MHox: a mesodermally restricted homeodomain protein that binds an essential site in the muscle creatine kinase enhancer

PETER CSERJESI¹, BRENDA LILLY¹, LAURA BRYSON¹, YAOQI WANG², DAVID A. SASSOON² and ERIC N. OLSON¹*

¹Department of Biochemistry and Molecular Biology, Box 117, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA
²Department of Biochemistry, Boston University School of Medicine, 80 East Concord Street, Boston, MA 02118, USA
*Corresponding author

Summary

Myogenic helix-loop-helix (HLH) proteins, such as myogenin and MyoD, can activate muscle-specific transcription when introduced into a variety of nonmuscle cell types. Whereas cells of mesodermal origin are especially permissive to the actions of these myogenic regulators, many other cell types are refractory to myogenic conversion by them. Here we describe a novel homeodomain protein, MHox, that binds an A+T-rich element in the muscle creatine kinase (MCK) enhancer that is essential for muscle-specific transcription and trans-activation by myogenic HLH proteins. MHox is completely restricted to mesodermally derived cell types during embryogenesis and to established cell lines of mesodermal origin. In contrast to most other homeobox genes, MHox expression is excluded from the nervous system, with the highest levels observed in limb bud and visceral arches. In adult mice, MHox is expressed at high levels in skeletal muscle, heart and uterus. The DNA-binding properties and pattern of MHox expression are unique among homeobox genes and suggest a role for MHox as a transcriptional regulator that participates in the establishment of diverse mesodermal cell types.

Key words: homeodomain, muscle specific transcription, muscle creatine kinase.

Introduction

The generation of differentiated cell types during development requires restriction in potential fates of multipotential stem cells and the selective activation of specific gene sets that encode the proteins required by individual cell types to perform specialized functions. The activation of tissue-specific gene expression has been shown to require combinatorial interactions among cell type-specific and widely expressed transcription factors that recognize distinct cis-acting DNA sequences within the control regions of tissue-specific genes. Skeletal muscle offers an attractive system for analyzing the mechanisms that govern tissue-specific gene transcription because myoblast differentiation is accompanied by the induction of a large set of muscle-specific genes, many of which have been cloned and characterized.

The commitment of multipotential stem cells to the myogenic lineage and subsequent induction of muscle-specific genes is a multistep process that is controlled in part by a family of myogenic regulatory factors, which includes MyoD (Davis et al., 1987), myogenin (Wright et al., 1989; Edmondson and Olson, 1989), myf5 (Braun et al., 1989) and MRF4 (Rhodes and Konieczny, 1989). These proteins share homology within a basic region and adjacent helix-loop-helix (HLH) motif and can activate muscle-specific genes when introduced into a variety of nonmuscle cell types (for reviews, see Olson, 1990; Weintraub et al., 1991). Cells of mesodermal origin, and in particular the multipotential mesodermal cell line C3H10T1/2 (10T1/2), are especially permissive to the actions of myogenic HLH proteins, whereas cells derived from other lineages show varying degrees of muscle gene expression in response to the myogenic regulators (Weintraub et al., 1989; Blau et al., 1985). Evidence suggests that the ability of a cell to express muscle-specific genes in response to myogenic HLH proteins is influenced by positive and negative factors that modulate the activities of these proteins (Weintraub et al., 1989; Schafer et al., 1990). The identities of such factors, however, remain largely unknown.

Among the factors that are required by myogenic HLH proteins for activation of muscle-specific transcription are the widely expressed Class A HLH proteins, which include E12, E47, and the HEB gene product (Murre et al., 1989a; Lassar et al., 1991; Hu et al., 1992). These proteins form heterooligomers with myogenic HLH proteins that bind with high affinity to a consensus sequence (CANNTG) known as an E-box, which is associated with most muscle-
specific genes. The regulatory elements responsible for muscle-specific transcription of the muscle creatine kinase (MCK) gene have been studied extensively. MCK transcription in skeletal muscle is dependent on a distal upstream enhancer that contains two E-boxes, which serve as binding sites for myogenenic HLH proteins (Lassar et al., 1989; Chakraborty et al., 1991a). Mutations within either of these sites dramatically diminish enhancer activity and prevent trans-activation by the myogenetic regulators (Buskin and Hauschka, 1989; Lassar et al., 1989; Brennan and Olson, 1990). Although E-boxes are essential for transcriptional activity of the MCK and numerous other muscle-specific enhancers, recent studies indicate that these sites alone are insufficient to account for full activity of these enhancers and that cooperativity among heterologous transcription factors is an important determinant of muscle-specific gene activation (Sartorelli et al., 1990; Lin et al., 1991; Gossett et al., 1989).

To identify regulatory factors that may cooperate with myogenenic HLH proteins to activate muscle-specific transcription, we have created a series of mutations within the MCK enhancer. In particular, we hoped to identify regulatory elements surrounding the E-box motifs that are essential for enhancer activity. Here we describe a novel homeodomain protein, MHox, which binds an A+T-rich element within the MCK enhancer core. The target sequence for MHox binding is required for muscle-specific transcription and trans-activation by myogenin and MyoD. MHox is expressed in a variety of mesodermally derived cell types during embryogenesis and in established cell lines of mesodermal origin. In adult mice, MHox expression is restricted primarily to skeletal muscle, heart and uterus. The pattern of MHox expression is unique among homeobox genes and suggests a role for MHox as a transcriptional regulator that acts in cell types of mesodermal origin at multiple stages of development.

