MOLECULAR AND CELLULAR BIOLOGY, Dec. 1988, p. 5339–5349
0270-7306/88/125339-11$02.00/0

Vol. 8, No. 12

Regulatory Myosin Light-Chain Genes of *Caenorhabditis elegans*

CLAUDIA CUMMINS AND PHILIP ANDERSON*

Department of Genetics, University of Wisconsin, 445 Henry Mall, Madison, Wisconsin 53706

Received 1 June 1988/Accepted 19 September 1988

We have cloned and analyzed the *Caenorhabditis elegans* regulatory myosin light-chain genes. *C. elegans* contains two such genes, which we have designated *mlc-1* and *mlc-2*. The two genes are separated by 2.6 kilobases and are divergently transcribed. We determined the complete nucleotide sequences of both *mlc-1* and *mlc-2*. A single, conservative amino acid substitution distinguishes the sequences of the two proteins. The *C. elegans* proteins are strongly homologous to regulatory myosin light chains of *Drosophila melanogaster* and vertebrates and weakly homologous to a superfamily of eucaryotic calcium-binding proteins. Both *mlc-1* and *mlc-2* encode abundant mRNAs. We mapped the 5′ termini of these transcripts by using primer extension sequencing of mRNA templates. *mlc-1* mRNAs initiate within conserved hexanucleotides at two different positions, located at −28 and −38 relative to the start of translation. The 5′ terminus of *mlc-2* mRNA is not encoded in the 4.8-kilobase genomic region upstream of *mlc-2*. Rather, *mlc-2* mRNA contains at its 5′ end a short, untranslated leader sequence that is identical to the *trans*-spliced leader sequence of three *C. elegans* actin genes.

Myosin is a ubiquitous eucaryotic protein that is found in both muscle and nonmuscle cells. In muscle cells, it is the principal structural component of thick myofilaments. Muscle contraction occurs when myosin-containing thick filaments slide past actin-containing thin filaments (29). This process is coupled to the hydrolysis of ATP and is regulated by the intracellular concentration of free calcium (for reviews, see references 9, 13, 21, 26). Native myosin is composed of two molecules of myosin heavy chains (MHCs) and two pairs of myosin light chains (MLCs). Myosin exhibits two distinct structural and functional domains, a rodlike tail and a globular head region. Heavy-chain rod segments align with each other along the length of thick filaments and are responsible for the precise packing of myosin in thick filaments. Globular heads protude from thick filaments, interact with actin, and hydrolyze ATP. MLCs are located within the globular heads, each head containing one molecule of regulatory and one molecule of alkali (also called essential) MLCs.

Regulation of the myosin ATPase by Ca^{2+} occurs by both myosin-based and actin-based control systems. Myosin-based control involves functions of the regulatory MLCs (for a review, see reference 1). For example, in vertebrate smooth muscle and in nonmuscle cells, phosphorylation of regulatory MLCs by a Ca^{2+}-dependent protein kinase allows activation of the myosin ATPase by actin. In molluscan muscle, direct binding of Ca^{2+} by the regulatory MLCs promotes actin activation of the myosin ATPase. In vertebrate skeletal and cardiac muscle, calcium regulation is actin based. Binding of Ca^{2+} to a troponin-tropomyosin complex induces a conformational change in thin-filament structure, such that the myosin ATPase is activated by actin. The regulatory MLCs of skeletal and cardiac muscle are phosphorylated, but this appears to play a minor, modulatory role in the contractile cycle (35, 50). The muscles of many invertebrates, including *C. elegans*, exhibit elements of both myosin-based and actin-based control (23, 36).

