Volume 17 Number 19 1989 Nucleic Acids Research

Characterization of a rat myosin alkali light chain gene expressed in ventricular and slow twitch skeletal muscles

Muthu Periasamy*, Raj Wadgaonkar[1], Chandra Kumar[2], Barbara Jill Martin and M.A.Q.Siddiqui[1]

Department of Physiology and Biophysics, University of Vermont College of Medicine, Given Medical Building, Burlington, VT 05405, [1]Department of Anatomy and Cell Biology, State University of New York, Brooklyn, NY 11203 and [2]Department of Tumor Biology, Schering Corporation, 60 Orange Street, Bloomfield, NJ 07003, USA

Received June 30, 1989; Revised and Accepted August 25, 1989

ABSTRACT

Mammalian cardiac muscle contains two myosin alkali light chains: 1) the atrial light chain ($MLC1_A$), and 2) the ventricular light chain ($MLC1_V$) predominantly expressed either in the atrium or in the ventricle. In this report we describe the isolation and characterization of the complete gene for rat $MLC1_V$. The rat $MLC1_V$ gene is ~6.5 kb long and the mRNA coding sequences are organized in 7 different exons. Comparison of this gene sequence with other known MLC1 gene sequences revealed that the exon−intron organization is highly conserved within the MLC1 gene family. The derived protein sequence of rat $MLC1_V$ showed a higher sequence homology with human ventricular (96%) $MLC1_V$ than with rat fast skeletal $MLC1_f$ (74%), suggesting functional similarities between different $MLC1_V$ proteins. S1 nuclease mapping and primer extension analysis demonstrated that this gene is expressed only in ventricular and slow twitch skeletal muscle tissues and is transcribed from the same promoter and transcription initiation site.

INTRODUCTION

Myosin, one of the major constituents of muscle fibers plays a critical role not only in muscle contraction, but also in the maintenance of cell shape and cellular movement in all eukaryotic cells. In muscle and non-muscle tissues, myosin is composed of two heavy chains and four light chains. The myosin light chains are classified into two types: the alkali (MLC1) and the dinitro-benzoic acid removable light chains (MLC2). The physiological role of alkali MLCs is not well understood, although recent studies have suggested that alkali myosin light chain (MLC1) plays a role in the interaction of the myosin head region with actin (1−3). It is nevertheless clear that distinct isoforms of MLC1 are present in different muscle types suggesting that they are associated with different contractile properties. At least six different tissue and/or developmental specific MLC1 isoforms have been identified at the protein level in higher vertebrates (4).

In vertebrates the different MLC1 isoforms are encoded by at least four different genes. Molecular cloning analysis of the various alkali light chains have demonstrated that the fast skeletal $MLC1_f$ and $MLC3_f$ isoforms are generated from a single gene by a mechanism of alternate splicing and promoter utilization (5−7). Further, the recent studies of Nabeshima et al., (8) showed that the alkali light chain found in smooth and non-muscle are very similar and probably originate from a single gene through alternate RNA splicing. Cloning of the atrial alkali light chain gene ($MLC1_A$) from mouse and human provided evidence that it is identical to the isoform expressed in fetal skeletal muscle (9−12). It is also clear from the studies recently reported (13,14) and the work described here that the alkali light chain expressed in the ventricle ($MLC1_V$) is identical to the one expressed in slow-twitch skeletal muscle fibers ($MLC1_S$) in the mouse, rat and human tissues.

In this paper we describe the isolation and characterization of the complete gene for the rat cardiac $MLC1_V$ and present the structure and sequence organization of the gene and its relationship to other alkali MLC1 genes. We also demonstrate that the mRNA expressed in ventricular and slow twitch skeletal muscle fibers is transcribed from the same promoter and transcription initiation site. A comparison of rat $MLC1_V$ sequence with other alkali light chains reveals features relevant to the structure, function and evolution of the MLC1 isoforms.

METHODS

Isolation of cDNA Clone pRLC 87

Plasmid pRLC 87 was isolated from a cDNA library constructed using poly (A+) RNA from rat heart ventricle. Double-stranded cDNA was produced according to Gubler and Hoffman (15) and inserted into λgt 10 as described in Huyngh et al. (16). Clones were identified by hybridization with the previously characterized $MLC1_V$ cDNA pRLC 267 (17). Of the isolated clones, pRLC 87 was identified as containing the largest insert ($\sim$850 bp) and its complete nucleotide sequence was determined.

Isolation of Rat Cardiac Ventricular $MLC1_V$ Gene Sequences

A partial Sau 3A rat genomic DNA library cloned into phage EMBL4 and a partial HaeIII library cloned into phage Charon 4A were used. The isolation of $MLC1_V$ genomic clones were performed as described by Blattner et al. (18) using nick translated rat ventricular myosin light chain cDNA pRLC 87 as a probe.

DNA Sequence Analysis

The sequencing of cDNA and genomic clones was performed by the method of Maxam and Gilbert (19) using end-labelled restriction fragments, and by the dideoxy nucleotide chain terminating method of Sanger et al. (20) after cloning subfragments into M13. Some of the genomic sequencing (exons 1 and 7) on M13 templates was performed using synthetic oligonucleotide primers derived from the cDNA sequence of pRLC 87.

S1 Nuclease Mapping Analysis

S1 nuclease mapping was performed using a modification of the Berk and Sharp (21) technique. For this purpose a 880 bp Pst 1/Bgl1 fragment was derived from the first exon of the $MLC1_V$ gene (Fig 1). The probe was 5′ end-labelled at the Bgl1 site with (γ^{32}P) ATP and polynucleotide kinase and liberated from the clone by cutting with Pst1. The labelled end corresponds to amino acid 22 in the first exon and the probe extends towards the 5′ end beyond the promoter (TATA) region and includes 5′ upstream elements. The double-stranded probe was hybridized in DNA excess to 30 μg of total RNA from heart and skeletal muscle tissues. The hybridization was in 25 μl of 80% deionized formamide/10 mM pipes (Sigma) pH 6.4, 0.4M NaCl/0.05% SDS/1mM EDTA for 20 hrs. at 48°C. S1 nuclease digestion was in 300 μl for 1 hr at 25°C with 150 units of enzyme (Boehringer Mannheim) in 200 mM NaCl/30mM sodium acetate pH 4.5/3mM $ZnSO_4$. The reactions were terminated with 10mM EDTA and precipitated with ethanol. The dried pellets were resuspended in 85% formamide and electrophoresed on a 6.5% polyacrylamide/urea sequencing gel.

