Volume 14 Number 7 1986 Nucleic Acids Research

Characterization of diverse forms of myosin heavy chain expressed in adult human skeletal muscle

Lino Saez and Leslie A.Leinwand

Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, NY 10461, USA

Received 15 January 1986; Accepted 4 March 1986

ABSTRACT

In an attempt to define myosin heavy chain (MHC) gene organization and expression in adult human skeletal muscle, we have isolated and characterized genomic sequences corresponding to different human sarcomeric MHC genes (1). In this report, we present the complete DNA sequence of two different adult human skeletal muscle MHC cDNA clones, one of which encodes the entire light meromyosin (LMM) segment of MHC and represents the longest described MHC cDNA sequence. Additionally, both clones provide new sequence data from a 228 amino acid segment of the MHC tail for which no protein or DNA sequence has been previously available. One clone encodes a "fast" form of skeletal muscle MHC while the other clone most closely resembles a MHC form described in rat cardiac ventricles. We show that the 3' untranslated region of skeletal MHC cDNAs are homologous from widely separated species as are cardiac MHC cDNAs. However, there is no homology between the 3' untranslated region of cardiac and skeletal muscle MHCs. Isotype-specific preservation of MHC 3' untranslated sequences during evolution suggests a functional role for these regions.

INTRODUCTION

Myosin is the major protein of the contractile apparatus. It is ubiquitous in eukaryotes where it converts chemical energy into mechanical force through the hydrolysis of ATP. Within the cell, myosin is organized as a pair of heavy chains of about 200,000 daltons each and two pairs of light chains (16-20,000 daltons each). The carboxy terminal region of the heavy chain forms an α helical tail which is responsible for the assembly of myosin into filaments (2). The globular amino terminal portion which is referred to as the head, binds the 2 pairs of light chains, contains ATPase activity and the actin-binding site (3). Electrophoretic and immunological studies have established the existence of multiple forms of myosin within different muscle types, nonmuscle tissues and developmental stages (4-6). Further lines of evidence suggest that the wide diversity of myosin heavy chain (MHC) proteins has functional significance. First, ATPase activity correlates with the contractile properties of individual muscles (7). Second, a recently

© IRL Press Limited, Oxford, England.

developed in vitro motility assay has made it possible to functionally distinguish different myosins by their velocity on actin cables (8). Thus, it is of interest to determine the genetic basis of MHC variation and to understand the regulation of the MHC genes. It is clear that complete determinations of the amino acid sequences of the various MHC forms will further the understanding of structure-function relationships in motility. It is also important to define the extent of diversity of MHC expression in different muscle types in order to correlate the program of myosin expression with the physiological properties of the muscles.

Molecular clones corresponding to some of these myosin forms have been obtained. The entire nematode unc-54 MHC gene has been sequenced and the protein encoded by it has been deduced from this sequence (9). MHC cDNAs from rat and rabbit cardiac ventricles, rat fetal, perinatal, and adult skeletal muscles and chicken embryonic skeletal muscle have also been reported (10-15). Sequence comparisons among the clones show them to be quite homologous, but each represents a distinct MHC form. None of these cloned vertebrate sequences which range in size from 0.6kb to 2.2kb represent full-length MHC cDNAs. In vertebrates, it has been possible to deduce a partial protein sequence for the 200 kilodalton (kd) MHC from cloned DNAs and to compare these sequences to partial protein sequence data from rabbit skeletal muscle (12-15). Alignment of all vertebrate MHC sequences obtained from DNA and protein sequences shows that about 30% of the molecule has not been sequenced. This includes 228 amino acids in the LMM of MHC as well as of about 400 amino acids in the subfragment-1 (see Figure 4). We present below sequence data from two different MHCs that provide the missing sequence information from the LMM region.

Vertebrate sarcomeric MHCs are encoded by highly conserved multigene families, consisting of at least 10 members (1,16,17). In mouse and human, skeletal MHC genes are localized to a single chromosome (18,19). The rat α and β ventricular cardiac MHC genes exist in a tandem arrangement in the genome, separated by less than 5 kb (20). Recent studies have shown that a mouse cardiac MHC gene and the human α and β cardiac MHC genes are unlinked to skeletal MHC genes (19; in preparation). One exception to this multigene organization is seen in Drosophila where it has been shown that a single gene encodes 3 MHC mRNAs (21,22).

We have concentrated our interest on the DNA sequence organization and tissue-specific expression in the human myosin heavy chain multigene family. We have isolated a number of genomic clones corresponding to different human

skeletal MHCs (1) and have shown that they map to a single chromosomal location (18). In this study we present complete DNA sequence analysis of two adult human skeletal muscle cDNA clones which encode the entire light meromyosin (LMM) region of two different MHCs. The longer of the two sequences also encodes part of the subfragment-2 region of MHC. The cDNA clones represent the products of two genes and encode MHCs that are quite divergent by comparison with each other. The skeletal muscle from which the cDNA library was constructed is composed of both fast and slow fibers, which are characterized by high and low ATPase activities, respectively. One of the cloned sequences studied here corresponds to a fast fiber form of MHC and the other is most homologous to a form that was originally described in rat cardiac ventricles (12). We find strong sequence conservation between these clones and among the 3' untranslated regions of tissue-specific forms of MHC from a wide variety of species.

MATERIALS AND METHODS

Isolation of Human MHC cDNA Clones

An adult human skeletal muscle cDNA library constructed in the Okayama-Berg vector system was a generous gift of Dr. L. Kedes, Stanford University Medical Center (23). The library was screened as described by Grunstein and Hogness (24). Recombinant DNA clones containing MHC sequences were identified using a 2.0 kilobase (kb) fragment of a human MHC genomic clone designated p10-3 previously described in Leinwand *et al*. (1). Cloned DNA probes were radiolabelled by nick translation (25) to specific activities of $>10^8$ cpm/µg.

DNA Analysis

DNA was cleaved with different restriction enzymes obtained from New England BioLabs under the conditions suggested by Maniatis (26). For sequencing purposes M13 subclones were generated where indicated in Figure 2.

RNA Isolation and Blot Hybridization

Total cellular RNA was isolated from frozen human muscle biopsies using the quanidinium-isothiocyanate procedure (27). Biopsy material was kindly provided by Dr. Paul Fisher, Columbia University, College of Physicians and Surgeons. 10 µg of RNA was size-fractionated on 1% agarose, 3% formaldehyde gels in 100 mM MOPS pH 7.4, 10 mM EDTA and transferred to nitrocellulose according to Derman *et al* (28). Hybridizations were carried out for 18 hours in 5X SSC, 1X Denhardt's, 10 mM $NaPO_4$ and 50 µg/ml denatured salmon sperm DNA at 65°C. Filters were washed in 2X SSC, 0.2% SDS at 65°C. Filters were exposed to X-ray films at -70°C with Dupont Cronex intensifying screens.

DNA Sequencing

DNA sequencing was carried out by the dideoxynucleotide method of Sanger and Coulson (29) or by the base-specific chemical method described by Maxam and Gilbert (30). For the latter method, in some cases, 5' protruding ends were created in blunt-end restriction fragments by treatment with Exo III (31). Fragments with 5' protruding ends were dephosphorylated using bacterial alkaline phosphatase (Worthington) and were subsequently 5' end-labeled with $\gamma^{32}P$ ATP (Amersham Corp.) using T_4 polynucleotide kinase (PL Biochemicals). The labelled fragments resulting from chemical modification and cleavage were fractionated on 8% and 12.5% (0.4 x 400 mm) or 6% (0.4 x 800 mm) polyacrylamide gels and were subsequently dried and autoradiographed at room temperature for 16 hours.

Computer Analysis of DNA and Protein Sequences

Graphic matrix analysis was carried out with the MBSP dot matrix program written at Albert Einstein College of Medicine. Per cent homologies were calculated according to an alignment program called NUCALN described by Wilbur and Lippman (32).

RESULTS

Isolation of human MHC cDNA clones

We were interested in obtaining sequence information for human MHC coding regions and also in determining the extent of MHC diversity in skeletal muscle. In order to obtain such information, 2000 clones of an adult human skeletal muscle cDNA library were screened with a DNA fragment from a previously described human MHC genomic clone (1). Forty positive clones were obtained, indicating that MHC cDNA clones that cross-hybridize with this COOH-proximal coding fragment are present at a frequency of 2% in this cDNA library. Given the large number of positive clones, it is important to distinguish those clones corresponding to distinct MHC forms from those representing overlapping or duplicate clones. Restriction endonuclease analysis with PstI + PvuII and partial DNA sequence analysis indicated that there are at least three and possibly five different MHC clones in this population whose insert sizes range from 0.6 to 2.7kb (data not shown). PstI was chosen because its recognition sequence (CTGCAG) encodes the dipeptide leucine-glutamine, a frequently occurring sequence in the rod portion of MHC (9, 15, 36). The clones all have multiple PstI sites, consistent with their identity as carboxy terminal portions of MHC (data not shown).