Genetic manipulation of MLC proteins will likely be required for a complete understanding of their in vivo functions. MLC genes from several genetically manipulable organisms, including *Drosophila melanogaster*, *Dictyostelium discoideum*, and the mouse, have been described (7, 16, 48, 52, 59). Such studies should lead to in vivo analyses of MLC function that will complement investigations in which mutant proteins are engineered by site-directed mutagenesis and are analyzed in vitro (51). We have begun to investigate the MLC genes of the nematode *C. elegans*. Genetic analysis of *C. elegans* is well developed (6, 25), and reliable methods for DNA transformation are now available (17). *C. elegans* is especially well suited to genetic analysis of muscle. Because the animal reproduces by self-fertilization, mutants having even severe muscle defects are, nevertheless, viable and fertile. Approximately two dozen *C. elegans* genes that are necessary for the synthesis, assembly, or function of muscle have been defined to date (for a review, see reference 63). The protein products of several of these genes are known, including actins, MHCs, and paramyosin (15, 34, 41, 42, 62, 64). The in vivo relationships and functions of these proteins are steadily being revealed by a combination of genetic, molecular, and cell biological analyses. Thus, *C. elegans* should be a favorable organism for the in vivo analysis of MLC function.

Much is known concerning the genes, subunit associations, tissue distributions, and functions of *C. elegans* MHC proteins (for a review, see reference 31), but relatively little is known about the MLCs or the genes that encode them. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis resolves two *C. elegans* MLC proteins (M_rs approximately 16,000 and 18,000) (22), but the degree of heterogeneity within each of these isoforms is currently unknown. In this paper we describe the family of *C. elegans* regulatory MLC genes, their complete DNA sequences, and an analysis of the mRNAs expressed from these genes.

MATERIALS AND METHODS

DNA isolation and filter hybridizations. Procedures for *C. elegans* growth, DNA extraction, and filter transfer hybridization by using nick-translated probes have been described previously (12, 57). Filter hybridizations using end-labeled oligonucleotide probes were performed as described by

* Corresponding author.

Meinkoth and Wahl (40). We calculated the hybridization temperature (T_H) for oligonucleotide probes as follows (40): T_H = {[2°C(no. of A · T base pairs)] + [4°C(no. of G · C base pairs)]} − 5°C. Oligonucleotides were end labeled by incorporation of [γ-^{32}P]ATP with T4 polynucleotide kinase (38). Plasmid DNA was prepared by the method of Hicks and Fink (27).

Construction and screening of a *C. elegans* genomic DNA library. *C. elegans* DNA from the wild-type Bristol (N2) strain was partially digested with *Mbo*I. The digestion products were size fractionated on NaCl gradients (55), and 15- to 20-kilobase fragments were collected. These fragments were ligated to *Bam*HI-digested lambda-EMBL3 DNA (18), and the products were packaged into virions (28). The genomic library was screened for phages that hybridized to clone pCDM20, a *D. melanogaster* regulatory MLC cDNA clone (48). We gel purified, radiolabeled, and used as a probe a 600-base-pair restriction fragment of pCDM20 that is derived exclusively from the cDNA insert. Plaque hybridizations were similar to those of Benton and Davis (5), except that hybridizations and washes were performed under conditions of reduced stringency (63°C; 5× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate], 10 mM Tris hydrochloride pH 7.5, 10× Denhardt solution, 100 mg of *Escherichia coli* carrier DNA per ml, 0.1% sodium dodecyl sulfate).

DNA sequence analysis. In preparation for DNA sequence analysis, restriction fragments were subcloned into M13 vectors (61, 66). We used the 3′ to 5′ exonuclease activity of T4 DNA polymerase to generate a series of overlapping deletions of these M13 clones (10) and determined their DNA sequences by the method of Sanger et al. (54). We used computer programs available from DNAStar, Inc., Madison, Wis., for storage and analysis of DNA and protein sequences. This software system provided access to release 54 of the GenBank Genetic Sequence Data Bank and release 15 of the National Biomedical Research Foundation Protein Identification Resource Database.