**Materials and methods**

**Cell culture**

10T1/2 and HeLa cells were maintained in Dulbecco’s modified Eagle’s medium (DMEM) containing 10% fetal bovine serum (FBS). C2 cells were maintained in DMEM containing 20% FBS (growth medium, GM) and were induced to differentiate by transfer to differentiation medium (DM) containing 2% horse serum.

**Library screening**

MHox clones were initially isolated by oligonucleotide screening (Vinson et al., 1988) of an oligo(dt)-primed λgt 11 cDNA expression library generated from differentiated C2 muscle cells. The concatenated MHox binding site GGTATATTAAACCAG was used to probe the library using the same buffer composition used for gel shift analysis. Positive phage clones from the primary screen were tested for sequence-specificity of DNA binding using a mutant MHox sequence GGTATCTAGAACCAG and a concatenated MEF-2 site oligonucleotide GATCGCTCTAAATATACCCCTGTCG. One of the sequence-specific clones (10Q) was used to screen a randomly primed λgt 11 library generated from differentiated C2 cells in order to obtain overlapping and longer clones.

**DNA sequencing and sequence analysis**

Phage inserts were subcloned into Bluescript II (Stratagene) for sequence analysis. Nested deletions were generated in both directions by EcoRI-Mung bean digestion and both strands were sequenced by Sequenase Version 2 sequencing kits (United States Biochemical).

**In vitro transcription and translation**

In vitro transcription of MHox mRNA and translation in a rabbit reticulocyte lysate were performed as described (Brennan and Olson, 1990). The MHox cDNA was contained in Bluescript (Stratagene) and was linearized with HindIII prior to in vitro transcription with T3 polymerase. For analysis of the clone 10Q translation product, the EcoRI insert from the original λ gt11 clone was subcloned into Bluescript. [35S]methionine-labeled translation products were analyzed by SDS-PAGE followed by fluorography. For gel mobility shift assays, parallel translations were performed in the absence of [35S]methionine, and products were tested for DNA-binding activity using an end-labeled probe corresponding to the A+T-rich element from the MCK enhancer.

**Preparation of nuclear extracts for gel mobility shift assays**

Nuclear extracts were prepared from cultured cells as described (Gossett et al., 1989), and amounts of protein were determined by Bradford assay. Ten micrograms of nuclear extract was used in each gel mobility shift assay. Oligonucleotide probes were annealed and end-labeled with 32P or were used directly as competitors. For competition experiments, double-stranded oligonucleotides were included at a 1000-fold excess over the labeled probe. The sequences of the oligonucleotides used in competition experiments are shown in Fig. 6B. Additional flanking sequences that did not affect MHox DNA binding were also included in the oligonucleotides, which were generally about 25 bp in length.

**Mutagenesis**

Site-directed mutagenesis of the MCK enhancer was performed on a single-stranded template in Bluescript as described (Brennan et al., 1991), and mutations were confirmed by DNA sequencing. Mutant enhancers were subcloned into the BamHI site of pCK246CAT (Sternberg et al., 1988).

**Preparation of GST-MHox fusion protein**

The GST-MHox fusion protein was prepared by ligation of an EcoRI insert from cDNA clone 10Q in-frame with GST. This fusion protein encodes amino acids 9-217 of MHox. Fusion protein was purified on glutathione agarose beads as described (Chakraborty et al., 1991b) and was estimated to be greater than 90% pure based on Coomassie blue staining following separation by SDS-PAGE.

**RNA isolation and northern analysis**

Isolation of total cellular RNA from cells and tissues was performed by the guanidinium isothiocyanate procedure as described (Edmondson and Olson, 1989). For northern analysis, RNA was separated on 1% agarose gels, transferred to nitrocellulose, and hybridized with the MHox cDNA labeled by random priming. Ethidium bromide staining confirmed that equivalent quantities of RNA were applied to each lane.

**Transfections**

Transfections were performed by calcium phosphate precipitation as described (Sternberg et al., 1988). 24 hours following transfection, cultures were transferred to fresh GM and were harvested 24 hours later or were transferred at that time to DM for an additional 48 hours. Ten micrograms of each CAT reporter was used
for transfections of C2 cells. For CAT assays, RSV-lacZ, which contains the Rous sarcoma virus long terminal repeat linked to beta-galactosidase, was used as an internal control to normalize for differences in transfection efficiency. For trans-activation assays, 10T1/2 cells were transiently transfected with 5 µg of MCK-CAT reporter plasmid containing the wild-type or mutant enhancer and 5 µg of expression vector encoding myogenin or MyoD. The vector used to express the myogenic regulators was EMSV (Davis et al., 1987), which contains the Maloney sarcoma virus long terminal repeat.

In situ hybridization

Staged embryos were obtained from timed breedings of CD-1 mice (Charles River) and the morning of vaginal plugs was counted as 0.5 days post coitum (p.c.). Embryos were removed from the surrounding decidua and fixed in freshly prepared cold 4% paraformaldehyde in PBS for 16 hours. Tissue was then slowly dehydrated and embedded in paraffin following standard procedures as described by Sassoon et al. (1988). To date, we have used in situ hybridization to analyze a number of important developmental gene transcripts during mouse embryogenesis and have refined the resolution to detect transcripts at the intracellular level (Sassoon et al., 1988). Serial sections (5-7 µm) are collected individually on glass microscope “subbed” slides (Gall and Pardue, 1971). The procedures used for section treatment, hybridization, and washings are based on those used by Wilkinson et al. (1987): 16°C. The 95% deionized formamide, 0.3 M NaCl, 20 mM Tris-HCl (pH 7.4) 5 mM EDTA, 10 mM NaPO₄ (pH 8). 10% dextran sulfate, 1 x Denhart’s solution, 50 µg/ml total yeast RNA with 75,000 counts/minute/µl 35S-labeled RNA probe under siliconized coverslips. Probes were generated using T7 polymerase on a Bluescribe recombinant and [α35S] UTP (>1000 Ci/mmol, New England Nuclear). Coverslips were then floated off in 5 x SSC (1 x SSC is 0.15 M NaCl, 15 mM sodium nitrate) 10 mM dithiothreitol (DTT), at 50°C followed by a stringent wash at 65°C in 50% formamide, 2 x SSC, 0.1 M DTT. Slides were rinsed in washing buffer and treated with RNaseA (20 µg/ml; Sigma) and washed at 37°C in 2 x SSC and 0.1 x SSC for 15 minutes, respectively. Slides were then dehydrated rapidly and processed for standard autoradiography using NTB-2 Kodak emulsion and exposed for 7-9 days at 4°C. Analysis was carried out using both light and darkfield optics on a Leitz Orthoplan microscope.