Primer Extension Analysis

An 18-mer synthetic oligonucleotide primer complementary to sequences located in the first exon position (+130 to +148) was used. The primer was 5′ end labelled to a specific activity of $1-3\times10^{-9}$ dpm/μg with (γ-^{32}P) ATP and T_4 polynucleotide kinase. One

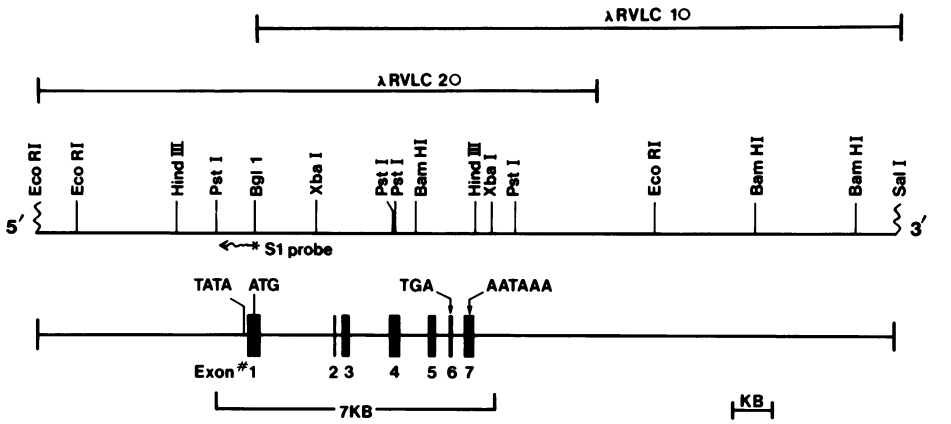

Figure 1. *Structure of rat MLC1$_V$ gene.* Partial endonuclease restriction map of the overlapping MLC1$_V$ genomic clones λRVLC 10, λRVLC 20 is shown in the 5′ to 3′ orientation. Cloning sites are indicated by (ζ). Diagram at the botton of the figure shows the organization of exons 1 to 7 in black boxes. The positions of TATA sequence, initiation codon ATG, stop codon TGA, and the polyadenylation signal (AATAAA) are marked. The genomic probe used for S1 nuclease mapping analysis (Pst1/Bgl1 fragment) is indicated by a horizontal arrow.

picomole of primer (1.2×10^7 dpm) was annealed with 30 μg total RNA in 30 μl of a solution containing 100 mM NaCl and 10 mM Tris/HCl, pH 7.4. Mixtures were heated to 75°C for 5 minutes and transferred to 37°C for 30 min. Twenty μl of reverse transcription buffer (250 mM Tris/HCl, pH 8.3, 250 mm KCl, 40 mM $MgCl_2$ and 40 mM DTT), 10 μl of deoxynucleotide mixture (10mM each) and 40 μl H_20 were then added. Synthesis of cDNA was started by the addition of 200 units of M-MLV reverse transcriptase (BRL Laboratories) and incubation was at 37°C for 1 hour. The reaction mixture was ethanol precipitated and resolved on a 6.5% polyacrylamide urea gel together with a sequencing ladder generated by the same primer using a M13 template containing the first exon and promoter sequences of the MLC1$_V$ gene.

RESULTS

Isolation of Rat Ventricular MLC1$_V$ cDNA Plasmid pRLC 87

We have previously described the isolation of a cDNA plasmid pRLC 267 encoding part of the rat MLC1$_V$ protein (17). The cDNA insert within this plasmid is relatively short (150 bp), and we have therefore isolated a second plasmid pRLC 87 using a λgt10 cDNA library from rat heart. Plasmid pRLC 87 was sequenced from both strands and found to contain an insert size of 831 bp. The cDNA sequence showed an open reading frame for 198 amino acids and a 237 nucleotide 3′ untranslated region with a canonical Poly (A) addition signal followed by a short Poly (A) tract (sequence data not shown). The cDNA did not contain the 5′ nontranslated region and the codons for the first two NH_2-terminal amino acids.