RNA methods. To obtain RNA samples from defined postembryonic stages, an unsynchronized culture of nematodes was separated according to size with Nitex nylon mesh filters (C.B.S. Scientific Co., Del Mar, Calif.) of 6-, 10-, 20-, 30, and 41-micron pore sizes. We inspected the animals retained on each filter and used their sizes to estimate their ages. We prepared RNA by dissolving nematodes in a urea buffer (7 M urea, 2% sodium dodecyl sulfate, 0.35 M NaCl, 0.001 M EDTA, and 0.01 M Tris hydrochloride [pH 8.0]) as described by Ross (53) with the following modifications. Vanadyl-ribonucleoside complexes (10 mM) were added to inhibit RNases (38), and 0.1% β-mercaptoethanol was added to dissolve the worm cuticles. Total cellular RNA was prepared from these lysates by standard procedures (38). For Northern (RNA) transfers, the RNA was electrophoresed through 1.4% agarose gels in 0.01 M NaH_2PO_4 (pH 7.0) after denaturation of the RNA with glyoxal and dimethylsulfoxide (38).

Primer extension sequencing. End-labeled oligonucleotides (20-mers; synthesized at the Beckman Research Institute of the City of Hope, Duarte, Calif.) were annealed to total RNA and extended with reverse transcriptase by using the method of Zaug et al. (67) as modified by Krause and Hirsh (32; M. Krause, personal communication). The oligonucleotide primers were annealed to the RNA by boiling for 3 min and slow cooling to 50°C. Extension reactions were carried out with avian myeloblastosis virus reverse transcriptase (Life Sciences, Inc., St. Petersburg, Fla.) at 50°C for 60 min. Extension products were analyzed on 8% polyacrylamide (8 M urea) sequencing gels.

RESULTS

Isolation of *C. elegans* regulatory MLC genomic clones. We identified the *C. elegans* regulatory MLC genes by hybridization to a muscle regulatory MLC gene probe of *D. melanogaster*. Genomic Southern blots indicated that pCDM20, a cDNA clone of the single *Drosophila* regulatory MLC gene (48), hybridized to *C. elegans* DNA under conditions of reduced stringency. We prepared a library of *C. elegans* genomic DNA inserted into lambda-EMBL3 and screened it for bacteriophages that hybridized to pCDM20 under the same conditions (see Materials and Methods). We screened 10 genome equivalents and isolated 23 hybridizing clones. The restriction maps of these clones defined three nonoverlapping chromosomal regions. Further mapping of the lambda clones, coupled with Southern hybridizations using the *Drosophila* probe, identified restriction fragments from each region that were responsible for hybridization to the *Drosophila* probe. Surprisingly, each of the three chromosomal regions contained two noncontiguous sites of hybridization. Thus, six different regions of the *C. elegans* genome hybridized to the *Drosophila* regulatory MLC probe.

To distinguish bona fide MLC genes from pseudogenes or sites of adventitious hybridization to the *Drosophila* probe, we sequenced each restriction fragment that hybridized to the *Drosophila* probe. We analyzed these DNA sequences for the presence of open reading frames and for any relationship of potential proteins to known MLC sequences. Four of the six hybridizing sites (representing 16 of 23 hybridizing lambda clones) exhibited either no open reading frames or no significant homology to MLC proteins. We have not analyzed these clones further. The remaining two sites (representing seven overlapping, hybridizing clones) defined the genes *mlc-1* and *mlc-2*. Their organization and DNA sequences are described below.

DNA sequences of *mlc-1* and *mlc-2*. We determined the complete nucleotide sequences of *mlc-1* and *mlc-2*. These genes are separated by 2.6 kilobases and are divergently transcribed. Figure 1 shows a restriction map of these genes and summarizes their structures. The DNA sequences of *mlc-1* and *mlc-2* are presented in Fig. 2 and 3, respectively. We sequenced both strands of all sequences shown. By inspecting the DNA sequences of *mlc-1* and *mlc-2*, we deduced the sizes and positions of their introns. Recognition of *C. elegans* introns is usually unambiguous, because most of them are very small (approximately 50 base pairs) and their splice sites are highly conserved (56; nematode information compiled by T. Blumenthal, personal communication). MLC proteins are strongly conserved during evolution, and we used this conservation of sequence to help identify MLC proteins and introns within our sequences. The introns of *mlc-1* and *mlc-2* are shown in Fig. 1 to 4. *mlc-1* and *mlc-2* each contain three small introns at equivalent positions in the protein; the introns range in size from 47 to 54 base pairs. *mlc-1* contains a fourth intron that is not present in *mlc-2*. This intron separates the ATG translational initiation codon from the remainder of the *mlc-1*-coding sequence. This feature is common to MLC genes of several organisms (44, 46, 48, 49, 52, 65; see Fig. 4).