The probe used for MHox was a cDNA fragment from position 471 to 732 subcloned into Bluescript minus (Stratagene). The plasmid was linearized with XhoI and antisense riboprobe was obtained using T3 polymerase.

Results

The MCK enhancer is composed of multiple regulatory elements that cooperate with the E-boxes to activate muscle-specific transcription

Previous studies have shown that the MCK 5’ enhancer lies between base pairs -1050 and -1350 upstream of the mouse MCK gene (Jaynes et al., 1988; Sternberg et al., 1988; Horlick and Benfield, 1989). Muscle specificity and high-level activity of this enhancer are dependent on two E-boxes located within the enhancer core (Buskin and Hauschka, 1988; Lassar et al., 1989; Brennan and Olson, 1990). These E-boxes serve as binding sites for heterooligomers formed between myogenin HLH proteins and Class A HLH proteins (Murine et al., 1989b; Chakraborty et al., 1991a). Mutagenesis of these sites results in a near-complete loss in transcriptional activity of the enhancer. To identify elements in the MCK enhancer that might cooperate with the E-boxes to enhance transcription, we created a series of deletion mutations of the MCK upstream region (Fig. 1A). These deletions revealed that the MCK enhancer contains a central core between base pairs -1204 and -1137. This enhancer core shows weak muscle-specific activity and is augmented by either of two peripheral activating regions located 5’ and 3’ to the core. The 5’ peripheral region contains a “CArG” box, which has been implicated in muscle-specific transcription of numerous contractile protein genes (Minty and Kedes, 1986), in addition to a consensus sequence for binding of AP-2 (Mitchell et al., 1987). The 3’ peripheral region, which has been characterized previously (Gossett et al., 1989; Cserjesi and Olson, 1991), contains a binding site for the myocyte-specific enhancer-binding factor MEF-2. Whereas these peripheral activating regions lack activity on their own, they synergize with the enhancer core to generate high levels of transcriptional activity.

The A+T-rich element in the MCK enhancer core binds multiple nuclear factors

To determine whether the A+T-rich element served as a binding site for factors that might cooperate with myogenin and MyoD to activate muscle-specific transcription, we performed gel mobility shift assays with a 32P-labeled oligonucleotide probe encompassing the A+T-rich element. As shown in Fig. 2, the A+T-rich probe gave rise to several DNA-protein complexes with nuclear extracts from 10T1/2
A. MCK Enhancer Core

B. MCK Enhancer Core

C. Transfected C2 cells

D. Transfected 10T1/2 cells
Fig. 1. The A+T-rich element in the MCK enhancer core is essential for muscle-specific transcription. (A) A schematic representation of the 5′ region of the mouse MCK gene is shown. Exons are indicated by black boxes (Jaynes et al., 1988; Sternberg et al., 1988). Numbers above the line designate kilobases of DNA relative to the transcription initiation site, designated 0. DNA fragments from the region between -1350 and -1048 were tested for enhancer activity by insertion into the BamHI site of the expression vector pCK246CAT, which contains the 246-bp MCK basal promoter immediately 5′ of CAT. The orientations of the DNA fragments with respect to the promoter are indicated + and – and the coordinates of each fragment are shown. Reporter genes were tested for activity following transient transfection into C2 myoblasts and myotubes as described in Materials and methods. CAT activities are expressed relative to the level of expression observed with 4800 bp of MCK 5′ flanking sequence linked to CAT. Values represent the average of at least three independent experiments and did not vary by greater than 20%. NT, not tested. Elements with homology to known regulatory elements are indicated with boxes. Detailed discussions of these elements have been presented (Jaynes et al., 1988; Sternberg et al., 1988; Buskin and Hauschka, 1989; Gossett et al., 1989). (B) The nucleotide sequence of the MCK enhancer core, as defined by deletion analysis in A is shown. The left and right E-boxes and the A+T-rich element are indicated. Sequence comparison between the MCK MEF-2 site, the A+T-rich element, and the MEF-2 consensus is shown. The 4-bp substitution mutation within the MEF-2 core. Nucleotides beneath the MEF2 consensus indicate changes in the MCK MEF2 site compatible with MEF2 binding. Lower case letters denote diminished binding of MEF2 with that nucleotide substitution. (C) C2 cells were transiently transfected with pCK246CAT containing the wild-type enhancer (-1350/ -1048) or the enhancer with the A+T-rich mutation shown in B (A+T-mutant). Levels of CAT activity were assayed in myoblasts and myotubes and were normalized to the level of beta-galactosidase activity from cotransfected RSV-lacZ. (D) 10T1/2 cells were transiently transfected with the expression vectors described in C and EMSV, EMSV-myogenin, or EMSV-MyoD. 48 hours following transfection, cultures were transferred to DM and CAT activity was assayed 48 hours later. Values are expressed as the amount of CAT activity under each condition relative to the amount observed in extracts from cells transfected with the wild-type MCK expression vector and EMSV-myogenin. Values represent the average of at least three independent experiments.