Isolation and characterization of the rat MLC1$_V$ gene

To determine the structural organization of the rat MLC1$_V$ gene and its evolutionary relationship with other members of the MLC1 multigene family, we have isolated the complete MLC1$_V$ gene. Plasmid pRLC 87 was used to screen two different rat genomic

```
ctgcagctgcctgctcagtccagacactcatacactagcacatacaaacacccagacataaatacacagag  -670
gccacatcacagacactcactcagacccgagcagacatacatgaatgggaacacctggaaatgggcccaca  -599
gagacaagacacagaacagtctcttcccctagacccaggcgggcactcacacaccctgcacggagacaacc  -528
agacgagacataccaatgtccagagacctcccacagtgacacacgccaaggacccctacttaaaaggagcc  -457
aggacacagatccacacatacgtccccttgctcttgatcaggggacctggaccacccaggtgcaaggactg  -386
tccttaggtacttctagccccaactatgatacatagcatggtacacacctttttttacctttacctgtctga  -315
ggctggcagagagattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgactgtctgtctgtc  -244
tgtgtgctcagagcccaggctgagcagtggcctctgtctccctctccctccctctctccaacaggtgactg  -173
ggaatgagcacactcccctcccccagcattccagggcctgaacaatgccctctccaggacccaaaatagcc  -102
                -85         -73
ccttgagcagagCCAATgccctttttatggccctgtccctattgtgcatggtaggggtggggctggggtcat  -31
       -22                    +1
gaggTATAT gagcaggatcaagcccagggt GGCAGTAAGGCTCACCTATCTCACTCTCTTCTCTCCTTC
TTCCGGCCTCTCTGTGATTACAGCTCCA ATG GCC CCC AAA AAG CCA GAG CCC AAG AAG GAC
                          aa[1]  M   A   P   K   K   P   E   P   K   K   D
                                                          primer 1
GAT GCC AAA ACC GCT GCC CCC AAA GCA GCT CCG GCC CCT GCG GCT GCA CCT GCG  EXON 1
 D   A   K   T   A   A   P   K   A   A   P   A   P   A   A   A   P   A
GCT GCA CCT GAG CCC GAA CGC CCT AAG GAA GCA GAG TTT GAT GCC TCC AAG ATT
 A   A   P   E   P   E   R   P   K   E   A   E   F   D   A   S   K   I
AAG gtgagtgtgggagctgagctgtagaagggacagagccaatgactgggtctcagagtgggaccaacca
 K [47]
tggcaggtgaggcctagacccgagctgaagctggcagtttctgtgcaagaggttaaacaagacctgtttgc
tttcggacatctgcgggaccaccgtgaagggat..~2.1kb..gtcttaagtagacagcataaggtcctgg
ggtggtagaggctgtgatcctcccatgagaacagctgacacagccagggtcctgagagctaggctgcacat
tgactcataattacttggtctttggtcttccttgcag ATT GAG TTC ACG CCT GAA CAG ATT    EXON 2
                                    [48] I   E   F   T   P   E   Q   I
GAA G gtgagcagggtcccgagctgagggatagaggtgggccaggggtctcaaatccaaatgagtgggg
 E  [56]
gtgggaccttggggagcatccagagtacactgtcctttcccacag AG TTC AAG GAG GCC TTC CAA
                                              [57] (E) F   K   E   A   F   Q
CTG TTT GAC CGC ACA CCT AAG GGC GAG ATG AAG ATC ACG TAC GGG CAG TGT GGG  EXON 3
 L   F   D   R   T   P   K   G   E   M   K   I   T   Y   G   Q   C   G
GAT GTC CTG CGG GCT CTG GGA CAG AAT CCT ACC CAG GCA GAG GTG CTC CGC GTC
 D   V   L   R   A   L   G   Q   N   P   T   Q   A   E   V   L   R   V
TTG GGG AAG CCC AAA CAG GAA G gtagtacatccggcccatttgcgccagcccaaggcagagtg
 L   G   K   P   K   Q   E [106]
ccatgtgggggcatcagatgtcatgggagggagactgccatggtgggggcatcagatgtcatgggaggg
agactgccatgtccggggcagtcagatgtcatgggaaccacactggacctcttctgattcctctcctgtt
gggatgaggatgtaacatggctata..~0.94kb..tggcaaaggtctatgtaaacccggtgtttctcccg
gtgtggtggcgtcatggtttgtacacagcacatccatgcacagtgctccagggaccatgggagaccaatgg
cccctcagggttaggcacaggggaagttgcacccatccattctctaacccaggacttgggaccatctgtct
gcag  AG CTC AAC TCC AAG ATG ATG GAT TTT GAA ACA TTC CTG CCC ATG CTG CAG
[107](E)  L   N   S   K   M   M   D   F   E   T   F   L   P   M   L   Q
CAC ATC TCC AAG AAC AAA GAC ACG GGC ACA TAT GAG GAC TTC GTG GAG GGG CTG  EXON 4
 H   I   S   K   N   K   D   T   G   T   Y   E   D   F   V   E   G   L
CGG GTC TTC GAC AAG GAG GGC AAC GGA ACT GTT ATG GGT GCA GAG CTC CGT CAT
 R   V   F   D   K   E   G   N   G   T   V   M   G   A   E   L   R   H
GTG CTG GCC ACG GTG G gtgaggacagcctcctcctctacccatctctacctctggggtgtcctag
 V   L   A   T   L [164]
ggaaaccctgtgtcacttggcc..~0.18kb..tctgtggccaaagaggggaaacactctcactctctgtg
ggccctgggtacatctcactgactccatcatcaggtggatctgatcacctacctcctccaagtgacttttc
ctggcagagagcatttattgagcgattgctgtgttcaaccctaccccggtttagggtttgcacacatgcct
gtatgtatgtgaagagccccaccatagctgggacttggccagtgaacagcacaactgtggcttgcctagca
ggcaggaagccctgggttggatcctcagcaccgtggcgcatgtgtttgggagatggaggtgggagaattag
aagttccaagatcgctcctagttatatagcccgtttgaggctagcccaggatataagaggtcctatcttaa
```

```
atataaaagagaaagaaaataaaaaagaaagaaggtttcagacgatagagcaggggaagtctctcagtgc
tttccacagtcatctgtcagggtaggtctcctcgagggccctctacagacctcaaaggg..~0.15kb..a
tggcaaactcaccacccagaaggattaaatagattccctccacttagtgagacttaacctcccaccggttg
ccag GT GAG AGG CTG ACA GAA GAC GAG GTA GAG AAA CTG ATG GCT GGT CAA GAA  EXON 5
[165](G) E   R   L   T   E   D   E   V   E   K   L   M   A   G   Q   E
GAC TCC AAT GGC TGC ATC AAC TAT GAA G gtgggaatcttctggggagaaggcaggggact
 D   S   N   G   C   I   N   Y   E [190]
cggctctgtgggggtgaagctccaggggaggggggacaa..~0.18kb..cccaactcagcaggtgggagt
ctcttggggaaaaggcacgggaccccattcgaggggtagggatcttccacccctggctgaacactgactc
ctcccatccttcctttccttttcag CA TTT GTG AAG CAC ATC ATG GCC AGC TGA GCGTC   EXON 6
                    [191] (A) F   V   K   H   I   M   A   S [199]
TCAG gtgagaaccccccactctttaaccttgttgctatgcccaccccatctcca..~0.2Kb..gcctctc
aggtgagaccccccactctttaaccttgttgctatgcccaccccatctccactcaccaactccggctttcat
ttggtgttctgctctcctgatccagtccttcctccgtaacctgaggctgcttttgcttcagGAAGCCCAGG
GCAGGCCAAGCTGGGGACATGCCATCTCCCAACCATAATGCTGACACCATGTCTGGAGCTTCGGGAAGGAA  EXON 7
AGGAGTGTCCCAGACTCCAGCACATGGTCGCAGATCACTCCGTGTGCGGTGGTCCTCCCTGGCCCCTGCTA
GGATGTGCTTGCCTACCCGTGGTACCCTTTCTGTCTCCACCCTGCGGCCTTATGAATAAATGATTTCCTTC
CTTCTGctctctctccacctgtccctctccccgccccttctctttctttccactcctttctggctcacctg
tccattcggtggcgttgagctagaggaagcccggctccccttcctttgttcttacc
```

