region (BR) (16, 72). They bind as dimers to DNA sites that generally share the consensus CANNTG (the E box) (8, 49), with each BR occupying half of the site (8, 19, 21, 44).

MyoD family members form homodimers relatively inefficiently (67) and instead appear to activate transcription of muscle-specific genes as heterodimers with the widely expressed E proteins (including E2A [E12 and E47], E2-2, and HEB [3]) (41). E proteins also form heterodimers with other tissue-specific bHLH proteins (49), to regulate still different sets of genes in various tissues such as erythrocytes (1), pancreas (51), and neurons (36), and they form homodimers which are essential for B-cell differentiation (4, 18, 39). The ability of these different protein complexes to regulate different sets of genes appears paradoxical, because their DNA binding specificities are often surprisingly overlapping. For example, certain sites that can be bound well by either MyoD-E or E-E complexes in vitro (8, 49) can be activated by only one or the other of these complexes in the cell (25, 77). It therefore appears either that other protein factors are involved in determining transcriptional specificity or that subtle differences in DNA binding specificity of the bHLH proteins must be biologically significant. These two mechanisms may not be mutually exclusive. Evidence suggests that MyoD may require a positive cofactor to activate transcription of appropriate genes (9, 16, 17, 47, 76) and that it can be prevented from acting at inappropriate sites by repressors that recognize bases overlapping the E box (25, 77). Remarkably, these proposed cofactor and repressor functions both appear to depend on the presence of particular MyoD BR residues, mutations of which may allow DNA binding but not transcriptional activity. X-ray crystallographic studies indicate that in the bound complex these MyoD residues are in intimate proximity to the DNA and that the MyoD BR assumes a binding conformation different from that of the E47 protein (19, 44), suggesting that this conformation may in turn constitute an important aspect of interactions between MyoD (or other bHLH proteins) and coactivator or repressor functions at the promoter.

In light of these observations, it is critical to study systematically the DNA sequence requirements for transcriptional activation by bHLH proteins. Preferred binding sites for MyoD and E proteins have been identified by random-sequence selection in vitro (8, 67, 79), but such assays cannot predict the ability of these proteins to support transcriptional activation. To this end, we have developed the first system that allows proteins to select functional DNA targets from random sequences in mammalian cells. In this new method, dubbed the selection of in vivo target elements (SITE) technique (Fig. 1A), oligonucleotides with random sequences at positions of interest were embedded in a muscle-specific promoter cassette, in this case the human cardiac α-actin (HCA) enhancer/promoter cassette, located upstream of a β-galactosidase (β-Gal) reporter gene. The resulting plasmid library was cotransfected with a MyoD expression vector into murine NIH 3T3 fibroblasts, cells that expressed β-Gal were then selected by fluorescence-activated cell sorting (FACS), and finally plasmid DNA was extracted from selected cells and amplified in *Escherichia coli* to allow this selection procedure to be repeated. Three such selection rounds yielded sequences through which MyoD can activate transcription in vivo. These sequences overlap with but are generally distinct from optimal MyoD or MyoD-E binding sites identified in vitro (8, 79). Instead, the selected sites are similar to various functional E boxes present in native muscle-specific promoters. Remarkably, the active sequences selected from the HCA promoter context do not necessarily activate transcription in the context of another muscle-specific promoter, such as the muscle creatine kinase (MCK) promoter. These findings indicate, surprisingly, that

* Corresponding author. Mailing address: Department of Biochemistry and Molecular Biology, USC School of Medicine, HMR 413, 2011 Zonal Ave., Los Angeles, CA 90033. Phone: (213) 342-1145. Fax: (213) 342-2764.

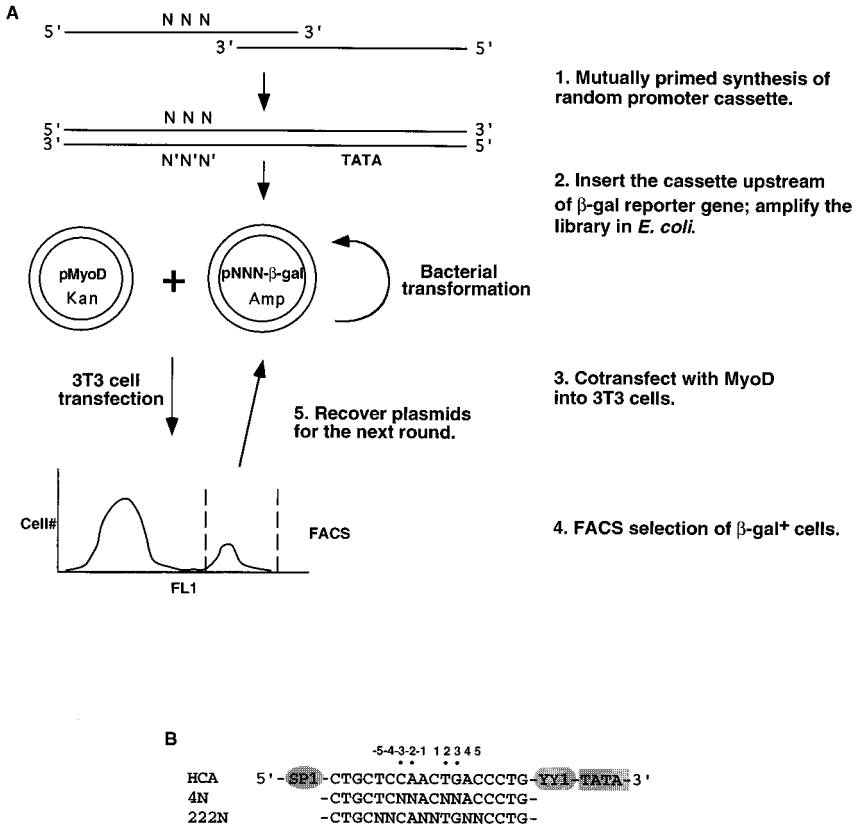


FIG. 1. Selection scheme. (A) An oligonucleotide was synthesized with random sequences (designated N) at positions of interest in the protein-binding site. A double-stranded cassette was generated by extension over a complementary segment and cut with appropriate restriction enzymes at both ends. The cassette was then ligated unidirectionally into a β-Gal reporter plasmid such that transcription of the β-Gal gene is under the control of the randomized enhancer/promoter cassette. The ligation mixture was electroporated into *E. coli* cells, the transformants were plated on LB-agarose-ampicillin plates and allowed to form colonies, which were then pooled, and the plasmids were extracted. (The size of the plasmid library was made larger than 4.6×4^n, n being the number of nucleotides made random, to ensure a probability of no less than 99% to encompass the complexity of the library 4^n [59].) The plasmid pool was cotransfected with a MyoD expression plasmid into 3T3 cells, and β-Gal$^+$ cells were collected by using a fluorescence-activated cell sorter, the stringency of selection being controlled by fluorescence gating. The plasmids in the collected cells were retrieved by the Hirt protocol, amplified by transformation back into *E. coli* cells, and used for the next round of selection. After the desired number of rounds of selection, the active enhancer sequences were determined by DNA sequencing and confirmed by a transfection test. (B) Core sequences of the random cassettes 4N and 222N. The 18-bp core sequences of 4N and 222N are based on the MyoD binding site of the HCA promoter (46). The consensus CA--TG motif is indicated by dots.

the optimal binding sites are not always capable of activating transcription and suggest that binding per se is not sufficient for activation, which may thus be very sensitive to promoter context and/or binding conformation.