cells, C2 myoblasts and myotubes, and HeLa cells. Complex-1, which migrated near the top of the gel, was observed with all extracts tested. The formation of this complex was inhibited efficiently by the unlabeled A+T-rich element and weakly by the MEF-2 site. Near the center of the gel, we observed a heterogeneous set of abundant complexes with extracts from 10T1/2 and C2 cells. These complexes were most abundant with nuclear extracts from 10T1/2 cells and were undetectable with extracts from HeLa cells. They were abolished in the presence of homologous competitor DNA but were unaffected by oligonucleotides corresponding to the MEF-2 site. As predicted from the previously derived consensus for MEF-2 binding (Cserjesi and Olson, 1991), the A+T-rich element bound MEF-2 weakly; the MEF-2 complex appeared above complex-1 with extracts from C2 myotubes (Fig. 2). The amount of binding of MEF-2 to this probe, however, is less than one-tenth the level observed with the MCK MEF-2 site. There was no competition for any of the above complexes by an oligonucleotide corresponding to the mutant A+T-rich element that impaired enhancer activity (data not shown). Together, these experiments demonstrate that the A+T-rich element in the MCK enhancer core is a weak binding site for MEF-2 and an as-yet-identified protein that forms complex-1, and a strong site for factors that are present in 10T1/2 and C2 cells but absent in HeLa cells.

Isolation of a cDNA encoding a novel homeodomain protein that binds the A+T-rich element

To investigate further the nature of the factors that interacted with the A+T-rich element, we screened a C2 myotube cDNA library in λ gt11 for recombinant proteins that could bind a labeled probe containing the multimerized A+T-rich element. Screening of 1.2 × 10⁶ independent phage plaques revealed four clones that encoded proteins able to recognize the A+T-rich element but not the MEF-2 site or the mutant A+T-rich element. By cross-hybridization and partial DNA sequencing, we determined that these clones overlapped. The longest of the clones, designated clone 10Q, was sequenced in its entirety and contained a single uninterrupted open reading frame of 208 amino acids in-frame with the beta-galactosidase coding sequences. To obtain the complete coding sequence of the A+T-rich binding factor, several additional cDNAs were isolated by screening a random-primed C2 myotube cDNA library using clone 10Q as a probe. Sequence analysis of multiple cDNAs revealed an open reading frame with the potential to encode a 217 amino acid polypeptide of predicted Mr 24,371 (Fig. 3A). The AUG at the beginning of the open reading frame was followed by a second in-frame AUG 48 amino acids downstream. A polypeptide initiating at this second AUG would possess a predicted Mr of 19,320. Both AUs were preceded by sequences that fit the predictions for translation initiation (Kozak, 1984).

The most striking feature of the deduced protein sequence of the A+T-rich-binding factor is the presence of a homeodomain near the center of the open reading frame (Fig. 3B). We have therefore designated this protein MHox, for muscle homeobox protein. The MHox homeodomain is closely related to the homeodomain of the mouse homeobox gene S8, which was identified as a partial cDNA sequence by degenerate PCR amplification from hematopoietic cell lines (Kongsuwan et al., 1988). Within the homeodomain, MHox and S8 differ at only two residues, both of which are conservative substitutions. At the nucleotide level, MHox and S8 are 79% identical within the homeodomain, with the divergent nucleotides primarily at the wobble positions of the codons, conserving amino acid sequence identity. This divergence clearly indicates that MHox and S8 do not arise by alternative splicing of a common transcript. The complete open reading frame for the mouse S8 protein has not yet been reported (Opstelten et al., 1991), but comparison of the existing sequence with that of MHox showed stretches of high homology carboxy terminal to the homeodomain (Fig. 3C). Within the 15 amino acids of sequence available for the region of S8 amino terminal to the homeodomain, there is conservation...
among the five basic residues immediately adjacent to the homeodomain, but no detectable homology amino terminal to these residues.

The sequences of their homeodomains suggest that MHox and S8 belong to a novel class of homeodomain proteins. Within the homeodomain, MHox shows greatest similarity to the paired family of Drosophila homeobox genes, which includes paired (prd) (Bopp et al., 1986), pax-3 (Goulding et al., 1991), and gooseberry-proximal and -distal (gsb-p, -d) (Cote et al., 1987), but it differs from these proteins by a glutamine-for-serine substitution at position 9 of the recognition helix (Fig. 3C). MHox also lacks a paired box, which is amino terminal to the homeodomain in members of the paired family (reviewed in Walther et al., 1991). The recognition helix of MHox is identical to that of Mix-1 (Rosa, 1989), Hox-7 (Robert et al., 1989) and Antennapedia (McGinnis et al., 1984). Since this portion of the homeodomain determines sequence specificity of DNA binding (Desplan et al., 1988), it is tempting to speculate that these proteins may exhibit similar DNA-binding specificities.

The region of MHox downstream of the homeodomain is rich in hydroxylated amino acids (serine, threonine, tyrosine) and proline and resembles activation domains in many other trans-activator proteins (Mermod et al., 1989; Schwarz et al., 1992). Within this region is a consensus sequence for phosphorylation by proline-dependent protein kinase (SP/TP) at residues 197 (Vulliet et al., 1989).