Both *mlc-1* and *mlc-2* contain putative polyadenylation signals (AATAAA sequences) at 602 and 90 base pairs, respectively, downstream of their translational termination

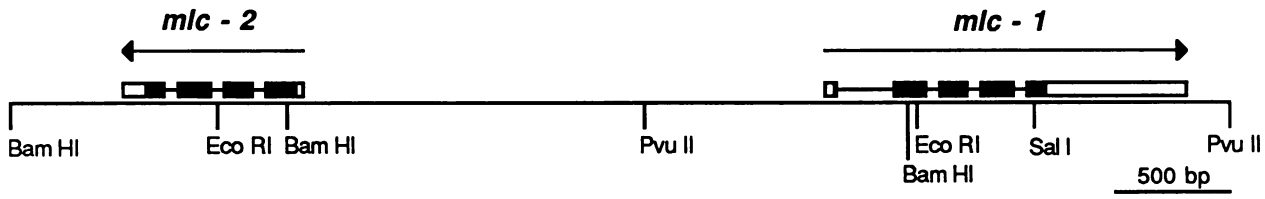

FIG. 1. Genomic organization of *mlc-1* and *mlc-2*. Structures of *mlc-1* and *mlc-2* are shown above a restriction map of the region. Boxed regions are exons. Protein-coding portions of the genes are closed; 5′ and 3′ untranslated regions are open. Arrows indicate the direction of transcription for each gene. bp, Base pairs.

codons. These sequences are likely to be the true polyadenylation signals, because the sizes of mRNAs that would result from their utilization are close to the sizes that we measured on Northern blots (see below).

The protein-coding regions of *mlc-1* and *mlc-2* are closely related to each other (96% nucleotide sequence identity; 99% amino acid sequence identity). The homology between *mlc-1* and *mlc-2* does not extend into their 5′ untranslated regions and extends for only 5 nucleotides 3′ of the translational termination codons. Introns of the genes are less conserved (57 to 87% nucleotide sequence identity when equivalent introns are compared). The inferred amino acid sequences of *mlc-1* and *mlc-2* are shown in Fig. 4; they are identical except for a single, conservative amino acid substitution (Ile to Val) at position 99 of the protein sequence.

We conclude that *mlc-1* and *mlc-2* are regulatory MLC genes because of their high degree of sequence homology to *Drosophila* regulatory MLC2 (48) (40% amino acid identity), rat skeletal MLC2 (46) (48% identity), chicken skeletal L-2 (39) (48% identity), and other regulatory MLCs (39 to 48% identity). These relationships are shown in Fig. 4. Regulatory MLCs are members of the troponin C superfamily of calcium-binding proteins (3). As expected, the protein sequences of *mlc-1* and *mlc-2* exhibit reduced, but significant, homology to a variety of calcium-binding proteins from other organisms, including alkali MLCs (23 to 27% amino acid identity), calmodulin (32 to 35% identity), troponin C (26 to 27% identity), and parvalbumin (19 to 31% identity).

