Volume 17 Number 15 1989 Nucleic Acids Research

Expression and DNA sequence analysis of a human embryonic skeletal muscle myosin heavy chain gene

Ilene Karsch-Mizrachi, Marilyn Travis[1], Helen Blau[1] and Leslie A.Leinwand*

Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, NY 10461 and [1]Department of Pharmacology, Stanford University School of Medicine, Stanford, CA 94305, USA

Received March 20, 1989; Revised and Accepted June 29, 1989 EMBL accession nos X15696-X15698 (incl.)

ABSTRACT

Vertebrate myosin heavy chains (MHC) are represented by multiple genes that are expressed in a spatially and temporally distinct pattern during development. In order to obtain molecular probes for developmentally regulated human MHC isoforms, we used monoclonal antibodies to screen an expression cDNA library constructed from primary human myotube cultures. A 3.4 kb cDNA was isolated that encodes one of the first MHCs to be transcribed in human skeletal muscle development. A portion of the corresponding gene encoding this isoform has also been isolated. Expression of this embryonic MHC is a hallmark of muscle regeneration after birth and is a characteristic marker of human muscular dystrophies. During normal human development, expression is restricted to the embryonic period of development prior to birth. In primary human muscle cell cultures, devoid of other cell types, mRNA accumulation begins as myotubes form, reaches a peak 2 days later and declines to undetectable levels within 10 days. The expression of the protein encoded by the embryonic skeletal MHC gene follows a similar time course, lagging behind the mRNA by approximately two days. Thus, expression of the human embryonic gene is efficiently induced and then repressed in cultured muscle cells, as it is in muscle tissue. The study of the regulation of a human MHC isoform with a central role in muscle development and in muscle regeneration in disease states is therefore amendable to analysis at a molecular level.

INTRODUCTION

Myosin heavy chain (MHC) is a major component of the contractile apparatus and is encoded by multigene families in vertebrates (for reviews see 1,2). In mammals, at least ten different MHC isoforms have been described from striated, smooth and non-muscle cells. These isoforms show expression that is spatially and temporally regulated during development (see 3,4, 5). The hexameric myosin molecule consists of two heavy chains (200 kdal each) and two pairs of light chains (16-20 kdal each). The heavy chain can be divided into two domains, the globular amino terminal head responsible for myosin light chain and actin binding and for ATP hydrolysis (6) and the α-helical carboxy terminal rod responsible for the ability of myosin to form filaments (7).

Analogous MHC isoforms from different species are more homologous to one another than are different isoforms from the same species (3,8,9, this report) and this homology extends into the 3′ untranslated region. The evolutionary conservation of MHC isoform sequences implies that they are functionally significant, but to a large degree, their distinct roles in sarcomere assembly or contractile function remain to be elucidated.

Expression of MHC isoforms is sequential and controlled by numerous signals including neuronal and hormonal triggers (10,11, 12). The fact that some developmental transitions occur in cultured muscle cells suggests that much of the information necessary for sequential

MHC expression is intrinsic to the cells (13,14). Human muscle cultures may be advantageous for examining the regulation of MHC expression due to the longer developmental cycle of primates compared to rodents and chickens. Because of its early appearance in skeletal muscle development, and its reappearance during muscle regeneration, the embryonic skeletal MHC isoform provides a particularly useful marker for charting the progression of myofiber formation and renewal (15,16,17).

In order to obtain molecular genetic probes encoding human MHC isoforms expressed in early development we constructed and screened a λgt11 expression cDNA library with monoclonal antibodies that recognize epitopes on the embryonic skeletal MHC molecule. Because pure populations of early differentiated cultured human myotubes were used as the source of mRNA the predominant MHC clone in this library (designated SMHCE) encodes the embryonic skeletal MHC. The entire myosin rod is encoded by this cDNA and as well as the epitope recognized by the well-characterized monoclonal antibody F1.652 (13). Gene-specific MHC DNA probes have been used to define the complement of MHC transcripts expressed in cultures of human myotubes. These studies demonstrate the power of combining immunological and molecular genetic data to characterize human developmental processes. Used in conjunction with monoclonal antibodies, DNA clones encoding the human embryonic skeletal MHC isoform will be particularly useful for studying the regulation of gene expression that accompanies human skeletal muscle development and the regeneration typical of various muscle diseases as exemplified by Duchenne muscular dystrophy (17).

MATERIALS AND METHODS

Cell Culture

Human skeletal muscle cells were isolated from the *vastus lateralis* muscle of a 5-year-old male by the procedure described in Blau and Webster (18). Pure populations of myoblasts were sorted in a fluorescence-activated cell sorter using a monoclonal antibody, 5.1H11, which recognizes a myoblast specific cell surface antigen (19,20). Cells were grown in mitogen-rich growth medium containing Ham F-10 nutrient mixture supplemented with 15% fetal bovine serum, 0.5% chicken embryo extract, 50 units/ml Penicillin G and 50 μg/ml Streptomycin Sulfate on tissue culture dishes coated with type IV collagen (Collaborative Research). After reaching confluence, the cells were exposed to mitogen-poor starvation DME with 1μM insulin for one day. Afterwards, the cultures were maintained in fusion medium (DME with 5% horse serum, 1 μM insulin and 10 μM cytosine-β-D-arabinofuranoside to allow for the survival of only non-dividing cells). Medium was replaced every 24 hours.

RNA Isolation

Total cellular RNA was isolated from myotubes 4 days after fusion for construction of the cDNA library by a modification of the guanidine hydrochloride procedure (21). All other RNA from cultured cells and tissue was isolated by the guanidinium-cesium chloride pelleting method (22).

cDNA Library Construction and Screening

A cDNA library was constructed in the λgt11 vector following standard procedures (23) using 10 μg of day 4 myotube poly $(A)^+$ mRNA. Phosphatased λgt11 arms and Gigapack Plus packaging extracts were purchased from Stratagene. Recombinant clones that contained striated muscle MHC sequences were isolated by plaque lifts and hybridization to

radiolabeled DNA probes made from conserved sequences in the tail regions of two adult human skeletal MHC cDNA sequences, pSMHCA and pSMHCZ (8). MHC cDNAs isolated by this method were screened for β-galactosidase-MHC fusion protein production with a panel of monoclonal antibodies (24), that recognize stage-specific MHC, using the screening procedure outlined by Huynh et al.(25).

Preparation of DNA Templates for Sequencing

The embryonic MHC cDNA insert, SMHCE, was cloned into pBluescript SK^- (Stratagene). Overlapping deletions were generated using the Erase-a-Base system (Promega Corp.). Preparation of template DNA and dideoxy reactions for sequencing using Sequenase modified T7 polymerase (U.S. Biochemical) were performed according to Kraft et al. (26). Sequences were analyzed and homology comparisons were made using the Staden package for sequence analysis.