Two clones were chosen for more detailed analysis because of their large

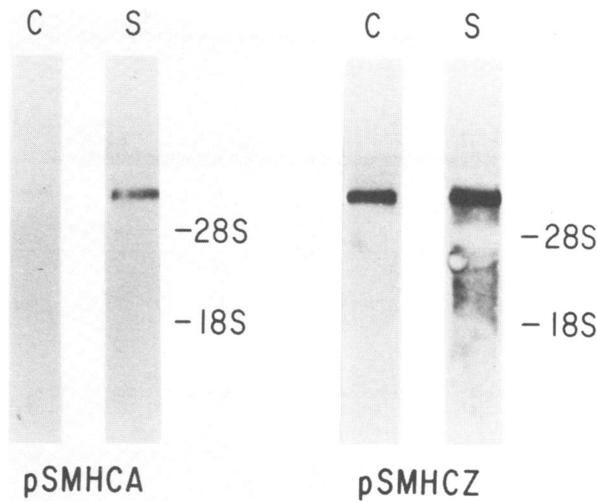
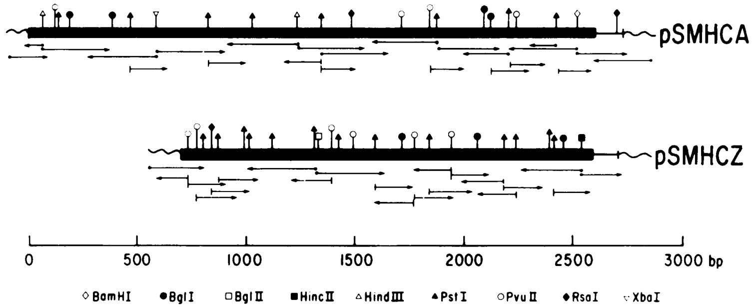


Figure 1 Hybridization of human MHC cDNA clones to cardiac and skeletal muscle RNAs.
DNAs from clones pSMHCA and pSMHCZ were radioactively labelled and hybridized to RNA from adult cardiac (C) and adult skeletal muscle (S). These RNA samples were fractionated on denaturing formaldehyde gels and transferred to nitrocellulose. 18S and 28S refer to migration of ribosomal RNA species.

insert sizes and because their restriction maps indicate they are likely to be the products of two genes. Inserts from these two clones (named pSMHCA and pSMHCZ) are 2.7kb and 2.0kb respectively, corresponding to less than half the length of total MHC mRNA. DNA from both clones hybridizes to a 7kb species of RNA from skeletal muscle (Figure 1). This size is the expected size of MHC mRNA. DNA from pSMHCA hybridizes nearly exclusively to skeletal muscle RNA and not to cardiac muscle RNA while DNA from pSMHCZ hybridizes equally well to RNA from both muscle sources. The differential hybridization has been quantitated by densitometric scans. The signal of pSMHCA hybridized to skeletal RNA is 31.7 times that of cardiac RNA. The signal of pSMHCZ hybridized to skeletal RNA is 1.32 times that of cardiac RNA. This implies that the same MHC gene may be expressed in more than one tissue. Similar results have been obtained in rat with expression of a β MHC in both skeletal and cardiac muscle (34). Differential hybridization of these human cDNAs to RNA from two types of muscle also suggests, but does not prove, that the clones encode different MHCs. Definitive proof requires sequence data which are presented below. In addition, sequence data allow identification of the MHC forms encoded by the clones by comparison with MHCs from other organisms.



