MOLECULAR AND CELLULAR BIOLOGY, Mar. 2004, p. 2132–2143 Vol. 24, No. 5
0270-7306/04/$08.00+0 DOI: 10.1128/MCB.24.5.2132–2143.2004

Quantitative Proteomic Identification of Six4 as the Trex-Binding Factor in the Muscle Creatine Kinase Enhancer

Charis L. Himeda,[1] Jeffrey A. Ranish,[2] John C. Angello,[1] Pascal Maire,[3] Ruedi Aebersold,[2] and Stephen D. Hauschka[1]*

Department of Biochemistry, University of Washington, Seattle, Washington 98195[1]; Institute for Systems Biology, Seattle, Washington 98103-8904[2]; and Departement Genetique, Developpement et Pathologie Moleculaire, Institut Cochin, INSERM 567, CNRS UMR 8104, Universite Paris V, 75014 Paris, France[3]

Received 18 July 2003/Returned for modification 9 September 2003/Accepted 5 December 2003

Transcriptional regulatory element X (Trex) is a positive control site within the *Muscle creatine kinase* (*MCK*) enhancer. Cell culture and transgenic studies indicate that the Trex site is important for *MCK* expression in skeletal and cardiac muscle. After selectively enriching for the Trex-binding factor (TrexBF) using magnetic beads coupled to oligonucleotides containing either wild-type or mutant Trex sites, quantitative proteomics was used to identify TrexBF as Six4, a homeodomain transcription factor of the Six/sine oculis family, from a background of ~900 copurifying proteins. Using gel shift assays and Six-specific antisera, we demonstrated that Six4 is TrexBF in mouse skeletal myocytes and embryonic day 10 chick skeletal and cardiac muscle, while Six5 is the major TrexBF in adult mouse heart. In cotransfection studies, Six4 transactivates the *MCK* enhancer as well as muscle-specific regulatory regions of *Aldolase A* and *Cardiac troponin C* via Trex/MEF3 sites. Our results are consistent with Six4 being a key regulator of muscle gene expression in adult skeletal muscle and in developing striated muscle. The Trex/MEF3 composite sequence ([C/A]ACC[C/T]GA) allowed us to identify novel putative Six-binding sites in six other muscle genes. Our proteomics strategy will be useful for identifying transcription factors from complex mixtures using only defined DNA fragments for purification.

The mouse *Muscle creatine kinase* (*MCK*) gene has served as a useful paradigm for understanding muscle gene regulation, as *MCK* expression is restricted to terminally differentiated skeletal and cardiac muscle. Several major regulatory regions have been identified in the *MCK* gene, including a 206-bp enhancer located from −1256 to −1050 bp upstream of the transcription start site. The *MCK* enhancer confers muscle-specific expression to reporter constructs in cell culture, transgenic mice, and virus-mediated gene transfer (1, 10, 20, 24, 25, 44, 47), and it contains at least seven control elements which were identified by footprinting, deletion-mutation analysis, and gel mobility shift assays (1, 3, 12, 29, 34). The relative positions and sequences of control elements (CArG/SRE, AP2, Trex, A/T-rich, left and right E-boxes, and MEF2) within the enhancer are highly conserved between mammalian species (22, 24, 50, 53). Many of the transcription factors associating with these control elements have been identified, but prior to this study the Trex-binding factor (TrexBF) was unknown.

The Trex site is not a good match to any known transcription factor-binding site using database searches; however, it is required for full transcriptional activity of the *MCK* enhancer in both skeletal and cardiac muscle, as assessed in cell culture and in transgenic mice (12, 36). In gel mobility shift assays, a Trex-specific binding activity (TrexBF) has been detected in nuclear extracts from a wide variety of cultured vertebrate cells, adult mouse tissues, and chicken embryo stages (C. L. Himeda, C. Fabre-Suver, and S. D. Hauschka, unpublished data). TrexBF from the majority of tissues and cell types tested migrates with the same apparent mobility in gel shift assays. Attempts to identify TrexBF by screening a mouse MM14 skeletal muscle cDNA library and human aortic smooth muscle expression library, as well as yeast one-hybrid screening of a HeLa cDNA library, were unsuccessful (C. Fabre-Suver, C. Rotermund, and S. D. Hauschka, unpublished data).

Here we have partially purified TrexBF from HeLa cells using DNA-affinity enrichment and have used a quantitative proteomics strategy (42) to identify TrexBF as Six4 from a background of ~900 copurifying proteins. We further confirm that Six4 is able to bind and transactivate gene expression from the *MCK* Trex site, as well as from MEF3 sites (see below) in the regulatory regions of other muscle-specific genes. Since the *MCK* Trex sequence does not precisely conform to the established MEF3 consensus, our studies modify and broaden the consensus binding motif for Six proteins, thereby facilitating the identification of MEF3/Trex control elements in other genes.

The MEF3 motif, while not originally identified in the *MCK* enhancer or promoter, is a highly conserved sequence (TCA GGTT) found in the regulatory regions of many muscle-specific genes and shown to be important in regulating *Myogenin*, *Aldolase A*, and *Cardiac troponin C* genes (6, 21, 43, 45, 46). MEF3 elements are recognized by Six proteins, mammalian homologues of the Sine oculis (So) family of homeodomain transcription factors. Six4 was the first identified So homologue, originally discovered as the ARE (Na/K-ATPase α1 subunit gene regulatory element) binding factor AREC3 (27, 49). The ARE site contains a sequence that matches the MEF3 motif, and both Six1 and Six4 are capable of binding and transactivating gene expression from MEF3 sites (45). In adult mouse tissues, *Six4* is expressed predominantly in skeletal muscle (27); however, Six4 protein has also been detected in the

* Corresponding author. Mailing address: Department of Biochemistry, Box 357350, University of Washington, Seattle, WA 98195. Phone: (206) 543-1797. Fax: (206) 685-1792. E-mail: haus@u.washington.edu.

developing somites, retina, nervous system, and lung, as well as in a variety of cultured cell lines (15, 37, 39).

This is the first report in which a mammalian transcription factor has been identified by quantitative proteomics. This technique should be widely applicable to the identification of a broad range of difficult-to-purify DNA-binding factors.

MATERIALS AND METHODS

Plasmids and antibodies. Reporter plasmids -enh80MCKCAT, (Trex-mt)-enh80MCKCAT containing the M1 mutation of the Trex site, 1256MCKCAT, and pUCSV2PAP have been described previously (1, 12). Reporter plasmids pM202CAT containing the rat *Aldolase A* promoter M and pM202CAT(mut C1) containing a MEF3 site mutation have been described (21), as have the reporter plasmid pHcTnC4500CAT/I1-500-S containing intron 1 and 4,500 bp of upstream sequences of human *Cardiac troponin C* (6) and the MEF2C expression construct (33). The MyoD expression construct pEMCIIs and the control vector pEMSVscribeα2 containing the simian virus 40 promoter and lacking the MyoD coding sequence were obtained from A. Lassar. The construct containing the full-length mouse Six4 cDNA in the pCR3 expression vector and the production of rabbit polyclonal antibodies to mouse Six4 have been described elsewhere (45). Rabbit polyclonal antibodies to human Six5 were described previously (15). Three tandem copies of the wild-type Trex site (GGACACCCGAGATGCCTG GTTA) from the mouse *MCK* enhancer and five tandem copies of the M1 mutant Trex site (GGACAGGTAACCACTCTGGTTA) were blunt-end ligated in the *Sal*I site of TKCAT (1) to generate $(Trex)_6$TKCAT and $(Trex\text{-}mt)_5$TKCAT. The human *Troponin I* slow upstream enhancer was excised from hTnIsUSE/-2000hSkeletalActin/CAT (gift of E. Hardeman) with *Bgl*II and *Hin*dIII and blunt-end ligated in the *Hin*dIII site of +enh358MCKCAT (47). The enhancer was then excised from this construct with *Hin*dIII and *Sph*I and sticky ligated in 358MCKCAT (45) to generate TropI.

Cell culture. HeLa cells adapted for growth in suspension (HeLa S) were grown in RPMI medium (Sigma) supplemented with 10% fetal bovine serum and antibiotics (100 U of penicillin/ml and 0.1 mg of streptomycin/ml) in 500-ml spinner flasks to a density of $\sim 10^6$ cells/ml. Approximately 5×10^8 cells were harvested every 3 to 4 days for nuclear extraction, and nuclear extracts were frozen at −70°C for ~5 weeks, until a total of $\sim 5 \times 10^9$ cells were obtained. Mouse MM14 skeletal myoblasts were grown on 100-mm collagen-coated tissue culture dishes in proliferation medium (Ham's F-10C supplemented with 15% horse serum and 2 ng of basic fibroblast growth factor/ml) as previously described (7). Myogenic differentiation was induced by switching to differentiation medium (Ham's F-10C supplemented with 1.5% horse serum and 1 μM insulin). Cells were maintained for 48 h in differentiation medium prior to harvesting. These cultures contained >90% terminally differentiated myocytes as assessed by immunostaining a parallel culture with the myosin-specific antibody MF-20. Ventricular myocardiocytes from 2- to 3-day-old Sprague-Dawley rats were isolated and cultured as described elsewhere (36). These cultures contained >90% myocardiocytes as assessed by immunohistochemistry with the MF-20 antibody.

Tissue dissections. Hearts were dissected from 5- to 8-week-old C57/BL6 mice, and atria were removed to reduce contamination by nonmyocardiocyte cell types. Fertilized White Leghorn chicken eggs (H and N International, Redmond, Wash.) were incubated at 38°C in humidified incubators. Embryos were staged according to the methods of Hamburger and Hamilton (18). Heart and quadriceps muscles were dissected; atria were removed from cardiac tissue, and skin and bones were removed from quadriceps. All dissected tissues were collected in Ham's F-10C medium on ice and rinsed several times in cold phosphate-buffered saline (PBS) prior to nuclear extraction.

Preparation of nuclear extracts. Crude nuclear extracts from cultured cells and embryonic tissues were prepared as described previously (9). Embryonic tissues were minced on ice before the release of nuclei. Crude nuclear extracts from adult mouse ventricle were prepared as described elsewhere (54). All nuclear extracts contained a cocktail of several protease inhibitors (product no. P8340; Sigma). Total protein in the extracts was quantitated by the Bradford method (2).