MATERIALS AND METHODS

Plasmid constructions. The reporter plasmid pool pNNN-β-gal (Fig. 1A) was constructed by ligating HCA promoter region −89 to +68 (46) containing each random nucleotide cassette (Fig. 1B), the *lacZ* reporter gene from the pNASSβ vector (Clontech), and a fragment from the pBluescript II KS (+) vector (Stratagene) that contained the ampicillin resistance gene and the ColE1 bacterial replication origin. Two E boxes (*Pvu*II sites) in the plasmid backbone were deleted to allow unambiguous analysis. The MyoD expression vector used for SITE experiments, pMyoD-Kan (Fig. 1A), was derived from the pEMSV-MyoD expression vector (16) by inserting the kanamycin resistance gene at the *Sca*I site of the ampicillin resistance gene, which inactivates the latter selection marker.

In plasmid 115MCK-β-gal, the MCK enhancer region (−1207 to −1093) and promoter region (−82 to +7) were used to drive the *lacZ* reporter gene in the same plasmid backbone as in pNNN-β-gal. The various E-box substitutions into the MCK context (Fig. 4) were created by mutagenesis using PCR (31) and confirmed by DNA sequence analysis.

Cell culture and transfection. NIH 3T3 cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% bovine calf serum (Gibco). Cells were cotransfected with 7.5 μg of β-Gal reporter plasmid and 7.5 μg of MyoD expression vector per 10-cm-diameter plate by calcium phosphate coprecipitation (12). Seventeen to 21 h posttransfection, the cells were switched to differentiation medium (Dulbecco's modified Eagle's medium containing 2% horse serum [Gibco] plus transferrin [Sigma; 10 μg/ml] and insulin [Sigma; 10 μg/ml]) for 2 days before they were harvested for FACS.

FACS selection of β-Gal$^+$ cells. Cells were washed three times with 1× phosphate-buffered saline (PBS), detached and dissociated in 5 mM EDTA in PBS, and passed through a 23-gauge needle to further dissociate doublets or bigger clumps. (EDTA was used instead of trypsin because plasma membrane damage caused by trypsin would increase leakage during and after the hypotonic shock.) The cells were resuspended in 50 μl of 1× PBS at 1×10^7 to 5×10^7 cells per ml, transferred to Falcon 2058 tubes, and warmed to 37°C. Flow cytometry analysis of β-Gal activity of the cells was performed as described previously (22, 53). Briefly, 50 μl of 2 mM fluorescein di-β-D-galactoside (FDG; Molecular Probes) in H_2O was added to the cells, and the mixture was incubated at 37°C for 1 min. At the end of the incubation, 10 volumes (1 ml) of ice-cold 1× PBS was immediately added, and the reaction mixture was placed on ice for 20 to 60 min. The reaction was terminated by addition of 2 mM phenylethyl-β-D-thiogalactoside (Molecular Probes). Prior to sorting, propidium iodide (Sigma) was added (5 μg/ml). Cells were analyzed and sorted at 4°C on a FACS machine (Vantage; Becton Dickinson). Fluorescein isothiocyanate (FITC)-positive cells (in which FDG had been converted to fluorescein) were collected into a 35-mm-diameter Falcon dish coated with 1% SeaPlaque agarose in 1× PBS.

Recovery of plasmids from FACS-selected cells. Plasmids in the sorted cells were extracted by the Hirt protocol (32). Electroporation of highly competent *E. coli* cells was used for efficient recovery of selected plasmids (29). Since the *lacZ* reporter plasmid and the MyoD expression vector were constructed to carry different bacterial selection markers (Fig. 1A), the reporter plasmids can be selectively amplified in *E. coli*. The bacterial transformants were pooled, and plasmids were extracted from them (Maxiprep; Qiagen). In the final round of the

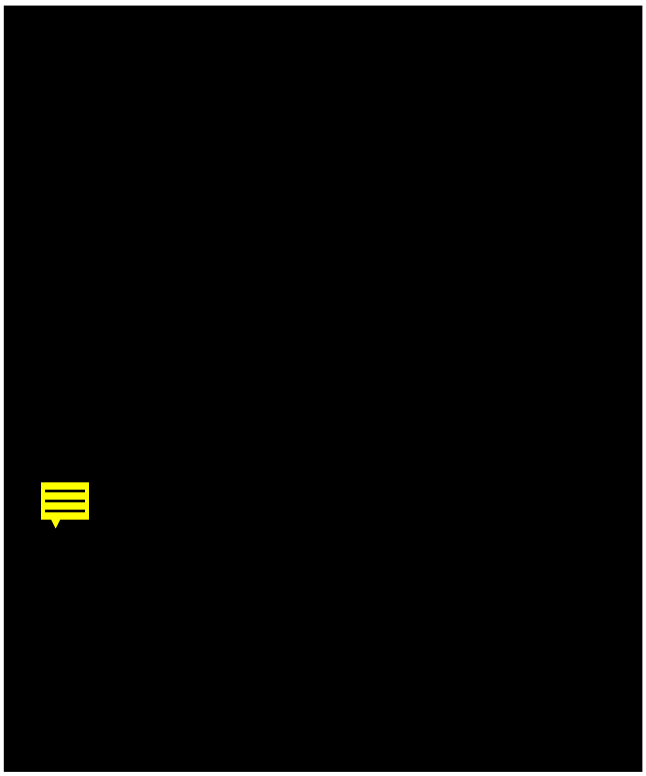


FIG. 2. 4N selection by SITE. (A and B) DNA sequencing gels showing random sequences at desired positions in the initial 4N plasmid pool (A) and emergence of CA and TG dinucleotides from the background after SITE (B). (C and D) X-Gal staining of cells cotransfected with the 4N template and MyoD expression vector before (C) and after (D) SITE.

222N selection, individual bacterial transformants were picked arbitrarily and plasmids were extracted (Miniprep; Promega).

Sequencing of recovered plasmids. DNA sequencing was done by the dideoxy method (61), using Sequenase (U.S. Biochemical).

β-Gal assays. β-Gal activities of transfected cells were monitored by either 5-bromo-4-chloro-3-indolylphosphate-β-D-galactopyranoside (X-Gal) staining of fixed cells (60) or *o*-nitrophenyl-β-D-galactopyranoside (ONPG) and Galacton-Plus (Tropix) hydrolysis by cell lysates, in accordance with standard protocols (59) or the manufacturer's instructions, respectively.