HUMAN SKELETAL pSMHCZ

```
                                                   30                                      60                                      90
CTG CTG GGT GTA GGT GAG CTT GCC TCG GGT CAG CTG GGG GAC CTC AAG AGG CAG CTG GAG GAG GAG GTT AAG GCG AAG AAC GCC CTG GCC
 L   L   G   V   G   E   L   A   S   G   Q   L   G   D   L   K   R   Q   L   E   E   E   V   K   A   K   N   A   L   A

                                                  120                                     150                                     180
CAC GCA CTG CAG TCG GCC CGG CAT GAC TGC GAC CTG CTG CGG GAG CAG TAC GAG GAG GAG ACG GAG GCC AAG GCG AGG CTG CAG GGC GAC
 H   A   L   Q   S   A   R   H   D   C   D   L   L   R   E   Q   Y   E   E   E   T   E   A   K   A   R   L   Q   G   D

                                                  210                                     240                                     270
CTT TCC AAG GCC AAC TCG GAG GTG GCC CAG TGG AGG ACC AAG TAT GAG ACG GAC GCC ATT CAG CGG ACT GAG GAG CTC GAG GAG GCC AAG
 L   S   K   A   N   S   E   V   A   Q   W   R   T   K   Y   E   T   D   A   I   Q   R   T   E   E   L   E   E   A   K

                                                  300                                     330                                     360
AAG AAG CTG GCC CAG CGC CTG CAG GAA GCT GAG GAG GCC GTG GAG GCT GTT AAT GCC AAG TGC TCC TCG CTG GAG AAG ACC AAG CAC CGG
 K   K   L   A   Q   R   L   Q   E   A   E   E   A   V   E   A   V   N   A   K   C   S   S   L   E   K   T   K   H   R

                                                  390                                     420                                     450
TTA CAG AAT GAG ATC GAG GAC TTG ATG GTG GAC GTA GAG CGC TCC AAT GCT GCT CTG CAG GCC CTA GAC AAG AAG CAG AGG AAC TTC GAC
 L   Q   N   E   I   E   D   L   M   V   D   V   E   R   S   N   A   A   L   Q   A   L   D   K   K   Q   R   N   F   D

                                                  480                                     510                                     540
AAG ATC CTA GCC GAG TGG AAG CAG AAG TAT GAG GAG TCG CAG TCG GAG CTG GAG TCC TCG CAG AAG GAG GCT CGC TCC CTC AGC ACA GAG
 K   I   L   A   E   W   K   Q   K   Y   E   E   S   Q   S   E   L   E   S   S   Q   K   E   A   R   S   L   S   T   E

                                                  570                                     600                                     630
CTC TTC AAA CTC AAG AAC GCG TAT GAG GAG TCC TTA GAA CAT CTG GAG ACC TTC AAG CGG GAG AAC AAA AAC CTG CAG GAG GAG ATC TCC
 L   F   K   L   K   N   A   Y   E   E   S   L   E   H   L   E   T   F   K   R   E   N   K   N   L   Q   E   E   I   S

                                                  660                                     690                                     720
GAC TTG ACT GAG CAG TTG GGT TCC AGC GGA AAG ACT ATC CAT GAG CTG GAG AAG GTC CGA AAG CAG CTG GAG GCC GAG AAG ATG GAG CTG
 D   L   T   E   Q   L   G   S   S   G   K   T   I   H   E   L   E   K   V   R   K   Q   L   E   A   E   K   M   E   L

                                                  750                                     780                                     810
CAG TCA GCC TTG GAG GAG GCC GAG GCC TCC TTG GAG CAC GAG GAG GGC AAG ATC CTC CGG GCC CAG CTG GTG TTC AAC CAG ATC AAG GCA
 Q   S   A   L   E   E   A   E   A   S   L   E   H   E   E   G   K   I   L   R   A   Q   L   V   F   N   Q   I   K   A

                                                  840                                     870                                     900
GAG ATC GAG CGG AAG CGG CAG GAG AAG GAC GAG GAG ATG GAA CAG GCC AAG CGC AAC CAC CTG CGG GTG GAA GAC TCG CTG CAG ACC TCC
 E   I   E   R   K   R   Q   E   K   D   E   E   M   E   Q   A   K   R   N   H   L   R   V   E   D   S   L   Q   T   S

                                                  930                                     960                                     990
TTG GAC GCA GAG ACA CGC AAC CGC AAC GAG GCC TTG AGG GTG AAG AAG AAG ATG GAA GGA GAC CTC AAT GAA ATG GAG ATC CAA CTC TCA
 L   D   A   E   T   R   N   R   N   E   A   L   R   V   K   K   K   M   E   G   D   L   N   E   M   E   I   Q   L   S

                                                 1020                                    1050                                    1080
CAC GCC AAC CGC ATG GCC GCC GAG GCC CAG AAG AAC TTA AGA GCC TCC CAG GAG CTT TTG AAG GAC ACC CAG ATT CAG CTG GAC GAT GCA
 H   A   N   R   M   A   A   E   A   Q   K   N   L   R   A   S   Q   E   L   L   K   D   T   Q   I   Q   L   D   D   A

                                                 1110                                    1140                                    1170
GTC CGT GCC AAC GAC GAC CTG AAG GAG AAC ATC GCC ATC GTG GAG CGG CGC AAC AAC CTG CTG CAG GCT GAG CTG GAG GAG TTG CGC GCC
 V   R   A   N   D   D   L   K   E   N   I   A   I   V   E   R   R   N   N   L   L   Q   A   E   L   E   E   L   R   A

                                                 1200                                    1230                                    1260
GTG GTG GAG CAG ACA GAG CGG TCC CGG AAG CTG GCG GAG CAG GAG CTG ATT GAG ACT AGT GAG CGG GTG CAG CTG CTG CAT TCC CAG AAC
 V   V   E   Q   T   E   R   S   R   K   L   A   E   Q   E   L   I   E   T   S   E   R   V   Q   L   L   H   S   Q   N

                                                 1290                                    1320                                    1350
ACC AGC CTC ATC AAC CAG AAG AAG AAG ATG GAC GCT GAC CTG TCC CAG CTC CAG GGA GAA GTG GAG GAG GCA GTG CAG GAG TGC AGG AAT
 T   S   L   I   N   Q   K   K   K   M   D   A   D   L   S   Q   L   Q   G   E   V   E   E   A   V   Q   E   C   R   N

                                                 1380                                    1410                                    1440
GCT GAG GAG AAG GCC AAG AAG GCC ATC ACG GAT GCC GCA ATG ATG GCA GAG GAG CTG AAG AAG GAG CAG GAC ACC TCA GCC CAC CTG GAG
 A   E   E   K   A   K   K   A   I   T   D   A   A   M   M   A   E   E   L   K   K   E   Q   D   T   S   A   H   L   E
```

```
                                             1470                                          1500                                          1530
CGC ATG AAG AAG AAC ATG GAA CAG ACC ATT AAG GAC CTG CAG CAC CGG CTG GAC GAA GCC GAG CAG ATC GCC CTC AAG GGC GGC AAG AAG
 R   M   K   K   N   M   E   Q   T   I   K   D   L   Q   H   R   L   D   E   A   E   Q   I   A   L   K   G   G   K   K

                                             1560                                          1590                                          1620
CAG CTG CAG AAG CTG GAA GCG CGG GTG CGG GAG CTG GAG AAT GAG CTG GAG GCC GAG CAG AAG CGC AAC GCA GAG TCG GTG AAG GGC ATG
 Q   L   Q   K   L   E   A   R   V   R   E   L   E   N   E   L   E   A   E   Q   K   R   N   A   E   S   V   K   G   M

                                             1650                                          1680                                          1710
AGG AAG AGC GAG CGG CGC ATC AAG GAG CTC ACC TAC CAG ACG GAG GAG GAC AGG AAA AAC CTG CTG CGG CTG CAG GAC CTA GTA GAC AAG
 R   K   S   E   R   R   I   K   E   L   T   Y   Q   T   E   E   D   R   K   N   L   L   R   L   Q   D   L   V   D   K

                                             1740                                          1770                                          1800
CTG CAG CTA AAG GTC AAG GCC TAC AAG CGC CAA GCC GAG GAG GCG GAG GAG CAA GCC AAC ACC AAC CTG TCC AAG TTC CGC AAG GTG CAG
 L   Q   L   K   V   K   A   Y   K   R   Q   A   E   E   A   E   E   Q   A   N   T   N   L   S   K   F   R   K   V   Q

                                             1830                                          1860                                          1890
CAC GAG CTG GAT GAG GCA GAG GAG CGG GCG GAC ATC GCC GAG TCC CAG GTC AAC AAG CTG CGG GCC AAG AGC CGT GAC ATT GGC ACG AAG
 H   E   L   D   E   A   E   E   R   A   D   I   A   E   S   Q   V   N   K   L   R   A   K   S   R   D   I   G   T   K

                                             1920                                          1950                                          1980
GGC TTG AAT GAG GAG TAG CTT TGC CAC ATC TTG ATC TGC TCA GCC CTG GAG GTG CCA GCA AAG CCC CAT GCT GGA GCC TGT GTA ACA GCT
 G   L   N   E   E   .

                                             2010
CCT TGG GAG GAA GCA GAA TAA AGC AAT TTT CCT TGA AGC CGA
```

Figure 2
Restriction maps of pSMHCA and pSMHCZ; nucleotide and deduced protein sequence of pSMHCZ.
Symbols for restriction endonuclease sites are indicated for each clone. Wavy lines correspond to vector sequences. Solid bars correspond to protein coding sequences. Solid lines correspond to 3' untranslated regions. Sequencing strategies are indicated by horizontal lines. •—> indicates sequencing by Maxam-Gilbert protocols and ⊢—> indicates sequencing by dideoxy protocols. Amino acids are represented by the single letter code. • indicates the termination codon.

Sequence Analysis of 2 Human Skeletal Muscle MHC cDNAs

pSMHCA and pSMHCZ were subjected to DNA sequence analysis. Their restriction maps with accompanying sequencing strategies are shown in Figure 2. Their maps are nonoverlapping. Complete DNA sequences and the single letter codes of the amino acids from the coding strands are presented in Figures 2 and 3. Each clone encodes the LMM portion of MHC, including the entire 3' untranslated region and a portion of the poly(A) tail. pSMHCA has a 2736 base pair (bp) insert of which 2628 bp encode the 876 COOH-terminal amino acids. pSMHCZ has a 2022 bp insert of which 1905 bp encode the 635 COOH-terminal amino acids. The relationship of the two human clones to the myosin molecule and all other published MHC cDNAs from other organisms is shown in Figure 4.

Comparison of the coding portions of the clones is presented below. Their 3' untranslated regions, including their utilization of different termination codons, are completely nonhomologous. The 3' untranslated region of pSMHCA is 105 nucleotides and that of pSMHCZ is 116 nucleotides. The two clones also have different poly(A) addition signals. pSMHCZ has the consensus AATAAA located 19 nucleotides from the poly(A) tail. Unlike most cDNAs which

