

# Human nonmuscle myosin heavy chain mRNA: Generation of diversity through alternative polyadenylation

(DNA sequence/chromosome mapping)

CLAUDIA G. SAEZ\*, JEANNE C. MYERS†, THOMAS B. SHOWS‡, AND LESLIE A. LEINWAND\*

\*Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, NY 10461; †Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104; and ‡Department of Human Genetics, Roswell Park Memorial Institute, Buffalo, NY 14263

Communicated by Harry Eagle, November 3, 1989

**ABSTRACT** Myosin is a ubiquitous eukaryotic contractile protein that generates the force responsible for such diverse cellular movements as muscle contraction and cytokinesis. Although there have been numerous studies of sarcomeric myosin heavy chain (MHC) genes, no molecular clones have been reported that encode mammalian nonmuscle MHC. This study presents the molecular genetic characterization of a human nonmuscle MHC that is expressed in fibroblasts, endothelial cells, and macrophages. Human nonmuscle MHC amino acids are weakly homologous (33%) to sarcomeric MHC but are ≈72% identical to smooth muscle MHC. In contrast to vertebrate sarcomeric MHCs, which generate diversity through the expression of members of a multigene family, an alternative polyadenylation site is used in the nonmuscle MHC gene to generate multiple transcripts that encode the same protein. We have mapped this gene to chromosome 22. It is thus unlinked to either of the sarcomeric MHC gene clusters on human chromosomes 14 and 17.

Myosin is a hexameric protein consisting of two heavy chains and two pairs of nonidentical light chains. Through its interaction with actin, myosin generates the force that results in many forms of cellular movement. Although best studied in muscle contraction, myosin is also involved in many forms of nonmuscle cell movement, including cytokinesis and cell locomotion (see refs. 1 and 2). It has been difficult to identify all of the functions for nonmuscle cell myosin. However, recent genetic experiments in *Dictyostelium* have shown that myosin heavy chain (MHC) mutants are defective in cytokinesis, impaired in chemotaxis, and show arrested development (3, 4). Biochemical studies have distinguished myosin isoforms as having high or low ATPase activities and molecular genetic studies have identified some of the differences in sequence that may contribute to functional diversity (see refs. 5 and 6).

In nonmuscle cells, myosin is found at relatively low levels and is not organized into a sarcomere-like structure with actin. While electron microscopic examination of myosin from nonmuscle cells demonstrates that it has the same general structure as that of muscle myosin (see refs. 7 and 8), it does not share the protease-susceptible sites found in striated muscle myosin. In contrast to the extreme sequence conservation seen between muscle and nonmuscle actin, sarcomeric and nonmuscle MHC are so divergent that most sarcomeric MHC DNA probes will not recognize RNA from nonmuscle sources (see ref. 9). There is also considerable sequence divergence demonstrated when sarcomeric and smooth muscle MHCs are compared (10, 11). It seems likely that the high degree of sequence divergence between sarco-

meric and nonmuscle myosin reflects their involvement in different contractile processes.

To learn more about the sequence requirements for nonmuscle motile processes, we have isolated human nonmuscle MHC cDNA clones.<sup>§</sup> We show that the use of an alternative polyadenylation site is responsible for the generation of at least two human nonmuscle MHC transcripts that encode the same protein. Genomic hybridization analysis suggests that there is a small number of human nonmuscle MHC genes. We have determined that the gene encoding the mRNAs studied here resides on chromosome 22 and is unlinked to the sarcomeric MHC gene clusters on human chromosomes 14 and 17.

## METHODS

**Enzymes and Biochemical Reagents.** Restriction enzymes were purchased from New England Biolabs. Random primer labeling kit, polynucleotide kinase, and the exonuclease III/S1 nuclease kit for the generation of nested deletion clones for sequencing were purchased from Promega Biotec. RNasin was purchased from Life Sciences, Saint Petersburg, FL. Radionucleotides ([ $\alpha$ -<sup>32</sup>P]dCTP, adenosine 5'-[ $\alpha$ -<sup>35</sup>S]-thio]triphosphate) were purchased from Amersham. Seakem GTG grade agarose (FMC) was used.

**Oligonucleotide Synthesis.** Oligonucleotides were synthesized on an Applied Biosystems synthesizer at the Albert Einstein College of Medicine shared facility and purified on 20% acrylamide gels prior to their use as primers in sequencing reactions.

**Subcloning and DNA Sequencing.** The MHC cDNAs were subcloned into pTZ19R (Pharmacia) or pSK (Stratagene). pNMHCM6 was digested with *Kpn* I and *Hind*III for the construction of nested deletions using the exonuclease III/S1 nuclease method (12). Deletions overlapped so that each base was determined a minimum of two times. Double-stranded sequencing was performed on minilysate DNA (13). The STADEN DNA sequence programs were used for computer analysis.

**cDNA Libraries and RNA Analysis.** The  $\lambda$ gt10 human macrophage cDNA library was generously provided by X. Fan (Albert Einstein College of Medicine) and was constructed from the U937 cell line (14). The pBR322 cDNA library was constructed from normal human fibroblast poly(A)<sup>+</sup> RNA (15). RNA for S1 nuclease analyses was prepared by the guanidine isothiocyanate method (16). RNA for Northern blot analysis was prepared according to Pihlajaniemi and Myers (17).

**RNA and DNA Hybridizations.** Genomic DNA was cleaved with restriction enzymes and analyzed by nitrocellulose blot

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MHC, myosin heavy chain; PCR, polymerase chain reaction.