Coupling of DNA to magnetic beads. Streptavidin-linked magnetic beads (M-280; Dynal) were coupled to biotinylated, double-stranded oligonucleotides containing either the wild-type Trex sequence from the human *MCK* enhancer (hTrex; (GGACACCCGAGACGCCCGGTTA) or a mutant Trex sequence (hTrex-mt; GGACACCCGACATGCCTGGTTA) according to the Dynal protocol at a concentration of 20 pmol of DNA/mg of beads. Efficiency of coupling was verified by agarose gel electrophoresis.

TrexBF enrichment. All enrichment steps were performed at 4°C. Crude nuclear extracts from $\sim 5 \times 10^9$ HeLa cells (~140 mg of total protein) were pooled, and equal amounts (~70 mg of each) were incubated with 25 mg of either hTrex or hTrex-mt beads, with a 100-fold molar excess of poly(dI-dC), with rotation for 2 h. The hTrex beads were washed with buffer [10 mM Tris-Cl (pH 7.4), 0.1 M NaCl, 0.1 mM EDTA, 0.5 mM Tris(2-carboxyethyl)phosphine] containing a 10-fold molar excess of hTrex-mt oligonucleotide competitor, and hTrex-mt beads were washed with buffer containing a 10-fold molar excess of hTrex competitor. Beads were washed four times for 5 min each. Bound proteins were eluted from the beads in 10 min with wash buffer adjusted to 0.3 M NaCl.

ICAT labeling and preparation of peptides for microcapillary reversed-phase liquid chromatography-tandem mass spectrometry (μLC-MS/MS). Proteins eluted from hTrex and hTrex-mt beads were diluted threefold with 20 mM Tris-HCl (pH 8.3), 1 mM EDTA to bring the salt concentration to 0.1 M and then concentrated in Microcon 3 devices (Amicon). Proteins were denatured by adding sodium dodecyl sulfate (SDS) to 0.3% and heating for 5 min at 100°C and then reduced with 5 mM tri-*n*-butylphosphine for 30 min at 37°C. Samples were diluted sixfold with 20 mM Tris-HCl (pH 8.3), 1 mM EDTA, and 7.2 M urea. Isotopically heavy and normal isotope-coded affinity tag (ICAT) reagents were added to hTrex and hTrex-mt samples, respectively, to 1.75 mM, and samples were incubated for 90 min at 22°C. Labeling reactions were quenched by adding dithiothreitol to 10 mM for 15 min at 37°C. Samples were combined and digested for 2 h at 37°C with 10 μg of endoproteinase Lys-C (Boehringer-Mannheim), which cleaves on the carboxy-terminal side of lysine residues. SDS and urea concentrations were reduced to 0.02% and 1.2 M, respectively, by adding 20 mM Tris-HCl (pH 8.3) and 1 mM EDTA, and the sample was digested overnight at 37°C with trypsin (sequencing grade modified; 1:20 [wt/wt]; Promega), which cleaves on the carboxy-terminal side of arginine and lysine residues. The sample was diluted to 15 ml with buffer A (5 mM KH_2PO_4 [pH 3], 25% CH_3CN), and the pH was adjusted to 3 with 10% trifluoroacetic acid. Peptides were fractionated by SCX high-performance liquid chromatography (2.1 by 200 mm PolySULFOETHYL A [PolyLC]) by running the following gradient: 0 to 15% buffer B (5 mM KH_2PO_4 [pH 3], 600 mM KCl, 25% CH_3CN) over 30 min, 15 to 60% buffer B in 20 min, and 60 to 100% buffer B in 15 min at 0.2 ml/min. A total of 53 fractions were collected. ICAT-labeled peptides were purified from unlabeled peptides by passing each fraction over monomeric avidin cartridges (ABI) and washing with 2× PBS (pH 7.2), 1× PBS (pH 7.2), and 50 mM NH_4HCO_3 (pH 8.3), 20% CH_3OH. Peptides were eluted with 0.4% trifluoroacetic acid in 30% CH_3CN and dried under reduced pressure, and each fraction was resuspended in 12 μl of 0.2% acetic acid.

μLC-MS/MS and data analysis. Avidin-purified fractions were pressure loaded onto 10-cm by 75-μm fused silica microcapillary reversed-phase columns (5-μm Magic C18 beads; Michrom Bioresources) equilibrated with 10% CH_3OH, 0.4% acetic acid, 0.005% heptafluorobutyric acid (HFBA). Peptides were resolved by running 80-min gradients of 10 to 40% buffer C (100% CH_3OH, 0.4% acetic acid, 0.005% HFBA) at 0.3 μl/min and analyzed by μLC-MS/MS using an LCQ ion trap mass spectrometer (ThermoFinnigan) as described previously (17). Cysteine-containing tryptic peptides were identified by searching MS/MS spectra against a human protein sequence database using SEQUEST as described elsewhere (11). Data were quantified and analyzed essentially as described previously (19) using the XPRESS and INTERACT computer programs, respectively. Data were filtered using a requirement for peptides tryptic at both N and C termini. SEQUEST Xcorr cutoff values used for +1, +2, and +3 charged ions were 1.5, 1.4, and 2.0, respectively. The algorithms PeptideProphet (28) and ProteinProphet (35) were used to determine the probability that peptide and protein assignments were correct. A probability cutoff value of 0.9 was used for each analysis. All peptide identifications and quantifications were confirmed by manual inspection of the data.

In vitro coupled transcription-translation. In vitro coupled transcription-translation of Six4 in pCR3 was performed using the Promega TNT T7 coupled reticulocyte lysate system according to the manufacturer's protocol.

Gel mobility shift assays. Gel shift assays were carried out as previously described (12) and with the modifications described below. Double-stranded oligonucleotide probes were end labeled with ^{32}P and purified on a 12% polyacrylamide gel. Briefly, 1 to 2 μg of nuclear extract or 3 μl of in vitro translation reaction mixture was mixed with 20 μg of bovine serum albumin in 10 mM Tris-Cl (pH 7.4), 0.1 M NaCl, 0.1 mM EGTA, 0.5 mM dithiothreitol, 0.3 mM $MgCl_2$, 0.5 mM phenylmethylsulfonyl fluoride, and 4% glycerol. Incubations with antisera or unlabeled oligonucleotide competitors were carried out at room temperature for 15 min prior to the addition of probe. The mixtures were incubated with 15 pg of probe/μl and 36 ng of poly(dI-dC)/μl for 15 min at room temperature, then loaded onto 4% polyacrylamide gels. Electrophoresis was carried out in 0.5× Tris-borate-EDTA (TBE) buffer at 200 V for ~1.5 h at 4°C.

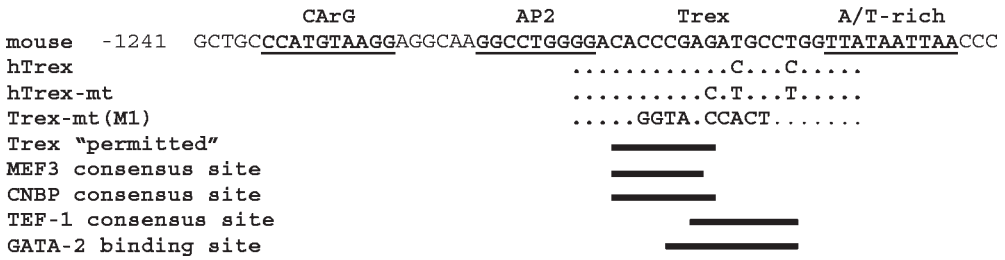


FIG. 1. Partial sequence of the mouse *MCK* enhancer. Control elements flanking the Trex site are underlined. Shown beneath the wild-type mouse Trex sequence are the wild-type hTrex sequence, the 3-bp mutation of hTrex (hTrex-mt), and the 9-bp mutation of the Trex site (M1) (12). The sequence within the Trex site important for TrexBF binding (based on individual base pair mutation) ("Trex permitted") and the sequences within the Trex site that are similar to TEF-1- and GATA-2-binding sites are indicated by thick black lines.

Gels were dried and analyzed by PhosphorImager (Molecular Dynamics). The sequences of the double-stranded oligonucleotides used for probes or competitors are as follows: hTrex, GGACACCCGAGACGCCCGGTTA; Trex-mt(M1), GGACAGGTAACCACTCCGGTTA; and MEF3, TGAATGTCAGGGGCTTCAGGTTCCCTA. The underlined bases in Trex-mt(M1) indicate changes from the wild-type hTrex sequence. The underlined bases in MEF3 indicate the MEF3 motifs in the *Aldolase A* promoter (45).

Transient transfections. Proliferating MM14 cells and primary neonatal rat myocardiocytes were transfected using a standard calcium phosphate technique (1) with 8 μg of chloramphenicol acetyltransferase (CAT) reporter plasmid and 2 μg of the pUCSV2PAP reference plasmid with or without 1 to 2 μg of Six4 expression plasmid. Four hours later, cells were glycerol shocked; MM14 cells were switched to differentiation medium (as described above), and myocardiocytes were switched to serum-free medium (DMEM/M199 [4:1] supplemented with 6 μg of insulin/ml and amphotericin B [Fungizone]). Cells were harvested 48 h after glycerol shock and analyzed for CAT and placental alkaline phosphatase (PAP) activity (1). For each construct tested, at least two plasmid preparations were used.