Copy number of *C. elegans* regulatory MLC genes. We believe that *mlc-1* and *mlc-2* constitute the complete family of regulatory MLC genes. *mlc-1* and *mlc-2* hybridized strongly to each other, even under conditions of high stringency. *mlc-1* and *mlc-2* did not hybridize to any other *C. elegans* genomic sequences, even under conditions of reduced stringency (Fig. 5). The probe designated TR#35 in Fig. 5 includes most of the *mlc-2* gene and some 3′-flanking sequence; it does not contain sequences from *mlc-1*. The probe designated TR#36 includes most of the *mlc-1*-coding sequence, but none from *mlc-2*. Both of these probes yielded an identical pattern of hybridizing restriction fragments on genomic Southern blots. The reduced stringency conditions used for these experiments were those that allowed hybridization between the *Drosophila* and *C. elegans* MLC genes.

***mlc-1* and *mlc-2* transcripts.** *mlc-1* and *mlc-2* each express an abundant mRNA. Figure 6B shows a Northern blot in which we prepared RNAs from defined postembryonic stages and hybridized them with a probe (clone TR#34) that detects both *mlc-1* and *mlc-2*. We detected two abundant mRNAs of approximately 1,150 and 680 bases. We observed no major differences in expression of these mRNAs from the L1 larval stage through adulthood. On the basis of the DNA sequences of the genes, *mlc-1* should encode the larger transcript and *mlc-2* should encode the smaller one.

To demonstrate this relationship directly, we repeated the Northern blots using gene-specific hybridization probes. Probe TR#38 is a clone that contains approximately 500 bases of homology to the 3′ untranslated region of *mlc-1*. Oligo-Mlc1.1 and oligo-Mlc2.1 are synthetic oligonucleotides that hybridize only to *mlc-1* and *mlc-2*, respectively (see legend to Fig. 6). We confirmed that these oligonucleotides are specific for their respective MLC genes by using them as hybridization probes on Southern blots of appropriate clones. When used as probes on Northern blots, the oligonucleotides generated weak but reproducible hybridization signals. TR#38 and oligo-Mlc1.1 hybridized to the larger mRNA, while oligo-Mlc2.1 hybridized to the smaller mRNA (Fig. 6C). We conclude that *mlc-1* and *mlc-2* each express a single-size class of mRNA.

We have considered the possibility that the ATG sequence beginning at position 1272 of *mlc-1* (located within the first intron; Fig. 2) represents a translational initiation codon for a second *mlc-1* mRNA. Use of a different *mlc-1* promoter or alternative processing of a *mlc-1* precursor RNA could produce an mRNA that is translated using this ATG sequence. To test this possibility, we used an oligonucleotide, designated oligo-Mlc1.2, that is complementary to the sequence from positions 1260 to 1279 in Fig. 2. Utilization of the ATG at position 1272 would require a mRNA that hybridizes to this oligonucleotide. Oligo-Mlc1.2 hybridized to *mlc-1* DNA on a Southern blot. When used as a probe on a Northern blot similar to that shown in Fig. 6C, oligo-Mlc1.2 did not hybridize to any RNAs. When used as a primer for mRNA sequencing (see below), oligo-Mlc1.2 was not extended by reverse transcriptase. *mlc-1* mRNAs that contain this region, therefore, are not detectable. We believe that the *mlc-1* mRNA diagrammed in Fig. 1 is the only one expressed from this gene.

5′ termini of *mlc-1* and *mlc-2*. Both *mlc-1* and *mlc-2* contain sequences upstream of their ATG initiator codons that closely resemble 3′ splice sites (Fig. 2 and 3; 47, 56). This suggested to us that these genes might contain introns within their 5′ untranslated regions. To determine the sites of MLC transcription initiation, we used gene-specific oligonucleotides as primers for sequencing the 5′ termini of *mlc-1* and *mlc-2* mRNAs. Our results are shown in Fig. 7.

mlc-1 does not contain an intron within its 5′ untranslated region. *mlc-1* exhibited two sites of transcription initiation, at positions −38 and −28 relative to the ATG translational initiator codon. Each of the initiation sites occurred within a short repeated sequence (5′-TCAGTT-3′), and at each of these sites, initiations occurred at both the C and A residues of the repeats. Like certain other eucaryotic genes, *mlc-1* does not contain a recognizable TATA box (4, 20) or CCAAT sequence (11) positioned appropriately upstream of its transcription initiation site.