Genomic DNA Analysis

Human genomic DNA was digested with Eco RI and 15 μg per lane was electrophoresed in 1% agarose-TAE (40 mM Tris-acetate, 2 mM EDTA) gel with recirculation of buffer. DNA was transferred to filters according to Southern (27) except that GeneScreen (New England Nuclear) was used in place of nitrocellulose. The DNA was fixed to the filter by exposure to UV light in a Stratalinker UV crosslinker (Stratagene). Random primed radiolabeled DNA probes were generated for hybridization using the oligolabeling kit (Pharmacia). The filters were prehybridized in 5× SSC, 1× Denhardt's solution, 50 mM $NaH_2PO_4 \cdot H_2O$ and 150 μg/ml heat denatured salmon sperm DNA at 65°C. The filters were hybridized in the same solution with the addition of dextran sulfate to a final concentration of 10% and heat denatured probe at a concentration of 2.0×10^5 cpm/ml. Following hybridization, the filters were washed at 65°C in 2× SSC, 0.2% SDS for 1 hour. The blots were exposed to Kodak X-Omat AR film with a DuPont Cronex intensifying screen at −70°C.

RNA Blot Hybridizations

For Northern blot analysis, 20 μg of RNA isolated from human muscle cell cultures were electrophoresed on 1% agarose gels containing 100 mM MOPS (3-[N-Morpholino] propanesulfonic acid) pH 7.4, 10 mM EDTA and 3% formaldehyde. Fractionated RNA was transferred to Gene Screen (NEN) overnight by capillary action in 10× SSC. For RNA dot blots, 2 μg of total RNA from fetal skeletal muscle (22 weeks), adult quadriceps muscle, adult cardiac tissue (provided by M. Yeager, Scripps Institute) and liver were resuspended in 1% formaldehyde and spotted onto Gene Screen that had been pre-wet in water, soaked in 2× SSC and air dried. Filters were baked for two hours in a vacuum oven prior to hybridization. A ^{32}P-UTP radiolabeled antisense RNA probe containing the sequences encoding the last 8 amino acids of the carboxy terminus and 3′ untranslated sequences of SMHCE was synthesized following the Riboprobe protocol outlined by Promega Corporation. The filters were prehybridized for at least two hours in 50% formamide, 5× SSPE (1× = 180 mM NaCl, 10 mM NaH_2PO_4, and 1 mM EDTA, pH 7.4), 2× Denhardts solution (1× =0.4% BSA, 0.4% Ficoll and 0.4% Polyvinylpyrrolodine), 0.2% SDS and heat denatured salmon sperm DNA (150 μg/ml). Filters were hybridized overnight in the same solution with the addition of 10% dextran sulfate and 1×10^6 cpm/ml of ^{32}P RNA at 55°C. Filters were washed in 1× SSPE for 1 hour at 65°C. A 1.0 kb insert from pRM-17 encoding 18S rDNA probe (28) and a 171 bp human cardiac α-actin 3′ untranslated region probe (29) were generated by random

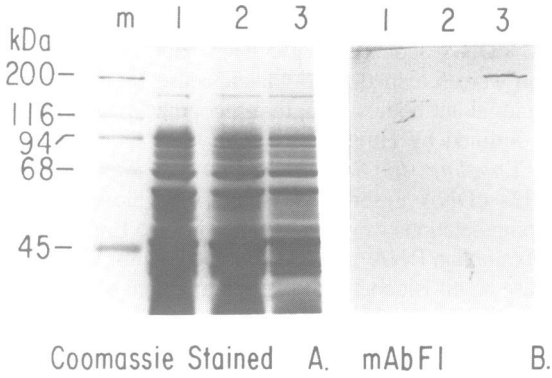


Figure 1. Expression of β-galactosidase-SMHCE Fusion Protein. Whole cell lysates of bacteria were separated on 7.5% SDS-PAGE gels in duplicate. One half of the gel was stained with Coomassie Brilliant Blue (A) and the other half was electroblotted onto nitrocellulose and reacted with monoclonal antibody F1.652 (B). In each panel, lane 1 contains the lysate from uninfected *E. coli* strain Y1090, lane 2, the lysate from Y1090 infected with λgt11-SMHCE cDNA clone, uninduced; and in lane 3, the infected lysate induced with 10 mM IPTG.

prime labeling using the oligolabeling kit (Pharmacia). The filters were prehybridized and hybridized under the same conditions as the DNA blots.

λgt11 Fusion Protein Synthesis

λgt11 fusion proteins were produced by the SMHCE cDNA clone in order to characterize the epitopes encoded by this MHC sequence according to the procedure outlined by Reinach and Fischman (30). *E. coli* strain Y1090 was grown to mid-log phase and infected with λgt11-SMHCE at a ratio of 10:1 (phage:bacteria). 10 mM IPTG (Isopropyl β-D-thiogalactoside) was added and the cells were grown for 45 minutes at 42°C. The cells were spun down, resuspended in SDS sample buffer (31) and boiled for five minutes.

Preparation of Whole Cell Lysates from Tissue Culture Cells

Myoblasts and myotubes grown on 60 mm tissue culture dishes were rinsed with PBS and subjected to trypsinization. The cells were pelleted, rinsed with PBS and resuspended in 100 μl SDS sample buffer (31) and boiled for five minutes.

Western Blot Analysis

Protein samples from bacterial and tissue culture cell lysates were analyzed by SDS-PAGE (SDS/polyacrylamide gel electrophoresis) in duplicate (31). Half of the gel was stained with Coomassie Brilliant Blue. The other half of the gel was transferred to nitrocellulose by electroblotting (32). The filters were incubated with hybridoma cell supernatant containing 1.2 μg/ml of monoclonal antibody F1.652 at 4°C overnight. The filters were incubated with a 1:1000 dilution of goat anti-Mouse IgG linked to horseradish peroxidase (Bio-Rad) as a second antibody for two hours. Antibody reactivity was detected by hydrogen peroxide and 4-chloro-1-napthol treatment.

RESULTS

Isolation and Sequence Analysis of a Human Embryonic MHC cDNA Clone.