<sup>§</sup>The sequence reported in this paper has been deposited in the GenBank data base (accession no. M31013).

hybridization as described (18). RNA hybridizations were carried out under similar conditions after electrophoresis in a  $1 \times$  Mops/formaldehyde system (18).

**S1 Nuclease Analysis.** S1 nuclease mapping was done by standard techniques (19). The *Nco* I site at nucleotide 4258 of pNMHCM2 were labeled with [ $\alpha$ - $^{32}$ P]dCTP and then digested with *Pvu* II to yield a fragment of 752 nucleotides. Thirty micrograms of RNA were used in the reactions with  $4 \times 10^4$  cpm of probe. Products were analyzed on 5% polyacrylamide/8.3 M urea gels.

**Anchored Polymerase Chain Reaction (PCR).** cDNA was synthesized from 2  $\mu$ g of macrophage RNA using 200 ng of oligonucleotide 1 (5'-CGCGGAATTCCCCGGGCGCGCT<sub>12</sub>-3') as a primer. The newly synthesized DNA was incubated with 50 mM KCl/10 mM Tris-HCl, pH 8.3/0.01% (wt/vol) gelatin/1.5 mM MgCl<sub>2</sub>/1  $\mu$ g each of oligonucleotides 1 and 2 (5'-TCAAGCTGGAAAGGCCAGCAG-3')/0.25 mM dNTPs/2.5 units of *Thermus aquaticus* DNA polymerase (Perkin-Elmer). PCR reactions were carried out in a Perkin-Elmer/Cetus DNA thermal cycler. Samples were denatured at 92°C, annealed at 54°C, and elongated at 74°C each for 1 min for a total of 30 cycles. The final products were electrophoresed, and the DNA was extracted from the gel and subcloned for double-stranded sequencing (13).

## RESULTS

**Isolation and DNA Sequence Analysis of Nonmuscle MHC cDNA Clones.** Three clones encoding human nonmuscle MHC were isolated from two cDNA libraries. The first clone, from fibroblasts (pNMHCF), was isolated by virtue of its presence in a size-selected cDNA library prepared from mRNA that was >5 kilobases (kb) and initially identified by the presence of multiple *Pst* I and *Pvu* II sites in the cloned sequence. The recognition sequences for *Pst* I and *Pvu* II encode the dipeptides Leu-Gln and Gln-Leu respectively, which have been shown to be characteristic of the sarcomeric myosin rod (20, 21). The relationship of pNMHCF to the myosin molecule is shown in Fig. 1. To obtain sequences encoding a greater proportion of the myosin molecule, the insert from pNMHCF was used as a probe to screen a human macrophage cDNA library. The additional clones that resulted from this screening (pNMHCM2 and pNMHCM6) are also shown in Fig. 1. Together, these clones encode 1247

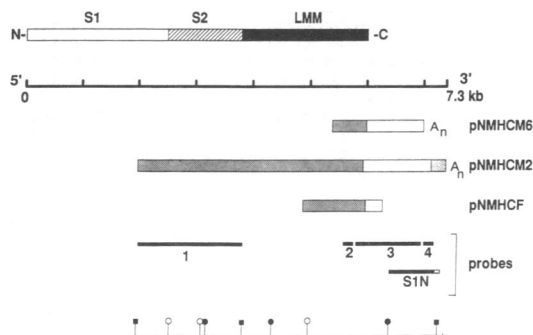


FIG. 1. Schematic representation and partial restriction map of human nonmuscle MHC cDNA clones. The MHC protein is shown above for orientation purposes. The cDNA clones isolated are shown below with positions of several restriction endonuclease cleavage sites. pNMHCF corresponds to a cDNA isolated from human fibroblasts. pNMHCM2 and -6 correspond to cDNAs isolated from a human macrophage cell line. Probes 1-4, which are used in hybridizations, are indicated by short horizontal bars. Probe S1N is the probe used in S1 nuclease analysis. Hatched areas of the cDNA clones correspond to protein coding sequence and the open areas to 3' untranslated regions. Stippled area of pNMHCM2 corresponds to the anchored PCR product. ■, *Eco*RI; ●, *Nco* I; ○, *Bgl* II.

amino acids of the MHC molecule and represent 4964 nucleotides.

Their complete DNA and deduced amino acid sequences are presented in Fig. 2. All three cDNA clones encode the same protein. Surprisingly, pNMHCM6 terminates in a poly(A) tail, 13 nucleotides downstream from AATAAA, while pNMHCM2 extends 200 nucleotides further 3' but does not terminate in a poly(A) tail. To identify the 3' end of the transcript represented by pNMHCM2, anchored PCR was carried out using as the primer in a reverse transcriptase reaction an oligonucleotide with oligo(T) at its 3' end (see *Methods*). The resulting PCR product was cloned and sequenced and found to extend pNMHCM2 by an additional 160 nucleotides, terminating in a poly(A) tail 14 nucleotides downstream of the sequence AATAAA. These data suggest that the two mRNA molecules derive from alternative poly(A) site selection in a single gene. The two transcripts have unusually long 3' untranslated regions of 1010 and 1370 nucleotides. The 3' untranslated regions of the sarcomeric MHC mRNAs are much shorter, ranging from 60 to 150 nucleotides (see ref. 22). Together the clones encode 121 amino acids of the subfragment 1 (S1) head and the entire 1126 amino acids of the rod. The proline residue marking the S1-rod junction is indicated in Fig. 2. The amino acids from the proline residue to the carboxyl terminus have features expected of a myosin rod; that is, they possess the typical periodicity of charge necessary for formation of the  $\alpha$ -helix (data not shown). The rod also has a nonhelical portion at its terminus, consisting of 34 residues. This nonhelical carboxyl terminus has also been described in vertebrate smooth muscle myosins (9, 10) and in a recently reported chicken nonmuscle MHC (23).