The MHox cDNA encodes a protein with the same DNA-binding properties observed for an A+T-rich binding factor in nuclear extracts

To investigate the properties of the MHox protein, RNA was transcribed from an MHox cDNA template encompassing the complete open reading frame (designated MHox - full length) and was translated in a rabbit reticulocyte lysate in the presence of [35S]methionine. Analysis of the in vitro translation products by SDS-PAGE revealed two predominant polypeptides of Mr ~32,000 and ~35,000, in a ratio of about 4:1, respectively (Fig. 4A). These sizes are slightly larger than the predicted Mr s of polypeptides that would be initiated from the first two in-frame AUGs. We refer to these two MHox protein products as MHox-long and MHox-short. Smaller products observed in the in vitro translation system occur variably and are likely to arise as a result of premature translation termination or initiation at AUG’s that are more 3’. To assess further the relative contributions of the two most 5’ AUGs to translation initiation, we transcribed and translated the clone 10Q cDNA (designated MHox-Δmet1), which lacks the more 5’ methionine. This gave rise to the 32,000 species, but not the 35,000 species, suggesting that MHox-long and MHox-short arise by differential translation initiation from the two in-frame AUGs.

To examine the DNA-binding properties of MHox, we performed gel mobility shift assays with a labeled probe encompassing the A+T-rich element and in vitro transla-
tion products obtained from the MHoX cDNA. Translation products derived from the cDNA containing both in-frame methionines (MHoX - full length) gave rise to two DNA-protein complexes (Fig. 4B), the more slowly migrating complex being the more abundant. With the MHoX-short species alone (MHoX-Δmet1), we observed only the faster migrating complex. We interpret these results to indicate that the two DNA-protein complexes reflect the binding of the two MHoX proteins. We noted that the ratio of the two complexes observed with the 32,000 and 35,000 species did not reflect the relative ratios of the proteins. Whereas the 32,000 species was more abundant in the in vitro translation reactions, the larger complex was more abundant in gel mobility shift assays, suggesting that the larger protein may bind DNA more avidly. The lack of a DNA-protein complex of mobility intermediate to the large and small proteins together suggests that MHoX binds DNA as a monomer. DNA binding by the in vitro translation products was abolished by the homologous competitor DNA, but not by the MEF-2 site (Fig. 4B).

By comparison of the relative migration of DNA-protein complexes generated with nuclear extracts and the MHoX in vitro translation products, it appears that the heterogeneous complexes of intermediate mobility observed with nuclear extracts from 10T1/2 and C2 cells represent the binding of MHoX (Fig. 4B; see also Fig. 2). The pattern of mobility-shifted complexes observed with nuclear extracts suggests that both forms of the MHoX protein are expressed in vivo, with the long form being predominant. Recent experiments have shown that transfection of HeLa cells with MHoX cDNA produces a similar pattern of DNA-protein complexes, indicating that MHoX is able to bind DNA in vivo.

Fig. 3. cDNA sequence and predicted open reading frame for MHoX. (A) The nucleotide sequence and deduced open reading frame for MHoX is shown. The homeodomain is underlined and the two in-frame AUGs that serve as sites for translation initiation are numbered 1 and 48. Nucleotides are numbered relative to the start of translation. (B) Amino acid homology between MHoX and S8. Lines and dots between the sequences designate identity and conservative substitutions, respectively. (C) Comparison of the homeodomain region of MHoX with those of other homeodomain proteins showing the most similarity. Sources of sequences are: S8 (Opstelten et al., 1991), Pax-3 (Goulding et al., 1991), prd (Bopp et al., 1986), gsb-p and gsb-d (Cote et al., 1987), otd (Finkelstein et al., 1990), Pax-6 (Walther and Gruss, 1991), Pax-7 (Jostes et al., 1990), Mix-1 (Rosa, 1989), Hox-7 (Robert et al., 1989), and Ant (McGinnis et al., 1984). A dash indicates amino acid identity.
with the full-length MHox cDNA gives rise to the heterogeneous complexes, whereas the MHox-Δmet1 cDNA gives rise only to the faster migrating species (data not shown), thereby confirming the identities of the proteins contained in these complexes.

Sequence-specificity of MHox DNA binding
To define further MHox DNA-binding properties, we expressed it as a bacterial fusion protein with glutathione-S-transferase (GST). Gel mobility shift assays with GST-MHox and the labeled A+T-rich element in the presence of a variety of A+T-rich oligonucleotide competitors showed that MHox can recognize a 9-bp sequence with A or T at every position (Fig. 5A). MHox, does not, however, bind indiscriminantly to A+T-rich sequences. The specificity of DNA sequence recognition is demonstrated by comparison of competitions with mutant 5 and MCK MEF-2, which differ by only a single A-to-T substitution. Whereas mutant 5 is an efficient competitor, the MEF-2 site does not compete for MHox DNA binding. A preliminary consensus sequence for the MHox DNA-binding site, based on these competition experiments, is shown in Fig. 5B. This consensus sequence contains tandem repeats of the conserved ATTA core sequence, which is present within the binding
MHox: a muscle homeodomain protein

MHox mRNA is enriched in muscle tissue

Analysis of MHox mRNA expression in a variety of established cell lines revealed its presence in 10T1/2 fibroblasts, as well as in C2 myoblasts and myotubes (Fig. 6A). The MHox cDNA hybridized to two major transcripts of approximately 3.6 and 4.0 kb. Additional smaller transcripts were observed upon longer exposure of northern blots to film (data not shown), suggesting that processing of MHox mRNA may be complex. The large sizes of the MHox transcripts indicate that additional sequence at the 5′ or 3′ ends is missing from the cDNA clones sequenced thus far. The MHox transcripts are considerably larger and more complex than the S8 mRNA, which exists as a single species of about 1.5 kb (Opstelten et al., 1991). MHox mRNA was more abundant in 10T1/2 cells than in C2 cells, and there appeared to be little or no apparent change in the level of MHox expression during myoblast differentiation. We do not detect MHox mRNA in HeLa cells (Fig. 6A), in agreement with the lack of the MHox complex in HeLa cell nuclear extracts (Fig. 2). MHox was also undetectable in HepG2 liver cells, F9 cells, or cos-1 cells (data not shown).