In order to obtain a molecular probe for one of the first MHC genes to be expressed in human skeletal muscle development, a human cultured skeletal muscle cell λgt11 expression cDNA library was constructed and screened for MHC sequences by hybridization to DNA probes and by monoclonal antibody reactivity. Previous studies with monoclonal antibody

```
E F Q K T K D E L A K S E A K R K E L E E K L V T L V Q E K N D L Q L Q V Q A E
GAATTCCAGAAAACCAAAGATGAACTCGCCAAGTCGGAGGCAAAAAGGAAGGAGCTAGAGGAAAAACTGGTGACTCTGGTCCAAGAGAAGAATGACCTGCAGCTCCAAGTACAAGCTGAA
      10        20        30        40        50        60        70        80        90       100       110       120
S E N L L D A E E R C D Q L I K A K F Q L E A K I K E V T E R A E D E E E I N A
AGCGAAAATTTGTTGGATGCTGAGGAAAGATGCGATCAGCTGATCAAAGCCAAATTCCAGCTCGAGGCCAAGATCAAGGAGGTGACAGAGAGAGCTGAAGATGAGGAGGAGATCAATGCT
     130       140       150       160       170       180       190       200       210       220       230       240
E L T A K K R K L E D E C S E L K K D I D D L E L T L A K V E K E K H A T E N K
GAGCTGACGGCCAAGAAGAGGAAACTGGAGGATGAATGCTCAGAGCTCAAGAAAGACATTGATGACCTTGAGTTGACCCTGGCCAAGGTTGAGAAGGAGAAGCATGCCACGGAGAACAAG
     250       260       270       280       290       300       310       320       330       340       350       360
V K N L T E E L S G L D E T I A K L T R E K K A L Q E A H Q Q A L D D L Q A E E
GTTAAAAACCTTACTGAGGAACTCTCCGGGTTAGATGAAACAATTGCAAAGTTAACCAGAGAGAAGAAGGCCCTCCAAGAGGCGCACCAGCAGGCCTTGGATGACCTCCAAGCTGAAGAA
     370       380       390       400       410       420       430       440       450       460       470       480
D K V N S L N K T K S K L E Q Q V E D L E S S L E Q E K K L R V D L E R N K R K
GACAAAGTCAATTCTTTGAACAAAACCAAGAGCAAACTGGAACAGCAAGTGGAAGACCTGGAAAGCTCCCTAGAACAAGAAAAAGAAGCTCCGCGTAGACCTGGAAAGGAACAAAAGGAAA
     490       500       510       520       530       540       550       560       570       580       590       600
L E G D L K L A Q E S I L D L E N D K Q Q L D E R L K K K D F E Y C Q L Q S K V
TTGGAAGGAGACTTGAAGCTTGCTCAAGAGTCCATATTAGATCTGGAGAATGACAAGCAACAGCTGGACGAAAGGCTCAAGAAGAAAGATTTTGAATATTGTCAACTTCAAAGCAAAGTG
     610       620       630       640       650       660       670       680       690       700       710       720
E D E Q T L G L Q F Q K K I K E L Q A R I E E L E E E I E A E R A T R A K T E K
GAAGATGAGCAGACACTGGGCCTCCAGTTTCAGAAGAAAATCAAAGAGTTGCAGGCTCGAATTGAGGAGCTGGAAGAGGAGATAGAGGCGGAGAGGGCCACCCGCGCGAAGACAGAGAAA
     730       740       750       760       770       780       790       800       810       820       830       840
Q R S D Y A R E L E E L S E R L E E A G G V T S T Q I E L N K K R E A E F L K L
CAGCGCAGCGACTATGCCCGGGAGCTGGAGGAGCTGAGCGAGCGGCTGGAGGAGGCGGGAGGCGTCACCTCCACGCAGATAGAGCTCAACAAGAAGCGGGAGGCGGAGTTCCTGAAGCTG
     850       860       870       880       890       900       910       920       930       940       950       960
R R D L E E A T L Q H E A M V A T L R K K H A D S V A E L G E Q I D N L Q R V K
CGCAGGGACCTGGAGGAGGCCACACTGCAGCACGAAGCCATGGTGGCCACGCTGAGGAAGAAGCATGCGGATAGTGTGGCCGAGCTTGGGGAGCAGATTGACAACCTGCAGCGGGTCAAG
     970       980       990      1000      1010      1020      1030      1040      1050      1060      1070      1080
Q K L E K E K S E F K L E I D D L S S S M E S V S K S K A N L E K I C R T L E D
CAGAAGCTGGAGAAGGAGAAGAGCGAGTTCAAGCTGGAGATCGATGACCTCTCCAGCAGCATGGAGAGTGTGTCGAAATCTAAGGCAAATCTGGAAAAAATCTGCCGAACCCTGGAGGAT
     1090      1100      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
Q L S E A R G K N E E I Q R S L S E L T T Q K S R L Q T E A G E L S R Q L E E K
CAGTTAAGTGAGGCCAGGGGCAAGAATGAGGAAATTCAGAGGAGCCTGAGCGAGCTGACCACACAGAAGTCTCGTTTGCAGACCGAGGCTGGTGAGCTGAGTCGTCAGCTGGAAGAAAAA
     1210      1220      1230      1240      1250      1260      1270      1280      1290      1300      1310      1320
E S I V S Q L S R S K Q A F T Q Q T E E L K R Q L E E E N K A K N A L A H A L Q
GAAAGCATAGTATCCCAACTTTCCAGGAGCAAGCAAGCCTTTACCCAGCAAACAGAAGAGCTCAAGAGGCAGCTGGAGGAAGAGAACAAGGCCAAGAACGCCCTGGCGCACGCCCTGCAG
     1330      1340      1350      1360      1370      1380      1390      1400      1410      1420      1430      1440
S S R H D C D L L R E Q Y E E E Q E G K A E L Q R A L S K A N S E V A Q W R T K
TCCTCCCGCCACGACTGTGACCTGCTGCGGGGAACAGTATGAGGAGGAGCAGGAAGGCAAAGCTGAGCTGCAGAGGGCGCTGTCCAAGGCCAATAGTGAGGTTGCCCAGTGGAGAACCAAA
     1450      1460      1470      1480      1490      1500      1510      1520      1530      1540      1550      1560
Y E T D A I Q R T E E L E E A K K K L A Q R L Q D S E E Q V E A V N A K C A S L
TACGAGACGGACGCCATCCAGCGCACAGAAGAGCTGGAGGAGGCCAAGAAAAAACTTGCTCAGCGCCTTCAAGATTCCGAGGAACAGGTTGAGGCAGTGAATGCTAAATGTGCTTCACTG