**Nonmuscle MHC Is Homologous to Smooth Muscle MHC but Not to Striated Muscle MHC.** Previous studies have shown that most striated muscle MHC molecular probes do not recognize MHC mRNA in nonmuscle cells (see ref. 9), demonstrating substantial sequence divergence and evolutionary distance between these MHC forms. Comparisons of the human nonmuscle MHC amino acid sequences with sarcomeric and smooth muscle MHC are shown in the form of graphic matrix analysis in Fig. 3. Strong amino acid sequence similarity (72% identity) is seen between human nonmuscle MHC and chicken smooth muscle MHC (10) (Fig. 3 *Left*). The similarity does not extend into the 3' untranslated regions of the mRNAs. There is barely detectable homology (33%) between human nonmuscle and a human striated muscle MHC (22) (Fig. 3 *Right*). There is 90% amino acid sequence identity with a chicken nonmuscle MHC (23). The human nonmuscle sequences were also compared to invertebrate MHCs such as those from yeast and *Dictyostelium*, but no direct sequence homology could be seen despite the conservation of the periodic features of the myosin rod (data not shown).

**Two Nonmuscle MHC Transcripts Result from Use of Alternative Polyadenylation Sites.** As shown in Fig. 1, two MHC cDNAs were isolated that were identical where they overlap, differing in length at their 3' ends. RNA hybridization and S1 nuclease analyses were carried out to investigate the origin and nature of the transcripts represented by the two cDNA clones. Probes 2, 3, and 4 (indicated in Fig. 1 as horizontal bars) were used in the hybridization studies shown in Fig. 4. A protein coding region probe (probe 2) recognizes a single 7.3-kb band in mouse and bovine poly(A)<sup>+</sup> RNA (lanes 1 and 2). However, in addition to the 7.3-kb RNA species, human fibroblast and endothelial cell RNA show an additional band of 7.1 kb (lanes 3 and 4). It seemed likely that the two transcripts seen in Northern blot analysis correspond to the two cDNA clones. Probe 3, which derives from 3' untranslated sequence shared by both cDNAs, hybridizes to both transcripts (lane 4). Probe 4, which is specific to the 3'

```

Q E F R Q R Y E I L T P N S I P K G F M D G K Q A C V L M I K A L E L D S N M L Y
CCAGGAGTTTCGGCAGAGATGAGATCTGACCTCAAACTCCATCCCAAGGTTTCATGGAGGGGAGCAGGCGTGGCTGCTCATGATAAAAGCCCTGGAGCTGCAGACAAATCTGTGA 120
R I G Q S K V F R Q F R A G V L A H L E E E R D L K I T D V I I G F Q A C R G Y L
CCGATTCGGCCAGAGCAAATCTTCCGTCGGGTTGCTGCGCCACCTGGAGGAGGAGCAGGATGAAGATCACCGAGCTCATATAGGGTCCAGGCTGCTGCAGGCGGTCTAC 240
A R K A F A K R A G T Q Q L T A M K V L Q R N C A A Y L K L R N W Q W L R F T K V
GGCCAGGAAGCATTGCCAAGCCGAGCAGCAGCTTACCGCCATGAAGTCTCCAGCGGAACTGGCTGCTTACCTGAGCTCGGGAAGCTGGCAGTGGTGGCGGCTCTCCACCAAGG 360
K [ ] L L Q V S R Q E E E M H A K E E E L V K V R E K Q L A A E N R L M E M E T L
CAAGCCCTGCTGAGCTGAGCCGAGGAGGAGATGATGGCCAGGAGGAGGAGCTGGTGAAGTCAAGAGAGCAGCTGGCTGCGGAGAACAGCTCATGAGATGGAGACGCT 480
Q S Q L M A E K L Q L Q E Q L Q A E T E L C A E A E E L R A R L L T A K K Q E L E
CGATTCAGCTTGGCAGAGAAATGGAGCTGCAGGAGCAGCTCCAGGAGAAACCGAGTGTGTGCCAGGCTGAGGAGCTCCGGGCGGCTGACCGCCAAAGAGCAGGAATTAGA 600