The pattern of MHox mRNA expression was also examined in adult mouse tissues by northern analysis (Fig. 6B). MHox transcripts were most abundant in skeletal muscle, heart and uterus, and were undetectable in most other tissues, with the exception of lung, where there was a trace of the transcripts.

MHox is expressed primarily in mesenchyme of developing mouse embryos

We examined the pattern of MHox expression using in situ hybridization on mouse embryo sections representing the entirety of postimplantation development (Fig. 7). MHox transcript was first detectable at about 9 days p.c. in the lat-
eral mesoderm and newly formed visceral arches. By 9.5 days p.c., strong expression of MHox was detected in the visceral arches and newly formed limb buds (see Fig. 7A-C). Expression in the somites was confined primarily to the dorsal dermamyotome (B), which is consistent with the strong expression of MHox in undifferentiated connective tissue and skeletal muscle. The dorsal aorta, which is primarily smooth muscle, also showed detectable levels of MHox (Fig. 7C). It has been previously reported that this structure expresses cardiac actin (Sassoon et al., 1988), whereas it never expresses the HLH myogenic factors. Expression remained highest in the facial and limb structures (7D-F, H), with lower levels of expression in the condensing vertebrae and ribs of the axial skeleton (7F,G). The intercostal muscles and diaphragm showed moderate levels of MHox expression by 15.5 days p.c., suggesting that MHox may exert a role in the axial muscles during late-fetal growth (7G). Although levels had generally decreased overall in the embryo by 15.5 days p.c., strong levels were still detected in the limbs (7H) and face (data not shown), with labeling being restricted to mesenchyme and peri-

chondrial tissues that had not yet differentiated in these structures. Whereas most homeobox genes are expressed in the central nervous system, we note a complete absence of MHox in brain and spinal cord. Thus, MHox expression is mesodermally restricted throughout development.

**Discussion**

Myogenic HLH proteins activate the MCK enhancer by binding to two E-boxes in the enhancer core (Lassar et al., 1989; Brennan et al., 1991; Chakraborty et al., 1991a). To determine whether these E-boxes were sufficient for transcriptional activation of the MCK enhancer and to identify target sequences for transcription factors that might cooperate with myogenic HLH proteins to activate muscle-specific transcription, we introduced mutations into regions of the MCK enhancer surrounding the E-boxes. This approach led to the identification of an A+T-rich element in the enhancer core that is essential for enhancer activation in muscle cells. This element serves as a binding site
for a novel homeodomain protein, MHox, that appears to be restricted to cells of mesodermal origin in vivo and in vitro.

MHox is a sequence-specific DNA binding protein

There are relatively few examples of vertebrate homeodomain proteins for which the target genes have been identified. Our results suggest that MCK may be a target gene for transcriptional activation by MHox. Consistent with this conclusion is the codistribution of MCK and MHox to adult skeletal muscle and heart. A role for MHox as an activator of the MCK enhancer is also supported by the loss in enhancer activity in skeletal muscle (this study) and cardiac myocytes (S. Amacher and S. Hauschka, personal communication).

Fig. 7. MHox expression in the mouse embryo detected by in situ hybridization. (A) Parasagittal view with dark-field illumination showing MHox expression in a 9.5 days post coitum embryo. Strong signal is detected over the visceral arches (a) and facial primordia. Note that at this stage and all other stages examined, no signal is ever detected in the neural tube (brain and spinal cord). (B) Parasagittal view with dark-field illumination showing MHox expression in a 9.5 days post coitum embryo. As shown in panel A, strong expression is noted in the visceral arches (a) and the frontonasal processes (fn). Also, expression in the somites (s) is observed and is confined to the dermamyotome (dorsal portion). (C) Parasagittal view with dark-field illumination showing MHox expression in a 10.5 days post coitum embryo. Again, strong expression is observed in the visceral arches and facial primordia. Labeling is also noted in the developing limb buds (lb). The dorsal aorta (da) is also labeled suggesting MHox expression in early smooth muscle. (D) Frontal (horizontal) view with dark-field illumination showing MHox expression in a 13.5 days post coitum embryo. Labeling is notable in the mandibles (m) which are derived from the visceral arches. In addition, labeling is noted in the developing eye (e) and the meninges (mn) of the brain (b). Again, no detectable levels of MHox are seen in the brain (b), telencephalon (brain). (E,F) Parasagittal view with dark-field illumination showing MHox expression in a 13.5 days post coitum embryo. Again, no detectable levels of MHox are seen in the brain (b), telencephalon (brain). (E,F) Parasagittal view with dark-field illumination showing MHox expression in a 13.5 days post coitum embryo. Again, no detectable levels of MHox are seen in the brain (b), telencephalon (brain). (E,F) Parasagittal view with dark-field illumination showing MHox expression in a 13.5 days post coitum embryo. Again, no detectable levels of MHox are seen in the brain (b), telencephalon (brain). (G) High magnification parasagittal view with dark-field illumination showing MHox expression in a 15.5 days post coitum embryo. The liver (lv) shows no detectable signal, whereas the perichondrial tissue surrounding the ribs (r), the intercostal muscles (ic) and the diaphragm (d), show moderate levels of MHox expression. (H) Horizontal section through the hind limb of a 15.5 days post coitum embryo. Strong labeling is noted in the distal tips of the phalanges of the digits as well as in the perichondrial tissue surrounding the developing skeleton (see arrows). At this stage, MHox is generally low in the embryo and is only detectable in the face and limbs and axial muscles.
munication) when the MHOx binding site is mutated. The MHOx binding site is conserved in the mouse (Jaynes et al., 1988; Sternberg et al., 1988), human (Trask et al., 1988), rat (Horlick and Benfield, 1989), and rabbit (Yi et al., 1991) MCK enhancers, further indicating its importance for transcriptional activity. It should be emphasized, however, that we have not yet directly demonstrated that MHOx functions as a transcriptional activator. Indeed, the existence of other factors that bind the MHOx site (Fig. 2) raises the formal possibility that proteins other than MHOx may be necessary for MCK gene expression and MHOx may play either a positive or negative role. In this regard, we have attempted to determine whether exogenous MHOx can potentiate the ability of myogenin and MyoD to trans-activate the MCK enhancer, but have observed no transcriptional potentiation in response to MHOx in 10T1/2, HeLa, or HepG2 cells (P.C., B.L., and E.O., unpublished observations). The lack of transcriptional activity of exogenous MHOx in 10T1/2 cells could potentially reflect saturating levels of endogenous MHOx protein in these cells. The lack of activity of exogenous MHOx in HeLa and HepG2 cells, which do not express endogenous MHOx, is difficult to interpret because myogenin and MyoD are inactive in these cell types (Weintraub et al., 1989; Schafer et al., 1990; Schwarz et al., 1992) and would therefore be unable to collaborate with MHOx to activate the MCK enhancer. Further analyses will be required to fully define the exact role of MHOx in the control of MCK transcription.