     1570      1580      1590      1600      1610      1620      1630      1640      1650      1660      1670      1680
E K T K Q R L Q G E V E D L M V D V E R A N S L A A A L D K K Q R N F D K V L A
GAGAAGACCAAGCAGAGGCTGCAAGGAGAGGTGGAGGATCTGATGGTTGATGTTGAAAGAGCCAATTCCTTGGCCGCCGCTCTGGACAAGAAGCAGAGGAACTTTGACAAGGTGTTGGCA
     1690      1700      1710      1720      1730      1740      1750      1760      1770      1780      1790      1800
E W K T K C E E S Q A E L E A S L K E S R S L S T E L F K L K N A Y E E A L D Q
GAGTGGAAGACAAAGTGTGAGGAGAGCCAAGCAGAGCTGGAGGCATCCCTGAAGGAGTCCCGCTCCTTGAGCACTGAGCTCTTCAAACTGAAAAATGCCTACGAGGAAGCCTTAGATCAA
     1810      1820      1830      1840      1850      1860      1870      1880      1890      1900      1910      1920
L E T V K R E N K N L E Q E I A D L T E Q I A E N G K T I H E L E K S R K Q I E
CTTGAAACTGTGAAACGGGAAAATAAGAACTTAGAGCAGGAGATAGCAGATCTCACAGAACAAATTGCTGAAAATGGCAAAACCATCCATGAACTGGAGAAATCAAGAAAGCAGATTGAG
     1930      1940      1950      1960      1970      1980      1990      2000      2010      2020      2030      2040
L E K A D I Q L A L E E A E A A L E H E E A K I L R I Q L E L T Q V K S E I D R
CTGGAAAAGGCTGATATCCAGCTGGCTCTCGAGGAAGCAGAGGCTGCTCTTGAGCATGAAGAAGCCAAGATCCTCCGAATCCAGCTTGAATTGACACAAGTGAAAATCAGAAATTGATAGA
     2050      2060      2070      2080      2090      2100      2110      2120      2130      2140      2150      2160
K I A E K D E E I E Q L K R N Y Q R T V E T M Q S A L D A E V R S R N E A I R L
AAGATTGCCGAGAAGGATGAAGAGATCGAGCAGCTGAAGAGGAACTACCAGAGAACAGTGGAAACCATGCAGAGCGCCCTGGACGCCGAGGTGCGGAGCAGGAATGAAGCCATCCGGCTC
     2170      2180      2190      2200      2210      2220      2230      2240      2250      2260      2270      2280
K K K M E G D L N E I E I Q L S H A N R Q A A E T L K H L R S V Q G Q L K D T Q
AAGAAGAAGATGGAGGGGGACCTGAATGAAATCGAGATCCAGCTGAGCCACGCCAACCGCCAGGCGGCGGAGACCCTCAAACACCTCAGGAGTGTCCAGGGACAGCTGAAGGATACGCAG
     2290      2300      2310      2320      2330      2340      2350      2360      2370      2380      2390      2400
L H L D D A L R G Q E D L K E Q L A I V E R R A N L L Q A E V E E L R A T L E Q
CTCCACCTGGATGATGCCCTCCGAGGCCAGGAGGACCTGAAGGAGCAGCTGGCGATTGTGGAGCGCAGAGCCAACCTGCTGCAGGCCGAGGTGGAGGAGCTGCGGGCTACTCTGGAGCAG
     2410      2420      2430      2440      2450      2460      2470      2480      2490      2500      2510      2520
T E R A R K L A E Q E L L D S N E R V Q L L H T Q N T S L I H T K K K L E T D L
ACGGAGAGGGCCCGGAAACTGGCGGAACAGGAGCTCCTGGACTCCAACGAGAGGGTGCAGCTGCTGCATACCCAGAACACCAGCCTCATCCACACCAAGAAGAAGCTGGAGACAGACCTC
     2530      2540      2550      2560      2570      2580      2590      2600      2610      2620      2630      2640
M Q L Q S E V E D A S R D A R N A E E K A K K A I T D A A M M A E E L K K E Q D
ATGCAGCTCCAGAGTGAGGTAGAAGATGCCAGCAGGGATGCAAGGAACGCTGAGGAGAAGGCCAAGAAGGCCATCACGGACGCTGCCATGATGGCGGAGGAGCTGAAGAAGGAGCAGGAC
     2650      2660      2670      2680      2690      2700      2710      2720      2730      2740      2750      2760
T S A H L E R M K K N L E Q T V K D L Q H R L D E A E Q L A L K G G K K Q I Q K
ACCAGCGCCCACCTTGAGCGGATGAAGAAGAACCTGGAACAGACGGTGAAGGACCTGCAGCATCGTCTAGATGAGGCCGAGCAGCTGGCGCTGAAGGGCGGGAAGAAGCAGATCCAGAAA
     2770      2780      2790      2800      2810      2820      2830      2840      2850      2860      2870      2880
L E T R I R E L E F E L E G E Q K K N T E S V K G L R K Y E R R V K E L T Y Q S
CTGGAGACCAGGATCCGAGAGCTGGAGTTTGAACTTGAGGGAGAGCAGAAGAAGAACACAGAGTCTGTTAAGGGCCTGAGGAAGTATGAGCGGAGGGTCAAGGAGCTGACGTACCAGAGT
     2890      2900      2910      2920      2930      2940      2950      2960      2970      2980      2990      3000
E E D R K N V L R L Q D L V D K L Q V K V K S Y K R Q A E E A D E Q A N A H L T
GAAGAGGACAGGAAGAATGTGCTGAGATTGCAGGATCTGGTGGATAAACTGCAAGTGAAAGTCAAGTCCTACAAGAGGCAGGCGGAGGAGGCTGATGAACAAGCCAATGCTCATCTCACC
     3010      3020      3030      3040      3050      3060      3070      3080      3090      3100      3110      3120
K F R K A Q H E L E E A E E R A D I A E S Q V N K L R A K T R D F T S S R M V V
AAATTCCGAAAGGCTCAGCATGAGCTGGAGGAGGCCGAGGAACGTGCGGATATCGCAGAATCTCAAGTCAACAAGCTCCGCGCTAAGACTCGAGACTTCACCTCCAGCAGGATGGTGGTC
     3130      3140      3150      3160      3170      3180      3190      3200      3210      3220      3230      3240
H E S E E *
CACGAGAGTGAAGAGTGAGCCAGCCCTTCTGGAGCAGGACAGAAGATATGCAAAATGTATATTTTCTTGATTCCTGACCATTGATACTTAATGTCCATGTGACTCTTTTTCACATGCAAT
     3250      3260      3270      3280      3290      3300      3310      3320      3330      3340      3350      3360
AAACTTTGCTTTGTTTCAAAAA
     3370      3380
```