E I C H D L E A R V E E E E E E R Y Q H L Q A E K K K M Q Q N I Q E L E E Q L E E
AGATTCGCCATGACCTAGAGGCTGGAGGAGGAGGAGCGCTACCAGCACCTCGAGCGGAGAGAAAGATGCAGCAGAACATCCAGGAGCTTGGAGCAGCTGGAGGA 720
E S A R Q K L Q L E K V T T E A K L K L E E E Q I I L E D Q N C K L A K E K
GGAGGAGCGCCGGCAGAACTGAGCTGGAGAGGTGACCACCGAGCCAGCTGAAAAGCTGGAGGAGGAGAGATCATCTGGAGGACCAAGCTGCAAGCTGGCCAAAGAAA 840
K L L E D R I A E F T T N L T E E E E E K S K S L A K L K N K H E A M I T D L E E E
GAACTGCTGGAAGCAGAAATAGCTGAGTCAACCACCTCACAGAAGGAGGAGAAATCAAGAGCTCGCCAACTCAAGAACAAGCATGAGCAGTACTGACTGCTGGAA 960
R L R R E E K Q R Q E L E K T T R R K L E G D S T D L S D Q I A E L Q I A E L Q I A E L
GGCCCTCCGAGGAGAGCAAGCAGCGAGAGGAGTGGAGAAGCCCGCGGAACTGGAGGAGACTCCACAGACTCAGCGACAGATCGCCGAGCTCCAGGCCAGATCGCGGAGCT 1080
K H Q L A L A K K E E E L Q A L A R V E E A A Q K N H A L K K I R E L E S Q I S
CAAGATCGAGCTGGCCAAAGAGGAGGAGCTCCAGCCGCTGGCCAGAGTGGAAAGAGAACTGGCCCAAGAACATGGCCCTCAAGAAAGATCCGGGAGCTGGAAATTCAGATCTC 1200
E L Q E D L E S E R A S R N K A E K Q K R D L G E E L E A L K T E L E D T L D S
TGAATCCAGGAAGACTGGAGTCTGAGCTGCTTCAGGAATAAAGCTGAGAAGCAGAAACCGGACCTTGGGGAAGAGCTAGAGGCGCTGAAAACAGAGTGGAGGACAGCTGGATT 1320
T A A Q Q E L R S K R E Q E V N I L K K T L E E E A K T H E A Q I Q E M R Q K H
CAGAGTCCGACAGCAGAGCTCAGTCAAACTGAGCAGGAGTGAACATCTGAAGAAAGCTGGAGGAGGAGCCAGAGCCACAGCGCCAGATCCAGGAGTGAAGGAGGAGCA 1440
S Q A V E E L A E Q L E Q T K R V K A N L E K A K Q T L E N E R G E L A M E V K
CTCAGAGCCGGTGGAGGAGCTGGCCGAGCAGTGGAGCAGAGAGCGGTTGAAAGCAAQCTGAGAAGGCAAGCAGACTCTGGAGAAGCAGCGGGGGAGCTGGCCAAAGCGTGA 1560
V L L Q G R D S E H K R K K V E A Q L Q E L Q V K F N E G E R V R T E L A D K
GGTCTGCTGAGGGCCGAAAGGACTCGGAGCACAAGCGCAAGAAAGTGGAGGCGAGCTGCAAGGACTGAGCTCAAGTCAACGAGGAGAGCGGGTCCGACAGAGCTGGCCGCA 1680
V T K L Q V L D N V T G L S Q S D S K S K L T K D F S A L E S Q L Q D T A
GGTCACCAAGCTGCAAGTGGAGTGGACAACGTGACCGGCTTTCAGCCAGTCCGACAGCAAGTCCAGCAAGCTCACCAGGACTTCTCCGGCTGGAGTCCAGCTGCAGGACTCA 1800
E L L Q E N E A R Q K L S L S T K L Q V E D E K N S F R E Q L E E E E E A K H
GAGCTGCTGAGGAGAGAACCGCAGAAAGCTGAGCTGAGCACAAGCTCAAGCAGGTTGAGAGCAGAGAAATTCCTTCCGGGAGCAGCTGGAGGAGGAGGAGGAGCCAAAG 1920
N L E K Q I A T L L H A Q V A D M K K M E D S V G C L E T A E E V K R K L Q L K D
CAACTCCAGGAAGACTCGCCACCTTCAATGCCAGTGAAGGCGCAGTGAAGAAAGAGTGGAGGAGCAGTGTGGGGTCCGCTGGAAACTGCTGAGGAGTGAAGGAGGAGCTCCAGGAGGA 2040
L E G L S Q R H E E K V A A Y D K L E K T K T R L Q Q E L D D L L V D L D H Q R
CTGGAGCCGCTGAGCCAGCGCAGGAGGAGTGGCCGCTTACGACAAGCTGGAGAAGCAGCAAGCAGCCGCTCGCAGGAGCTGGACGACTGCTGGTGGACTGGACCCAGCCG 2160
Q S A C N L E K K Q K K F D Q L L A E E K T I S A R Y A E E R D R A E A E R E
CCAGAGCGGTGCAACTGGAGAAGAAGCAGAAAGTTGACCAAGCTTCTGGCCAGGAGAAAGACTCTGCAAGATGAGCAGGAGCGCCAGCGGGCTGAGCGAGGCGCCGAG 2280
K E T A L S L A L E A M E Q K A E L E R L N K Q F R T E M E D L M S S K
GAAGGACCAAGGCTGTCTGCTGCGCCGGCGCTGGAGAGGACTGGAGCAGAAAGCGGAGCTGGAGCGGCTCAACAGCAGTCCGACGAGATGGAGGACTTATGAGCTCAA 2400
D D V G K S V R L E K S K R A L E Q Q V E E M K T Q L E E L E D E L Q A T E A D
GGATGATGGCCAGAGTGTCCAGGCTGGAGAAGTCAAGCGCGCCCTAGACAGCAGGTTGGAGGAGATGAGAGCCAGCTGGAAGAGCTGGAGGAGGAGCTGCAGCCACCGAAGA 2520
A K L R L E V L N L Q A M K A Q F E R D L Q G R D E Q S E E K K K Q Q L V R Q V R E