RESULTS AND DISCUSSION

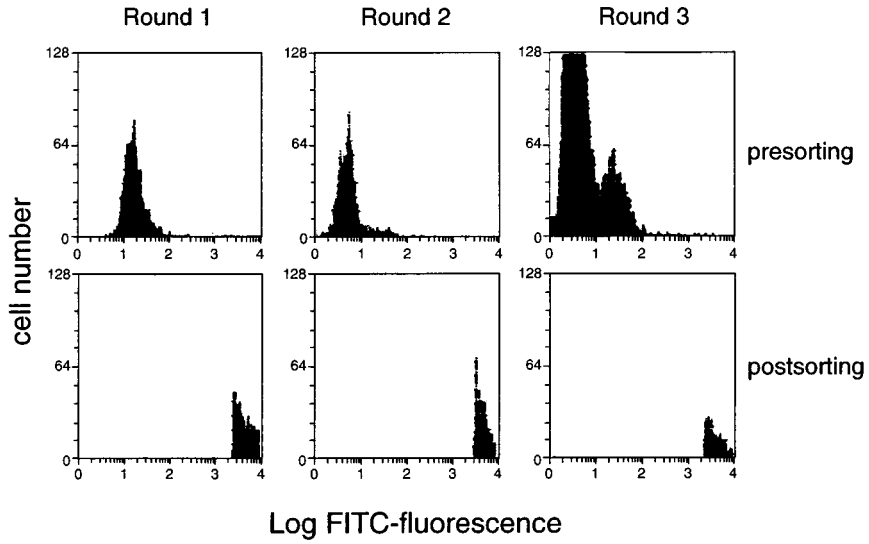
The CANNTG motif is selected by MyoD to activate transcription. A transient-transfection approach was used both to generate MyoD-expressing cells and to select MyoD-responsive sequences. This strategy was chosen because stable expression of MyoD inhibits the cell cycle (15, 27, 28, 66) and because myoblast cells that constitutively express MyoD may fuse, thus complicating selection of individual cells. The selection templates (Fig. 1B) were designed on the basis of the HCA promoter (46), which is specific to skeletal and cardiac muscle and contains a single E box that is required for MyoD-dependent HCA promoter activity in skeletal muscle (62, 65). To test the feasibility and fidelity of the SITE approach, we first used the 4N promoter template (Fig. 1B), in which the CA and TG positions of the E box are random. When NIH 3T3 cells were cotransfected with the 4N pool and a MyoD expression vector, β-Gal activity was close to the basal level (Fig. 2C), suggesting that only a very small portion (if any) of the random-sequence combinations were activated by MyoD. In contrast, after two rounds of selection, the pool β-Gal activity was dramatically increased (Fig. 2D), and the DNA sequence of the recovered plasmids revealed selection for the CA and TG dinucleotides (compare Fig. 2B with Fig. 2A), thus confirming both their importance and the validity of the selection system.

DNA sequence requirement for MyoD transactivation through the CANNTG motif. To identify E boxes through which MyoD can activate transcription, we used the 222N template (Fig. 1B), in which base pairs of the HCA promoter internal to and flanking the CANNTG consensus are of random sequence. In preliminary studies, we determined by quantitative β-Gal assay (59) that the wild-type HCA sequence can generate signals at least 50-fold over the level in nontransfected cells or cells transfected with a mutant construct. Consequently, we set the FACS gates for FITC-positive cells at levels well above basal levels in order to allow selection of cells with only the highest levels of β-Gal activity, which probably contain multiple copies of the most active sequences. In the first round, the cells that were cotransfected with the 222N population and MyoD turned the solutions slightly yellow (as a result of FITC leakage) during the loading of FDG, suggesting that the 222N pool contained a considerable portion of active sequence. However, this activity was clearly much lower than that observed for the wild-type sequence. Three rounds of increasingly stringent FACS selection yielded a sequence pool with activity that was significantly higher than that of the starting population, as is apparent from the appearance of a new peak in the FACS spectrum (Fig. 3A). This dramatic improvement in pool activity is also indicated by much more leaking of visible FITC from the cells during the FDG loading procedure in the final round.

When individual plasmids recovered from this pool were assayed for transcriptional activation by cotransfected MyoD, most (15 of 17) were active, some (4 of 17) even more so than the wild-type HCA promoter. Of this last most active group, two (sequences 5 and 15) were represented multiple times (Fig. 3B). None of these selected sequences were active without cotransfection of the MyoD expression vector or could be activated by expression of the E proteins alone. Although a background of inactive plasmids (2 of 26) was still present in the selected pool, these results confirm that the selection procedure enriched significantly for MyoD-responsive sequences (Fig. 3B).

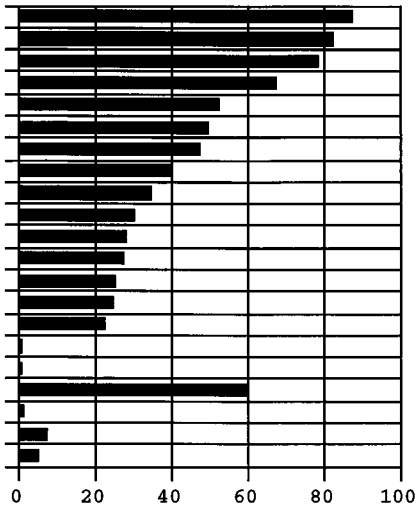
Differences between in vivo activation and in vitro DNA binding. Only a subset of the possible E-box CANNTG hexamer core site motifs were represented among the selected molecules that were responsive to MyoD, none of which contained the CACGTG hexamers to which MyoD or MyoD-E complexes are known not to bind (7). Of the 10 sequences that were at least half as active as the wild-type HCA promoter, one contained the wild-type CAACTG hexamer core sequence, four contained a CAAGTG hexamer, and five contained a CAGGTG sequence. Previously, binding-site selections performed in vitro demonstrated that MyoD homodimers and MyoD-E47 heterodimers prefer sites that contain the CAGCTG and CAGGTG motifs, respectively, and identified their preferred sequences at positions flanking these elements (Fig. 3D) (8). It is remarkable that only a subset of the sequences selected here in vivo contained a CAGGTG motif and that relatively few (4 of 15) of the active plasmids contained the base preferred by MyoD at position −4 or +4. However, the majority of the sequences selected by this method are identical or very similar to E boxes in the native promoters of the transcriptional regulators MyoD, myogenin, and MRF4, as well as in the promoters of many muscle structural protein genes. For instance, the hexamer core CAAGTG, which was not identified through in vitro binding selection but was recovered repeatedly by SITE, has been observed to date only in the

A



B

LABEL	FREQUENCY	SEQUENCE	ANECDOTE
		.. ..	
#26	1	GTCAGGTGTG	mAchR-γ R & HCA half-sites
#5	8	ATCAAGTGCA	MyoD DRR1
#18	1	ACCACCTGCA	Desmin ME3 half-site
#15	3	ATCAAGTGCT	MyoD PRR, Myogenin E1 half-site
#40	1	GGCAGGTGTC	MyoD DRR3, MCK R
#28	1	AGCAGGTGTG	MCK R, sTnC, MLC 1/3A, Myogenin E1
#27	1	TGCAAGTGAG	MyoD PRR & DRR1 core
#9	1	GGCAGTTGAA	Myogenin E1, HCA
#30	1	GGCAACTGGA	MRF4 E2
#36	1	GTCAGGTGGC	mAchR-γ R & HCA half-sites
#8	1	GACAGATGGG	mAchR-γ L, MRF4 E1, AchR-δ E1, Desmin ME3
#17	1	GTCAAGTGGG	MyoD PRR
#10	1	CCCACATGTG	MCK L
#19	1	GTCAGGTGAG	mAchR-γ R, HCA & #27 half-sites
#29	1	ATCAGGTGAA	both half-sites unfavorable for E
#6	1	AACAATTGGT	background
#22	1	CTCATTTGAA	background
w.t.	N/A	TCCAACTGAC	w.t. HCA
mut	N/A	TTGGTCCTGA	mutant HCA
gc	N/A	GTCAGCTGAC	in vitro MyoD homodimer site
aa	N/A	AACAGGTGTG	in vitro optimal M-E binding site