Our results indicate that the MHOx site in the MCK enhancer is a weak site for MEF-2, as well as at least one other unidentified nuclear factor. MEF-2 has recently been shown to belong to the MADS box gene family, which includes serum response factor (SRF) and several transcriptional activators from yeast and plants (see Pollock and Treisman, 1991 and references therein). The ability of MEF-2 to bind the MHOx site is intriguing given the demonstrated interactions between homeodomain proteins and MADS box-containing proteins in yeast. In budding yeast, genes involved in mating type switching are regulated in part by MCM1, which contains a MADS box similar to that of SRF (for review, see Herskowitz, 1989). MCM1 binds a target sequence upstream of the mating type genes and recruits different accessory proteins to the DNA resulting in activation or repression of transcription. Among the proteins recruited by MCM1 is the homeodomain protein MATa1. Studies are underway to further investigate the significance of MEF-2 binding to the MHOx site and whether there is a physical or functional interaction among these proteins analogous to that observed with MCM1 and MATa1.

While the MHOx binding site appears to be essential for transcriptional activation of the MCK enhancer, the presence of MHOx in 10T1/2 cells and a variety of other mesenchymal cells in vivo, which do not express MCK, shows that MHOx is unable to induce the myogenic program or activate the MCK enhancer alone. Previous studies, which showed that the E-boxes in the MCK enhancer were essential for enhancer activity (Buskin and Haushka, 1989; Lassar et al., 1989; Brennan and Olson, 1990), also demonstrate that MHOx alone cannot activate transcription from this enhancer and reveal an interdependence between myogenic HLH proteins and MHOx (or other factors that bind the same site) for transcriptional activation of the MCK enhancer. This apparent cooperativity between the MHOx binding site and the adjacent E-boxes in the MCK enhancer is the first example of a possible regulatory interaction between homeodomain proteins and myogenic HLH proteins and raises the possibility that other cell type-specific enhancers may be controlled by similar combinations of transcription factors from these two families. Indeed, sites resembling the MHOx binding site are present in several other muscle promoters and enhancers (P.C., B.L. and E.O., unpublished observations).

Gel mobility shift assays using a variety of A+T-rich DNA probes indicate that MHOx exhibits a high degree of specificity in DNA sequence recognition. We do not yet know the full spectrum of sequences recognized by MHOx, but competition experiments suggest that MHOx recognizes a 9-bp element containing A or T at every position, with up to seven nucleotides being invariant. The MHOx recognition sequence contains the ATTA motif, which is found in the binding sites of other homeodomain proteins (Desplan et al., 1988) and is remarkably similar to the binding site for the homeodomain protein HNF-1 (GTTATNATTAAC), which is involved in activation of liver-specific genes (for review, see Mendel and Crabtree, 1991). HNF-1 from liver cell nuclear extracts can, in fact, bind the MCK MHOx site (P.C. and E.O., unpublished results). However, since myogenin and MyoD cannot activate the MCK enhancer in transfected HepG2 liver cells, which express high levels of HNF-1 (P.C. and E.O., unpublished results), it appears that HNF-1 is unable to collaborate with myogenic HLH proteins to activate muscle transcription.

Myogenic HLH proteins rely on other regulatory factors for activation of myogenesis

Previous studies have shown that the ability of myogenin or MyoD to activate muscle-specific transcription is dependent on the cell background, suggesting that the actions of these myogenic regulators are influenced by other cellular factors (Weintraub et al., 1989; Schafer et al., 1990; see also Blau et al., 1985). With the exception of other HLH proteins, which interact directly with members of the MyoD family to modulate their activity (Benezra et al., 1990; Lassar et al., 1991), the identities of factors that determine whether a cell can activate muscle-specific genes in response to myogenic HLH proteins are largely unknown. Our results, which show that MHOx is absent from cell types such as HeLa and HepG2, which are refractory to the actions of myogenin and MyoD, suggests that MHOx may contribute to the establishment of a cell background permissive for muscle gene activation. MHOx could potentially modulate the activity of the myogenic regulators by collaborating with them directly to activate muscle-specific genes, as may be the case with MCK, or it could act indirectly by regulating expression of other genes that influence the myogenic program.