Figure 2. *Nucleotide and derived amino acid sequence of the rat MC1$_V$ gene.* DNA sequences corresponding to exons are in capital letters; intron sequences are in lower case letters. The deduced amino acid sequences (aa1 − 199) are indicated below the nucleotide sequences. The numbers within the introns indicate the approximate distance between the sequences shown. Nucleotide position +1 is assigned to the transcription initiation site (cap site) as determined by S1 nuclease mapping and primer extension analysis. The putative CarG box element at position −85 (27) and the GT dinucleotide stretch at position −300 are also underlined.

libraries as described in METHODS. Out of 12 different positive plaques, two genomic clones, λRVLC 10 and λRVLC 20, were chosen for further characterization as these two clones showed hybridization to the 5′- and 3′- coding regions of the cDNA. The two genomic clones cover a total length of 21.5 kb genomic DNA which encompasses the total MLC1$_V$ gene (Fig. 1).

To determine the precise location of exons and introns, a combination of Southern blot, S1 nuclease mapping and DNA sequence analysis was performed on the genomic clones. Selected fragments which hybridized to cDNA clone pRLC 87 were sequenced according to Maxam−Gilbert (19) and by the dideoxy chain termination method (20) after subcloning the DNA into M13. As illustrated in Fig. 2, the sequences coding for the 199 amino acids and the 3′ and 5′ untranslated sequences are organized in 7 different exons which spans approximately 6.5 kB DNA. The most striking feature of the gene is the conservation of exon/intron organization in relation to other known vertebrate MLC1 genes. The organization of coding and noncoding regions and the location of introns within the coding regions of the MLC1$_V$ gene are virtually identical with the other MLC1 genes, from the rat MLC1$_f$/3$_f$ (5), mouse MLC1$_A$ (10), chicken MLC1$_V$ (22) and human MLC1$_V$ (23). The first exon in the rat MLC1$_V$ contains the complete 5′-untranslated region and codons for 47 NH_2-terminal amino acids. This exon differs from other MLC1 genes in the number of codons, 42 in MLC1$_V$ of human (23) and 41 in chicken MLC1$_V$ (22), 40 in MLC1$_A$ from mouse (10), 37 in MLC1$_f$ of chicken (6) and mouse (7), and 38/40 in MLC1$_f$ of rat (5). All the other amino acid coding exons in the above-mentioned genes are identical in size specifying exactly the same number of amino acids. The feature of a mini exon (28 nts) which encodes 9 amino acids (exon 2 in MLC1$_V$) is conserved in all the striated MLC1 genes. It is duplicated as two mini exons in MLC1$_f$/3$_f$ gene (exons