RESULTS

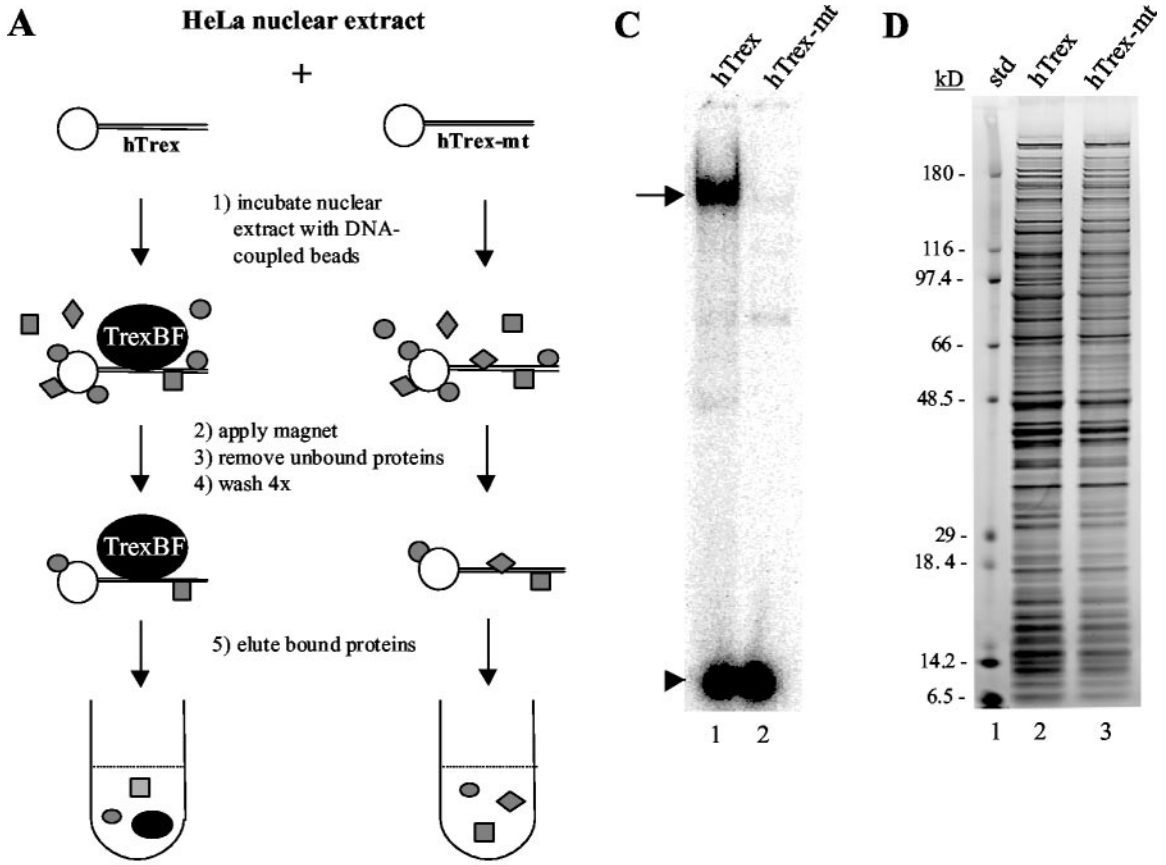
Enrichment of TrexBF from HeLa cells. Although TrexBF was initially discovered in mouse muscle cells, human HeLa cells were used as a TrexBF source because they can easily be grown to high density in suspension and because they contain relatively large amounts of TrexBF, as assessed by gel shift studies (data not shown). HeLa cells were also advantageous because the most complete vertebrate protein database is available for human proteins. Because the Trex site is critical for high-level activity of the *MCK* enhancer in striated muscle, it was important to compare TrexBF from HeLa cells with TrexBF from mouse myocytes. We found that in gel shift assays, oligonucleotide competitors containing various mutations in the Trex site have similar effects on TrexBF complex formation using either HeLa cell or mouse MM14 skeletal myocyte nuclear extracts (data not shown). This, in addition to the identical mobility of the TrexBF complex from both HeLa cells and MM14 myocytes, suggested that TrexBF is the same in both cell types.

In pilot purification studies using a DNA-affinity isolation strategy (16), it was difficult to purify TrexBF away from other proteins in the nuclear extracts without losing significant amounts of TrexBF activity. To circumvent this problem, we used a DNA-affinity technique in which streptavidin-linked magnetic beads were coupled to biotinylated, double-stranded oligonucleotides containing either the wild-type Trex site from the human *MCK* enhancer (hTrex) or the same sequence differing at 3 bp (hTrex-mt). This minimal mutation was used to keep the levels of nonspecific protein binding similar between the two oligonucleotides. The sequences of hTrex and hTrex-mt and overall purification strategy are shown in Fig. 1 and 2A and B.

Eluted proteins from both sets of beads were analyzed by gel shift assays for recovery of TrexBF and by SDS-polyacrylamide gel electrophoresis followed by silver staining for total protein composition (Fig. 2C and D). Based on pilot studies using hTrex beads, the estimated recovery of TrexBF was ~40% and the overall enrichment was ~80-fold. As expected, TrexBF activity was recovered from hTrex beads, but not hTrex-mt beads (Fig. 2C). However, staining for total protein revealed more than 50 easily discernible copurifying proteins in each sample and no unique bands in the eluents from hTrex versus hTrex-mt beads (Fig. 2D). Therefore, we assumed that TrexBF was obscured by higher-abundance, contaminating proteins of similar molecular masses.

Identification of TrexBF candidates using quantitative proteomics. The use of cysteine-reactive ICAT reagents to identify and quantify proteins from complex mixtures has been described elsewhere (17, 42). There are several major advantages to this approach. (i) Differential incorporation of stable isotopes in two samples allows for the determination of the relative abundance of proteins in the two samples. This permits the identification of proteins that are enriched in one sample despite a high background of unenriched proteins. (ii) Selection of cysteine-containing peptides reduces sample complexity, which can increase the number of proteins detected. (iii) Due to advantages i and ii and the high resolution of μLC-MS systems, extensive purification is not necessarily required to identify low-abundance proteins.

A schematic representation of the quantitative proteomics strategy is shown in Fig. 3A, and a detailed description is provided in the figure legend. Overall, 1,904 unique cysteine-containing peptide sequences with a PeptideProphet (28) probability of at least 0.9 were detected, representing 893 unique proteins or protein groups with a ProteinProphet (35) probability of at least 0.9. The vast majority of these peptides were present in approximately equal amounts in the heavy- versus normal-tagged sample, indicating that they were recovered in equal abundance from hTrex and hTrex-mt beads. Only three proteins showed >2-fold enrichment in the hTrex versus hTrex-mt samples (Table 1). A single peptide with a 2.7-fold enrichment was identified for annexin a7, a protein implicated in membrane fusion and in secretion (4). Based on its known function and cell localization, annexin a7 was considered an



B

	Time:	Temp:	[NaCl]:	competitor:
Binding	2 h	4° C	0.1 M	100x poly(dIdC)
Washing (4x)	5 min	4° C	0.1 M	10x hTrex-mt - for hTrex beads 10x hTrex - for hTrex-mt beads
Elution	10 min	4° C	0.3 M	--

FIG. 2. Selective enrichment of TrexBF. (A and B) Experimental strategy and conditions used for selective enrichment of TrexBF. Equal amounts of crude HeLa nuclear extract were incubated with beads coupled to either the wild-type human Trex site (hTrex) or a mutant Trex site (hTrex-mt). Grey shapes indicate nuclear proteins. (C) Gel shift assays showing binding to the hTrex probe of proteins eluted from hTrex beads (lane 1) and hTrex-mt beads (lane 2). TrexBF (indicated by the arrow) was recovered from hTrex beads, but not hTrex-mt beads. The arrowhead indicates free probe. (D) Silver-stained SDS-polyacrylamide gel electrophoresis showing the distribution of total protein eluted from hTrex beads (lane 2) and hTrex-mt beads (lane 3). For panels C and D, ~0.1% of total eluted protein was loaded in each lane.

unlikely TrexBF candidate. Other nonnuclear proteins (of approximately equal abundance in the hTrex versus hTrex-mt samples) were also detected, a result which can probably be attributed to contamination of nuclear extracts with cytoplasmic proteins during the extraction procedure.

Six peptides with a 2.1- to 10.7-fold enrichment were assigned to cellular nucleic acid binding protein (CNBP), a ubiquitous 19-kDa transcription factor (Table 1). Interestingly, five of these peptides are also common to an 18.4-kDa protein with high sequence similarity to CNBP. CNBP has been shown to regulate a number of genes, including those involved in cholesterol biosynthesis (14, 32, 41). Since gel shift experiments utilizing single-base-pair mutagenesis of the Trex site indicate that the sequence permitted for TrexBF binding is fairly degenerate, consisting of [C/A/T]A[C/T][C/T][C/T/G]GA[G/A/T] (C. L. Himeda and S. D. Hauschka, unpublished data), it was important to determine how closely the CNBP consensus binding site (CTGGGGTG) matches the sequence(s) recognized by TrexBF. The reverse CNBP consensus sequence is similar to the Trex "permitted" sequence, as well as to the *MCK* enhancer Trex site, differing only at a single base (Table 2). However, neither single- nor double-stranded oligonucleotide competitors containing the CNBP consensus sequence had any effect on the Trex-specific complex in gel shift assays, suggesting that CNBP is not part of TrexBF (data not shown).