TGCCAAAGCTCCGGTTGAGGTCNACTGCAAGCTGCAAGGCGCAAGTTCAGCGCGGAGCTCGAGGCGGAGCAGGAGGAGAGAAAGAGAGCAGCTGGTCCAGCAGGTCCGCG 2640
M E A E L E D E R K Q R S M A V A A R K K L E M D L K D L E A H I D S A A N K N R
GATGGAGGACTGAGGAGCAGAGGAGAGCAGCGCTGATGGCAGTGGCGCGGAGGAGGAGCTGGAGTGGACTGGAAGCCTGGAGGCGCAGCTGAGCTGAGGAGGAGGAGCG 2760
D E A I K Q L R K L Q A Q M K D C M R E L D D T R A S R E E I L A Q A R E N E K
GGCAGAGCCTCAAACTGCGGAGGCTCGAGCGCCAGATGAAGGACTGCATCGCCGAGTGGATGACACCCCGCCCTCTCGTGGAGGAGTCTGGCCAGGCGCAAGAGAGAG 2880
K L K S H E A E M I Q L Q E E L A A E R A K R Q A Q Q E R D E L A E I A N S
GAAGCTGAGAGCATTGAGGCGGAGATGATTCAGTTCAGGAGGAACTGGCAGCGGAGCGTGGCAAGCGCCAGCGCCAGGAGCGGAGGAGGAGGAGTGGTGGTGGTGGT 3000
S R K G A L A L E A R K R R L E A R I A Q L E E E L E E Q G N T R L E I N D R L K
CAGCGGCAAGGAGCCCTGGGTTAGAGGAGAGCGGCTGAGAGGCGCCGCTCCGCAAGCTGGAGGAGGAGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3120
K A N L Q I D I N A D L N R G H A Q K N E N A R Q K N E L R Q K L E R K V K
GAAGCGCACTGCAAGTGCAGATCAAGCGCCAGCTGAACTGGAGCGCGGAGCCAGAGAAACGAGAATGCTCGCAGGAGCTGGAAGCCAGAACAGGAGCTTAAAGTCAA 3240
L Q E H E G T V K S K Y K A S I T A L E A K I A Q L E E Q L D N E T K E R Q A A
GCTGAGGAGTGGAGGCGACTGCAAGTCAAGTCAAGGCTCCATCCAGCCCTCGAGGCGCAAGATGACACAGCTGGAGGAGCAGCTGGACAAAGCAGGAGGAGGAGCGCAG 3360
C K Q V R R T E K K L K D V L Q V D L Q D E R R N A E Q Y K D Q A D K A S T R L K
CTGCAAAAGGTTGCTGGAGCAGAAAGCTGAAAGATGTGCTGCTGCAAGTGGATGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3480
Q L K R Q L E E A E E A Q R A N A S S R R K L Q R E L E D A T E T A D A M N R E
GCAGCTCAAGCGGCGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3600
V S S L K N K L R R G D L P F V V P R R M A R K G A G D G S D E V E D G K A D G
AGTCAGCTCCCTAAAGAACAACTCAGGCGCGGAGCTGCCGTTTCGTCGCCCGCGAATGCCCGGAAAGGCGCGGGGATGGCTCCGACGAGAGGTAGATGGCAAAGCGGATGG 3720
A E A K P A E *
GGCTGAGGCGCAAAGCTCGCGAAATAGGCTCTTCTCCTGAGCCTGAGATGGATGGACAGACAGACACAGCTCCCTTCCAGACCCCGAGCAGCGCTCTCCCCACTCTTGGGA 3840
CTGCTGTAACATGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3960
AGATCTGATACCAAAGACAGAGCGCCGCGGCGAGGAGGAGCAGCAGCAGCTCTCACGCTCTCTTCCAAAAGACAAAGATGTTGAGGCGAGCAGGCGAGCGCCCGGGGAGG 4080
CCAGAGCTTTCTATGATATTTTCTTCCAGACTGAGGCTTTTGGTGTGGAGCTCCCGCAGTGTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4200
GCTGTTGCAATCACAGCTGGTACCTCACAGACTGCGCCCTTGGGCTTCCACTGCTTGGCGGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4320
TGGGGTGGGGGAGGAGTGGATTCCTTCACTGCTCCAGCAGCGCCACTGCTGCTTCTTCTGATTTAAATGCTCAAGTGAATGCCCCCTTCCCTTCTTACCGAGGAGCAG 4440
CTGCTTCCACAGCAAGCTGTGCGGGTCAAGCTGGAAAGGCGAGCAGCTCCAGTGGCTTCCGGAAGCTTGGGGACCAAATATACTTAATGGTTAAGGAGCTTGTCCCAAG 4560
TCTGACAGGAGGCTTAGAGGGCGAGCGCTCCGAGCGGATTTGCTGCTACTTAAAGCTGCGGCGGAGGCTGTTTACCTGCACCTTGACTAGATATTTAAATCTGCC 4680
ACCTGCAGGATTTTTGAAAGCAAAATAGGTTTTTCTTCTTCTGTAATGATGATAAATGATGATGATAAATGATGATAAATGATGATAAATGATGATAAATGATGATAA 4800
TGGTACACTGGTGGCGAATTACTTGTATTTCTTAAGTGTGTATATGCTGCATGGAGACTACGGGCAAGAGGGGCAATTTTTTTTTTAAAGGAAACAACTCAAAATCATGA 4920
AGTGATATAAAGCTGCATATGCTTCAAAAGCTGTAATTCAGTCCAGTGTGCTGCTACCAAGGAGTGAGTGAACACCCACCTACCCCTTTTTTATATAAATAAAGTGCCTGAG 5040
ATGTTGTCAGCTGACCACTACAGTGTGTTACAGATGTTTTACACTGAGCACAATAAAGAGAACATGTGCTCAAAAAAAA 5131