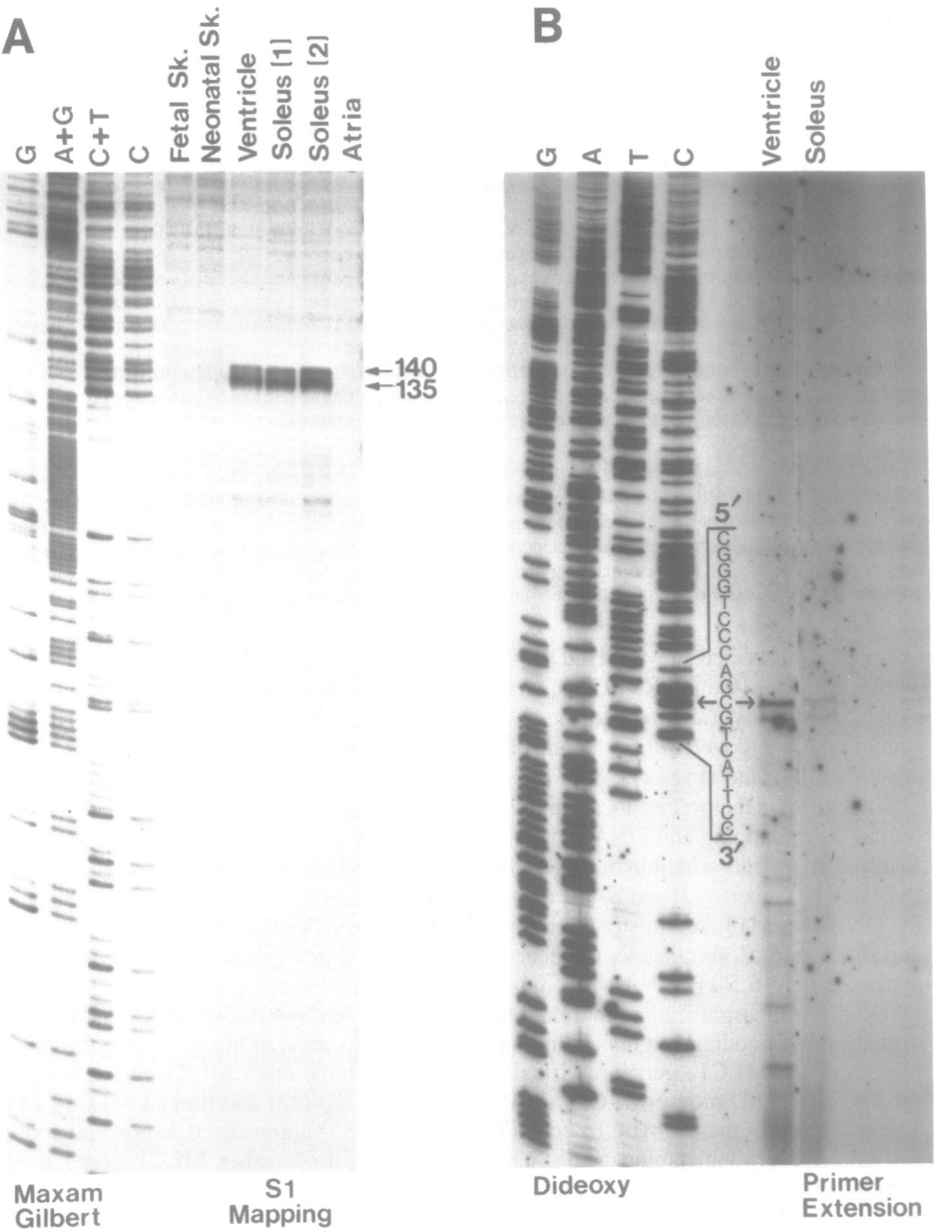

Figure 3. *S1 nuclease mapping and primer extension analysis.*
A) S1 nuclease mapping was performed using a double-stranded genomic probe (880 bp, Pst 1/Bgl 1 fragment) derived from the first exon and promoter region (see figure 1 for map). The probe was 5′ end-labelled at the Bgl 1 site, corresponding to amino acid 22. After hybridization with various RNA samples the S1 nuclease-digested product was analyzed together with a Maxam and Gilbert sequencing ladder of the same probe. A protected

3 & 4). All the exons that end in split codons, always at G (exons 2,3,4,5) are also conserved between the three MLC1 genes expressed in vertebrate striated muscle. Even the corresponding exons (exons 2,3,4) of *Drosophila melanogaster* MLC1 gene (24) end with split codons.

The 3′ untranslated region is characteristically split between the last two exons in all MLC1 genes (exons 6,7 in rat $MLC1_V$, human $MLC1_V$, chicken $MLC1_V$, and exons 8,9 in rat $MLC1_f/3_f$). Thus the striking homology found in the sequence organization of MLC1 genes suggests that the MLC1 isoforms have not only originated from the same ancestral gene but have undergone little change in their organization.

Identification of the Cap Site

The site of initiation of mRNA transcription was defined first by primer extension analysis using an oligonucleotide synthetic primer corresponding to nucleotides +130 to +148 bp located in the first exon (Fig. 2). Primer extension analysis using total RNA from rat ventricle and soleus are shown in Fig. 3B. A major extension product of 148 nucleotides was obtained with ventricular and soleus muscle RNAs. The sequencing ladder derived from the same primer using a M13 template containing the first exon and the promoter (TATA) region shows the genomic DNA sequence corresponding to the cap site. These experiments locate the transcription initiation site at a position 101 bp upstream of the AUG translation initiation codon. The cap site was further verified by S1 nuclease mapping, in which a probe derived from the middle of the first exon and extending towards the 5′ direction was used. When hybridized with RNA from ventricular and soleus muscle tissues, the probe yielded a protected fragment of ~140 bp, which corresponds to the same nucleotide position as observed by primer extension analysis (Fig. 3A). Thus, based on these results and on the sequence analysis of the genomic DNA, the $MLC1_V$ mRNA transcription initiation site has been assigned to a position 101 upstream from the translation initiation codon AUG.

Analysis of the $MLC1_V$ Promoter and Upstream Putative Regulatory Sequences

To identify the promoter and upstream regulatory elements which may be involved in the tissue specific regulation of the $MLC1_V$ gene and those which share homology with other MLC1 genes, we have sequenced the putative promoter and 740 nucleotides of the 5′ upstream region. The consensus TATA and a CCAAT box element are located at −22 and −84 bp respectively (Fig. 2). A characteristic feature of the rat $MLC1_V$ promoter is the presence of a GT dinucleotide stretch which is 40 bp long (starting at −302), a feature that might favor Z DNA conformation (25). Similar dinucleotide stretches are also found in the mouse $MLC1_V$ upstream sequences, but are located at −750 position (26). The rat and mouse $MLC1_V$ genes show considerable sequence homology in the 5′ flanking region with almost 100% conservation of sequence in the −1 to −150 segment near the promoter (TATA). In addition the rat $MLC1_V$ promoter contains one CArG box [cctttẗatgg] element located at position −80 as has been found in other MLC1 genes (26). The CArG elements are shown to be critical for cardiac actin gene expression and deletion of these elements suppresses promoter activity (27).

fragment of 135−140 bp long was seen only with RNA from ventricle and soleus (1=neonatal, 2=adult). B) Primer extension analysis was performed using a ^{32}P labelled synthetic oligonucleotide corresponding to position +130 to +148 in the first exon (See Fig. 2). The extended products were analyzed together with M13 sequencing reactions using the same primer on a template (2 kb) containing exon 1 and the entire 5′ promoter region.