C

(a)	#5	ATCAAGTGCA
	MyoD DRR1	ACCAAGTGTA
(b)	#15	ATCAAGTGCT
	MyoD PRR	GCCAAGTGCT
	Myogenin E1	GCCAACTGCT (-)
(c)	#17	GTCAAGTGGG
	MyoD PRR	GCCAAGTGCT
	#27	TGCAAGTGAG
(d)	#26	GTCAGGTGTG
	AchR-γ R	GTCAGCTGTT
	HCA	GTCAGTTGGA
	#36	GTCAGGTGGC
(e)	#18	TGCAGGTGGT
	Desmin E3	TGCAGCTGTC
(f)	#40	GGCAGGTGTC
	MyoD DRR3	GGCAGGTGCT
	MCK R	AGCAGGTGTT (-)
	Desmin E2	GGCAGCTGTT
(g)	#28	AGCAGGTGTG
	MLC 1/3A	AGCAGGTGCA
	MCK R	AGCAGGTGTT (-)
	sTnC	AGCAGGTG (-)
	Myogenin E1	AGCAGTTGGC
(h)	#8	GACAGATGGG
	mAchR-γ L	GACAGATGGG (-)
	MRF4 E1	CCCAGATGGC (-)
	MyoD DRR2	GACAGGTGGA (-)
	AchR-δ E1	GACAGGTGCT (-)
	cAchR-γ R	GACAGCTGTC
	Desmin E3	GACAGCTGCA (-)
	qTnI IRE	GACAGCTGCA
(i)	#9	GGCAGTTGAA
	Myogenin E1	AGCAGTTGGC
	HCA	GTCAGTTGGA
(j)	#30	TCCAGTTGCC (-)
	MRF4 E2	CCCAGTTGTT
(k)	#10	CCCACATGTG
	MCK L	GCCACATGTC (-)

D

MyoD*:	AA**CA**GC**TG**TT
	Gt aC
Myogenin†:	AA**CA**GC**TG**TT
	T
E47*:	NT**CA**CC**TG**Aa
E47-MyoD*:	NT**CA**CC**TG**TN
E12-MyoD*:	cT**CA**CC**TG**TT
	C
E-myogenin‡:	TG**CA**CC**TG**TT

FIG. 3. SITE-selected sequences and transcriptional activities from the 222N template. (A) FACS sorting of the 222N pool. The sort gates for β-Gal activity were set such that the ratios of FITC fluorescence intensity of the gate over that of the bulk of β-Gal-negative cells were 150, 500, and 700 for the first, second, and third rounds, respectively. Note that the fluorescence intensity axis is logarithmic. Dead cells (propidium iodide bright) were eliminated through a second fluorescence gating. Cell debris, doublets, and larger clumps were gated out by forward- and side-scattering settings. The upper panels are histograms of the FITC intensities of all cells; the lower panels show the sorted cells while they were accumulating over an arbitrary period during the sorting. The cells containing FITC levels higher than the sort gate were collected, and the plasmid DNA was extracted and amplified in *E. coli* cells in order to do the next round of SITE. (B) Transcriptional activation by sequences selected from the third round of SITE. The plasmids recovered from the third round of the selection were transfected individually with the MyoD expression vector into 3T3 cells. β-Gal activity was measured from the cell lysates, using ONPG as the substrate. Each datum point was the average of at least three independent transfections. The data were plotted as the fold activation by MyoD over that of pEMSVscribe (MyoD$^-$ control vector). The Anecdote column lists known muscle-specific enhancer/promoters that contain E boxes with sequences identical or very similar to those selected by SITE. Only E boxes that were tested to be functional in the literature are listed. Note that in sequence 5, although the 3′ flanking sequence CA can in theory be grouped with nucleotides immediately downstream to form another E box (CACCTG site), this second E box is nonfunctional (data not shown). mAchR-γ, mouse acetylcholine receptor γ subunit; DRR1, distal regulatory region 1; PRR, proximal regulatory region; sTnC, smooth muscle troponin C; MLC, myosin light chain; w.t., wild type; mut, mutant. (C) Alignment of selected E-box sequences to known enhancer/promoter sequences in muscle-specific genes. (−), the sequence displayed is antisense to the template strand. (D) Summary of in vitro DNA binding preferences of myogenic bHLH and E proteins. ∗, data from reference 8; †, data from reference 24; ‡, data from reference 80. A line drawn over an uppercase letter indicates a base that is absent in that position, and a line over a lowercase letter indicates a decrease in use of that base.

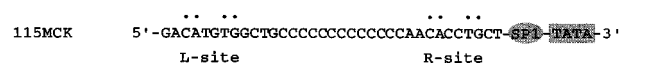
MyoD promoter. Among them, sequence 5 resembles the MyoD distal regulatory region 1, which is important for transactivation of the MyoD promoter (69), and sequence 15 is very similar to the E box in the MyoD proximal regulatory region, which is indispensable for both distal regulatory region transactivation (69) and autoregulation (80) by MyoD. Although the disproportionate representation of sequences 5 and 15 suggests that the arbitrarily chosen number of cells from which plasmids were recovered may have contained sequences of limited complexity, a variety of different sequences that indeed differ from the preferred in vitro consensus were selected.

Perhaps the most striking aspect of the selected MyoD-responsive sequences is that they closely resemble functional E boxes that are present in the promoters of muscle-specific genes (Fig. 3B and C) (5, 13, 14, 23, 26, 38, 40, 42, 43, 57, 63, 78) but not those found in the target genes of nonmuscle bHLH proteins (see references in reference 8). This resemblance is apparent not only in their core hexamers but also in flanking sequences, many with identical pentamer half-sites (Fig. 3C). In their natural promoter context, they are also capable of mediating transcriptional activation by the myogenic bHLH proteins, which apparently function as heterodimers with E proteins. These muscle-specific sites can, in general, be bound in vitro by MyoD and the other myogenic bHLH proteins and by heterodimers of these proteins with E proteins. The MyoD-responsive sequences selected here by SITE can similarly be bound by MyoD-E heterodimers (although the in vitro binding affinities do not parallel the in vivo transcriptional activity [data not shown]), and when we examined them in the cotransfection assay, we found that they mediate activation by a tethered MyoD-E47 dimer (52) to levels no higher than those characteristic of cotransfected MyoD (data not shown). The simplest interpretation of these findings is that these selected sites are direct targets of heterodimers of cellular E proteins with the cotransfected MyoD or with other myogenic bHLH proteins that may be activated by MyoD (see below).

The SITE data thus support the notion that DNA binding alone is not sufficient for a bound MyoD (or other myogenic bHLH protein)-E heterodimer to be transcriptionally active. To confirm that this SITE experiment did not somehow simply miss functional sequences that more closely resemble optimal MyoD and MyoD-E in vitro binding sites, we tested such sites (Fig. 3B) in the HCA cassette for transactivation by MyoD. In this assay, surprisingly, an optimal in vitro MyoD-E site was inactive, and the activity of a preferred MyoD homodimer site was very low (Fig. 3B). On the other hand, SITE recovered some transcriptionally active plasmids bearing a sequence that was strongly selected against by in vitro methods, e.g., T at the −4 position of the E47 protein half-site (8, 67). These data indicate that the E boxes that are most responsive to MyoD in vivo do not necessarily correspond to the in vitro binding preferences of MyoD or MyoD-E complexes.