```

FIG. 2. DNA and predicted amino acid sequence of human nonmuscle MHC. DNA sequence is presented in a 5' to 3' orientation. Amino acids use the single-letter designation. The proline marking the S1/S2 junction is indicated by a box. The termination codon is marked by an \*. The polyadenylation signals used by both transcripts are underlined.

untranslated region of the longer cDNA, hybridizes only to the higher molecular weight (7.3 kb) transcript (lane 5). Even on a longer exposure of this lane, the 7.1-kb transcript was not apparent (data not shown).

S1 nuclease analysis of RNA was carried out to examine the sequence diversity of MHC transcripts. A 3'-end-labeled probe originating at the *Nco* I site at nucleotide 4258 in the sequence and terminating at a *Pvu* II site in the vector, totaling 752 nucleotides, was used in reactions with RNA from a human macrophage cell line, U937 (14), and from HeLa cells. It is predicted that if the two transcripts seen on the Northern blots correspond to the cDNAs isolated, two fragments of 706 and 506 base pairs (bp) should be protected.

The results are shown in Fig. 5. The two expected protected fragments are found (lanes 2 and 3) and represent the major species seen. In addition, a third species of 606 bp is seen whose identity is unknown but could represent either incomplete nuclease digestion or a third transcript.

**A Human Nonmuscle MHC Gene Is on Chromosome 22.** To investigate the number of human nonmuscle MHC genes, human genomic DNA was hybridized to four probes, which are indicated by horizontal bars in Fig. 1. Probes 1 and 2 derive from the coding region of the MHC molecule and recognize two or three bands in each of three restriction digests (Fig. 6). The more weakly hybridizing bands are likely to be smooth muscle MHC genes. Probes 3 and 4 derive

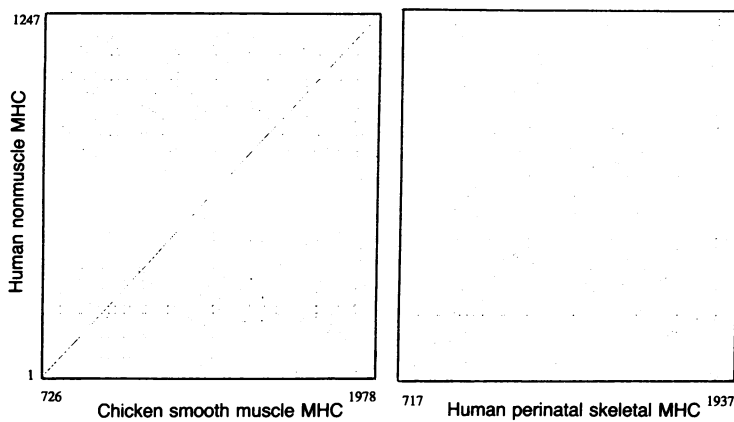


FIG. 3. Graphic matrix comparative analysis of human nonmuscle with smooth muscle and skeletal muscle MHC protein sequences. Comparisons were done using the STADEN DIAGON program. Each dot signifies three identical amino acids in a row. The chicken smooth muscle sequence is presented by Yanagisawa *et al.* (10) and the human perinatal sequence is presented by Feghali and Leinwand (22). Numbering of muscle MHC sequences is from the first methionine.

entirely from the 3' untranslated regions of pNMHCM6 and pNMHCM2, respectively. Probe 4 contains sequences that are found exclusively in pNMHCM2. Probes 3 and 4 recognize the same single *Hind*III and *Xba* I fragments, while probe 4 recognizes one additional *Eco*RI fragment. The most likely explanation of these results is that a single MHC gene encodes both mRNAs and that the longer portion of pNMHCM2 resides partially on a separate exon. The resolution of the absolute gene number awaits further study, but it is clear from this analysis that there are fewer nonmuscle MHC genes than there are sarcomeric MHC genes (24, 25), or that the members of the nonmuscle gene family are highly divergent and do not cross-hybridize under our conditions.

To determine whether this nonmuscle MHC gene is linked to either of the two sarcomeric MHC gene clusters on chromosomes 14 and 17 (26, 27), and, if not, to determine its chromosomal location, genomic DNA from a panel of human-mouse somatic cell hybrids was hybridized to probe 3. Results from these hybridizations are presented in Table 1. By correlating the hybridization pattern of the human MHC gene with the karyotype and isozyme composition of the cells in the panel of hybrids, it is possible to conclude that the nonmuscle MHC gene is located on human chromosome 22 and is therefore unlinked to the two sarcomeric MHC clusters. A cell line containing a translocation involving chromosome 22 (REX-26) allows a regional assignment to be made to 22pter → 22q13.

## DISCUSSION

Vertebrate muscle-specific contractile protein gene expression has been characterized by an amazing diversity of gene products. In the case of myosin light chain, tropomyosin, and the troponins, this diversity is the result of complex patterns of alternative splicing (for a review, see ref. 28). In the case of MHC and actin, the diversity results from expression of members of multigene families (as examples, see refs. 29 and

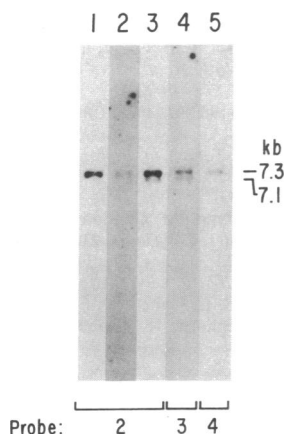


FIG. 4. Analysis of cellular RNA for expression of nonmuscle MHC mRNA. Radiolabeled DNA probes 2-4 (as indicated in Fig. 1) were hybridized to 0.5  $\mu$ g of poly(A)<sup>+</sup> RNA from mouse embryo fibroblasts (lane 1), bovine smooth muscle (lane 2), human foreskin fibroblasts (lane 3), and human umbilical vein endothelial cells (lanes 4 and 5).