Figure 2. Complete Nucleotide and Derived Amino Acid Sequence of the Human Embryonic Skeletal Muscle cDNA, SMHCE. The complete protein coding and 3′ untranslated sequence is shown with the deduced amino acid sequence indicated by single letter code. The termination codon is denoted by a *. The gene specific RNA probe that was used in the RNA hybridizations was derived from nucleotides 3147–3376.

Table 1 Sequence Comparisons of Human Embryonic MHC with Other MHC Isoforms.

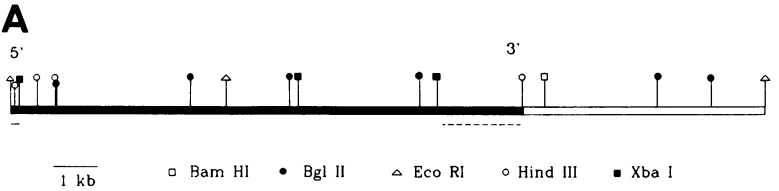
Isoform	% Homology Coding Nucleotides	% Homology 3'Noncoding Nucleotides	% Homology Amino Acids	Conservative Total A.A changes
Rat Embryonic[33]	89.0%	79.8%	97.0%	28/32
Human Perinatal[9]	78.5%	N.D.[a]	83.8%	82/108
Human Adult Fast[8]	77.7%	57.7%	80.8%	96/122
Human Adult Slow[8]	73.8%	N.D.[a]	77.2%	117/149

Comparison of the nucleotide and amino acid sequences of SMHCE over its entire length (3.4 kb, 1083 amino acids) with the analogous full length rat embryonic MHC cDNA sequence. The 3' 2.0 kb and 635 carboxy amino acids of SMHCE was also compared with analogous regions of three human isoforms: perinatal MHC, an adult fast skeletal and adult slow skeletal isoform.
[a]not detectable.

staining had demonstrated that at 4 days post-fusion, primary human muscle cultures are capable of synthesizing at least three MHC isoforms: embryonic, neonatal and slow skeletal (unpublished observations). Human DNA probes are available for two of these three MHC forms: perinatal and slow skeletal as well as for adult fast skeletal MHC (8,9). We screened 5×10^5 clones from this cDNA library first with a radioactive DNA probe corresponding to a highly conserved region of the sarcomeric MHC rod (8). Sixty MHC clones were selected for further analysis. Hybridization of these cDNA clones to MHC gene-specific probes showed that 10% of them correspond to perinatal skeletal MHC and 3% to adult fast skeletal MHC. 78% of these 60 cDNA clones represent a single MHC sequence (as determined initially by restriction endonuclease digestion) that do not correspond to adult fast, slow or perinatal skeletal MHC (data not shown). The remaining 9% of the MHC clones have not yet been identified.

A cDNA clone from the predominant MHC class (named SMHCE with a 3.4 kb insert) was further characterized by reactivity with 12 monoclonal antibodies raised against developmentally regulated human MHC isoforms. The fusion protein encoded by the λgt11 cDNA clone reacts with antibody F1.652 (13,17,24), which is known to recognize embryonic skeletal muscle myosin. Figure 1 shows the immunoreactivity of F1.652 with the 220 kilodalton fusion protein encoded by the λgt11 SMHCE cDNA clone. The size of this protein product is consistent with the 116 kdal size of β-galactosidase plus a cDNA peptide translated by a 3.4 kb mRNA.

In order to derive the sequence for SMHCE, the insert was subcloned into plasmid pSK^- and nested deletions were generated. SMHCE encodes all but 8 amino acids of the MHC rod as well as 123 nucleotides of 3' noncoding sequences and the poly (A) tail. The entire nucleotide and derived amino acid sequences for pSMHCE are shown in Figure 2. Table 1 shows the results of the comparison of the nucleotide and protein sequences of SMHCE with the analogous rat embryonic MHC gene sequence and three human MHC cDNA sequences (8,9). SMHCE is most homologous to the gene encoding the rat embryonic MHC (33). The human and rat embryonic isoforms are 89% homologous at the nucleotide level and 97% homologous at the amino acid level over the length of pSMHCE. This high degree of homology extends into the 3' untranslated sequences which are 80% homologous. The rat and human proteins are virtually identical structurally, since the vast majority (88%) of the differences in protein sequence are conservative amino acid changes. When the sequence of SMHCE is compared with the available 2.0 kb of sequence encoding



B

```
 E  F  Q  K  T  K  D  E  L  A  K  S  E  A  K  R  K  E  L  E
GAATTCCAGAAAACCAAAGATGAACTCGCCAAGTCGGAGGCAAAAAGGAAGGAGCTAGAG
        10         20         30         40         50         60
 E  K  L  V  T  L  V  Q  E  K  N  D  L  Q  L  Q  V  Q  A
GAAAAACTGGTGACTCTGGTCCAAGAGAAGAATGACCTGCAGCTCCAAGTACAAGCTgta
        70         80         90        100        110        120
cgtgtttagcaatctttttttttttttaactcttctagatttgattatttatttagaaaaa
       130        140        150        160        170        180
ttgagttattggtggggatttttgcttttcagttgaaatgcttcataaagaacttcatat
       190        200        210        220        230        240
atagcgaaatactggaatgtcgcaaaccttatatcattcaacgaagttattattcttttt
       250        260        270        280        290        300
     E  S  E  N  L  L  D  A  E  E  R  C  D  Q  L  I  K
ttttaagGAAAGCGAAAATTTGTTGGATGCTGAGGAAAGATGCGATCAGCTGATCAAAG
       310        320        330        340        350        360
A  K  F  Q  L  E  A  K
CCAAATTCCAGCTCGAGGCCAAGA
       370        380
```

C

```
ccgctgaacacacacagcatccgacagtaggtcgctgggtcttcagcgtttggggcttgc
        10         20         30         40         50         60
tctttcggagggccagtttctcaggctatggggtgacttcatgtgtgcaatcgtgcgttt
        70         80         90        100        110        120
tcccaaggtccaaaaaatatcctgagtcccactgaagtaggcaagcaactctgatccagc
       130        140        150        160        170        180
cggcagaagtggtgggtgggaatgtgtctctgccacccgatctgactccaggtgcctact
       190        200        210        220        230        240
          M  V  V  H
ttcttcctagATGGTGGTCCAC
       250        260
```

Figure 3. Restriction Map and Partial Nucleotide Sequence of λHMHC2. (A) Restriction map of the λHMHC2 insert. The shaded area corresponds to the transcribed region as determined by hybridization with a radiolabeled SMHCE insert. The 3′ end of the gene has been mapped to a 1.9 kb Xba I-Hind III restriction fragment. (B) DNA sequence of the region indicated by the solid line was generated using a synthetic oligonucleotide derived from sequences near the 5′ end of the cDNA as a primer. Protein coding sequence is shown in upper case letters with the deduced amino acid sequence above in single letter code. Intervening sequence is shown in lower case letters. (C) DNA sequence obtained from the region indicated by the dashed line was determined using a synthetic oligonucleotide derived from sequences of the 3′ untranslated region of the cDNA. The coding and noncoding sequences are determined as above.