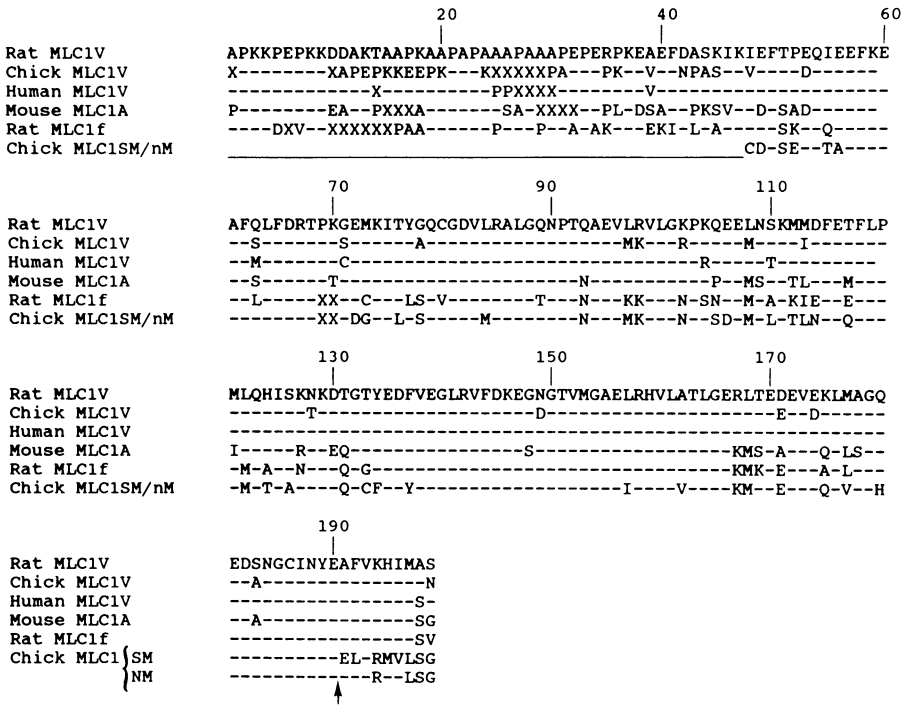

Figure 4. *Direct comparison of the complete amino acid sequence of the rat MLC1$_V$ with the chicken MLC1$_V$, human MLC1$_V$, mouse MLC1$_A$, rat MLC1$_f$, and chicken MLC SM/nM.* Standard one-letter amino acid symbols are used and numbered for the rat MLC1$_V$ sequence. Only nonidentical amino acids are shown, and homology with the rat MLC1$_V$ sequence is indicated by (−). For maximal alignment, gaps (XX) are introduced. Arrow indicates the point of divergence between smooth and non-muscle myosin alkali light chain (12).

Rat MLC1$_V$ Protein Shows a High Degree of Sequence Homology with other Alkali Light Chain Isoforms

Figure 4 shows a comparison of the amino acid sequences derived from the rat ventricular MLC1$_V$ gene with known sequences of human ventricular (11,12), chicken cardiac (22,33), mouse embryonic/atrial (10), rat fast skeletal (5), and chicken smooth/non-muscle (8). It can be seen that MLC1 isoforms expressed in different striated muscle fibers share a high degree of sequence homology. The human MLC1$_V$ protein has the highest sequence homology (96%) to the rat MLC1$_V$. The rat MLC1$_V$ protein shows 82% homology with the chicken MLC1$_V$ protein, 77% homology with the mouse MLC1$_A$ and 74% with the rat MLC1$_f$ protein sequence. This observed amino acid sequence conservation extends even between striated and smooth/non-muscle light chains. The smooth/non-muscle (sm/nm) alkali light chain isoform also shows ~70% sequence homology when compared with amino acids 48−199 of rat MLC1$_V$. There is, however, a significant degree of sequence divergence in the NH_2-termini (aa1−47 in MLC1$_V$) of all the MLC1 proteins (Fig. 1). The three MLC1 isoforms (MLC1$_f$, MLC1$_V$ and MLC1$_A$) expressed in striated muscle fibers also exhibit heterogeneity in the amino-terminus length with a high content of proline and alanine residues. The observed heterogeneity at the NH_2-terminus might have some

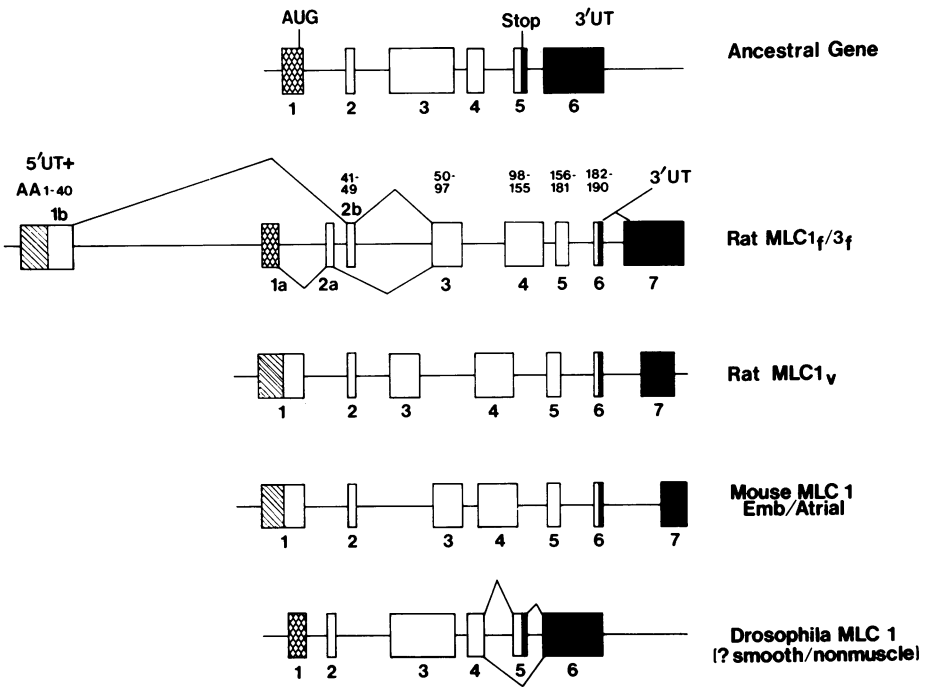

Figure 5. *Structural and proposed evolutionary relationship between various MLC1 genes.* The structure of the ancestral gene is derived from information based on the organization of Drosophila MLC gene. The coding exon sequences are shown in closed boxes, 3′ noncoding in black boxes, and 5′ noncoding sequences in stripes or hatches. The size and position of exons are approximate and not drawn to scale. Amino acid numbering is only given for $MLC1_f/3_f$ gene and the number of amino acids coded by the first exon varies from gene to gene. The differential splicing pathways for smooth/non-muscle and Drosophila genes are indicated by joining the appropriate exons.

functional relevance, since the NH_2-terminus peptide of MLC1 has been shown to interact with actin and is implicated in modulating acto-myosin interactions (2,3).