Role of promoter context in MyoD transactivation through different E boxes. One possible explanation for differences in vivo and in vitro is that in vivo repressors might bind to certain E-box sequences and prevent MyoD from binding to these sites and activating transcription. Alternatively, differences among interacting proteins on these promoters might affect the conformation of the DNA and/or of the protein-DNA complex in a way that prevents particular E-box sequences from acting as an effective enhancer. To distinguish between these possibilities, we compared the abilities of different E boxes to activate transcription in the HCA promoter and the MCK enhancer/promoter. The 3,300-bp enhancer/promoter of MCK contains two E boxes that are targets of MyoD activation (37, 74). A 115-bp segment including these E boxes that retains all of the muscle-specific activity of the enhancer (35) was used for the E-box substitution experiment (Fig. 4A), the results of which are shown in Fig. 4B. The sequence 5 E-box decamer, a strong binding site in vitro and a strong activation site in the HCA promoter, was only moderately active when replacing the

A



B

Constructs		Fold Activation		
E box	Sequence(X)	HCA(X)	MCK(L-X)	MCK(2X)
HCA	TCCAACTGAC	60	17	57
MCK-R	AACACCTGCT	ND	51	70
#5	ATCAAGTGCA	82	9	3
aa	AACAGGTGTG	6	107	220
#6	AACAATTGGT	1	13	121

FIG. 4. Comparison of transcriptional activities of different E boxes in the MCK and HCA promoter contexts. (A) Molecular structure and enhancer sequence of the MCK enhancer and promoter. Only the two E boxes (L and R sites) are shown. (B) Comparison of transcriptional activities of different E boxes in the MCK and HCA promoter contexts. X's stand for different E-box sequences; HCA (X) denotes X replacing the E box in the HCA promoter, MCK (L-X) denotes X replacing the R site in the MCK enhancer, and MCK (2X) denotes X replacing both the L and R sites in the MCK enhancer. Each datum point was derived from transfection of two independent clones. β-Gal activity was measured from cell lysates by using Galacton-Plus (Tropix) as the substrate.

right (R) site of the MCK enhancer. Its activity was even lower if both left (L) and R sites were replaced. On the other hand, the in vitro optimal binding site (construct aa), although inactive in the HCA context, was the most active in the MCK context. It is interesting that the HCA E box, a poor binding site in vitro, acts as a strong activation site both in its native context and when replacing the L and R sites simultaneously in the MCK enhancer. Thus, these data for the MCK promoter confirmed the notion that the ability of an E-box enhancer to support transcriptional activation cannot be predicted from its affinity for the protein factors. These results demonstrate that promoter context is an important determinant of activation properties of different E-box sequences.

It is interesting that a majority (20 of 26) of the sites that were selected bear a MyoD half-site on the 5′ end in proximity to the Sp1 site on the HCA promoter (Fig. 1B and 3B). However, it is unlikely that Sp1 could be the determinant of context specificity because the MCK enhancer also contains an Sp1 site. Sp1 sites are also found adjacent to MyoD half-sites in the promoters of the acetylcholine receptor δ subunit (AchR-δ), myogenin, troponin, and MyoD genes, although the importance of these Sp1 sites has not been uniformly evaluated. Thus, it is not obvious why the activities of some E boxes are so different in the HCA and MCK promoters. Our results thus form a basis for further investigation of *cis*-acting elements that are involved in determining promoter context specificity.

Application of the SITE method. The SITE technique described here has, for the first time, used the power of random-sequence selection (6, 54, 64, 68) in metazoan cells to identify functional regulatory sites. It can yield sequences that are responsive to any transcription factor, or to a protein that activates such a factor, provided that either the cloned gene or a cell line expressing the factor is available. The reiterative selection process permits recovery of a large number of sites that mediate a range of transcriptional activities, with more stringent selection yielding more active sites. This method is likely to yield biologically significant sequences that might not be identified by in vitro methods, because they are selected in the presence of the full complement of cellular factors under physiological conditions and can thus provide rational guidance for designing reporter constructs. As such, the in vivo selection method provides an approach complementary to the in vitro method for discovering target sequences for regulatory proteins. A more important benefit of this strategy is that it allows a systematic comparison of the functional requirements of proteins under different circumstances and conditions and can thus illuminate critical regulatory mechanisms. For example, a comparison of the MyoD-responsive sequences identified in this study with the preferred in vitro binding sites of MyoD suggests the existence of particular sequence requirements for activation. The versatility of the SITE approach will now make it possible to undertake similar investigations of activation by MyoD (and its BR variants) in the context of other promoters and associated factors and to identify similarly sequences that respond to other bHLH proteins.

The SITE technique should be adaptable to a variety of other experimental situations. For example, repression sequences could be selected, provided that a screening step first identified cells that contain the reporter library DNA. It should be similarly possible to select native regulatory sequences from a library of genomic DNA fragments or, if transcribed positions were randomized, to identify sequences involved in RNA processing or translational control. If positions within coding sequences were randomized, a selection could identify the amino acid residues that are encoded by the corresponding positions which allow a transcription factor to function, as has been done in bacteria (34) and in *Saccharomyces cerevisiae* (56). In principle, the number of positions that can be randomized in the SITE technique should be limited only by the number of cells that can be handled. For example, eight random positions will generate 4^8 (about 6.6×10^4) different plasmid molecules. If one molecule were introduced per cell, screening of a library of this size would require on the order of 3×10^5 cells (59), a number readily attainable by transient transfection. However, the efficiency of screening is actually higher, because each cell takes up multiple plasmids and because even at a high input copy number, positive cells can be identified by FACS, which can detect as few as 5 β-Gal enzyme molecules per cell (53), is rapid ($1/10^6$ cells per h), and can be applied to different markers (e.g., the cell surface molecule CD4 [45] or green fluorescent protein [GFP] [11]). Multiple markers can be used either alone or simultaneously to examine different sets of regulatory sequences. Alternatively, the technique can be applicable to other transient selection markers (e.g., puromycin resistance [71]).

Specificity of MyoD-mediated transcriptional activation. Although the sequences identified here are responsive to MyoD, it is possible that these sequences actually are or include the direct targets of another myogenic bHLH protein. In cultured cells, expression of a myogenic bHLH protein can activate the corresponding endogenous gene along with other members of this group, with the particular myogenic genes activated varying among different lines (2, 33, 70). In the 3T3 cells used in this study, MyoD generally does not activate the endogenous MyoD, myf-5, or MRF4 gene but can activate the myogenin gene. In the mouse, myf-5 and MyoD determine the myoblast cell fate (58), while myogenin is required to activate the full spectrum of sarcomeric genes (30, 50), but in cells cultured from myogenin$^{-/-}$ mice, MyoD can also activate these genes (50). Accordingly, we cannot determine whether in our experiments myogenin might have also contributed to the activation of sequences selected. However, it is unlikely that only myogenin (but not MyoD) is the direct activator, because when the selected sequences were cotransfected with a myogenin expression vector, they were activated to levels only half as high as those achieved by MyoD. As these sites are bound by MyoD-E heterodimers in vitro and activated similarly by MyoD and MyoD-E tethered dimers in vivo (data not shown), and since the DNA binding requirements of MyoD in vitro and those of myogenin with the nuclear extracts are virtually identical (Fig. 3D) (8, 79), the most likely conclusion appears to be that the selected regulatory sites may be the direct targets of heterodimers of either MyoD or myogenin with endogenous cellular E proteins.

The group of sequences that were identified in this experiment share some attributes with, but are not identical to, the preferred in vitro binding sites for MyoD (or myogenin) or MyoD-E complexes. In the HCA promoter, these bHLH complexes might bind cooperatively with neighboring proteins, and these interactions might have contributed enough binding energy to have relaxed the sequence requirements for activation. However, two lines of evidence argue against this interpretation. First, an optimal site was not capable of supporting activation in this context (Fig. 3B). Second, the selected sequences were actually less similar to the in vitro preferences than to E boxes found in the native promoters of muscle-specific genes. These native E boxes have been shown to be important for auto- and/or cross-activation of the myogenic bHLH genes or for activation of muscle structural genes by these bHLH proteins, suggesting that the sequences selected here by SITE are biologically significant.