A single peptide with a 2.4-fold change in abundance was

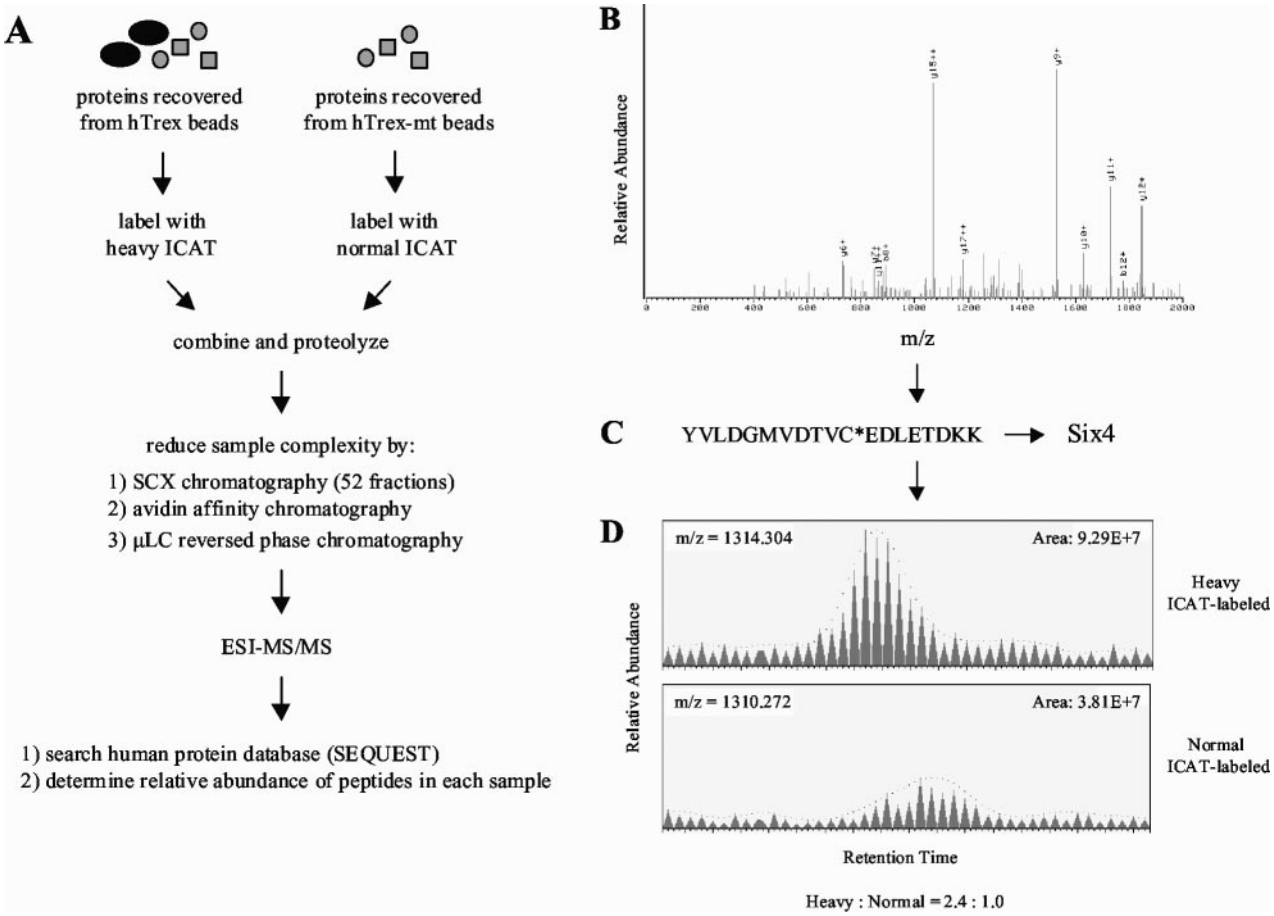


FIG. 3. Quantitative proteomics strategy used to identify TrexBF candidates. (A) Protein samples recovered from hTrex and hTrex-mt beads were labeled with isotopically heavy (deuterium) and normal (hydrogen) forms of the ICAT reagent, respectively. The labeled protein samples were then combined and proteolyzed with endoproteinase Lys-C and trypsin. Sample complexity was then reduced by sequential chromatography. Peptides were fractionated by strong cation exchange (SCX) chromatography, after which ICAT-labeled, cysteine-containing peptides were isolated from each of the 52 fractions by avidin affinity chromatography. Peptides in each fraction were further resolved and analyzed by μLC-electrospray ionization (ESI)-MS/MS. In this last step, both the sequence identity and relative abundance of the peptides were determined by automated multistage mass spectrometry. The search algorithm SEQUEST was used to identify cysteine-containing tryptic peptides from a human protein database. The relative abundance of peptides recovered from hTrex versus hTrex-mt beads was determined from the ratio of the heavy and normal signal intensities for each peptide pair. (B and C) Identification of Six4. A peptide ion with m/z = 1,314.3 was selected for collision-induced dissociation (CID), and the resulting fragmentation spectrum is shown in panel B. Fragment ions in the spectrum represent mainly single-event preferential cleavage of peptide bonds, resulting in sequence information from the N and C termini of the peptide. The CID spectrum was used to search a human protein database using the SEQUEST algorithm. SEQUEST identified the ICAT-labeled peptide sequence shown in panel C, which is derived from the Six4 protein, as the best match to the CID spectrum (PeptideProphet probability = 0.9916). Recorded ion fragments that matched theoretical ion fragments from the ICAT-labeled Six4 peptide are labeled (y = C-terminal ion fragments; b = N-terminal ion fragments; numbers indicate the number of amino acid residues; + and ++ indicate charge state). (D) Relative quantification of the Six4 peptide in hTrex versus hTrex-mt samples. Single-ion chromatograms were reconstructed for the isotopically heavy (m/z = 1,314.3) and normal (m/z = 1,310.3) peptide pair using XPRESS software, and the relative abundance was calculated after summing the signal intensities for each peptide over their respective elution times. The Six4 peptide is 2.4-fold more abundant in the isotopically heavy sample (isolated from hTrex beads) than in the isotopically normal sample (isolated from hTrex-mt beads).

detected and assigned to Six4, a 98-kDa homeodomain transcription factor of the Six/So family (Table 1). The detection, identification, and relative quantification of this peptide are shown in Fig. 3B, C, and D. The Six proteins recognize MEF3 motifs (TCAGGTT), which have been identified in the regulatory regions of several muscle-specific genes (21, 40, 43, 45, 46). The reverse MEF3 motif matches the Trex permitted sequence; however, it is not a perfect match to the mouse or human *MCK* enhancer Trex sequence, differing at two of seven bases (Table 2). This finding was significant, because the Trex site was not identifiable as a MEF3 element in transcription factor binding site database searches.

Six4 is TrexBF in mouse skeletal myocytes and embryonic day 10 (E-10) chick skeletal and cardiac muscle, whereas Six5 is the major TrexBF in adult mouse heart. To determine whether TrexBF in MM14 skeletal myocyte nuclear extract is able to recognize the consensus MEF3 motif, we tested this motif as a competitor in gel shift assays. An oligonucleotide containing the MEF3 site from the *Aldolase A* promoter competed for TrexBF binding to the *MCK* Trex probe, whereas the

TABLE 1. Peptides enriched >2-fold in hTrex versus hTrex-mt DNA-affinity purified samples

Protein[a]	Peptide[a]	Heavy/normal[b] ratio	Probability score[c]
Annexin a7	**C**YQSEFGRDLEK	2.7:1	0.9986
CNBP	**C**GESGHLAK	2.1:1	0.9994
	D**C**DLQEDA**C**YN**C**GR	2.6:1	0.9997
	GFQFVSSSLPDI**C**YR	2.9:1	0.9999
	CYS**C**GEFGHIQK	7:1	0.9996
	CGESGHLAR	9.2:1	1.0000
	CGETGHVAIN**C**SK	10.7:1	0.9998
Six4	YVLDGMVDTV**C**EDLETDKK	2.4:1	0.9916

[a] Peptides were analyzed by μLC-ESI-MS/MS, and protein identifications were assigned using the SEQUEST algorithm to search a human protein sequence database. ICAT-labeled cysteine residues are indicated in bold.

[b] The relative abundance of each peptide in heavy (isolated from hTrex beads) versus normal (isolated from hTrex-mt beads) ICAT-labeled samples was calculated using XPRESS and is expressed as a ratio. For CNBP, the three peptides with ratios of 2.1 to 2.9 were derived from the N-terminal portion of CNBP, whereas the three peptides with ratios of 7 to 10.7 were derived from the C-terminal portion of CNBP.

[c] PeptideProphet probability scores are shown for each peptide.

Trex mutant sequence (M1) had little effect (Fig. 4A, lanes 1 to 4). To address the question of whether TrexBF in MM14 myocytes is Six4, we tested the ability of Six4 antiserum to affect the formation or mobility of the TrexBF complex. Antiserum to mouse Six4 completely prevented the formation of the Trex-specific complex from MM14 nuclear extracts, whereas nonimmune serum had no effect (Fig. 4A, lanes 5 and 6), indicating that TrexBF in skeletal myocytes is, indeed, Six4. Furthermore, Six4 synthesized in vitro bound to the Trex probe, forming a complex that migrated with the same apparent mobility as TrexBF from MM14 nuclear extract (Fig. 4A, lane 7). Interestingly, the *Aldolase A* MEF3 site competed more efficiently than the *MCK* Trex site for Six4 binding to the *MCK* Trex probe, although both competitors diminished the Six4 signal compared to M1 (Fig. 4A, lanes 8 to 10). As expected, Six4 antiserum, but not nonimmune serum, prevented the formation of this complex (Fig. 4A, lanes 11 and 12). Control reticulocyte lysate translation reactions did not contain any Trex-binding activity (data not shown).

To determine the identity of TrexBF in cardiac muscle, we performed gel shift assays using nuclear extract from adult mouse heart ventricle. Two Trex-specific complexes that migrated with very similar mobilities were observed (Fig. 4B, lanes 1 to 3). (The minor, slower-migrating complex is most easily seen in lane 5.) Interestingly, Six4 antiserum prevented formation of the minor, slower-migrating complex, whereas it had no effect on the major, faster-migrating complex (Fig. 4B, lane 4). Although *Six4* mRNA is weakly expressed in adult mouse heart (27), the previously characterized MEF3-specific binding complex in adult mouse cardiac muscle is not Six4 (45). In light of this, it is possible that the minor, slower-migrating Six4 complex we observed was due to contamination by other cell types during the nuclear extraction from mouse ventricle.

In contrast to Six4, Six5 has been described as a 71-kDa factor with a fairly ubiquitous expression profile (39). Interestingly, antiserum to human Six5 (generous gift of C. Thornton) affected the major, but not the minor, Trex-specific complex in adult heart (Fig. 4B, compare lane 5 to lanes 1 and 4). Nonimmune serum had no effect on either complex (Fig. 4B, lane 6). These results suggest that Six5 is the major Trex-binding activity in adult mouse myocardiocytes.

To determine whether Six4 is the Trex-binding activity in developing striated muscle, we tested the polyclonal antiserum to mouse Six4 in gel shift assays using nuclear extracts from E-10 chicken skeletal and cardiac muscle. Six4 antiserum prevented formation of the TrexBF complex in nuclear extracts from both striated muscle types (Fig. 4C, lanes 4 and 9), whereas both nonimmune serum and Six5 antiserum had no effect (Fig. 4C, lanes 5 and 10, and data not shown). These data indicate that TrexBF in embryonic chicken striated muscle is Six4, and they suggest a role for Six4 in the regulation of both skeletal and cardiac muscle genes during development.

Six4 transactivates gene expression from the *MCK* Trex site. To determine whether Six4 is capable of transactivating gene expression via the Trex site, we performed cotransfection experiments in MM14 skeletal myocytes. Cells were transfected with a reporter construct containing either six multimerized copies of the wild-type Trex site [$(Trex)_6$TKCAT] or five copies of the M1 mutant Trex site [$(Trex\text{-}mt)_5$TKCAT] with or without cotransfected Six4 expression vector. Importantly, Six4 upregulated activity from the multimerized wild-type Trex construct by nearly threefold but had no effect on the mutant construct (Fig. 5A), demonstrating that Six4 is capable of transactivating gene expression via the Trex site.