The $MLC1_V$ mRNA is Expressed in Ventricular and Slow Twitch Muscle Fibers and Transcribed from the Same Promoter Region

Recent studies indicate that the MLC1 isoform expressed in the ventricular tissue and slow twitch skeletal muscle fibers were probably the same (13,14). However, in these studies, only northern blot analysis was used to confirm their identity. In order to unequivocally demonstrate that the $MLC1_V$ mRNA expressed in these two muscle tissues is identical and are transcribed from the same promoter region, we performed S1 nuclease mapping analysis (Fig. 3A), using a 5′ end-labelled DNA probe (a 880 bp, Pst 1/Bgl 1 fragment) derived from the $MLC1_V$ genomic clone (Fig. 1). This fragment contains codons for the first 22 N-terminal amino acids, the entire 5′ untranslated region and the putative promoter TATA and upstream sequences (−750 bp). Thus, the probe served the dual purpose of analyzing mRNA expression as well as mapping of the transcription initiation site in the tissues examined. The S1 nuclease mapping analysis with total RNA from both skeletal and cardiac tissues, showed that the $MLC1_V$ mRNA expression is restricted to the ventricle and slow twitch muscle fibers as revealed by the protected fragment of size between

135−140 nt. The $MLC1_V$ mRNA is not expressed in the fast skeletal and atrial muscle (Fig. 3A). Further, the data shows that both ventricular and soleus muscle RNA produce an identical pattern of S1 nuclease protection. The heterogeneity in the protected fragment (135−140 nt) is probably due to an S1 nuclease digestion artifact or due to mRNA that is partially degraded at its 5′ termini. These S1 mapping analyses taken together with primer extension data (Fig. 3B) indicates that the $MLC1_V$ mRNA is transcribed from the same promoter and transcription initiation site in both ventricular and slow twitch skeletal muscle fibers.

DISCUSSION

The $MLC1_V$ Gene Belongs to a Conserved Multigene Family

We have isolated and characterized the complete gene encoding the myosin alkali light chain expressed in rat ventricle and slow-twitch skeletal muscle fibers. The $MLC1_V$ gene is relatively small when compared to the $MLC1_f/3_f$ gene (~22 KB long) and shows a simple organization. The entire gene is 6.5 kb long and the mRNA coding sequences are organized in 7 different exons. The organization of exons in the $MLC1_V$ gene is very similar to the recently characterized MLC1 genes from mouse, chicken and human, with a remarkable conservation in exon size, number, exon/intron junctions and coding capacity. One notable difference between the $MLC1_V$ and the $MLC1_f/3_f$ genes is in the size of the first intron, which in the $MLC1_V$ gene is ~2.5 kb whereas it is over ~10 kb in $MLC1_f/3_f$ genes of rat, mouse and chicken. It is possible that a strict conservation of intron locations and their sizes is essential for $MLC1_f/3_f$ genes since they contain two differentially regulated promoters and give rise to two differentially spliced RNA products. However, this large intron has been eliminated with the evolution of $MLC1_V$ like genes and only a single promoter is present (Fig. 5). Despite the highly conserved nature of the overall structure and organization of the MLC1 genes, the sequences found at the promoter and 5′ and 3′ nontranslated regions are gene specific. The ventricular $MLC1_V$ gene promoter sequence shows little homology with other MLC1 genes, including the atrial $MLC1_A$ which is expressed transiently in the fetal ventricular muscle. A consensus sequence element (CCTTTTATAG) present in the alkali MLC1 genes (26) which is conserved in sequence and position, is also present at −85 bp in the rat $MLC1_V$ gene. This sequence CCTTTTATGG also matches the 'CArG' box element identified originally in the cardiac alpha-actin gene promoter sequence. This sequence has been found to bind nuclear protein factors and positively regulate transcription (27,28). In addition, the $MLC1_V$ gene described here exhibits a long stretch of alternating pyrimidine/purine (20 GT) nucleotides, a feature favoring Z DNA formation that may be important for gene regulation (25). Recent studies have indicated that Z DNA forming sequences cause elevated recombination frequencies and in addition may play an important role in recombination between members of gene families (29−31). The exact functional role of these sequences, however, remains to be elucidated.

The Evolution of Vertebrate Alkali Myosin Light Chain Genes

The structure of the three alkali light chain genes ($MLC1_f$, $MLC1_A$, $MLC1_V$) expressed in vertebrate striated muscle has already been determined. Similarily, the primary structure of alkali light chains expressed in vertebrate smooth and non-muscle tissues has been determined (8). The smooth/non-muscle alkali MLC isoforms are structurally similar to invertebrate alkali light chains (24,32) since they do not contain the region corresponding to amino acid 1−40 found in most MLC1 proteins. Although the structure of the

smooth/non-muscle gene is not completely known, the preservation of alternate splicing at the carboxy terminus indicates that this gene might be structurally similar to the Drosophila (alkali) MLC gene (24). Based on the available structural information of the alkali MLC_1 genes in invertebrates and in vertebrate smooth/non-muscle tissues, we propose that the ancestral type MLC gene might resemble the Drosophila MLC gene in organization (Fig. 5). However, the evolution of alternate splicing, a feature found in Drosophila and smooth/non-muscle alkali light chain genes, might have been added during metazoan evolution, a feature no longer preserved in striated MLC1 genes.