mlc-2 mRNA exhibited a very different structure at its 5′ end. The sequence of the 5′ terminus of *mlc-2* mRNA diverged from the *mlc-2* genomic sequence at position −22

FIG. 2. DNA sequence of *mlc-1*. The sequence of a 3.2-kilobase *Pvu*II-*Bgl*II restriction fragment is shown. We sequenced both DNA strands for the entire fragment. The *Pvu*II site at the beginning of this sequence is the same one that is at the beginning of the sequence in Fig. 3. The sequences in Fig. 2 and 3 together include the entire intergenic region between *mlc-1* and *mlc-2*. The predicted amino acid sequence is shown below the DNA sequence. ***, Stop codon. The probable poly(A) addition signal and two hexanucleotide repeats that include the sites of transcription initiation are underlined. Sequences and coordinates of two regions from the *unc-54* gene that are similar to regions in *mlc-1* are shown above the DNA sequence (see Discussion).

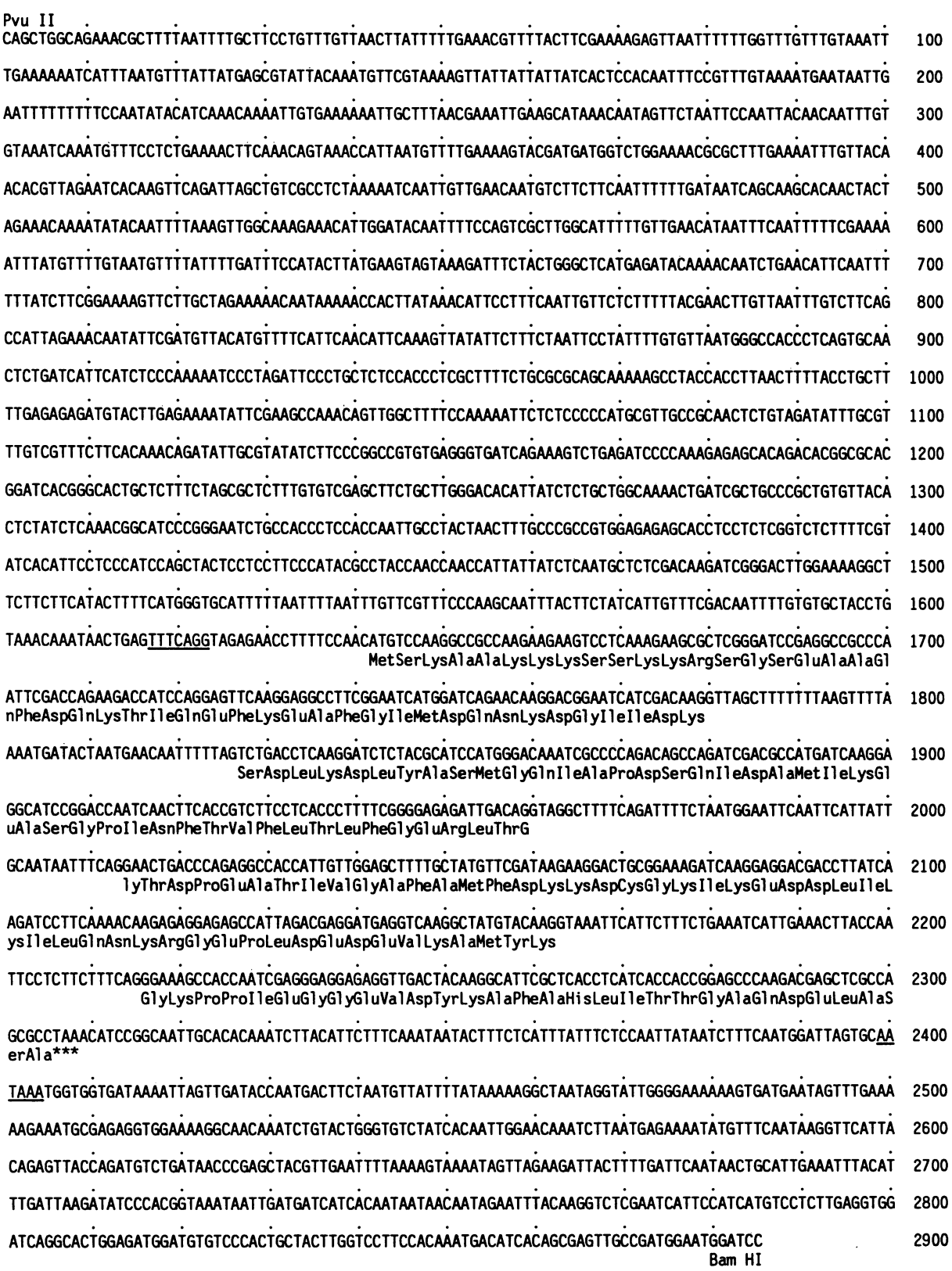

FIG. 3. DNA sequence of *mlc-2*. The sequence of a 2.9-kilobase *Pvu*II-*Bam*HI restriction fragment is shown. We sequenced both DNA strands for the entire fragment. The *Pvu*II site at the beginning of this sequence is the same one that is at the beginning of the sequence in Fig. 2. The sequences in Fig. 2 and 3 together include the entire intergenic region between *mlc-1* and *mlc-2*. The predicted amino acid sequence is shown below the DNA sequence. ***, Stop codon. The probable poly(A) addition signal and a 3′ splice site upstream of the ATG initiator codon are underlined.