HUMAN SKELETAL pSMHCA

```
30 60 90
AAA TTG GCT CAA GAA TCC GCA ATG GAT ATA GAA AAT GAC AAA CAA CAA CTT GAT GAA AAG CTT AAA AAG AAA GAG TTT GAA ATG AGC GGT
 K   L   A   Q   E   S   A   M   D   I   E   N   D   K   Q   Q   L   D   E   K   L   K   K   K   E   F   E   M   S   G

120 150 180
CTG CAA AGC AAG ATT GAA GAT GAA CAA GCC CTT GGT ATG CAG CTG CAG AAG AAA ATC AAG GAG TTA CAA GCC CGC ATT GAG GAG CTG GAG
 L   Q   S   K   I   E   D   E   Q   A   L   G   M   Q   L   Q   K   K   I   K   E   L   Q   A   R   I   E   E   L   E

210 240 270
GAG GAA ATC GAG GCA GAG CGG ACC TCC CGG GCC AAA GCA GAG AAG CTG CGC TCT GAT CTC TCC CGG GAG CTG GAG GAG ATC AGT GAG AGG
 E   E   I   E   A   E   R   T   S   R   A   K   A   E   K   L   R   S   D   L   S   R   E   L   E   E   I   S   E   R

300 330 360
CTG GAA GAA GCC GTG GGG GCC ACC TCG ACC CAG ATT GAG ATG AAC AAG AAG CGG GAA GCT GAG TTC CAG AAA ATG CGC AGG GAC CTG GAG
 L   E   E   A   V   G   A   T   S   T   Q   I   E   M   N   K   K   R   E   A   E   F   Q   K   M   R   R   D   L   E

390 420 450
GAG GCC ACC CTA CAG CAT GAG GCC ACG GCG GCC ACC CTG AGG AAG AAG CAT GCA GAT AGT GTG GCC GAG CTT GGG GAG CAG ATT GAC AAC
 E   A   T   L   Q   H   E   A   T   A   A   T   L   R   K   K   H   A   D   S   V   A   E   L   G   E   Q   I   D   N

480 510 540
CTG CAG CGA GTG AAG CAG AAG CTG GAG AAG GAG AAG AGT GAG ATG AAG ATG GAG ATC GAT GAC CTT GCT AGT AAC ATG GAG ACT GTC TCC
 L   Q   R   V   K   Q   K   L   E   K   E   K   S   E   M   K   M   E   I   D   D   L   A   S   N   M   E   T   V   S

570 600 630
AAA GCC AAG GGA AAC CTT GAA AAG ATG TGC CGC GCT CTA GAA GAT CAA CTG AGT GAA ATT AAG ACC AAG GAA GAG GAG CAG CAG CGG CTG
 K   A   K   G   N   L   E   K   M   C   R   A   L   E   D   Q   L   S   E   I   K   T   K   E   E   E   Q   Q   R   L

660 690 720
ATC AAT GAC CTC ACA GCA CAG AGA GCG CGC CTG CAA CAG AAT CAG GTG GAA TAT TCA CGC CAG CTA GAT GAA AAG GAC ACA CTA GAA ACA
 I   N   D   L   T   A   Q   R   A   R   L   Q   Q   N   Q   V   E   Y   S   R   Q   L   D   E   K   D   T   L   E   T

750 780 810
CAG CTC TCG AGG GGC AAA CAA GCC TTT ACT CAA CAG ATT GAG GAA CTG AAA AGG CAA CTT GAA GAG GAG ATA AAG GCC AAG AGT GCC CTG
 Q   L   S   R   G   K   Q   A   F   T   Q   Q   I   E   E   L   K   R   Q   L   E   E   E   I   K   A   K   S   A   L

840 870 900
GCA CAT GCC CTG CAG TCC TCC CGC CAT GAC TGT GAC CTG CTG CGG GAA CAG TAT GAG GAG GAG CAG GAA GCC AAG GCC GAG CTA CAG AGA
 A   H   A   L   Q   S   S   R   H   D   C   D   L   L   R   E   Q   Y   E   E   E   Q   E   A   K   A   E   L   Q   R

930 960 990
GCA ATG TCC AAG GCC AAC AGT GAG GTT GCC CAG TGG AGG ACC AAA TAT GAG ACA GAT GCC ATC CAG CGC ACA GAG GAG CTG GAG GAG GCC
 A   M   S   K   A   N   S   E   V   A   Q   W   R   T   K   Y   E   T   D   A   I   Q   R   T   E   E   L   E   E   A

1020 1050 1080
AAG AAG AAG CTG GCT CAG CGT CTG CAG GAT GCT GAG GAA CAT GTA GAA GCT GTG AAT GCC AAA TGT GCT TCC CTT GAG AAG ACG AAG CAG
 K   K   K   L   A   Q   R   L   Q   D   A   E   E   H   V   E   A   V   N   A   K   C   A   S   L   E   K   T   K   Q

1110 1140 1170
AGG CTC CAG AAT GAA GTT GAG GAC CTC ATG ATT GAT GTT GAG AGG ACA AAT GCT GCC TGT GCC GCC CTG GAC AAA AAG CAA ACC AAC TTT
 R   L   Q   N   E   V   E   D   L   M   I   D   V   E   R   T   N   A   A   C   A   A   L   D   K   K   Q   T   N   F

1200 1230 1260
GAT AAG ATC CTG GCA GAA TGG AAA CAG AAG TGT GAA GAA ACT CAT GCT GTT CTT GAA AGC TTT CAA AAG GAA TCC CGC TCA CTC AGC ACA
 D   K   I   L   A   E   W   K   Q   K   C   E   E   T   H   A   V   L   E   S   F   Q   K   E   S   R   S   L   S   T

1290 1320 1350
GAA CTA TTT AAG ATT AAG AAT GCT TAT GAG GAA TCT TTA GAC CAA CTT GAA ACC TTG AAA CGG GAA AAT AAG AAT CTG CAG CAG GAG ATT
 E   L   F   K   I   K   N   A   Y   E   E   S   L   D   Q   L   E   T   L   K   R   E   N   K   N   L   Q   Q   E   I

1380 1410 1440
TCT GAT CTC ACT GAA CAG ATT GCA GAA GGA GGA AAG CGC ATC CAT GAA CTG GAA AAA ATA AAG AAG CAA GTT GAG CAA GAA AAG TCT GAA
 S   D   L   T   E   Q   I   A   E   G   G   K   R   I   H   E   L   E   K   I   K   K   Q   V   E   Q   E   K   S   E

1470 1500 1530
CTT CAG GCT GCC TTA GAG GAG GCA GAG GCA TCT CTT GAA CAT GAA GAG GGA AAG ATC CTG CGC ATC CAG CTT GAG GTT AAC CAA GTC AAG
 L   Q   A   A   L   E   E   A   E   A   S   L   E   H   E   E   G   K   I   L   R   I   Q   L   E   V   N   Q   V   K

1560 1590 1620
TCT GAG GTT GAT AGG AAA ATT GCT GAA AAA GAT GAG GAA ATT GAC CAG ATG AAG AGA AAC CAC ATT AGA ATC GAG GAG TCC ATG CAG AGC
 S   E   V   D   R   K   I   A   E   K   D   E   E   I   D   Q   M   K   R   N   H   I   R   I   E   E   S   M   Q   S

1650 1680 1710
ACA CTG AAT GCT GAG ATC AGG AGC AGG AAT GAT GCC ATT AGG CTC AAG AAG AAG ATG GAG GGA GAC CTC AAT GAA ATG GAA ATC CAG CTG
 T   L   N   A   E   I   R   S   R   N   D   A   I   R   L   K   K   K   M   E   G   D   L   N   E   M   E   I   Q   L

1740 1770 1800
AAC CAT GCC AAC CGC ATG GCT GCT GAG GAC CTG AGG AAC TAT CAG AAC ACC CAA GCC ATC CTC AAG GAT ACC CAG CTC CAC CTA GAT GAT
 N   H   A   N   R   M   A   A   E   D   L   R   N   Y   Q   N   T   Q   A   I   L   K   D   T   Q   L   H   L   D   D

1830 1860 1890
GCT CTC CGG AGC CAA GAG GAC CTG AAG GAA CAG CTG GCT ATG GTG GAG CGC AGA GCC AAC CTG CTG CAG GCT GAG ATC GAG GAA CTA CGA
 A   L   R   S   Q   E   D   L   K   E   Q   L   A   M   V   E   R   R   A   N   L   L   Q   A   E   I   E   E   L   R
```

```
                                         1920                                    1950                                    1980
GCC ACT CTG GAA CAG ACG GAG AGG AGC AGG AAA ATC GCA GAA CAG GAG CTC CTG GAT GCC AGT GAA CGT GTT CAG CTC CTG CAC ACC CAG
 A   T   L   E   Q   T   E   R   S   R   K   I   A   E   Q   E   L   L   D   A   S   E   R   V   Q   L   L   H   T   Q

                                         2010                                    2040                                    2070
AAC ACC AGC CTG ATC AAC ACC AAG AAG AAG CTG GAG ACA GAC ATT TCC CAA ATC CAG GGA GAG ATG GAA GAC ATC ATC AAG GAA GCC CGC
 N   T   S   L   I   N   T   K   K   K   L   E   T   D   I   S   Q   I   Q   G   E   M   E   D   I   I   K   E   A   R

                                         2100                                    2130                                    2160
AAT GCA GAA GAG AAG GCC AAG AAG GCC ATC ACT GAT GCT GCC ATG ATG GCT GAG GAG CTG AAG AAG GAA CAG GAC ACC AGC GCC CAT CTG
 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   T   S   A   H   L

                                         2190                                    2220                                    2250
GAG CGG ATG AAG AAG AAC TTG GAA CAG ACG GTG AAG GAC CTG CAG CAT CGT CTG GAT GAG GCT GAG CAG CTG GCC CTG AAG GGT GGG AAG
 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

                                         2280                                    2310                                    2340
AAG CAG ATC CAG AAA CTG GAG GCC AAG GTT CGT GAA CTT GAA GGT GAA GTT GAA AGT GAA CAG AAG CGC AAT GTT GAA GCT GTC AAG GGT
 K   Q   I   Q   K   L   E   A   K   V   R   E   L   E   G   E   V   E   S   F   Q   K   R   N   V   E   A   V   K   G

                                         2370                                    2400                                    2430
CTA CAC AAA CAT GAG AGA AAA GTG AAG GAA CTC ACT TAC CAA ACT GAG GAA GAC CGC AAG AAT ATT CTC AGG CTG CAG GAC CTG GTG GAC
 L   H   K   H   E   R   K   V   K   E   L   T   Y   Q   T   E   E   D   R   K   N   I   L   R   L   Q   D   L   V   D

                                         2460                                    2490                                    2520
AAG CTG CAA GCA AAG GTG AAA TCC TAC AAG AGA CAA GCT GAA GAA GCG GAG GAA CAA TCC AAC GTC AAC CTC TCC AAA TTC CGG AGG ATC
 K   L   Q   A   K   V   K   S   Y   K   R   Q   A   E   E   A   E   E   Q   S   N   V   N   L   S   K   F   R   R   I

                                         2550                                    2580                                    2610
CAG CAC GAG CTG GAG GAG GCC GAG GAA AGG GCT GAC ATT GCT GAG TCC CAG GTC AAC AAG CTG AGG GTG AAG AGC AGG GAG GTT CAC ACA
 Q   H   E   L   E   E   A   E   E   R   A   D   I   A   E   S   Q   V   N   K   L   R   V   K   S   R   E   V   H   T

                                         2640                                    2670                                    2700
AAA ATC ATA AGT GAA GAG TAA TTT ATC TAA CTG CTG AAA GGT GAC CAA AGA AAT GCA CAA AAT GTG AAA ATC TTT GTC ACT CCA TTT TGT
 K   I   I   S   E   E   .

                                         2730
ACT TAT GAC TTT TGG AGA TAA AAA ATT TAT CTG CCA
```