To address the question of whether Six4 can transactivate gene expression from the Trex site within the more natural context of the *MCK* enhancer and basal promoter, we performed similar experiments using *MCK* enhancer-promoter constructs. MM14 skeletal myocytes were transfected with -enh80MCKCAT containing the *MCK* enhancer linked to the basal *MCK* promoter or with the equivalent construct containing the M1 mutation of the Trex site [(Trex-mt)-enh80MCKCAT] plus or minus the Six4 expression vector. Six4 upregulated activity from the wild-type, but not the mutant, construct by ~50% (Fig. 5B). When the same experiment was performed in primary neonatal rat myocardiocytes, Six4 upregulated activity from the wild-type, but not the mutant, construct by ~4-fold (Fig. 5C). These results demonstrate that Six4 is able to transactivate gene expression from the Trex site within the context of the *MCK* enhancer and basal promoter in both skeletal and cardiac myocytes. However, the limited transactivation of the *MCK* enhancer achieved in skeletal myo-

TABLE 2. Sequence comparisons between TrexBF candidate binding sites and the *MCK* Trex site

Sequence	Base(s)							
CNBP (reverse)[a]	**C**	**A**	**C**	**C**	**C**	**C**	**A**	**G**
Trex permitted	**C**/A/T	**A**	**C**/T	**C**/T	**C**/T/G	G	**A**	**G**/A/T
MCK Trex	**C**	**A**	**C**	**C**	**C**	G	**A**	**G**
MEF3 (reverse)[b]	**A**	**A**	**C**	**C**	**T**	**G**	**A**	
Trex permitted	C/**A**/T	**A**	**C**/T	**C**/T	C/**T**/G	**G**	**A**	**G**/A/T
MCK Trex	C	**A**	**C**	**C**	C	**G**	**A**	**G**

[a] Comparison between the CNBP (reverse) consensus sequence, the Trex permitted sequence (based on individual base pair mutation), and the Trex sequence from the *MCK* enhancer. Matching bases are indicated in bold, and mismatches with the candidate binding site are underlined.

[b] Comparison between the MEF3 (reverse) consensus sequence (the element recognized by Six proteins), the Trex permitted sequence, and the Trex sequence from the *MCK* enhancer.

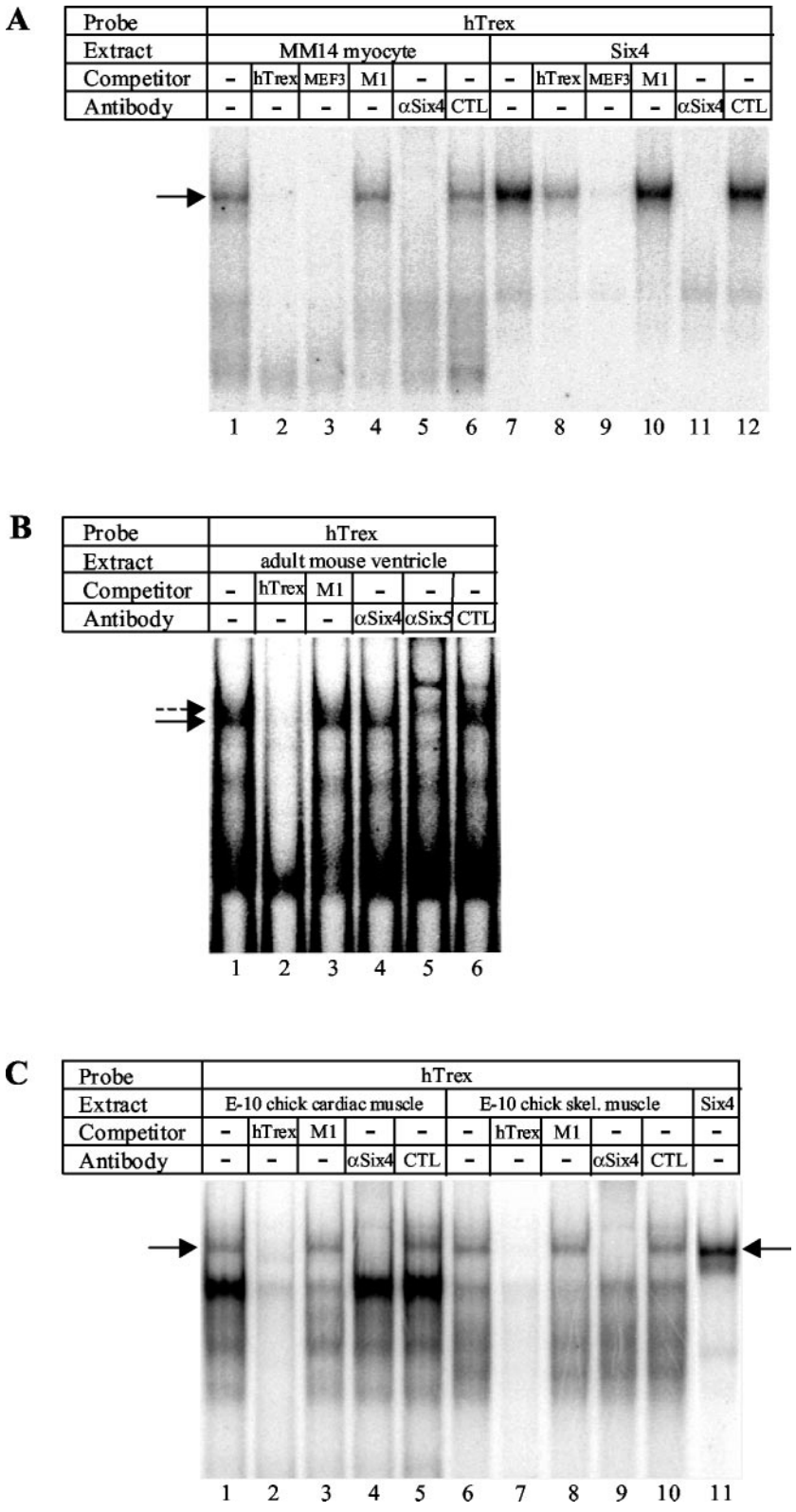


FIG. 4. Six4 is TrexBF in MM14 skeletal myocytes and E-10 chick skeletal and cardiac muscle, whereas Six5 is the major TrexBF in adult mouse heart. (A) Labeled hTrex probe was mixed with 1 μg of MM14 myocyte nuclear extract (lanes 1 to 6) or 1 μl of Six4 in vitro translation reaction mixture (lanes 7 to 12) and analyzed via gel shift assays. The Trex-specific complex in MM14 nuclear extract and in Six4 translation reaction mixture (lanes 1 and 7, indicated by the arrow) was competed away or decreased by wild-type hTrex oligonucleotide competitor (lanes 2 and 8) and by the MEF3 site from the *Aldolase A* promoter (lanes 3 and 9), but not by the M1 Trex mutant competitor (lanes 4 and 10). Antiserum to Six4 (α-Six4) prevented formation of the TrexBF complex in MM14 nuclear extract and in the Six4 translation reaction mixture (lanes 5 and 11), whereas

cytes suggests that endogenous Six4, or a cofactor, may be limiting in these cells.

To determine how the transactivation ability of Six4 compares with that of other transcription factors whose binding sites are also present in the *MCK* enhancer, we carried out similar transactivation studies with MEF2C and MyoD. Interestingly, while Six4 upregulated activity by ~50% from the 1,256-bp *MCK* regulatory region, MEF2C had no significant effect on activity and MyoD caused a slight decrease in activity (Fig. 5D). This was surprising, since the MEF2 site and right E-box in the *MCK* enhancer are both critical for high-level activity in cultured skeletal myocytes (1). Although MEF2 and MyoD have the ability to synergize in transactivating target genes (33), coexpression of all three factors (Six4, MEF2C, and MyoD) did not lead to a significant increase in activity (Fig. 5D). The minimal effects of overexpressing MEF2 and MyoD may be due to limiting amounts of the endogenous factors in MM14 skeletal myocytes or to indirect effects on other regulators of *MCK*.

Six4 transactivates gene expression from MEF3 sites in the regulatory regions of other muscle-specific genes. To determine whether Six4 is able to transactivate other muscle-specific genes containing MEF3 sites, we performed cotransfection experiments in MM14 skeletal myocytes. Importantly, Six4 upregulated activity by ~3.5-fold from a reporter construct containing the *Aldolase A* muscle-specific promoter, whereas it had little effect on the equivalent construct bearing a MEF3 site mutation in the *Aldolase A* promoter (Fig. 5E). Six4 has also been reported to upregulate expression from the MEF3 site in the *Myogenin* promoter (38). Interestingly, Six4 upregulated activity by only ~30% from a reporter construct containing intron 1 of *Cardiac troponin C*, which contains a MEF3 site, as well as 4,500 bp of upstream sequences (Fig. 5F). By contrast, Six4 did not upregulate expression from a reporter construct containing the *Troponin I* slow upstream enhancer with the 358-bp *MCK* promoter, neither of which contains MEF3 sites (Fig. 5G). The absence of a positive transactivation response from this test construct indicates that elevated Six4 does not cause a global activation of muscle genes; rather, the response may be limited to genes containing MEF3/Trex control elements.

The *Aldolase A* MEF3 site is a higher-affinity Six4-binding site than the *MCK* Trex site (Fig. 4A), which may explain the increased ability of Six4 to activate the *Aldolase A* versus *MCK* construct. Furthermore, because the *Aldolase A* promoter is smaller (~500 bp) and much less active in MM14 skeletal myocytes than either the 1,256 bp upstream of *MCK* or the *Cardiac troponin C* regulatory regions, the relative impact of overexpressing Six4 may be higher on *Aldolase A* than on *MCK* or *Cardiac troponin C*.