635 carboxy-terminal amino acids of three other human muscle MHC cDNAs, their sequences are ~ 75% homologous on the nucleotide level and ~ 80% homologous on the protein level. There is little homology between the 3′ untranslated regions of the human isoforms.

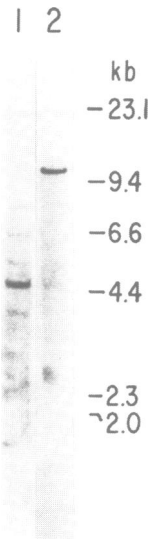


Figure 4. Hybridization of Human Genomic DNA with probes derived from SMHCE. 15μg of DNA was digested with Eco RI, electrophoresed in Tris-Acetate and transferred to GeneScreen. Lane 1 was hybridized to a random primed probe (nucleotides 1−610) that encodes a conserved portion of the MHC molecule and therefore recognizes numerous bands. Lane 2 was hybridized to a gene-specific probe corresponding to nucleotides 3147−3376. Exposure was for 6 hours at −70°C with a DuPont Cronex intensifying to Kodak XAR-5 film.

Identification of the Gene Encoding Human Embryonic MHC.

A human embryonic skeletal muscle MHC genomic clone was previously isolated which has a 17.1 kb insert contained on two EcoRI fragments (34). λHMHC2 has been identified as the gene encoding SMHCE by hybridization with a gene-specific radiolabeled oligonucleotide probe from the 3′ untranslated region of the cDNA (data not shown). This genomic clone contains sequences encoding the entire 3.4 kb of SMHCE over 11.4 kb

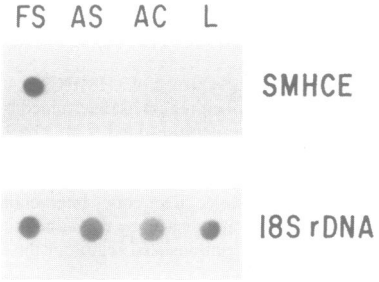


Figure 5. RNA Dot Blot Analysis Showing SMHCE Expression. 5 μg of total RNA isolated from human fetal skeletal (FS), adult skeletal (AS), adult cardiac (AC) and liver (L) was spotted onto a filter and hybridized to a gene-specific probe from the 3′ untranslated region of SMHCE. The same filter was hybridized after erasing to 18S rDNA to control for variability in loading. Exposure was for 4 hours at −70°C with a DuPont Cronex intensifying screen to Kodak XAR-5 film.

of genomic sequence including the EcoRI site at the 5′ end of the cDNA which is encoded by the gene and is not artifactual as a result of an EcoRI linker. Partial sequence analysis of the genomic clone primed with oligonucleotides synthesized to sequences complementary to sequences near the 5′ end of the cDNA clone (nucleotides 200−223) and in the 3′ untranslated region (nucleotides 3257−3274) confirm the identity of the gene. These sequences, as well as a partial restriction map of the genomic clone, are shown in Figure 3. The positions of the two introns in the human embryonic MHC gene that have been analyzed are identical to introns within the rat embryonic MHC gene (33). The coding sequences at the 5′ end of λHMHC2 correspond to exons 22 and 23 (out of 41) (Figure 3, Panel B). The sequences of the human and rat intron at this position show 59% homology to each other. The human intron is two nucleotides shorter than the corresponding intron in the rat gene. While the position of this intron is also conserved in the chicken embryonic MHC gene (35), its size is not conserved between mammalian (191−193 nucleotides) and chicken (913 nucleotides) embryonic MHC genes. The sequence from the 3′ end of the gene (Figure 3, Panel C) contains the last coding exon (exon 41) and part of an intron that is 65% homologous to the corresponding intron in the analogous rat gene.

Expression of the Human Embryonic MHC Gene.

Embryonic skeletal MHC protein is expressed only during fetal development and its expression declines at birth (15,36, 37). In order to establish the pattern of expression of the gene at the mRNA level, RNA was isolated from four tissues and hybridized to a radiolabeled gene-specific RNA probe from the 3′ end of the cDNA (nucleotides 3147−3376). The probe used was gene-specific since it recognized only a single band in an Eco RI digest of human genomic DNA while a coding region probe (nucleotides 1−610) recognized multiple bands (Figure 4). This probe reacted only with the RNA isolated from fetal skeletal muscle, and not from adult skeletal muscle, adult cardiac muscle or liver (Figure 5).

The temporal pattern of SMHCE mRNA and protein expression was investigated in tissue culture. A clonal population of pure myoblasts originally isolated from the *vastus lateralis* of a 5 year old male was grown to confluence and then fused by changing the culture medium of the cells to a mitogen poor medium. RNA was isolated from myoblasts at low (50%) and high (>80%) confluence and from myotubes during a time course that spanned from two to ten days after changing the cells into low serum, differentiation medium. At high confluence, a small percentage of the cells had started to fuse even though they were in proliferation promoting media. There were no multinucleated myotubes seen in the low confluence myoblasts. These RNAs were hybridized to the gene specific radiolabeled RNA probe from the 3′ untranslated region of SMHCE (Figure 6A, Panel 1). No SMHCE mRNA is seen in low confluence myoblasts but is apparent at very low levels in high confluence cells. Expression of SMHCE mRNA is maximal at 2 days post fusion and declines to barely detectable levels by 10 days post fusion. The same filters were also hybridized to α-cardiac actin to show that these cells are still expressing muscle specific proteins after SMHCE expression has decreased (Figure 6A, Panel 2). Protein extracts were made from cells at the same time points and Western blot analysis was carried out using the monoclonal antibody, F1.652, which recognizes a specific epitope in the rod encoded by SMHCE. The pattern of SMHCE protein expression follows RNA accumulation patterns, with an approximate 24 hour delay (Figure 6B). There is no embryonic MHC seen in myoblasts either at low or at high confluence. Protein expression peaks at 4 days after fusion and cannot be detected at 10 days post fusion. Thus, the