Figure 3 Nucleotide and deduced protein sequence of pSMHCA. Amino acids are represented by the single letter code • indicates the termination codon.

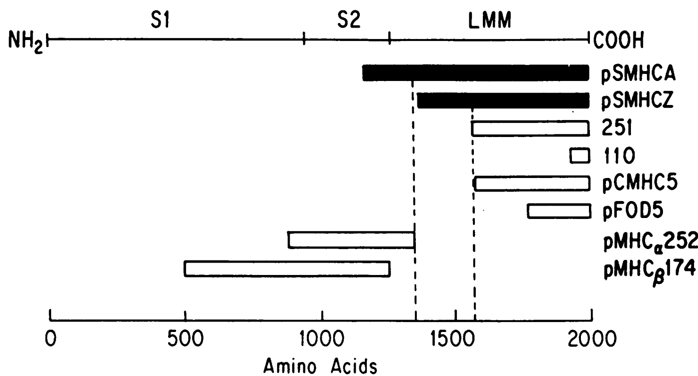
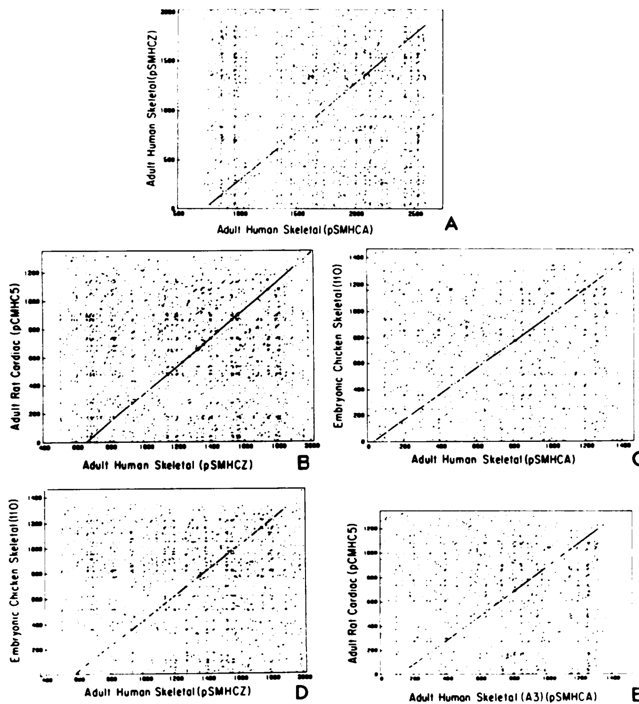


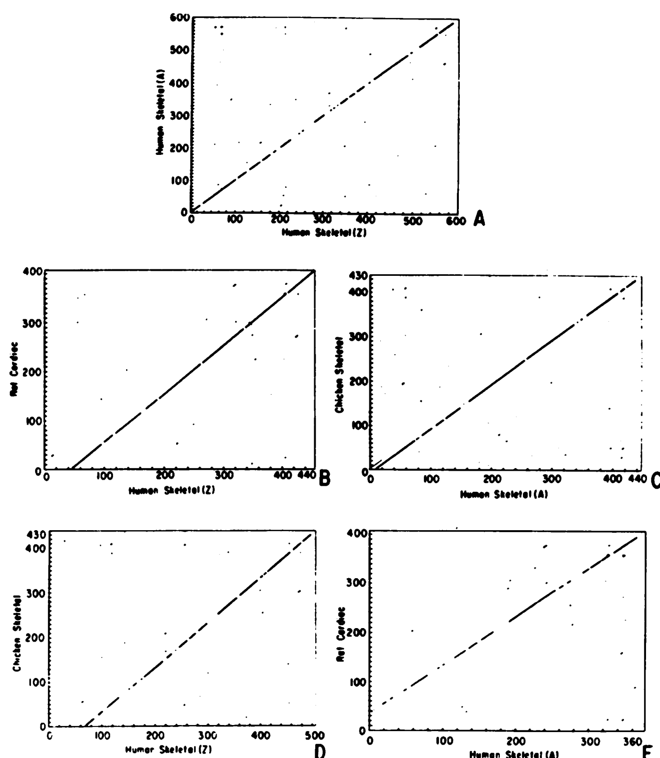
Figure 4 Alignment of MHC cDNA clones with the MHC protein molecule. The MHC protein is drawn at the top of the diagram with domains indiciated. Bars correspond to the MHC cDNA clones presented in this report and previous publications. Their identities are: pSMHCA and pSMHCZ (this report); 251 and 110 (embryonic chicken skeletal (15)); pCMHC5 (adult rat cardiac (12)); pFOD5 (perinatal rat skeletal (14)); pMHCα252 and pMHCβ174 (rabit cardiac (13)). Dotted lines indicate the protion ofthe human clones that provide new data.

contain AATAAA, pSMHCA has the sequence GATAAA which is located 15 nucleotides from the poly(A) tail. The sequence determination was confirmed on both strands and with both chemical sequence and dideoxynucleotide methods since it represents a departure from the consensus. The significance of this base change is not known.

Comparisons of Human, Rodent and Chicken MHC cDNAs

In order to identify the MHC forms represented by the 2 human MHC clones and to evaluate the evolutionary conservation of these forms, graphic matrix analyses of MHC DNA and deduced protein sequences were carried out. pSMHCA and pSMHCZ were compared to each other and to previously described MHC sequences derived from physiologically characterized muscle fibers. The DNA analyses were conducted at a 75% stringency level with a window of 10 where a match of 8 of 10 bases is required for placement of a dot. The protein comparisons were carried out with a window of 5 amino acids where a match of 4 of 5 is required for placement of a dot. These analyses are presented in Figure 5 (Panels A-E) for DNA comparisons and Figure 6 (Panels A-E) for amino acid comparisons. Several features of the results are striking. pSMHCA and pSMHCZ are less homologous to each other than each is to MHCs from other





Figures 5 and 6   Simple graphic matrix analyses of DNA and protein sequences of MHC cDNA clones.
Comparisons include pSMHCA and pSMHCZ against each other (Panel A) and against an embryonic chicken fast fiber skeletal clone, 110 described in (15) (Panels C and D); and a rat β ventricular cardiac clone, pCHMHC 5 described in (12) (Panels B and E). In DNA comparisons, the stringency was 75% with a window of 10 nucleotides. For protein sequence comparisons, the stringency was 75% with a window of 5 amino acids.