DISCUSSION

Using a selective enrichment procedure followed by quantitative proteomics, we have identified Six4 as a factor that binds and transactivates the Trex site in the *MCK* enhancer. In a previous study, this technique was used to identify both known and potentially new components of the yeast RNA polymerase II preinitiation complex (42); however, this is the first report that a mammalian transcription factor has been identified by this method. Importantly, the use of ICAT reagents in conjunction with mass spectrometry allowed us to identify TrexBF candidates within a high background of copurifying proteins. Out of 893 unique proteins or protein groups detected, only 3 were present in >2-fold-higher abundance in the hTrex versus hTrex-mt sample. One of these (annexin a7) was considered an unlikely TrexBF candidate based on its known function and cell localization, and the remaining two candidates (CNBP and Six4) were tested further.

Six peptides were detected from CNBP (five of these were also common to a CNBP-like protein), with heavy/normal ratios ranging from 2.1 to 10.7. Interestingly, three peptides with ratios of 2.1 to 2.9 are all derived from the N-terminal portion of CNBP, whereas three peptides with ratios ranging from 7 to 10.7 are derived from the C-terminal portion of the protein. This variation in relative abundance is likely due to the presence of two forms of CNBP and/or the presence of CNBP and a CNBP-like protein that are present in different quantities. Initially, CNBP (or its homologue) appeared to be a strong candidate for TrexBF based on the number of peptides, their relatively high abundance in hTrex versus hTrex-mt samples, and the close similarity between the Trex site and the CNBP consensus sequence. However, the inability of single- and double-stranded CNBP binding sites to compete for TrexBF in gel shift assays suggested that CNBP is not part of the TrexBF complex. A possible complexity to this conclusion is if CNBP has a relatively low affinity for the Trex site, it is possible that no detectable binding takes place in the 15-min incubation period used in gel shift assays, whereas enough CNBP bound to the Trex site during the 2-h incubation used in the enrichment procedure to be detectable by mass spectrometry. This possibility could not be excluded by gel shift assays because probes migrate faster than proteins; therefore, the probe is always moving away from weakly binding factors that dissoci-

nonimmune serum (CTL) did not prevent complex formation (lanes 6 and 12). Oligonucleotide competitors were used at a 100-fold molar excess for lanes 2 to 4 and a 200-fold molar excess for lanes 8 to 10. Antibodies were used at a 1:100 dilution. (B) Labeled hTrex probe was mixed with 3 μg of nuclear extract from adult mouse heart ventricle and analyzed via gel shift assays. There are two Trex-specific complexes in lane 1, a major complex (indicated by the solid arrow) and a minor complex (indicated by the dotted arrow), which were competed away by hTrex (lane 2), but not M1 (lane 3). α-Six4 prevented formation of the minor complex, but not the major complex (lane 4), whereas antiserum to Six5 (α-Six5) affected the major, but not the minor, complex (lane 5). Both Trex-specific complexes were unaffected by the addition of nonimmune serum (lane 6). (C) Labeled hTrex probe was mixed with 10 μg of nuclear extract from E-10 chicken heart ventricle (lanes 1 to 5), 2.5 μg of nuclear extract from E-10 chicken skeletal muscle (lanes 6 to 10), or 0.4 μl of Six4 in vitro translation reaction mixture (shown for a reference; lane 11) and analyzed via gel shift assays. The Trex-specific complex in cardiac and skeletal muscle (lanes 1 and 6, respectively, indicated by arrows) was competed away by wild-type hTrex oligonucleotide competitor (lanes 2 and 7), but not by the M1 Trex mutant competitor (lanes 3 and 8). α-Six4 prevented formation of the TrexBF complex (lanes 4 and 9), whereas nonimmune serum (CTL) did not (lanes 5 and 10). For panels B and C, oligonucleotide competitors were used at a 100-fold molar excess and antibodies were used at a 1:100 dilution.

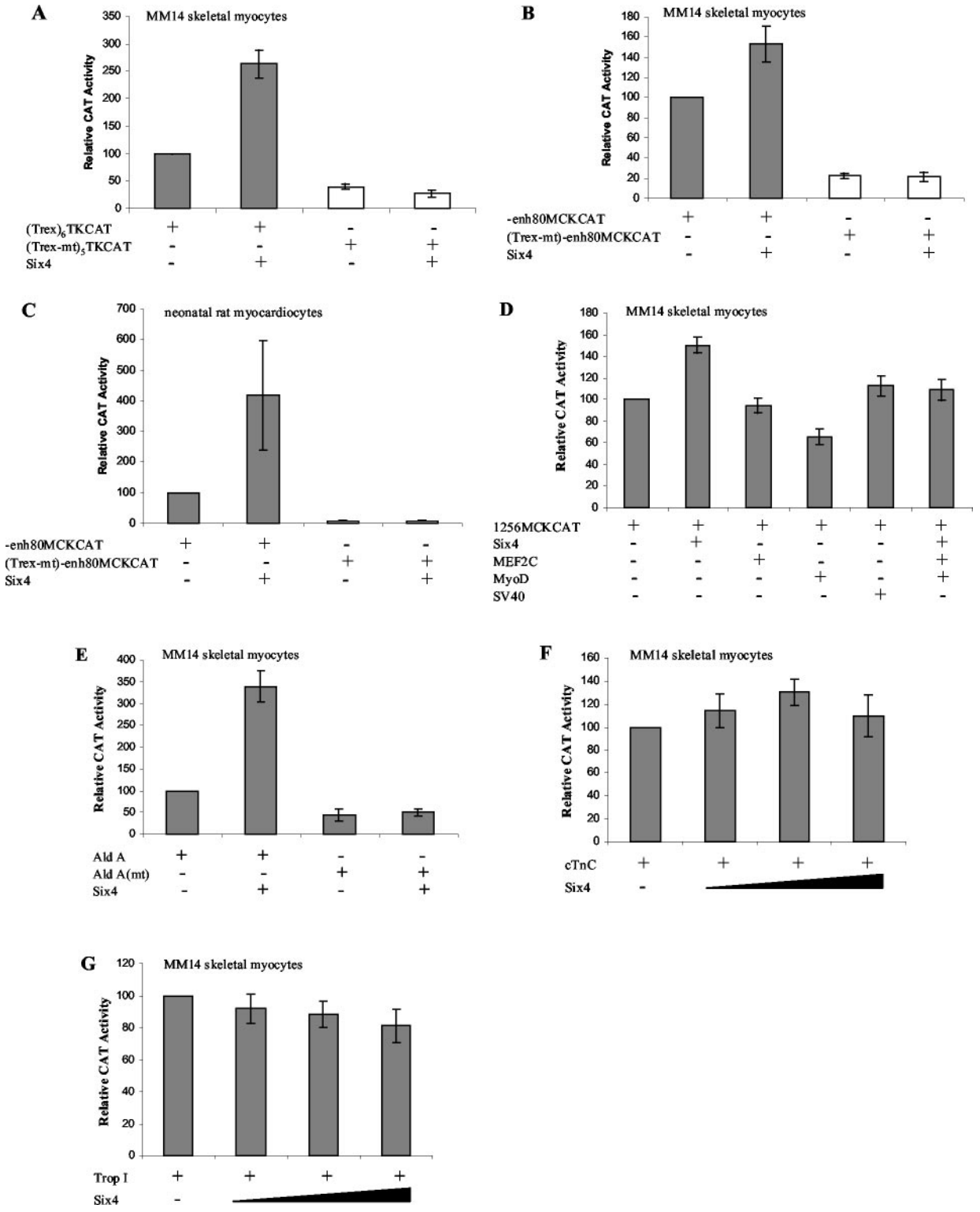


FIG. 5. Six4 transactivates gene expression from the Trex site in the *MCK* enhancer and from MEF3 sites in the regulatory regions of other muscle-specific genes. (A) MM14 skeletal myocytes were transfected with constructs containing the CAT reporter under the control of the thymidine kinase (TK) minimal promoter and either six copies of the wild-type Trex site [$(Trex)_6$TKCAT] or five copies of the M1 mutant Trex site [$(Trex\text{-}mt)_5$TKCAT] ± 2 μg of the Six4 expression construct (Six4). For $(Trex)_6$TKCAT ± Six4, $P = 3.5 \times 10^{-9}$ (Student's *t* test). (B) MM14 skeletal myocytes were transfected with a CAT reporter construct containing the wild-type *MCK* enhancer linked to the 80-bp minimal *MCK* promoter (-enh80MCKCAT) or the equivalent construct containing the M1 mutation of the Trex site [(Trex-mt)-enh80MCKCAT] ± 1 μg of Six4. For -enh80MCKCAT ± Six4, $P = 1.1 \times 10^{-6}$ (Student's *t* test). (C) Neonatal rat myocardiocytes were transfected with -enh80MCKCAT or (Trex-mt)-enh80MCKCAT ± 1 μg of Six4. For -enh80MCKCAT ± Six4, $P = 1.6 \times 10^{-3}$ (Student's *t* test). (D) MM14 skeletal myocytes were transfected with a CAT reporter construct containing the 1,256 bp upstream of *MCK*, which includes the enhancer and promoter (1256MCKCAT)

ate during the 90-min electrophoresis period. Such a progressive loss of bound factors would be greatly reduced in an affinity purification. Whether CNBP plays a role in regulating *MCK* expression via the Trex site remains to be determined.

The homeodomain transcription factor Six4 was identified as the other TrexBF candidate. The human Six4 protein contains three cysteine-containing tryptic peptides that are within the limits of mass detection; however, in this study, only one cys-containing tryptic peptide (with a heavy/normal ratio of 2.4) was detected for Six4. It is likely that the other peptides were not detected due to the complexity of the sample, i.e., the other Six4 peptides may have been obscured by peptides of similar *m*/*z* values from more abundant proteins.

Using gel shift experiments with Six-specific antisera, we have demonstrated that Six4 is TrexBF in MM14 skeletal myocytes and E-10 chick skeletal and cardiac muscle and that Six5 is the major TrexBF in adult heart. In addition, we have shown that Six4 is able to transactivate gene expression from the Trex site in the *MCK* enhancer as well as from MEF3 sites in the regulatory regions of other muscle-specific genes.

Six4 contains a ~200-amino-acid transactivation domain; however, it is unknown whether Six4 requires a coactivator to affect *MCK* expression. Eya proteins, the vertebrate homologues of Eyes absent in *Drosophila*, have been identified as coactivators of the Six proteins (38). *Eya* and *Six* share overlapping expression patterns in the somites and myotome during development (15, 52); Six4 has also been shown to induce nuclear translocation of Eya (38). Eya was not detected in the HeLa nuclear extract by our proteomics strategy and was probably not part of the Six4/TrexBF complex in gel shift assays, since this complex exhibited the same apparent mobility as Six4 synthesized in reticulocyte lysate. Nonetheless, it is important to determine whether Eya factors play a role in coactivating Six4 and Six5 bound to the *MCK* enhancer in skeletal and cardiac muscle.