30). Prior to this report, the only example of multiple transcripts resulting from a single MHC gene has been seen in the *Drosophila* indirect flight muscle where developmentally regulated MHCs are expressed from a single gene (31). In the case of the *Drosophila* MHC gene, the alternatively spliced transcripts encode different proteins. In contrast, the human nonmuscle MHC transcripts encode the same protein. Their biological significance is unclear and requires further investigation but may be linked to regulation. In the four human cell types examined, there was no apparent variation in the relative levels of each of the two mRNA species. This is also true in the variable 3' ends of the murine dihydrofolate reductase gene (32). The presence of a third protected fragment in S1 nuclease protection experiments may represent an additional transcript from the same gene. An interesting finding was that neither the bovine nor mouse cells examined expressed multiple transcripts that could be resolved by Northern blot analysis.

Genomic Southern analysis indicates that there is a small number of human nonmuscle MHC genes. This is in contrast to the observed 10-15 sarcomeric MHC genes believed to encode distinct proteins (24, 25). It is presumed that each of the sarcomeric MHC gene products provides some distinct function to the muscle in which it is expressed. Given what appears to be one or a small number of nonmuscle MHC genes, it will be interesting to determine whether different contractile processes such as cytokinesis vs. cell locomotion require distinct highly divergent MHC proteins. The analysis presented here demonstrates that the same nonmuscle MHC gene is expressed in fibroblasts, endothelial cells, and mac-

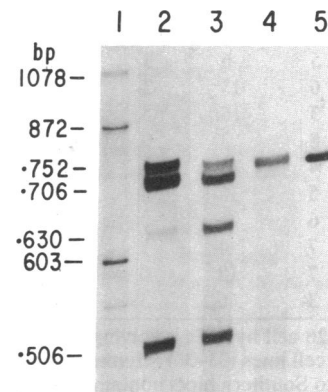


FIG. 5. S1 nuclease analysis of human RNA for nonmuscle MHC expression. An end-labeled probe (S1N denoted in Fig. 1) was hybridized to 30  $\mu$ g of RNA from human macrophages (lane 2), HeLa cells (lane 3), yeast (lane 4), and no RNA (lane 5) and subjected to S1 nuclease treatment prior to electrophoresis on a 5% polyacrylamide gel. Lane 1 corresponds to molecular weight markers. Probe and protected fragments are denoted by .

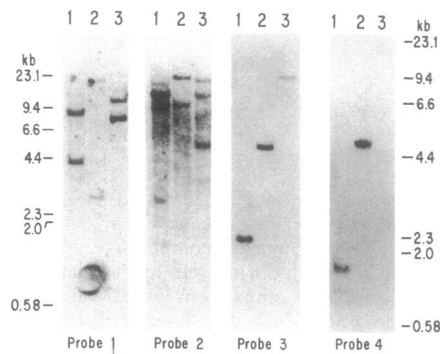


Fig. 6. Genomic DNA hybridizations to human nonmuscle MHC DNA probes. Lanes 1–3, *EcoRI*, *HindIII*, and *Xba I* digests, respectively, of human genomic DNA. Each panel was hybridized to one of four radioactively labeled DNA fragments indicated in Fig. 1. Markers are a  $\lambda$ *HindIII* digest. The filter hybridized to probe 4 represents a different gel run under similar conditions.

rophages, suggesting that nonmuscle contractile processes require fewer myosin proteins than those of striated muscle.

DNA sequence analysis and chromosome mapping suggest that the evolutionary relationship between sarcomeric and nonmuscle MHC genes is quite distant. Sarcomeric MHCs are all quite homologous to each other, ranging from 75% up

Table 1. Segregation of a DNA probe for human nonmuscle MHC with human chromosomes in *EcoRI*-digested human–mouse hybrid cell DNA

Chromosome	Hybrid DNA				% discordancy
	Concordant hybrids, no.		Discordant hybrids, no.		
	+/+	-/-	+/-	-/+	
1	4	16	3	1	17
2	5	11	3	7	38
3	4	8	4	7	48
4	4	9	4	9	50
5	4	9	4	9	50
6	3	15	5	3	31
7	5	9	2	9	44
8	5	7	3	11	54
9	4	16	4	1	20
10	7	5	1	12	52
11	6	12	2	6	31
12	5	6	3	12	58
13	4	12	4	6	38
14	5	6	3	12	58
15	6	13	2	5	27
16	3	16	5	2	27
17	8	3	0	14	56
18	4	10	4	8	46
19	5	16	3	2	19
20	6	6	2	12	54
21	7	4	1	14	58
22	7	18	0	0	0
X	3	6	3	10	59

Compiled from 26 cell hybrids involving 10 unrelated human cell lines and 4 mouse cell lines (33–36). Human nonmuscle MHC probe 3 was hybridized to Southern blots containing *EcoRI*-digested DNA from the human–mouse hybrids. Concordant hybrids have either retained or lost the human MHC bands, but not a specific chromosome, or the reverse. Percent discordancy indicates the degree of discordant segregation for a marker and a chromosome. A 0% discordancy is the basis for chromosome assignment.

to 94%, even across diverse species (see ref. 22). As shown in Fig. 3, nonmuscle MHC does not have any significant sequence homology to sarcomeric MHC. This suggests that the sarcomeric MHC gene family evolved after the divergence of striated muscle MHC and nonmuscle MHC genes. In addition, the gene encoding nonmuscle MHC is chromosomally unlinked to the two sarcomeric MHC gene clusters, which also points to an evolutionary distance between these gene families. Future studies should make it possible to define those regions of the MHC molecule that are functionally significant and to supply information about the sequence requirements for nonmuscle contractile functions.