organisms. This shows that tissue-specific MHCs between different species are more homologous to each other than are different MHC forms within an organism. The two human skeletal clones are 71.8% homologous at the DNA level. This high degree of sequence diversity was unexpected given earlier observations by others that two embryonic chicken skeletal MHC clones were 89% homologous (15); two adult rat ventricular cardiac MHC clones were 85% homologous (12); and 2 rabbit β cardiac sequences were 93% homologous (13). pSMHCA is more homologous to an embryonic chicken skeletal MHC clone (80.7%) (Panel C) than it is to pSMHCZ, the human MHC with which it is co-expressed (Panel A). Comparison of pSMHCA with a rat cardiac MHC form shows

```
d e f g a b c    d e f g a b c    d e f g a b c    d e f g a b c
+       +        +       +        +       +        +       +

            - -  - - - - - - -    - - - - - - -    - - - - - - - pSMHCZ
            K L  A Q E S A M D    I E N D K Q Q      L D E K L K K pSMHCA

- - - - - - -    - - - - - - -    - - - - - - -    - - - - - - -
K E F E M S G    L Q S K I E D    E Q A L G M Q    L Q K K I K E

- - - - - - -    - - - - - - -    - - - - - - -    - - - - - - -
L Q A R I E E    L E E E I E A    E R T S R A K    A E K L R S D

- - - - - - -    - - - - - - -    - - - - - - -    - - - - - - -
L S R E L E E    I S E R L E E    A V G A T S T    Q I E M N K K

- - - - - - -    - - - - - - -    - - - - - - - -  - - - - - - - pSMHCZ
R E A E F Q K    M R R D L E E    A T L Q H E A T  A A T L R K K pSMHCA

- - - - - - -    - - - - - - -    - - - - - - -    - - - - - - -
H A D S V A E    L G E Q I D N    L Q R V K Q K    L E K E K S E

- - - - - - -    - - - - - - -    : - - - - - -    - - - - - - -
M K M E I D D    L A S N M E T    V S K A K G N    L E K M C R A

- - - - - - -    - - - - - - -    - - - - - - -    - - - - - - -
L E D Q L S E    I K T K E E E    Q Q R|L I N D    L T A Q R A R

- - - - - - -    - - - - - - -    - - - - - - -    L L G V G E L
L Q Q N Q V E    Y S R Q L D E    K D T L E T Q    L S R G K Q A

A S G Q L G D    L K R Q L E E    E V K A K N A    L A H A L Q S pSMHCZ
F T Q * I E E    * * * * * * *    * I * * * S *    * * * * * * * pSMHCA

A R H D C D L    L R E Q Y E E    E T E A K A R    L Q G D L S K
S * * * * * *    * * * * * * *    * Q * * * * E    * * R A M * *

A N S E V A Q    W R T K Y E T    D A I Q R T E E  L E E A K K K
* * * * * * *    * * * * * * *    * * * * * * * *  * * * * * * *

L A Q R L Q E    A E E A V E A    V N A K C S S    L E K T K H R
* * * * * * D    * * * H * * *    * * * * * A *    * * * * * Q *

L Q N E I E D    L M V D V E R    S N A A L Q A    L D K K Q R N
* * * * V * *    * * I * * * *    T * * * C A *    * * * * * T *

F D K I L A E    W K Q K Y E E    S Q S E L E S    S Q K E A R S pSMHCZ
* * * * * * *    * * * * C * *    T H A V * * *    F * * * S * * pSMHCA

L S T E L F K    L K N A Y E E    S L E H L E T    F K R E N K N
* * * * * * *    I * * * * * *    * * D Q * * *    L * * * * * *

L Q E E I S D    L T E Q L G S    S G K T I H E    L E K V R K Q
L * Q * * * *    * * * * I A E    G * * R * * *    * * * I K * *

L E A E K M E    L Q S A L E E    A E A S L E H    E E G K I L R
V * Q * * S *    * * A * * * *    * * * * * * *    * * * * * * *

A Q L V F N Q    I K A E I E R    K R Q E K D E E  M E Q A K R N
I * * E V * *    V * S * V D R    * I A * * * * *  I D * M * * *

H L R V E D S    L Q T S L D A    E T R N R N E    A L R V K K K pSMHCZ
* I * I * E *    M * S T * N *    * I * S * * D    * * I R L * * pSMHCA

M E G D L N E    M E I Q L S H    A N R M A A E    A Q K N L R A
* * * * * * *    * * * * * N *    * * * * * * *    D L R * Y Q N

S Q E L L K D    T Q I Q L D D    A V R A N D D    L K E N I A I
T * A I * * *    * * L H * * *    * L * S Q E *    * * * Q L * M

V E R R N N L    L Q A E L E E    L R A V V E Q    T E R S R K L
* * * * A * *    * * * * I * *    * * * T L * *    T * * * * * I

A E Q E L I E    T S E R V Q L    L H S Q N T S    L I N Q K K K
* * * * * L D    A * * * * * *    * * T * * * *    * * * T * * *

M D A D L S Q    L Q G E V E E    A V Q E C R N    A E E K A K K pSMHCZ
L E T * I * *    I * * * M * D    I I K * A * *    * * * * * * * pSMHCA

A I T D A A M    M A E E L K K    E Q D T S A H    L E R M K K N
* * * * * * *    * * * * * * *    * * * * * * *    * * * * * * *
```

```
M E Q T I K D   L Q H R L D E   A E Q I A L K G   G K K Q L Q K
L * * * V * *   * * * * * * *   * * * L * * *     * * * * I * *

L E A R V R E   L E N E L E A   E Q K R N A E     S V K G M R K
* * * K * * *   * * G * V * S   * * * * * V *     A * * * L H *

S E R R I K E   L T Y Q T E E   D R K N L L R     L Q D L V D K
H * * K V * *   * * * * * * *   * * * * I * *     * * * * * * *

L Q L K V K A   Y K R Q A E E   A E E Q A N T     N L S K F R K pSMHCZ
* * A * * * S   * * * * * * *   * * * * S * V     * * * * * * R pSMHCA

V Q H E L D E   A E E R A D I   A E S Q V N K     L R A K S R D
I * * * * E *   * * * * * * *   * * * * * * *     * * V * * * E

I G T K G L N   E E
V H * * I I S   E E

+     +         +     +         +     +           +     +
d e f g a b c   d e f g a b c   d e f g a b c     d e f g a b c
```

Figure 7 Comparison of the amino acid sequences of pSMHCA and pSMHCZ. Alignment was made with the amino acids grouped into 7 residue repeat units. a-e correspond to the postions in this unit where a and d are frequently occupied by uncharged residues (9). Dashes indicate the portion of pSMHCA which is longer than that of pSMHCZ. Asterisks (*) represent homologous amino acids and substitutions are indicated. The residues contained in the box correspond to new MHC sequence provided in this study.

considerable sequence divergence (72.6% homology) (Panel E). These analyses make it possible to tentatively identify pSMHCA as a fast human skeletal form. It should be noted that the graphic matrix comparisons cover only the regions of overlap. Since both human clones are much longer than other tail region clones described (see Figure 4), portions of their sequence can only be compared to each other.

pSMHCZ is most homologous to a rat MHC described as a β cardiac ventricular cDNA (88.1% homology) (Panel B), and is much less homologous to a chicken fast MHC cDNA clone (72.8%) (Panel D). As expected, graphic matrix analysis of the MHC proteins encoded by these cDNAs reveals stronger homology than that seen at the DNA level. The identification of pSMHCA and pSMHCZ is more obvious by this analysis (Figure 6). The sequence differences between pSMHCA and pSMHCZ are still very high when protein sequences are compared (23.7% divergence). These comparisons provide a dramatic example of the evolutionary maintenance of myosin isotypes and suggest that there is functional significance to the various forms. In the present study, we have documented the diversity of myosin forms in one human skeletal muscle at the sequence level.

The Structure of the MHCs Encoded by pSMHCA and pSMHCZ

The derived protein sequences of the two human MHC clones contain an abundance of α-helix forming amino acids and are devoid of proline residues.

Like other vertebrate MHCs, neither sequence has a nonhelical segment at its carboxy terminus as has been seen for nematode (9) and Dictyostelium (35) myosin termini. The portion of the MHC protein encoded by pSMHCA and pSMHCZ clones includes the entire light meromyosin region. pSMHCA also contains 96 amino acids of the S2 region of MHC. Comparison of the amino acids encoded by pSMHCA and pSMHCZ is shown in Figure 7. The amino acids are grouped into the 7-residue repeat unit which has been shown to exist for proteins that form a coiled coil such as MHC (9,37) (see below). Asterisks indicate homologous amino acid residues and substitutions are indicated by the appropriate letters. There are short regions where the two sequences are identical and other regions with many changes, but it is not yet possible to assign functional significance to those regions of change and homology. In general, the two MHC sequences utilize synonomous codons (data not shown).