We have selected E-box sequences that are able to support

transcriptional activation by muscle-specific bHLH protein MyoD. The data indicate that only a subset of MyoD binding sites allow activation of transcription in the HCA promoter context and that certain E boxes function differently in different muscle-specific promoters. These findings demonstrate that promoter context plays an important role in transcriptional activation by MyoD through E boxes. Promoter context could affect transcriptional activation in several ways. For example, the neighboring protein factors could influence the conformation of the promoter DNA and/or of the protein-E box complex; different proteins could also recruit distinct associated factors to the promoters. In addition, our findings are reminiscent of the positive-control MyoD BR mutations (17, 76), which also appear to permit DNA binding but not activation. Significantly, these same BR residues not only appear to influence the conformation of the protein-DNA complex (19, 44) but also have been implicated in interactions with proposed regulatory factors (9, 16, 17, 47, 76). The findings described here indicate that some, or perhaps all, of the potential regulatory interactions require specific DNA sequences, which themselves could influence the conformation of the bound MyoD-E complex. It remains to be determined whether such conformational changes might provide distinctive targets for coactivators or repressors or would derive from such interactions.

ACKNOWLEDGMENTS

We thank members of the Weintraub and Kedes laboratories for advice and support; S. Fiering for initial help and advice with FACS; A. Berger for assistance with numerous cell sortings; V. Sartorelli for the HCA promoter; B. Wold for the MyoD-E47 expression vector; A. Jimenez and A. Lassar for pac cDNA and expression vector; M. Emerman for CD4 cDNA; M. Chalfie for GFP cDNA; N. Rosenthal for sharing information; and D. Turner, S. Handeli, J. Gogos, Y. Zhuang, M. Horwitz, K. Munir, and R. Schreck for helpful discussions and encouragement. J.H. is grateful to J. T. Huang, M. H. Wu, F. Zhang, S. Lew, H.W., T.K.B., and L.K. for faith, understanding, and support. We thank M. Groudine, H. Sucov, and B. Wold for critical reading of the manuscript.

This work was supported in part by grants from NIH (to H.W. and to L.K.), the U.S. Army Breast Cancer Research Program (to T.K.B.), and HHMI (to H.W.).

REFERENCES

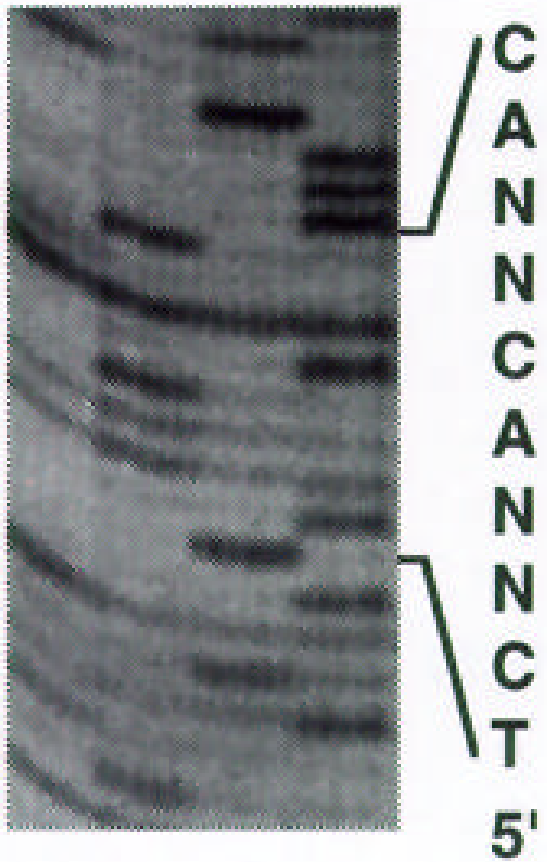
1. **Aplan, P. D., K. Nakahara, S. H. Orkin, and I. R. Kirsch.** 1992. The SCL gene product: a positive regulator of erythroid differentiation. EMBO J. **11:**4073–4081.
2. **Aurade, F., C. Pinset, P. Chafey, F. Gros, and D. Montarras.** 1994. Myf5, MyoD, myogenin and MRF4 myogenic derivatives of the embryonic mesenchymal cell line C3H10T1/2 exhibit the same adult muscle phenotype. Differentiation **55:**185–192.
3. **Bain, G., S. Gruenwald, and C. Murre.** 1993. E2A and E2-2 are subunits of B-cell-specific E2-box DNA-binding proteins. Mol. Cell. Biol. **13:**3522–3529.
4. **Benezra, R.** 1994. An intermolecular disulfide bond stabilizes E2A homodimers and is required for DNA binding at physiological temperatures. Cell **79:**1057–1067.
5. **Black, B. L., J. F. Martin, and E. N. Olson.** 1995. The mouse MRF4 promoter is trans-activated directly and indirectly by muscle-specific transcription factors. J. Biol. Chem. **270:**2889–2892.
6. **Blackwell, T. K.** 1995. Selection of protein binding sites from random nucleic acid sequences. Methods Enzymol. **254:**608–618.
7. **Blackwell, T. K., J. Huang, A. Ma, L. Kretzner, F. W. Alt, R. N. Eisenman, and H. Weintraub.** 1993. Binding of Myc proteins to canonical and noncanonical DNA sequences. Mol. Cell. Biol. **13:**5216–5224.
8. **Blackwell, T. K., and H. Weintraub.** 1990. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science **250:**1104–1110.
9. **Brennan, T. J., T. Chakraborty, and E. N. Olson.** 1991. Mutagenesis of the myogenin basic region identifies an ancient protein motif critical for activation of myogenesis. Proc. Natl. Acad. Sci. USA **88:**5675–5679.
10. **Buckingham, M. E.** 1994. Muscle: the regulation of myogenesis. Curr. Opin. Genet. Dev. **4:**745–751.
11. **Chalfie, M., Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher.** 1994. Green fluorescent protein as a marker for gene expression. Science **263:**802–805.
12. **Chen, C., and H. Okayama.** 1987. High-efficiency transformation of mammalian cells by plasmid DNA. Mol. Cell. Biol. **7:**2745–2752.
13. **Cheng, T. C., M. C. Wallace, J. P. Merlie, and E. N. Olson.** 1993. Separable regulatory elements governing myogenin transcription in mouse embryogenesis. Science **261:**215–218.
14. **Christensen, T. H., H. Prentice, R. Gahlmann, and L. Kedes.** 1993. Regulation of the human cardiac/slow-twitch troponin C gene by multiple, cooperative, cell-type-specific, and MyoD-responsive elements. Mol. Cell. Biol. **13:**6752–6765.
15. **Crescenzi, M., T. P. Fleming, A. B. Lassar, H. Weintraub, and S. A. Aaronson.** 1990. MyoD induces growth arrest independent of differentiation in normal and transformed cells. Proc. Natl. Acad. Sci. USA **87:**8442–8446.
16. **Davis, R. L., P. F. Cheng, A. B. Lassar, and H. Weintraub.** 1990. The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. Cell **60:**733–746.
17. **Davis, R. L., and H. Weintraub.** 1992. Acquisition of myogenic specificity by replacement of three amino acid residues from MyoD into E12. Science **256:**1027–1030.
18. **Desiderio, S.** 1995. Transcription factors controlling B-cell development. Curr. Biol. **5:**605–608.
19. **Ellenberger, T., D. Fass, M. Arnaud, and S. C. Harrison.** 1994. Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. Genes Dev. **8:**970–980.
20. **Emerson, C. P.** 1990. Myogenesis and developmental control genes. Curr. Opin. Cell Biol. **2:**1065–1075.
21. **Ferre-D'Amare, A. R., G. C. Prendergast, E. B. Ziff, and S. K. Burley.** 1993. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. Nature (London) **363:**38–45.
22. **Fiering, S. N., M. Roederer, G. P. Nolan, D. R. Micklem, D. R. Parks, and L. A. Herzenberg.** 1991. Improved FACS-Gal: flow cytometric analysis and sorting of viable eukaryotic cells expressing reporter gene constructs. Cytometry **12:**291–301.
23. **Fujisawa, S. A., K. Hanaoka, M. Hayasaka, Y. T. Hiromasa, and Y. Nabeshima.** 1993. Upstream region of the myogenin gene confers transcriptional activation in muscle cell lineages during mouse embryogenesis. Biochem. Biophys. Res. Commun. **191:**351–356.
24. **Funk, W. D., and W. E. Wright.** 1992. Cyclic amplification and selection of targets for multicomponent complexes: myogenin interacts with factors recognizing binding sites for basic helix-loop-helix, nuclear factor 1, myocyte-specific enhancer-binding factor 2, and COMP1 factor. Proc. Natl. Acad. Sci. USA **89:**9484–9488.
25. **Genetta, T., D. Ruezinsky, and T. Kadesch.** 1994. Displacement of an E-box-binding repressor by basic helix-loop-helix proteins: implications for B-cell specificity of the immunoglobulin heavy-chain enhancer. Mol. Cell. Biol. **14:**6153–6163.
26. **Gilmour, B. P., G. R. Fanger, C. Newton, S. M. Evans, and P. D. Gardner.** 1991. Multiple binding sites for myogenic regulatory factors are required for expression of the acetylcholine receptor gamma-subunit gene. J. Biol. Chem. **266:**19871–19874.
27. **Gu, W., J. W. Schneider, G. Condorelli, S. Kaushal, V. Mahdavi, and G. B. Nadal.** 1993. Interaction of myogenic factors and the retinoblastoma protein mediates muscle cell commitment and differentiation. Cell **72:**309–324.
28. **Halevy, O., B. G. Novitch, D. B. Spicer, S. X. Skapek, J. Rhee, G. J. Hannon, D. Beach, and A. B. Lassar.** 1995. Correlation of terminal cell cycle arrest of skeletal muscle with induction of p21 by MyoD. Science **267:**1018–1021.
29. **Hanahan, D., J. Jessee, and F. R. Bloom.** 1991. Plasmid transformation of *Escherichia coli* and other bacteria. Methods Enzymol. **204:**63–113.
30. **Hasty, P., A. Bradley, J. H. Morris, D. G. Edmondson, J. M. Venuti, E. N. Olson, and W. H. Klein.** 1993. Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene. Nature (London) **364:**501–506.
31. **Higuchi, R.** 1989. Using PCR to engineer DNA, p. 61–70. *In* H. A. Erlich (ed.), PCR technology. Stockton Press, New York.
32. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. **26:**365–369.
33. **Hollenberg, S. M., P. F. Cheng, and H. Weintraub.** 1993. Use of a conditional MyoD transcription factor in studies of MyoD trans-activation and muscle determination. Proc. Natl. Acad. Sci. USA **90:**8028–8032.
34. **Hu, J. C., E. K. O'Shea, P. S. Kim, and R. T. Sauer.** 1990. Sequence requirements for coiled-coils: analysis with lambda repressor-GCN4 leucine zipper fusions. Science **250:**1400–1403.
35. **Huang, J., and H. Weintraub.** Unpublished data.
36. **Jan, Y. N., and L. Y. Jan.** 1993. HLH proteins, fly neurogenesis, and vertebrate myogenesis. Cell **75:**827–830.
37. **Jaynes, J. B., J. E. Johnson, J. N. Buskin, C. L. Gartside, and S. D. Hauschka.** 1988. The muscle creatine kinase gene is regulated by multiple upstream elements, including a muscle-specific enhancer. Mol. Cell. Biol. **8:**62–70.
38. **Jia, H. T., H. J. Tsay, and J. Schmidt.** 1992. Analysis of binding and acti-