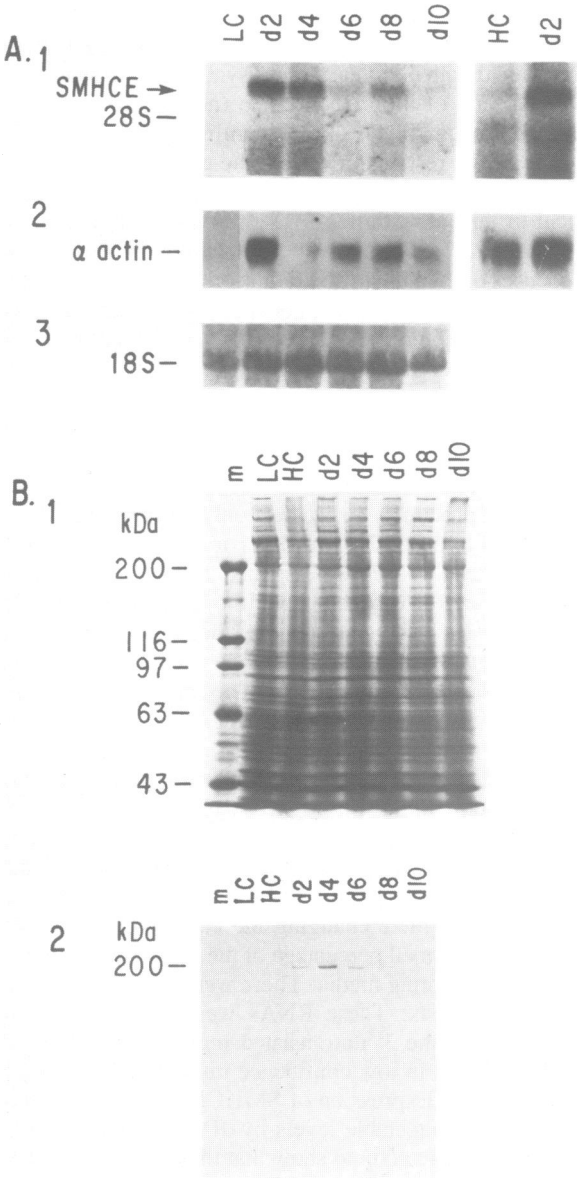


Figure 6. Expression of SMHCE in Skeletal Muscle Cell Cultures. 20 μg of total RNA isolated from low confluence cultured myoblasts (LC), high confluence myoblasts (HC) and myotubes 2 (d2), 4 (d4), 6 (d6), 8 (d8), and 10 (d10) days after fusion were size fractionated on a 1% agarose gel. (A) The filter was hybridized to (Panel 1) a radiolabeled RNA probe from the 3′ end of SMHCE, (Panel 2) a random-primed DNA fragment from a human α cardiac actin cDNA (29) and (Panel 3) a random-primed 18S rDNA probe to control for RNA loading. (B)

expression of the gene encoding SMHCE is induced and then declines in cultured cells in a manner that mimics its temporal regulation *in vivo*.

DISCUSSION

We have isolated a human MHC cDNA and partial genomic clone that encode an embryonic stage-specific MHC isoform that is expressed exclusively in skeletal muscle. We have identified the cDNA through DNA sequence analysis, monoclonal antibody reactivity and RNA hybridization patterns. The carboxy terminal 14−17 amino acids of sarcomeric MHCs are highly divergent among developmentally regulated isoforms within an organism and can be used to identify any given isoform (9,38). When these carboxy terminal amino acids of the rat and human embryonic MHC isoforms are compared, they are 100% homologous. The last eight amino acids and 3′ untranslated sequences are contained on a separate exon in all mammalian MHC genes that have been isolated and sequenced (3,32,39,40,41), including the human embryonic MHC gene presented in this report. This conservation of genomic organization may be explained by functional constraints on maintaining distinct MHC isoform sequences. The functional role of these sequences has not been elucidated, but their conservation during evolution implies their significance. Since these sequences are in the myosin rod, such function may reside in the ability of myosin isoforms to self-associate to form homogeneous thick filaments (42). A role in sarcomere assembly may be particularly relevant to myosins expressed early in myogenesis such as the embryonic MHC described here.

The cell culture system used in this study when used in conjunction with a molecular DNA probe encoding the embryonic skeletal MHC isoform should allow the analysis of the induction and repression of expression of the human embryonic skeletal MHC gene. Both embryonic and perinatal MHC genes are expressed in early development, peak before birth and decline at some point after birth. The earliest time of expression of these genes has not yet been determined, but *in situ* hybridization of embryonic tissue should make this determination possible. Developmental changes in expression of both genes are nerve independent and portions of this process have been shown to occur in myogenic cell lines (13,14) as well as in human primary cultures (this report). However, the decline in expression of the embryonic gene with time has not always been observed in myogenic cell lines (14) Therefore, human primary cultures may more accurately reflect the *in vivo* regulation of MHC gene expression than do some myogenic cell lines and primary cultures from other organisms. Accordingly, it should now be possible to define those sequences of MHC genes that are responsible for their developmental regulation by introduction of regulatory regions of these genes into primary muscle cultures.

The embryonic MHC protein is diagnostic of muscle regeneration and is expressed aberrantly after birth in several human genetic muscle disorders such as Duchenne muscular dystrophy and infantile muscular atrophy (15,17). In addition, muscle tumors such as rhabdomyosarcoma express the embryonic skeletal MHC isoform. However, in the nongenetic atrophic muscle diseases, it appears that muscles recycle through their developmental pattern, starting with SMHCE. Thus, SMHCE expression is not irreversibly

shows a Coomassie brilliant blue stained (Panel 1) and a immunoblot (Panel 2) of a 7.5% SDS/PAGE gel (31) containing whole cell lysates of myoblasts and myotubes isolated at the same time points as the RNAs. The immunoblot (Panel 2) was reacted with mAb F1.652 in order to show the time course of expression of SMHCE protein.

inactivated. We hope to further investigate the regulation of SMHCE expression to gain insight into both the process of muscle development and these diseased states. Different slow and fast muscle fiber types cease expression of the embryonic MHC gene at different times development and different fiber types activate and express the gene at different times during muscle regeneration (17). Therefore, the F1.652 antibody together with the DNA clone described here can serve as sensitive markers for identifying muscle fiber types during development and regeneration as well as documenting the regulatory mechanisms underlying very early changes in MHC gene expression.

ACKNOWLEDGEMENTS

We would like to thank R. Feghali for assistance with RNA analysis. I. K.-M. is supported by NIH5T32CA09060. The research was supported by NIH GM29090 to L.A.L. and a grant from the Muscular Dystrophy Association and NIH HD18179 to H.B. L.A.L. is an American Heart Association Established Investigator. H.B. is the recipient of an RCDA from the NIH.