Our results are consistent with a role for Six4 as a transcriptional regulator of *MCK*, and they expand the known MEF3 consensus sequence to include the *MCK* Trex site. Interestingly, the Trex element is also very similar to the consensus TEF-1/M-CAT binding site ([A/G]CATNC[C/T][T/A]) (13), differing only by a single base. TEF-1s are ubiquitous nuclear factors with multiple isoforms, including some concentrated in skeletal and cardiac muscle (8, 48). TrexBF from skeletal myocytes is not competed away by consensus TEF-1 sites or affected by the addition of TEF-1 antisera (12). Furthermore, mutating the Trex site to a consensus TEF-1 site in the context of the *MCK* enhancer produces a twofold decrease in enhancer activity (C. Fabre-Suver and S. D. Hauschka, unpublished data). Taken together, these results indicate that TrexBF is not TEF-1.

The *MCK* Trex site also corresponds closely to a random site-selected GATA-2-binding sequence ([G/C/T][A/G/C]NGAT[A/G]G[G/C/T][G/C/T]) (31), with only a single base mismatch. In previous experiments, different consensus GATA sequences were able to compete efficiently for TrexBF binding in gel shift assays, and *GATA-2* was found to be expressed in MM14 skeletal myocytes by reverse transcription-PCR (Himeda and Hauschka, unpublished). However, a polyclonal antibody to GATA-2 (gift of S. Orkin) had no effect on the TrexBF complex in gel shift assays; furthermore, purified glutathione *S*-transferase–GATA-2 fusion protein (GATA-2 cDNA provided by J. D. Engel) was unable to bind the Trex site, indicating that TrexBF is not GATA-2 (Himeda and Hauschka, unpublished).

More recently, a Trex permitted sequence ([C/A/T]A[C/T][C/T][C/T/G]GA[G/A/T]) has been determined via gel shift experiments utilizing single-base-pair mutagenesis of the Trex site (Himeda and Hauschka, unpublished). Two major points can be inferred from this: (i) the sequences within the *MCK* Trex site that are similar to both MEF3 and CNBP binding sites lie entirely within this region, whereas the sequences that are similar to GATA and TEF-1 binding sites are distinct from this region (Fig. 1) and, thus, this finding underscores the importance of narrowing down a novel element to the critical bases required for binding; and (ii) this sequence is fairly degenerate, suggesting that Six4 may bind to a spectrum of sequences including, but not limited to, the established MEF3 consensus. This is especially important because the *MCK* Trex site was not identifiable as a MEF3 element by transcription factor binding site database searches. It is important to note, however, that this sequence may not be an entirely accurate assessment of the DNA binding specificity of Six factors, since only single base pair changes were tested, i.e., it is possible that simultaneous changes of several nucleotides would disclose other permitted sequences. While the exact DNA-binding requirements for Six factors have yet to be determined, sequence database searches using the Trex/MEF3 composite sequence ([C/A]ACC[C/T]GA) revealed its presence in the regulatory regions of many muscle-specific genes. Trex/MEF3 sequences are found twice in the mouse α-*Myosin heavy chain* promoter, once in the mouse β-*Myosin heavy chain* promoter, and once in the human *Skeletal muscle* α-*actin* promoter. Four putative Trex/MEF3 sites are present in the rat *m1 Muscarinic acetylcholine receptor* (*AChR*) promoter, one is present in the rat *Neuronal AChR* β promoter, and two are found in the rat *Nicotinic AChR* δ promoter. Importantly, like the *MCK* Trex

alone or with 2 μg of either Six4, MEF2C expression vector (MEF2C), MyoD expression vector (MyoD), a control vector containing the simian virus 40 promoter (SV40), or with 0.67 μg each of Six4, MEF2C, and MyoD. For 1256MCKCAT ± Six4, $P = 2.4 \times 10^{-9}$ (Student's *t* test). (E) MM14 skeletal myocytes were transfected with a CAT reporter construct containing the *Aldolase A* muscle-specific promoter (Ald A) or the equivalent construct containing a MEF3 site mutation [Ald A (mt)] ± 2 μg of Six4. For Ald A ± Six4, $P = 2.4 \times 10^{-8}$ (Student's *t* test). (F) MM14 skeletal myocytes were transfected with a CAT reporter construct containing intron 1 and 4,500 bp of upstream sequences of *Cardiac troponin C* (cTnC) alone or with 0.5, 1, or 2 μg of Six4. For cTnC ± Six4, $P = 5.7 \times 10^{-4}$ (Student's *t* test). (G) MM14 skeletal myocytes were transfected with a CAT reporter construct containing the *Troponin I* slow enhancer upstream of the 358-bp *MCK* promoter (Trop I) alone or with 1, 2, or 4 μg of Six4. For each panel, data are plotted as the mean value and standard deviation of the CAT/PAP ratio, and the activity of the wild-type construct without Six4 was set at 100.

site, none of these putative control elements is a perfect match to the originally established MEF3 consensus.

The ability of the same transcription factor to recognize different sequences is not unprecedented; TEF-1, for example, binds the unrelated GTIIC and SphI/II elements of the simian virus 40 enhancer (54), as well as an A/T-rich sequence in the β-myosin heavy chain promoter (26). The ability of different factors to bind overlapping sites in the same element has also been well documented, as in the case of serum response factor and Yin Yang 1, which compete for binding to the CArG element in muscle-specific genes (23, 30). It is likely that this flexibility contributes to the fine-tuned regulation of gene transcription in different cell types and at different stages of development.

While our studies have identified Six4 as the most likely TrexBF, it is conceivable that other factors, such as CNBP, bind the Trex site in response to different physiological conditions. For example, in different anatomical muscles or under different environmental conditions, the signal transduction pathways that regulate Six4, CNBP, or factors binding adjacent *MCK* enhancer control elements might alter the expression, subcellular localization, protein processing and modifications, or DNA-binding affinities of one or more of these factors. Under certain conditions, CNBP might exhibit preferential association with the Trex site, thereby providing an additional layer of complexity to *MCK* transcriptional regulation. Analogous modifications have been proposed for factors binding to muscle gene promoter CArG/SRE motifs (5). Moreover, the severalfold differences between expression levels of *MCK* transgenes containing Trex site mutations in different anatomical muscles are consistent with complex physiological regulation of Six4-mediated transcriptional activity (36).

This study demonstrates the power and utility of quantitative mass spectrometry for the identification of site-specific DNA-binding factors in partially purified samples. When coupled with appropriate functional assays, this technology should prove to be extremely valuable for the identification of a wide range of DNA-binding proteins and associated complexes.

ACKNOWLEDGMENTS

We thank C. Bjornson, D. Kimelman, Q.-G. Nguyen, R. Palmiter, and E. Wang for many constructive discussions and for critical comments on the manuscript. We thank J. Eng, C. Fabre-Suver, Q.-G. Nguyen, K. Vaughn, and E. Yi for technical assistance. We are grateful to A. Weiner for providing a starter culture of HeLa S cells, C. Thornton for providing Six5 antisera, E. Olson for the MEF2C expression plasmid, A. Lassar for the MyoD expression construct, L. Kedes for pHcTnC4500CAT/I1-500-S, K. Hidaka and T. Mukai for pM202CAT and pM202CAT(mut C1), E. Hardeman for hTnIsUSE/-2000hSkeletal-Actin/CAT, J. D. Engel for providing GATA-2 cDNA, and S. Orkin for GATA-2 antisera.

This work was supported by National Institutes of Health grants AR18860 and R24 64387 to S.D.H. and predoctoral training grant GM07270 to C.L.H., as well as by the Muscular Dystrophy Association and federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, under contract no. NO1-HV-28179.