vating functions of the chick muscle acetylcholine receptor gamma-subunit upstream sequence. Cell. Mol. Neurobiol. **12:**241–258.
39. **Kadesch, T.** 1992. Helix-loop-helix proteins in the regulation of immunoglobulin gene transcription. Immunol. Today **13:**31–36.
40. **Lassar, A. B., J. N. Buskin, D. Lockshon, R. L. Davis, S. Apone, S. D. Hauschka, and H. Weintraub.** 1989. MyoD is a sequence-specific DNA binding protein requiring a region of myc homology to bind to the muscle creatine kinase enhancer. Cell **58:**823–831.
41. **Lassar, A. B., R. L. Davis, W. E. Wright, T. Kadesch, C. Murre, A. Voronova, D. Baltimore, and H. Weintraub.** 1991. Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. Cell **66:**305–315.
42. **Li, H., and Y. Capetanaki.** 1994. An E box in the desmin promoter cooperates with the E box and MEF-2 sites of a distal enhancer to direct muscle-specific transcription. EMBO J. **13:**3580–3589.
43. **Lin, H., K. E. Yutzey, and S. F. Konieczny.** 1991. Muscle-specific expression of the troponin I gene requires interactions between helix-loop-helix muscle regulatory factors and ubiquitous transcription factors. Mol. Cell. Biol. **11:** 267–280.
44. **Ma, P. C., M. A. Rould, H. Weintraub, and C. O. Pabo.** 1994. Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. Cell **77:**451–459.
45. **Maddon, P. J., D. R. Littman, M. Godfrey, D. E. Maddon, L. Chess, and R. Axel.** 1985. The isolation and nucleotide sequence of a cDNA encoding the T cell surface protein T4: a new member of the immunoglobulin gene family. Cell **42:**93–104.
46. **Minty, A., and L. Kedes.** 1986. Upstream regions of the human cardiac actin gene that modulate its transcription in muscle cells: presence of an evolutionarily conserved repeated motif. Mol. Cell. Biol. **6:**2125–2136.
47. **Molkentin, J. D., B. L. Black, J. F. Martin, and E. N. Olson.** 1995. Cooperative activation of muscle gene expression by MEF2 and myogenic bHLH proteins. Cell **83:**1125–1136.
48. **Murre, C., P. S. McCaw, and D. Baltimore.** 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell **56:**777–783.
49. **Murre, C., P. S. McCaw, H. Vaessin, M. Caudy, L. Y. Jan, Y. N. Jan, C. V. Cabrera, J. N. Buskin, S. D. Hauschka, A. B. Lassar, H. Weintraub, and D. Baltimore.** 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell **58:**537–544.
50. **Nabeshima, Y., K. Hanaoka, M. Hayasaka, E. Esumi, S. Li, I. Nonaka, and Y. Nabeshima.** 1993. Myogenin gene disruption results in perinatal lethality because of severe muscle defect. Nature (London) **364:**532–535.
51. **Nelson, C., L. P. Shen, A. Meister, E. Fodor, and W. J. Rutter.** 1990. Pan: a transcriptional regulator that binds chymotrypsin, insulin, and AP-4 enhancer motifs. Genes Dev. **4:**1035–1043.
52. **Neuhold, L. A., and B. Wold.** 1993. HLH forced dimers: tethering MyoD to E47 generates a dominant positive myogenic factor insulated from negative regulation by Id. Cell **74:**1033–1042.
53. **Nolan, G. P., S. Fiering, J. F. Nicolas, and L. A. Herzenberg.** 1988. Fluorescence-activated cell analysis and sorting of viable mammalian cells based on beta-D-galactosidase activity after transduction of Escherichia coli lacZ. Proc. Natl. Acad. Sci. USA **85:**2603–2607.
54. **Oliphant, A. R., and K. Struhl.** 1988. Defining the consensus sequences of *E. coli* promoter elements by random selection. Nucleic Acids Res. **16:**7673–7683.
55. **Olson, E. N.** 1992. Proto-oncogenes in the regulatory circuit for myogenesis. Semin. Cell Biol. **3:**127–136.
56. **Pu, W. T., and K. Struhl.** 1993. Dimerization of leucine zippers analyzed by random selection. Nucleic Acids Res. **21:**4348–4355.
57. **Rosenthal, N., E. B. Berglund, B. M. Wentworth, M. Donoghue, B. Winter, E. Bober, T. Braun, and H. H. Arnold.** 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.
58. **Rudnicki, M. A., P. N. Schnegelsberg, R. H. Stead, T. Braun, H. H. Arnold, and R. Jaenisch.** 1993. MyoD or Myf-5 is required for the formation of skeletal muscle. Cell **75:**1351–1359.
59. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
60. **Sanes, J. R., J. L. Rubenstein, and J. F. Nicolas.** 1986. Use of a recombinant retrovirus to study post-implantation cell lineage in mouse embryos. EMBO J. **5:**3133–3142.
61. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
62. **Sartorelli, V., K. A. Webster, and L. Kedes.** 1990. Muscle-specific expression of the cardiac alpha-actin gene requires MyoD1, CArG-box binding factor, and Sp1. Genes Dev. **4:**1811–1822.
63. **Simon, A. M., and S. J. Burden.** 1993. An E box mediates activation and repression of the acetylcholine receptor δ-subunit gene during myogenesis. Mol. Cell. Biol. **13:**5133–5140.
64. **Singer, V. L., C. R. Wobbe, and K. Struhl.** 1990. A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. Genes Dev. **4:**636–645.
65. **Skerjanc, I. S., and M. W. McBurney.** 1994. The E box is essential for activity of the cardiac actin promoter in skeletal but not in cardiac muscle. Dev. Biol. **163:**125–132.
66. **Sorrentino, V., R. Pepperkok, R. L. Davis, W. Ansorge, and L. Philipson.** 1990. Cell proliferation inhibited by MyoD1 independently of myogenic differentiation. Nature (London) **345:**813–815.
67. **Sun, X. H., and D. Baltimore.** 1991. An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers. Cell **64:**459–470.
68. **Szostak, J. W.** 1992. In vitro genetics. Trends Biochem. Sci. **17:**89–93.
69. **Tapscott, S. J., A. B. Lassar, and H. Weintraub.** 1992. A novel myoblast enhancer element mediates MyoD transcription. Mol. Cell. Biol. **12:**4994–5003.
70. **Thayer, M. J., S. J. Tapscott, R. L. Davis, W. E. Wright, A. B. Lassar, and H. Weintraub.** 1989. Positive autoregulation of the myogenic determination gene MyoD1. Cell **58:**241–248.
71. **Vara, J., F. Malpartida, D. A. Hopwood, and A. Jimenez.** 1985. Cloning and expression of a puromycin N-acetyl transferase gene from Streptomyces alboniger in Streptomyces lividans and Escherichia coli. Gene **33:**197–206.
72. **Voronova, A., and D. Baltimore.** 1990. Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains. Proc. Natl. Acad. Sci. USA **87:**4722–4726.
73. **Weintraub, H.** 1993. The MyoD family and myogenesis: redundancy, networks, and thresholds. Cell **75:**1241–1244.
74. **Weintraub, H., R. Davis, D. Lockshon, and A. Lassar.** 1990. MyoD binds cooperatively to two sites in a target enhancer sequence: occupancy of two sites is required for activation. Proc. Natl. Acad. Sci. USA **87:**5623–5627.
75. **Weintraub, H., R. Davis, S. Tapscott, M. Thayer, M. Krause, R. Benezra, T. K. Blackwell, D. Turner, R. Rupp, and S. Hollenberg.** 1991. The myoD gene family: nodal point during specification of the muscle cell lineage. Science **251:**761–766.
76. **Weintraub, H., V. J. Dwarki, I. Verma, R. Davis, S. Hollenberg, L. Snider, A. Lassar, and S. J. Tapscott.** 1991. Muscle-specific transcriptional activation by MyoD. Genes Dev. **5:**1377–1386.
77. **Weintraub, H., T. Genetta, and T. Kadesch.** 1994. Tissue-specific gene activation by MyoD: determination of specificity by cis-acting repression elements. Genes Dev. **8:**2203–2211.
78. **Wentworth, B. M., M. Donoghue, J. C. Engert, E. B. Berglund, and N. Rosenthal.** 1991. Paired MyoD-binding sites regulate myosin light chain gene expression. Proc. Natl. Acad. Sci. USA **88:**1242–1246.
79. **Wright, W. E., M. Binder, and W. Funk.** 1991. Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site. Mol. Cell. Biol. **11:**4104–4110.
80. **Zingg, J. M., A. G. Pedraza, and J. P. Jost.** 1994. MyoD1 promoter autoregulation is mediated by two proximal E-boxes. Nucleic Acids Res. **22:**2234–2241.