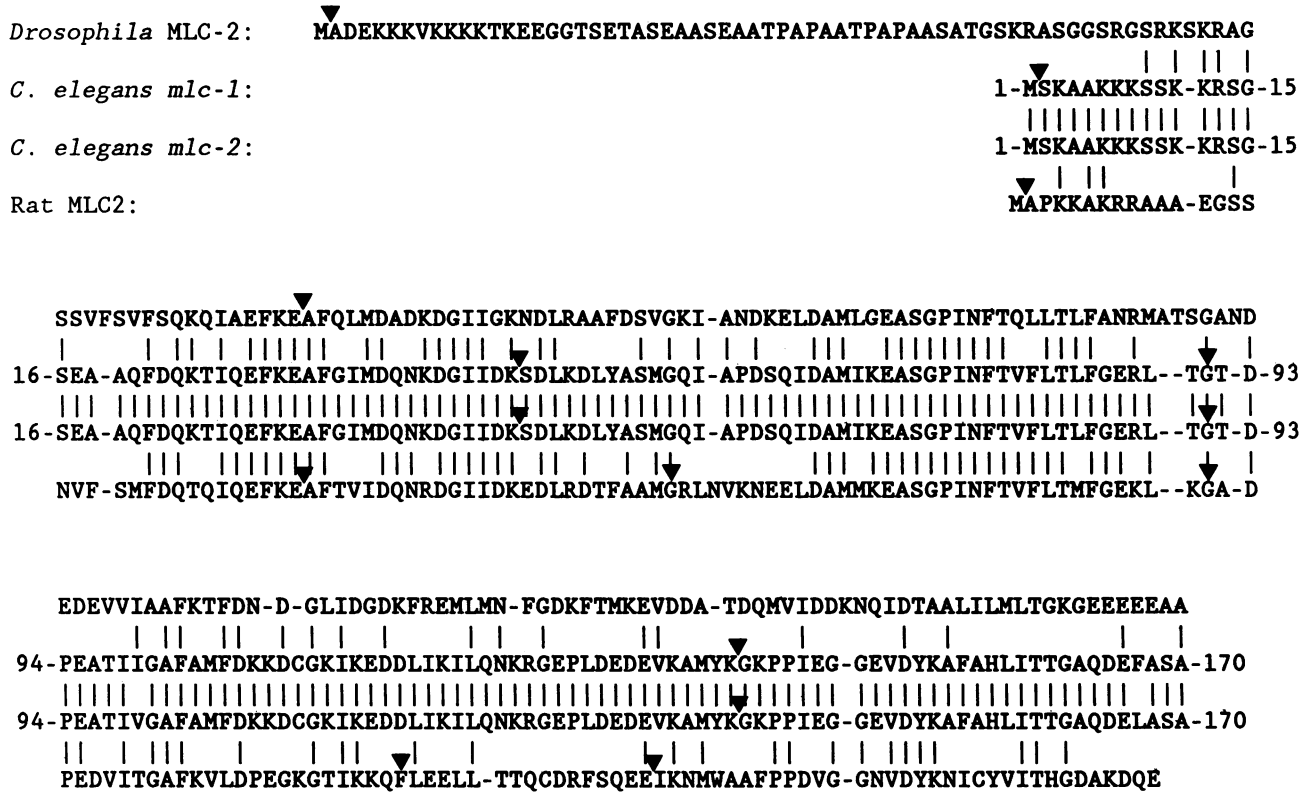

FIG. 4. Comparison of regulatory MLCs. *C. elegans mlc-1* and *mlc-2* are compared with each other and with *Drosophila* MLC-2 (48) and rat MLC2 (46). Arrowheads indicate the positions of introns relative to the amino acid sequences.

relative to the ATG initiator codon. In the genomic sequence, this position exactly coincides with a 3′ splice site consensus sequence. Our sequence of the *mlc-2* mRNA extended for 21 or 22 nucleotides beyond the splice site and then terminated. These 5′-terminal nucleotides are not present within the 4.8-kilobase region upstream of *mlc-2* that we have sequenced. However, the sequence is identical to the *trans*-spliced leader sequence that is found at the 5′ termini of three different *C. elegans* actin mRNAs (32).

DISCUSSION

We have identified two regulatory MLC genes in *C. elegans*. These genes, designated *mlc-1* and *mlc-2*, are separated by 2.6 kilobases and are divergently transcribed. The DNA sequences of both *mlc-1* and *mlc-2* predict proteins of about 18,600 daltons. This agrees well with the measured size (18,000 daltons) for the larger of two MLC size classes of proteins identified in *C. elegans* (22). The molecular sizes predicted from the DNA sequences may be overestimates, since certain MLCs are modified posttranslationally by removal of N-terminal amino acids (24, 65). The smaller of the two size classes of *C. elegans* MLC proteins (16,000 daltons; 22) is likely to be the alkali MLC(s).