REFERENCES

1. **Amacher, S. L., J. N. Buskin, and S. D. Hauschka.** 1993. Multiple regulatory elements contribute differentially to muscle creatine kinase enhancer activity in skeletal and cardiac muscle. Mol. Cell. Biol. **13:**2753–2764.
2. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72:**248–254.
3. **Buskin, J. N., and S. D. Hauschka.** 1989. Identification of a myocyte nuclear factor that binds to the muscle-specific enhancer of the mouse muscle creatine kinase gene. Mol. Cell. Biol. **9:**2627–2640.
4. **Caohuy, H., M. Srivastava, and H. B. Pollard.** 1996. Membrane fusion protein synexin (annexin VII) as a Ca^{2+}/GTP sensor in exocytotic secretion. Proc. Natl. Acad. Sci. USA **93:**10797–10802.
5. **Chen, C. Y., and R. J. Schwartz.** 1997. Competition between negative acting YY1 versus positive acting serum response factor and tinman homologue Nkx-2.5 regulates cardiac alpha-actin promoter activity. Mol. Endocrinol. **11:**812–822.
6. **Christensen, T., and L. Kedes.** 1999. The myogenic regulatory circuit that controls cardiac/slow twitch troponin C gene transcription in skeletal muscle involves E-box, MEF-2, and MEF-3 motifs. Gene Expr. **8:**247–261.
7. **Clegg, C. H., T. A. Linkhart, B. B. Olwin, and S. D. Hauschka.** 1987. Growth factor control of skeletal muscle differentiation: commitment to terminal differentiation occurs in G_1 phase and is repressed by fibroblast growth factor. J. Cell Biol. **105:**949–956.
8. **Davidson, I., C. Fromental, P. Augereau, A. Wildeman, M. Zenke, and P. Chambon.** 1986. Cell-type specific protein binding to the enhancer of simian virus 40 in nuclear extracts. Nature **323:**544–548.
9. **Dignam, J. D., R. M. Lebovitz, and R. G. Roeder.** 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. **11:**1475–1489.
10. **Donoviel, D. B., M. A. Shield, J. N. Buskin, H. S. Haugen, C. H. Clegg, and S. D. Hauschka.** 1996. Analysis of muscle creatine kinase gene regulatory elements in skeletal and cardiac muscles of transgenic mice. Mol. Cell. Biol. **16:**1649–1658.
11. **Eng, J. K., A. L. McCormack, and J. R. Yates.** 1994. An approach to correlate tandem spectral data of peptides with amino acid sequences in protein databases. J. Am. Soc. Mass Spectrom. **5:**976–989.
12. **Fabre-Suver, C., and S. D. Hauschka.** 1996. A novel site in the muscle creatine kinase enhancer is required for expression in skeletal but not cardiac muscle. J. Biol. Chem. **271:**4646–4652.
13. **Farrance, I. K. G., J. H. Mar, and C. P. Ordahl.** 1992. M-CAT binding factor is related to the SV40 enhancer binding factor, TEF-1. J. Biol. Chem. **267:**17234–17240.
14. **Flink, I. L., and E. Morkin.** 1995. Alternatively processed isoforms of cellular nucleic acid-binding protein interact with a suppressor region of the human β-myosin heavy chain gene. J. Biol. Chem. **270:**6959–6965.
15. **Fougerousse, F., M. Durand, S. Lopez, L. Suel, J. Demignon, C. Thornton, H. Ozaki, K. Kawakami, P. Barbet, J. S. Beckmann, and P. Maire.** 2002. Six and Eya expression during human somitogenesis and MyoD gene family activation. J. Muscle Res. Cell. Motil. **23:**255–264.
16. **Gabrielsen, O. S., and J. Huet.** 1993. Magnetic DNA affinity purification of yeast transcription factor. Methods Enzymol. **218:**508–525.
17. **Gygi, S. P., B. Rist, S. A. Gerber, F. Turecek, M. H. Gelb, and R. Aebersold.** 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol. **17:**994–999.
18. **Hamburger, V., and H. L. Hamilton.** 1992. A series of normal stages in the development of the chick embryo. Dev. Dyn. **195:**231–272.
19. **Han, D. K., J. Eng, H. Zhou, and R. Aebersold.** 2001. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. Nat. Biotechnol. **19:**946–951.
20. **Hauser, M. A., A. Robinson, D. Hartigan-O'Connor, D. A. Williams-Gregory, J. N. Buskin, S. Apone, C. J. Kirk, S. Hardy, S. D. Hauschka, and J. S. Chamberlain.** 2000. Analysis of muscle creatine kinase regulatory elements in recombinant adenoviral vectors. Mol. Ther. **2:**16–25.
21. **Hidaka, K., I. Yamamoto, Y. Arai, and T. Mukai.** 1993. The MEF-3 motif is required for MEF-2-mediated skeletal muscle-specific induction of the rat aldolase A gene. Mol. Cell. Biol. **13:**6469–6478.
22. **Horlick, R. A., and P. A. Benfield.** 1989. The upstream muscle-specific enhancer of the rat muscle creatine kinase gene is composed of multiple elements. Mol. Cell. Biol. **9:**2396–2413.
23. **Itoh, S., Y. Katoh, H. Konishi, N. Takaya, T. Kimura, M. Periasamy, and H. Yamaguchi.** 2001. Nitric oxide regulates smooth-muscle-specific myosin heavy chain gene expression at the transcriptional level-possible role of SRF and YY1 through CArG element. J. Mol. Cell. Cardiol. **33:**95–107.
24. **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.
25. **Johnson, J. E., B. J. Wold, and S. D. Hauschka.** 1989. Muscle creatine kinase sequence elements regulating skeletal and cardiac muscle expression in transgenic mice. Mol. Cell. Biol. **9:**3393–3399.
26. **Karasseva, N., G. Tsika, J. Ji, A. Zhang, X. Mao, and R. Tsika.** 2003. Transcription enhancer factor-1 binds multiple muscle MEF2 and A/T-rich elements during fast-to-slow skeletal muscle fiber type transitions. Mol. Cell. Biol. **23:**5143–5164.
27. **Kawakami, K., H. Ohto, K. Ikeda, and R. G. Roeder.** 1996. Structure, function and expression of a murine homeobox protein AREC3, a homologue of Drosophila sine oculis gene product, and implication in development. Nucleic Acids Res. **24:**303–310.
28. **Keller, A., A. I. Nesvizhskii, E. Kolker, and R. Aebersold.** 2002. Empirical

statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. **74:**5383–5392.
29. **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.
30. **Martin, K. A., A. Gualberto, M. F. Kolman, J. Lowry, and K. Walsh.** 1997. A competitive mechanism of CArG element regulation by YY1 and SRF: implications for assessment of Phox1/MHox transcription factor interactions at CArG elements. DNA Cell Biol. **16:**653–661.
31. **Merika, M., and S. H. Orkin.** 1993. DNA-binding specificity of GATA family transcription factors. Mol. Cell. Biol. **13:**3999–4010.
32. **Michelotti, E. F., T. Tomonaga, H. Krutzsch, and D. Levens.** 1995. Cellular nucleic acid binding protein regulates the CT element of the human c-myc protooncogene. J. Biol. Chem. **270:**9494–9499.
33. **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.
34. **Mueller, P. R., and B. Wold.** 1989. In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. Science **246:**780–786.
35. **Nesvizhskii, A. I., A. Keller, E. Kolker, and R. Aebersold.** 2003. A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. **75:**4646–4658.
36. **Nguyen, Q.-G. V., J. N. Buskin, C. L. Himeda, C. Fabre-Suver, and S. D. Hauschka.** 2003. Transgenic and tissue culture analyses of the muscle creatine kinase enhancer Trex control element in skeletal and cardiac muscle indicate differences in gene expression between muscle types. Transgenic Res. **12:**337–349.
37. **Niiya, A., H. Ohto, K. Kawakami, and M. Araki.** 1998. Localization of Six4/AREC3 in the developing mouse retina; implications in mammalian retinal development. Exp. Eye Res. **67:**699–707.
38. **Ohto, H., S. Kamada, K. Tago, S. I. Tominaga, H. Ozaki, S. Sato, and K. Kawakami.** 1999. Cooperation of Six and Eya in activation of their target genes through nuclear translocation of Eya. Mol. Cell. Biol. **19:**6815–6824.
39. **Ohto, H., T. Takizawa, T. Saito, M. Kobayashi, K. Ikeda, and K. Kawakami.** 1998. Tissue and developmental distribution of Six family gene products. Int. J. Dev. Biol. **42:**141–148.
40. **Parmacek, M. S., H. I. Ip, F. Jung, T. L. Shen, J. F. Martin, A. J. Vora, E. N. Olson, and J. M. Leiden.** 1994. A novel myogenic regulatory circuit controls slow/cardiac troponin C gene transcription in skeletal muscle. Mol. Cell. Biol. **14:**1870–1885.
41. **Rajavashisth, T. B., A. K. Taylor, A. Andalibi, K. L. Svenson, and A. J. Lusis.** 1989. Identification of a zinc finger protein that binds to the sterol regulatory element. Science **245:**640–643.
42. **Ranish, J. A., E. C. Yi, D. M. Leslie, S. O. Purvine, D. R. Goodlett, J. Eng, and R. Aebersold.** 2003. The study of macromolecular complexes by quantitative proteomics. Nat. Genet. **33:**349–355.
43. **Salminen, M., S. Lopez, P. Maire, A. Kahn, and D. Daegelen.** 1996. Fast-muscle-specific DNA-protein interactions occurring in vivo at the human aldolase A M promoter are necessary for correct promoter activity in transgenic mice. Mol. Cell. Biol. **16:**76–85.
44. **Shield, M. A., H. S. Haugen, C. H. Clegg, and S. D. Hauschka.** 1996. E-box sites and a proximal regulatory region of the muscle creatine kinase gene differentially regulate expression in diverse skeletal muscles and cardiac muscle of transgenic mice. Mol. Cell. Biol. **16:**5058–5068.
45. **Spitz, F., J. Demignon, A. Poreu, A. Kahn, J.-P. Concordet, D. Daegelen, and P. Maire.** 1998. Expression of myogenin during embryogenesis is controlled by Six/*sine oculis* homeoproteins through a conserved MEF3 binding site. Proc. Natl. Acad. Sci. USA **95:**14220–14225.
46. **Spitz, F., M. Salminen, J. Demignon, A. Kahn, D. Daegelen, and P. Maire.** 1997. A combination of MEF3 and NFI proteins activates transcription in a subset of fast-twitch muscles. Mol. Cell. Biol. **17:**656–666.
47. **Sternberg, E. A., G. Spizz, W. M. Perry, D. Vizard, T. Weil, and E. N. Olson.** 1988. Identification of upstream and intragenic regulatory elements that confer cell-type-restricted and differentiation-specific expression on the muscle creatine kinase gene. Mol. Cell. Biol. **8:**2896–2909.
48. **Stewart, A. F. R., S. B. Larkin, I. K. G. Farrance, J. H. Mar, D. E. Hall, and C. P. Ordahl.** 1994. Muscle-enriched TEF-1 isoforms bind M-CAT elements from muscle-specific promoters and differentially activate transcription. J. Biol. Chem. **269:**3147–3150.
49. **Suzuki-Yagawa, Y., K. Kawakami, and K. Nagano.** 1992. Housekeeping Na, K-ATPase alpha 1 subunit gene promoter is composed of multiple *cis* elements to which common and cell type-specific factors bind. Mol. Cell. Biol. **12:**4046–4055.
50. **Trask, R. V., A. W. Strauss, and J. J. Billadello.** 1988. Developmental regulation and tissue-specific expression of the human muscle creatine kinase gene. J. Biol. Chem. **263:**17142–17149.
51. **Xiao, J. H., I. Davidson, H. Matthes, J. M. Garnier, and P. Chambon.** 1991. Cloning, expression, and transcriptional properties of the human enhancer factor TEF-1. Cell **65:**551–568.
52. **Xu, P. X., I. Woo, H. Her, D. R. Beier, and R. L. Maas.** 1997. Mouse Eya homologues of the Drosophila eyes absent gene require Pax6 for expression in lens and nasal placode. Development **124:**219–231.
53. **Yi, J.-M., K. Walsh, and P. Schimmel.** 1991. Rabbit muscle creatine kinase: genomic cloning, sequencing, and analysis of upstream sequences important for expression in myocytes. Nucleic Acids Res. **19:**3027–3033.
54. **Zahradka, P., D. E. Larson, and B. H. Sells.** 1989. RNA polymerase II-directed gene transcription by rat skeletal muscle nuclear extracts. Exp. Cell Res. **185:**8–20.