We would like to thank R. Feghali for assistance with RNA and genomic DNA analysis. L.A.L. is an American Heart Association Established Investigator. This research is supported in part by National Institutes of Health Grants GM29090 to L.A.L. and GM20553 and HL41882 to J.C.M.

- Mabuchi, I. & Okuno, M. (1977) *J. Cell Biol.* **74**, 251–261.
- Kiehart, P. P., Mabuchi, I. & Inoue, S. (1982) *J. Cell Biol.* **94**, 165–178.
- DeLozanne, A. & Spudich, J. A. (1987) *Science* **236**, 1086–1091.
- Knecht, D. & Loomis, W. (1987) *Science* **236**, 1081–1086.
- Emerson, C. P. & Bernstein, S. I. (1987) *Annu. Rev. Biochem.* **56**, 695–726.
- Warrick, H. M. & Spudich, J. A. (1987) *Annu. Rev. Cell Biol.* **3**, 379–421.
- Burridge, K. & Bray, D. (1975) *J. Mol. Biol.* **99**, 1–14.
- Citi, S. & Kendrick-Jones, J. (1987) *BioEssays* **7**, 155–159.
- Medford, R., Wydro, R., Nguyen, H. & Nadal-Ginard, B. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5749–5753.
- Yanagisawa, M., Hamada, Y., Katsuragawa, Y., Imamura, M., Mikawa, T. & Masaki, T. (1987) *J. Mol. Biol.* **198**, 142–157.
- Nagai, R., Larson, D. M. & Periasamy, M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 1047–1051.
- Henikoff, S. (1984) *Gene* **28**, 351–359.
- Kraft, R., Tardiff, J., Krauter, K. & Leinwand, L. (1988) *Biotechniques* **6**, 544–547.
- Fan, X., Stark, G. & Bloom, B. (1989) *Mol. Cell. Biol.* **9**, 1922–1928.
- Meyers, J. C., Chu, M.-L., Faro, S., Clark, W., Prockop, D. J. & Ramirez, F. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 3516–3520.
- Chirgwin, J. M., Przybyla, A., MacDonald, R. & Rutter, W. (1979) *Biochemistry* **18**, 5294–5299.
- Pihlajaniemi, T. & Myers, J. C. (1987) *Methods Enzymol.* **145**, 213–222.
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Cold Spring Harbor, NY).
- Berk, A. & Sharp, P. (1977) *Cell* **12**, 721–732.
- Kavinsky, C. J., Umeda, P. K., Sinha, A., Elzinga, M., Tong, S., Zak, R., Jakovic, S. & Rabinowitz, M. (1983) *J. Biol. Chem.* **258**, 5196–5205.
- Saez, L. & Leinwand, L. A. (1986) *Nucleic Acids Res.* **14**, 2951–2969.
- Feghali, R. & Leinwand, L. (1989) *J. Cell Biol.* **108**, 1791–1797.
- Shohet, R. V., Conti, M. A., Kawamoto, S., Preston, Y., Brill, D. & Adelstein, R. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7726–7730.
- Wydro, R. A., Nguyen, H., Gubits, R. & Nadal-Ginard, B. (1983) *J. Biol. Chem.* **258**, 670–678.
- Leinwand, L. A., Saez, L., McNally, E. & Nadal-Ginard, B. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3716–3720.
- Leinwand, L. A., Fournier, R. E. K., Nadal-Ginard, B. & Shows, T. B. (1983) *Science* **221**, 766–769.
- Saez, L., Gianola, K., McNally, E., Feghali, R., Eddy, R., Shows, T. & Leinwand, L. (1987) *Nucleic Acids Res.* **15**, 5443–5459.
- Breitbart, R. E., Andreadis, A. & Nadal-Ginard, B. (1987) *Annu. Rev. Biochem.* **56**, 467–495.
- Izumo, S., Nadal-Ginard, B. & Mahdavi, V. (1986) *Science* **231**, 597–600.
- Engel, J., Gunning, P. & Kedes, L. (1982) *Mol. Cell Biol.* **2**, 674–684.
- Rozek, C. & Davidson, N. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2128–2132.
- Setzer, D., McGrogan, M. & Schimke, R. (1982) *J. Biol. Chem.* **257**, 5143–5147.
- Shows, T. B. (1983) in *Isozymes: Current Topics in Biological and Medical Research*, eds. Rattazzi, M. C., Scandalios, M. G. & Whitt, G. S. (Liss, New York), Vol. 10, pp. 323–339.
- Shows, T. B., Brown, J. A., Haley, L. L., Byers, M. G., Eddy, R. L., Cooper, E. S. & Goggin, A. P. (1978) *Cytogenet. Cell Genet.* **21**, 99–104.
- Shows, T. B., Eddy, R., Haley, L., Byers, M., Henry, M., Fujita, T., Matsui, H. & Taniguchi, T. (1984) *Somat. Cell Mol. Genet.* **10**, 315–318.
- Shows, T. B., Sakaguchi, A. Y. & Naylor, S. L. (1982) in *Advances in Human Genetics*, eds. Harris, H. & Hirschhorn, K. (Plenum, New York), Vol. 12, pp. 341–452.