*To whom correspondence should be addressed

REFERENCES

1. Emerson Jr.,C.P. and Bernstein,S.I. (1987) Ann. Rev. Biochem., **56**, 695–726.
2. Warrick,H.M. and Spudich,J.A. (1987) Ann. Rev. Cell Biol., **3**, 379–421.
3. Wieczorek,D.F., Periasamy,M., Butler-Browne,G.S., Whalen,R.G. and Nadal-Ginard,B. (1985) J. Cell Biol., **101**, 618–629.
4. Nagai,R., Larson,D.M. and Periasamy,M. (1988) Proc. Natl. Acad. Sci. USA., **85**, 1047–1051.
5. Leinwand,L.A., Saez,C., and Myers,J. (1988) J. Cell. Biochem., **Supp 12C**, 329 (abstr.).
6. Lowey,S., Slayter,H.S., Weeds,A.G. and Barber,H. (1969) J. Mol. Biol., **42**, 1–29.
7. Philpott,D.E. and Szent-Gyorgyi,A.G. (1954) Biochim. Biophys. Acta., **15**, 165–173.
8. Saez,L.J. and Leinwand,L.A. (1986) Nucleic Acids Res., **14**, 2951–2969.
9. Feghali,R. and Leinwand,L.A. (1989) J. Cell Biol., **108**, 1791–1797.
10. Cerney,L.C. and Bandman,E. (1986) J. Cell Biol., **103**, 2153–2163.
11. Mahdavi,V., Izumo,S. and Nadal-Ginard,B. (1987) Circ. Res., **60**, 804–814.
12. Whalen,R.G., Bugaisky,L.B., Butler-Browne,G. and Pinset,C. (1984) In Eppenberger,H.M. and Perriard J.C. (eds.), Developmental Processes in Normal and Diseased Muscle. Basel, Karger, Vol. 9, pp. 155–164.
13. Silberstein,L., Webster,S.G., Travis,M. and Blau,H.M. (1986) Cell, **46**, 1075–1081.
14. Weydert,A., Barton,P. Harris,A.J., Pinset,C. and Buckingham,M. (1987) Cell, **49**, 121–129.
15. Fitzsimons,R.B. and Hoh,J.F.Y. (1981) J. Neurol. Sci., **52**, 367–384.
16. Schiaffino,S., Gorza,L., Satore,S., Saggin,L. and Carli,M. (1986) Exp. Cell Res., **163**, 211–220.
17. Webster,C., Silberstein,L., Hays, A.P. and Blau,H.M. (1988) Cell, **52**, 503–513.
18. Blau,H. and Webster,C. (1981) Proc. Natl. Acad. Sci. USA., **78**, 5623–5627.
19. Walsh,F.S., Quinn,C.A., Yasin,R., Thompson,E.J. and Hurko,O. (1982) In Pearson,M.L. and Epstein,H.F. (eds.), Muscle Development: Molecular and Cellular Control. Cold Spring Harbor Laboratory Press, Cold Spring Harbor. pp. 535–542.
20. Webster,C., Pavlath,G. Parks,D., Walsh,F. and Blau,H.M. (1988) Exp. Cell Res., **174**, 252–265.
21. Childs, G., Maxson,R. and Kedes,L. (1979) Dev. Biol., **73**, 153–173.
22. Glisin,V., Crkvenjakov,R. and Byus,C. (1974) Biochemistry, **13**, 2633–2637.
23. Leonard,D.G., Ziff,E.B. and Greene,L.A. (1987) Mol. Cell. Biol., **7**, 3156–3167.
24. Silberstein,L. and Blau,H.M. (1986) In Emerson,C., Fischman,D.A., Nadal-Ginard,B. and Saddiqui,M.A.Q. (eds.), Molecular Biology of Muscle Development. Alan Liss Inc., NY. pp. 253–262.
25. Huynh T.V., Young,R.A. and Davis,R.W. (1985) In Glover,D.M. (ed.), DNA Cloning-A Practical Approach. IRL press, Oxford, Vol. I, pp. 49–78.
26. Kraft,R., Tardiff,J., Krauter,K.S. and Leinwand,L.A. (1988) BioTechniques, **6**, 544–549.
27. Southern,E.M. (1975) J. Mol. Biol., **98**, 503–517.

28. Arnheim,N., Krystal,M., Schmickel,R., Wilson,G., Ryder,O. and Zimmer,E. (1980) Proc. Natl. Acad. Sci. USA., **77**, 7323–7327.
29. Gunning,P., Mohun,T., Ng,S-Y., Ponte,P. and Kedes,L. (1984) J. Molec. Evol., **20**, 202–221.
30. Reinach,F.C. and Fischman,D.A. (1985) J. Mol. Biol., **181**, 411–422.
31. Laemmli,U.K. (1970) Nature., **227**, 680–685.
32. Towbin,H., Staehelin,T. and Gordon,J. (1979) Proc. Natl. Acad. Sci. USA., **76**, 4350–4354.
33. Strehler,E.E., Strehler-Page,M.-A., Perriard,J.-C., Periasamy,M. and Nadal-Ginard,B. (1986) J. Mol. Biol, **190**, 291–317.
34. Leinwand,L.A., Saez,L., McNally,E. and Nadal-Ginard,B. (1983) Proc. Natl. Acad. Sci. USA., **80**, 3716–3720.
35. Molina,M.I., Kropp,K.E., Gulick,J. and Robbins,J. (1987) J. Biol. Chem., **262**, 6478–6488.
36. Bandman,E., Matsuda,R. and Strohman,R.C. (1982) Dev. Biol., **93**, 508–513.
37. Whalen,R.G., Sell,S.M., Butler-Browne,G.S., Schwartz,K., Bouveret,P. and Pinset-Harstrom,I. (1981) Nature, **292**, 805–809.
38. Mahdavi,V., Periasamy,M. and Nadal-Ginard,B. (1982) Nature, **297**, 659–664.
39. Mahdavi,V., Chambers,A.P. and Nadal-Ginard,B. (1984) Proc. Natl. Acad. Sci. USA., **81**, 2626–2630.
40. Periasamy,M., Wydro,R.M., Strehler-Page,M.-A., Strehler,E.E. and Nadal-Ginard,B. (1985) J. Biol. Chem., **260**, 15856–15862.
41. Saez,L.J., Gianola,K.M., McNally,E.M., Feghali,R., Eddy,R., Shows,T.B. and Leinwand,L.A. (1987) Nucleic Acids Res., **15**, 5443–5459.
42. Taylor,L.D. and Bandman,E. (1989) J. Cell Biol., **108**, 533–542.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.