The coiled coil structure of the MHC rod is responsible for the assembly of myosin into thick filaments (2). This portion of the heavy chain can be formulated into groups of seven amino acids in which the first and fourth amino acid positions are frequently occupied by non-polar residues (9,36). This sequence organization serves to maximize interactions between the hydrophobic amino acid side-chains of the two heavy chain subunits. Previous studies have shown such a repeating unit exists in isolated MHC cDNA clones and the nematode unc-54 gene (9,15,37). pSMHCA and pSMHCZ also show such periodicity (Figure 7). The continuity of the 7- and larger 28- residue repeat units is interrupted by insertions of single amino acids known as skip residues. This occurs at four positions in pSMHCA and two positions in pSMHCZ. The two positions in pSMHCZ coincide with two of the four positions in pSMHCA. These skips are thought to be accomodated by localized changes in the pitch of the helix (36). However, the skip residues may have functional significance given the conservation of their positions in several MHC cDNA clones including pSMHCA and pSMHCZ.

3' Untranslated Region Conservation Among Tissue-Specific MHCs

The 3' untranslated regions of eukaryotic mRNAs have no demonstrated function. In most cases, these sequences are widely divergent between organisms and within members of a multigene family despite a high degree of coding region sequence conservation (38). The rate of sequence divergence in 3' untranslated regions of mRNAs is thought to be equal to the rate of divergence of DNA on which there is no selective pressure (38). The 3' untranslated regions of members of highly conserved multigene families can frequently be used to distinguish individual members of a multigene family

3'UNTRANSLATED REGION HOMOLOGIES

```
Human(A)  TAA  TTTATCTAA   CTGCT   G AAAGGTGACCAAAGAAATGCACAAAATGTGAAAAT
RatI      TAAGGCACCTCTGA   CTGCT   G AA   TGACCGAAGAAAGGCACAAAATGTGAA GC
RatII     TAG     CTC AATTCCTTCTGTTG AAAGGTGACAGAAGAAAT CACACAATGTGACGTT
RatIII    TAAACGCATCTTGAGGAGGGCC   GCCAAG TGGCTAAGGAAAGGCACAGAATGTGC TGC
ChickI    TAGATGCCTC AAGCGGT       GCAAAG TG AAATAGAATTGCACAAAATGTGAAATT
ChickII   TGAAGATATCATCTGACA       GCAAAG TGGCCTGAGGAGTGCACAAAATGTGTGAA CCCTC
Quail     TAGATGCCTCC AG   TGGT    GCAAAG TG AAAGAGAATTGCACAAAATGTGAAATT
----------------------------------------------------------------------------
Human(Z)  TAG  CTTTGCCACATCTT    GATC  TGCTCAGCCCTGGAGGTGCCAGCAAAGCCCC
Rat V     TAGATCTT GCTCTACCCAACCCTAAGGATG    CC     TGTG        AAGCCCT
RatIV     TAA CCT  GTCCAGCA      GAAAG                          AGCCTC


Human(A)       CTTTG  TCACT CCATTTTGTAATTA TGAC
RatI           CTTTGG TCATG CCCCCATGTGATTC T      TTTAA        TCCTA
RatII          CTTTG  TCACTGTCCTGTATA                          TC     AAA
RatIII         C TTGGGTCGCTTGCTGGGTCGCTTGCCTCTCGTGTTTACTT  T TCTCCCACT
ChickI         CTA    TCACTT TGATTTGTAATTACTGCTTAGTT  CTTCAACTATCTAGATAAT
ChickII   TGTTGCTA    TCACTTATAATTTATCTTTA    TAACCAC CT CAATG TCTAGA
Quail          CTAT   TCACTT TGAT TGTGATTAC GCT AGTT  CTTCATCAATC     AAT
----------------------------------------------------------------------------
Human(Z)  ATG  CTGGAGCCTGTGTAACAGCTCCTT  GGGA
Rat V     GAGACCTGGAGCCTTTGAAACAGCACCTTCAGGCA
RatIV            GCCGTT        GCCATC


Human(A)            GGA GATAAAAAAA         ATTTA       TCTGC     CAAn
RatI      TTGTAA    GG  AATAAA       GAG         CC  CAAGTTCTTCAAGCAn
RatII               GGA AATAAA CTGCAGATA   ATTTT                 GCAn
RatIII               G  AATAAAACCACAACTC   ATTGT                AATTAn
ChickI       TAATATTTAGATAATAAAAATTGTAGAG  ATTTT CC  CAT         GGAn
ChickII              G  AATAAAGACCATAG     ATTCCTCT  GCAT        ATAn
Quail       GTAATGTTT GATAATAAA ATTGTAGAG  ATTTT CC              ATGAn
----------------------------------------------------------------------------
Human(Z)      GGAAGC AG  AATAAA   GCA      ATTTT CCTTGAAGC       CGAn
Rat V          GAA   ACACAATAAA   GCA      ATTTT CCTTCAAG        CCAn
Rat IV               CCACAATAAATACGAATGTTCGATTTG CCT             GCAn
```

Figure 8 Comparison of the 3' untranslated regions of skeletal and cardiac MHC cDNA clones.
Skeletal MHC clones compared include pSMHCA (this report), two chick embryonic skeletal muscle cDNAs, 110 (I) and 251 (II) (15), adult rat skeletal (rat I), adult rat skeletal (rat II), fetal rat skeletal (III) all described in (10) and adult quail breast (42). Cardiac clones compared include pSMHCZ (this report) and two adult rat cardiac MHC cDNAs (rat IV and V) (12). The first three nucleotides of each sequence are termination codons.

within an organism, but not between species. One notable exception to this is the recent observation that members of the highly conserved actin gene family show conservation of the 3' untranslated regions of isotypes between organisms (39).

We compared the 3'untranslated regions of pSMHCA and pSMHCZ to each other and, as expected, found no homology (see Figures 2 & 3). However, when the two human MHCs described here were compared with previously described MHCs from other organisms, significant homology was seen among the 3' untranslated

regions of all described skeletal muscle MHCs. Among the three cDNAs described from cardiac muscle, two (including pSMHCZ by virtue of the RNA and homology data presented here) show one region of strong homology. Figure 8 shows these comparisons. The 3' untranslated region of pSMHCA is homologous to the analogous region of six skeletal MHC clones from rat, chicken, and quail, from both adult and embryonic tissues. This homology does not extend throughout the entire region, but over about 40 nucleotides in the approximate center of the region, indicated in the box. Over this region, the homology ranges from 60% up to 85%. Interestingly, such extensive homology exists between pSMHCZ (which we believe to be homologous to a β cardiac form) and only one of two rat cardiac venticular MHC clones.

DISCUSSION

In the current study we report the first sequences of human myosin heavy chains and the first complete sequence of the LMM region of vertebrate MHC. The clones used to generate these data derive from adult human skeletal muscle and are the products of two genes. By comparison with previously described sarcomeric MHCs from chicken and rat, we have been able to identify these two clones as a fast fiber form and a form which is likely to be coexpressed in cardiac ventricles. Examination of the structural features of the proteins encoded by pSMHCA and pSMHCZ shows them to exhibit the characteristic patterns of the tail region of MHC. These include the 7- and 28-amino acid repeat units where the first and fourth amino acids in the 7 residue unit are occupied by uncharged residues. An additional conserved feature is the propensity to form an α helix and the complete absence of proline residues. Sequence analysis of the two human MHC clones points to several interesting features of MHC gene expression. pSMHCA and pSMHCZ encode MHCs that are quite different from each other (23% divergent). This is a higher degree of diversity than previously reported when MHCs isolated from one tissue were compared. One study of MHC expression in adult rat skeletal muscle differs from our results by showing that two different MHC cDNA clones from the same tissue have identical amino acids at the carboxy terminus (9). Similar to our findings is the study in which a rat actin species was shown to be expressed in both skeletal and cardiac tissues by sequence analysis (40). S1 nuclease analysis of RNA from multiple muscle types with a rat cardiac MHC DNA probe suggests coexpression of one MHC in both cardiac and skeletal muscle (34). This latter result is similar to what we show in this manuscript for pSMHCZ, but our work provides DNA sequence analysis as identification.