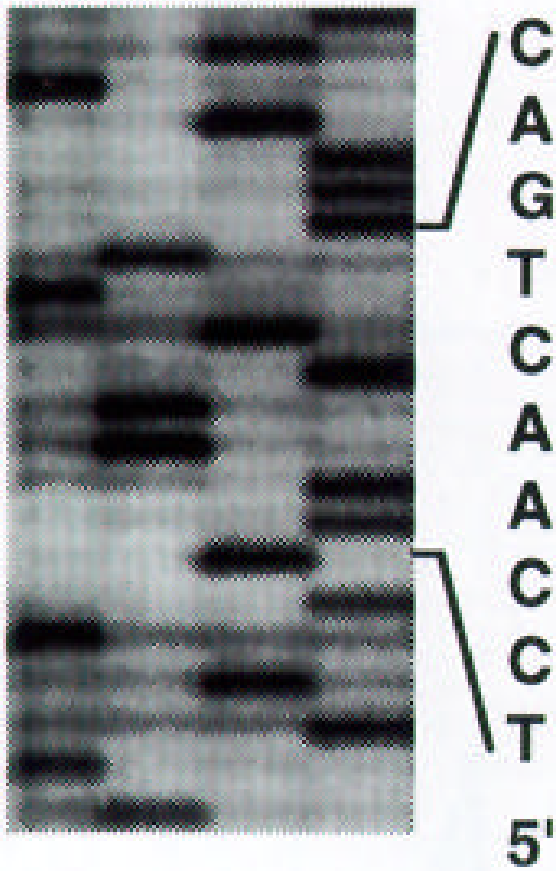
presorting

postsorting

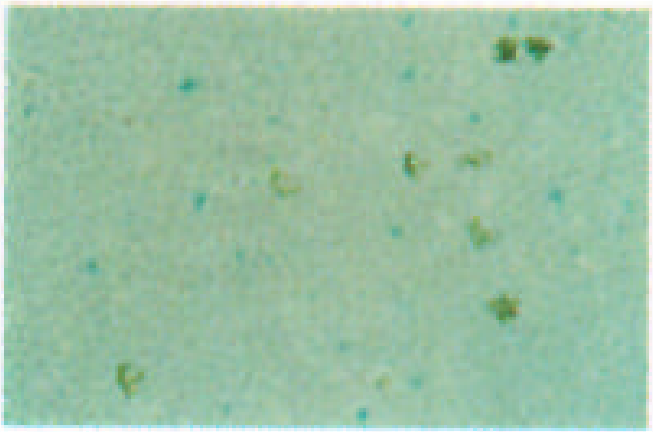
A



B



C



D