MHOx expression is restricted to mesenchymal cells during embryogenesis and is enriched in muscle tissue of adult

Homeobox genes have been studied most extensively in Drosophila, where they are involved in commitment of cells
to specific developmental pathways and play an important role in pattern formation (reviewed in Levine and Hoey, 1988). In vertebrates, the majority of homeobox genes are expressed in a rostral-to-caudal gradient and also participate in morphogenesis and establishment of the body plan (reviewed in Shashikant et al., 1991). The spatial pattern of MHOX expression during development is distinct from that of most other homeobox genes and suggests a novel role for MHOX in mouse development. MHOX appears to be expressed exclusively in mesodermal derived cell types, with the highest levels of expression in the embryonic mesenchyme and precartilage (perichondrial) elements of the face and limbs. MHOX is also observed in the dorsal aorta of the 10 days p.c. embryo. This suggests that MHOX plays a role in developing smooth muscle. Unlike the majority of homeobox genes, MHOX is never observed in the developing central nervous system (brain and spinal cord). MHOX expression is also absent from most internal organs. Thus, MHOX most likely acts as a regulator of mesodermal development in the embryo. The distribution of MHOX is highly reminiscent of S8 distribution (Ostptlen et al., 1991), as well as the msh class of HOX genes which includes HOX-7.1 (Hill et al., 1989; Robert et al., 1989). The distribution of MHOX, S8 and HOX-7.1 also differs from the Antennapedia class of homeobox genes (Hox-1-6), which display rostral-caudal boundaries in their expression in the embryo. In particular, the expression of MHOX in the visceral arches appears highest in the distal mesenchyme. This expression can be correlated to regions of high proliferation in these structures (Summerbell et al., 1973).

The pattern of MHOX expression during development is consistent with a potential role as a regulator of MCK expression. Developmental in situ analysis of MCK expression in the mouse has revealed that MCK is first detected four to five days after the initial appearance of MyoD1 and myogenin (Lyons et al., 1991). Thus, although the myogenic factors are capable of driving expression of reporter genes linked to MCK regulatory elements in cultured cells, the regulation of MCK expression is clearly more complex in utero. MCK transcripts are observed in muscle by 13.5 days of embryonic development and undergo a dramatic rise in levels by 15.5 days which coincides with the onset of MHOX expression in skeletal muscle (D.A. Sassoon, unpublished data). We suggest, therefore, that MHOX may participate in the more complex regulation of MCK during development.

Whereas several homeobox genes have been identified in skeletal muscle of Drosophila (Dohrmann et al., 1990; Bodmer et al., 1990; Barad et al., 1991), MHOX is to our knowledge the only homeobox gene described thus far which is expressed in skeletal muscle and heart of mammals. Its presence in uterus, as well as embryonic aorta, also suggests that MHOX is found in smooth muscle. Although the three muscle cell types express many of the same muscle-specific genes and are all derived from mesoderm, they have different embryonic origins. Skeletal muscle first appears in mouse at about embryonic day 8.5 in the somite myotome (Sassoon et al., 1989); whereas cardiac muscle appears several days earlier and is derived from lateral plate mesoderm. Smooth muscle contributes to numerous tissues, including uterus, stomach, intestine and the vasculature. Its embryonic origins are less well understood and are likely to vary depending on the tissue to which it contributes. It is interesting that MHOX mRNA is highly abundant in uterus, which is composed predominantly of smooth muscle, but it is undetectable in intestine and stomach, which also contain smooth muscle. Whether this reflects the existence of different types of smooth muscle cells or heterogeneity of cell types in these different tissues remains to be determined.

The apparent restriction of MHOX to mesenchymal cells during development and to muscle tissues of the adult suggests that MHOX represents a lineage marker for cells of mesodermal origin and that its expression and function become progressively restricted during ontogeny. Future analyses of the control of MHOX expression should yield insight into mechanisms involved in lineage-specific regulation of gene expression and of the role of MHOX in developmental patterns of mesodermal gene transcription.

This work was supported by grants AR40339 and AR39849 from the National Institutes of Health and the Muscular Dystrophy Association to E.N.O., and The Council for Tobacco Research - U.S.A., Inc. to D.A.S. B.L. is an NIH Predoctoral Fellow. E.N.O. is an Established Investigator of the American Heart Association. Technical assistance of J. Toma and S. Jasser and the secretarial assistance of Kathy Tucker is appreciated. L. Garrett, X. Luo and R. Burgess participated in initial characterization of MHOX cDNAs and are also gratefully acknowledged.

References
Chakraborty, T., Brennan, T. J., Li, L., Edmondson, D. and Olson, E. N. (1991b). Inefficient homo-oligomerization contributes to the
dependence of myogenin on E2A products for efficient DNA binding. 

Carter, A., Preuss, H., Hall, J., Schuh, R., Kienlin, A., Seifert, E. and 
genes encode different spatially restricted transcripts in the embryo.
*EMBO J.* 6, 2793-2801.

Lassar, A. B., Buskin, J. N., Lockshon, D., Davis, R. L., Apone, S., 


MHox: a muscle homeodomain protein


(Accepted 1 May 1992)

Note added in proof: We have recently become aware of a human homologue of MHox. phox enhances the DNA-binding activity of serum response factor (SRF) and potentiates transcriptional activation by SRF. Grueneberg, Natesan, Alexandre and Gilman, *Science* (in press).