The two cDNA clones characterized in our studies were selected from a population of twelve MHC clones. We have used oligonucleotides from the 3' untranslated regions of pSMHCA and pSMHCZ as hybridization probes against the twelve MHCs and have shown that 2/12 are identical or overlapping with pSMHCA and 3/12 are identical or overlapping with pSMHCZ (manuscript in preparation). These and other data suggest a minimum of three and more likely at least five different MHCs expressed in adult human skeletal muscle tissue. In adult human skeletal muscle a total of five isomyosins associated with two types of fibers has been described (41). The relative proportion of those isoforms shows considerable variation in each individual. Determination of the full extent of MHC diversity in adult skeletal muscle awaits more extensive sequence analysis. These experiments are currently in progress.

Several evolutionary relationships can be drawn from these studies. The homology between the human skeletal pSMHCZ sequence and the rat ventricular MHC is much greater than that between two human clones that are coexpressed. Analogously, pSMHCA is more related to a chicken skeletal sequence than to pSMHCZ. These data provide further evidence that myosin isotypes have been maintained through evolution. This, in turn suggests that MHC forms provide physiological diversity to the muscles in which they are expressed. The sequence homology between evolutionarily divergent species extends into the 3' untranslated region. It is intriguing that both actin and muscle MHC show conservation of 3' untranslated region across a broad range of species while most other genes do not. It may be that the 3' untranslated regions have a function in tissue-specific expression of MHC. It is interesting that we do not see conservation among the 3' untranslated regions of all three cardiac MHCs. Further investigation is required to define possible function(s) of these regions.

While pSMHCA and pSMHCZ are quite divergent at both DNA and protein levels, the structural features of their encoded proteins as well as of other myosins (9,15) have been maintained. They include the absence of proline residues in LMM, the α helical nature of the tail region and the periodicity of uncharged residues in the 7- and 28-residue repeats. Recently, a molecular clone corresponding to nonmuscle MHC from Dictyostelium has been characterized. Despite the absence of sequence homology to muscle MHC, these same structural features have been maintained (35). It remains to be demonstrated that sequence diversity in the rod portion of the myosin molecule is functionally significant. However, the evolutionary conservation of these forms has strong functional implications.

ACKNOWLEDGEMENTS

We wish to acknowledge the excellent secretarial assistance of Donna Jackson and Laurie Vitagliano. Thanks go to the weekly Molecular Biology group for many helpful suggestions (S. Emmons, G. Childs, K. Krauter, S. Hawley and J. Lenz). This work was funded in part by grants from the National Institutes of Health (GM29090) and the American Heart Association. L. L. is an American Heart Association Established Investigator and L.S. is supported by an award from Dupont Chemicals.

REFERENCES

1. Leinwand, L., Saez, L., McNally, E. and Nadal-Ginard, B. (1982) Proc. Natl. Acad. Sci. U.S.A. **80**,3716-3720.
2. Cohen, C. and Szent-Gyorgyi, A.G. (1957) J. Amer. Chem. Soc. **79**,248.
3. Lowey, S., Slayter, H., Weeds, A., and Baber, H. (1969) J. Mol. Biol. **42**,1-29.
4. Whalen, R.G., Schwartz, K., Bouvret, P. and Gros, F. (1979) Proc. Natl. Acad. Sci. USA **76**,5197-5201.
5. Gauthier, G.F. and Lowey, S. (1977) J. Cell Biol. **74**,760-779.
6. Huszar, G. and Elzinga, M. (1972) J. Biol. Chem. **247**,745-753.
7. Wagner, P.D. and Giniger, E. (1981) Nature **292**,560-562.
8. Sheetz, M.P. and Spudich, J.A. (1983) Nature **303**,31-35.
9. McLachlan, A.D. and Karn, J. (1982) Nature **229**,226-231.
10. Nadal-Ginard, B., Medford, R., Nguyen, H., Penasamy, M., Wydro, R. Homig, D., Gubits, R., Garfinkel, L., Weiczorek, D., Bekesi, E. and Mahdavi, V. in Molecular and Cellular Control of Muscle Development (M. Pearson and H. Epstein, eds.) Cold Spring Harbor Press, N.Y. pp. 143-168.
11. Medford, R., Wydro, R., Nguyen, H. and Nadal-Ginard, B. (1980) Proc. Natl. Acad. Sci. U.S.A. **77**,5749-5753.
12. Mahdavi, V., Periasamy, M. and Nadal-Ginard, B. (1982) Nature **297**,659-665.
13. Kavinsky, C.J., Umeda, P.K., Levin, J.E., Sinha, A.M., Nigro, J.M., Jakovcic, S. and Rabinowitz, M. (1984) J. Biol. Chem. **259**,2775-2781.
14. Periasamy, M., Wieczorck, D.F. and Nadal-Ginard, B. (1984) J. Biol. Chem. **259**,13573-13578.
15. Kavinsky, C.J., Umeda, P.K., Sinha, A.M., Elzinga, M., Tong, S.W., Zak, R., Jakovcic, S. and Rabinowitz, M. (1983) J. Biol. Chem. **258**,5196-5205.
16. Wydro, R., Nguyen, H., Gubits, R., and Nadal-Ginard, B. (1983) J. Biol. Chem. **258**,670-678.
17. Robbins, J., Freyer, G., Chisholm, D., and Gilliam, T.C. (1981) J. Biol. Chem. **257**,549-556.
18. Leinwand, L., Fournier, R.E.K., Nadal-Ginard, B. and Shows, T.B. (1983) Science **221**,766-769.
19. Weydert, A., Baubas, P., Lazaridis, I., Barton, P., Garner, I., Leader, D.P., Bonhomme, F., Catalan,J., Simon, D., Guenet, J.L., Gros, F. and Buckingham, M.E. (1985) Proc. Natl. Acad. Sci. **82**,7183-7187.
20. Mahdavi, V., Chambers, A. and Nadal-Ginard, B. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**,2626-2630.
21. Rozek, C.E. and Davidson, N. (1983) Cell **32**,23-24.
22. Bernstein, S., Mogami, K., Donadi, J. and Emerson, C. (1983) Nature **302**,393-397.

23. Gunning, P., Ponte, P., Okayama, H., Engel, J., Blau, H. and Kedes, L. (1983) Mol. Cell Biol. **3**,787-795.
24. Grunstein, M. and Hogness, D. (1975) Proc. Natl. Acad. Sci. U.S.A. **72**,3961-3965.
25. Rigby, P.W., Dieckman, M., Rhodes, C., and Berg, P. (1977) J. Mol. Biol. **113**,237-251.
26. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
27. Childs, G., Maxon, R. and Kedes, L. (1979) Devel. Biol. **73**,153-173.
28. Derman, E., Krauter, K., Walling, L., Weinberger, C., Ray, M. and Darnell, J.E., Jr. (1981) Cell **23**,731-739.
29. Sanger, F., Nichlen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**,5463-5467.
30. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**,560-564.
31. Donelson, J.E. and Wu, R. (1972) J. Biol. Chem. **247**,4661-4668.
32. Wilbur, W.J. and Lippman, D.J. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**,726-730.
33. Capony, J.P. and Elzinga, M. (1981) Biophysical Journal **33**,148a.
34. Lompre, A.M., Nadal-Ginard, B. and Mahdavi, V. (1984) J. Biol. Chem. **259**,6436-6446.
35. DeLozanne, A., Lewis, M., Spudich, J. and Leinwand, L. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**,6807-6810.
36. Parry, D.A.D. (1981) J. Mol. Biol. **153**,459-466.
37. McLachlan, A.D. and Karn, J. (1983) J. Mol. Biol. **164**,605-626.
38. Efstradiadis, A., Posakony, J., Maniatis, T., Lawn, R., O'Connell, C., Spritz, R., DeRiel, J., Forget, B., Weissman, S., Slighton, J., Blechl, A., Smithies, O., Baralle, F., Shoulders, C. and Proudfoot, N. (1980) Cell **21**,653-668.
39. Ponte, P., Ng, S.Y., Engel, J., Gunning, P. and Kedes, L. (1984) Nucleic Acids Research **12**,1687-1696.
40. Gunning, P., Ponte, P., Blau, H. and Kedes, L. (1983) Mol. Cell. Biol. **3**,1985-1997.
41. Fitzsimons, R.B., and Hoh, J.F.Y. (1981) J. of Neurological Sciences. **52**,367-384.
42. Hastings, K.E.M. and Emerson, C.P., Jr. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**,1553-1557.