It is likely that the first striated MLC1 gene to be evolved is the $MLC1_f/3_f$ gene as it embodies the prototype ancestral gene and the present day MLC1 gene (Fig. 5). This gene probably arose through duplication of the mini exon region (exon 2 in the ancestral gene) and the addition of a far upstream promoter and a newly acquired first exon. However, this newly evolved gene retains the capacity to function from the original promoter ($MLC3_f$ promoter) and produces a protein with a short amino terminus. The $MLC1_V$ gene described here and the $MLC1_A$ gene share the same archetypal features with the $MLC1_f/3_f$ gene, but they lack the $MLC3_f$ specific exons 2 and 3. Thus, the complex $MLC1_f/3_f$ gene might have undergone duplication and further rearrangement in which the $MLC3_f$ promoter and its specific exons 2 and 3 were deleted. However, this rearrangement requires a major deletion in the first intron and the precise retention of the cognate mini exon (exon 4 in $MLC1_f/3_f$ gene) corresponding to the $MLC1_f$ first exon. The $MLC1_V$ and $MLC1_A$ genes probably represent the most recently evolved MLC1 genes. Furthermore, the remarkable conservation in the structure and organization of these two genes suggest that they probably arose by a single duplication event and have acquired distinct regulatory elements, with different patterns of developmental and tissue specific expression.

ACKNOWLEDGEMENTS

We would like to thank Drs. C. Saidapet and P. Delaney for their participation in cDNA cloning and C. Kelly for DNA sequence analysis. We thank Dr. P. Babij for comments on the manuscript and J. Lovelette for expert secretarial assistance. This work was supported by a grant from NIH (P01 HL 28001-01) to M.P. M. Periasamy is an established investigator of the American Heart Association.

*To whom correspondence should be addressed

REFERENCES

1. Henry, G.D., Winstanely, M.A., Dalgarno, D.C., Scott, G.M., Levine, B.A. and Trayer, IP (1985) Biochem. Biophys. Acta 830, 233–243.
2. Trayer, H.P. and Trayer, I.P. (1985) FEBS Letters 180, 170–174.
3. Labbe J.P., Audemard, E., Bertrand, R. and Kassab, R. (1986) Biochemistry 25, 8325–8330.
4. Barton, P.J.R. and Buckingham, M. (1985) Biochem. J. 231, 249–261.
5. Periasamy, M., Strehler, E.E., Garfinkel, L., Gubits, R.M., Ruiz-Opazo, N. and Nadal-Ginard, B. (1984) J. Biol. Chem. 259, 13595–13604.
6. Nabeshima, Y., Fujii-Kuriyama, Y., Muramatsu, M. and Ogata, K. (1984) Nature 308, 333–338.
7. Robert, B., Daubas, P., Akimenko, M.A., Cohen, A., Garner, I., Guenet J.L. and Buckingham, M. (1984). Cell 39, 129–140.
8. Nabeshima, Y., Nabeshima, Y., Nonomura Y. and Fujii-Kuriyama, Y. (1987) J. Biol. Chem. 262, 10608–10612.
9. Barton, P.J.R., Robert, B., Fiszman, M.Y., Leader, D.P. and Buckingham, M.E. (1985b) J. Muscle Res. Cell Motil. 6, 461–475.
10. Barton, P.J.R., Robert, B., Cohen, A., Garner, I., Sassoon, D., Weydert, A. and Buckingham, M.E. (1988)

J. Biol. Chem. 263, 12669–12676.
11. Arnold, H.H., Lohse, P., Seidel, U. and Bober, E. (1988) Eur. J. Biochem. 178, 53–60.
12. Kurbayashi, M., Komuro, I., Tsuchimaki, H., Takaku, F. and Yazaki, Y. (1988) J. Biol. Chem 263, 13930–13936.
13. Barton, P.J.R., Cohen, A., Robert, B., Fiszman, M.Y., Bonhomme, F., Guenet, J.L., Leader D.P. and Buckingham, M.E. (1985a) J. Biol. Chem. 260, 8578–8584.
14. McNally, E.M., Buttrick, P.M. and Leinwand, L.A. (1989) Nucleic Acids. Res. 17, 2753–2767.
15. Gubler, U. and Hoffman, B.J. (1983) Gene 25:263–269.
16. Huynh, T.V., Young, R.A. and Davis, R.W. (1985) In Glover, D.M. (ed) *DNA Cloning–A Practical Approach*, IRL Press, Washington D.C., Vol I, pp 49–78.
17. Kumar, C., Saidapet, C., Delaney, P., Mendola, Ch. and Siddiqui, MAQ (1988) Circ. Res. 62, 1093–1097.
18. Blattner, F.R., Blechl, A.E., Dennison-Thompson, K., Faber, H.E., Richards, J.E., Slightom, J.L., Tucker, P.W. and Smithies, O. (1978) Science 202, 1279–1283.
19. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
20. Sanger, F., Nichlen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74:5463–5467.
21. Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721–732.
22. Nabeshima, Y., Nabeshima, Y., Kawashima, M., Nakamura S. and Fuji-Kuriyama Y. (1988) J. Mol. Biol. 204, 497–505.
23. Fodor, W.L., Darras, B., Seharaseyon, J., Falkenthal, S., Francke, U. and Vanin, E.F. (1989) J. Biol. Chem. 264, 2143–2149.
24. Falkenthal, S., Parker, V.P., and Davidson, N. (1985). Proc. Natl. Acad. Sci. USA 82, 449–453.
25. Rich, A., Nordheim A. and Wang, A.H.J. (1984) Ann. Rev. Biochem. 53, 791–846.
26. Cohen, A., Barton, P.J.R., Garner, I., Alonso S. and Buckingham, M.E. (1988) Nucleid Acid. Res. 16, 10037–10052.
27. Miwa, T. and Kedes, L. (1987) Mol. Cell. Biol. 7, 2803–2813.
28. Boxer, L.M., Prywes, R., Roeder, R.G., and Kedes, L. (1989) Mol. and Cell. Biol. 9, 515–522.
29. Flanagan, J.G., Lefran, C.M.P. and Rabbits, T.H. (1984) Cell. 36, 681–688.
30. Stringer, J. (1985) Mol. Cell. Biol. 5, 1247–1259.
31. Hellman, L., Steen, M.L., Sundvall, M. and Pettersson, V. (1988) Gene 68, 93–100.
32. Goodwin, E., Szent-Gyorgy, A.G. and Leinwand, L.A. (1987) J. Biol. Chem. 262, 11–52–11056.
33. Maita T., Umegane, T., Kato, Y., and Matsuda, G. (1980) Eur. J. Biochem. 107, 565–575.

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