C. elegans regulatory MLCs are strongly homologous to regulatory MLCs from other eucaryotes (Fig. 4) and weakly homologous to other members of the troponin C superfamily of calcium-binding proteins (3). This superfamily of proteins is believed to have a common ancestor that contained four calcium-binding domains (19, 33). Through mutation, the regulatory light chains lost the ability to bind Ca^{2+} by all but the first of these domains (2, 8, 39). *C. elegans*, *Drosophila*, and rat regulatory MLCs exhibit two regions of especially strong amino acid sequence conservation (Fig. 4). The first region, extending from positions 25 to 55 of the *C. elegans* sequences, corresponds to the first, functional calcium-binding domain. The second region, extending from positions 61 to 89 of the *C. elegans* sequence, corresponds to the ancestral calcium-binding domain 2. The high degree of conservation within this second domain, despite its loss of calcium binding, suggests that it is functionally important. This region may be involved in binding to MHCs (19).

mlc-1 and *mlc-2* are the only regulatory MLC genes that we detected in the *C. elegans* genome. Of six *C. elegans* genomic regions that hybridized under conditions of reduced stringency to the *Drosophila* probe, only *mlc-1* and *mlc-2* encode bona fide MLC proteins. *mlc-1* and *mlc-2*, furthermore, did not hybridize to any other *C. elegans* genomic sequences except each other, even under conditions of reduced stringency (Fig. 5). It is possible, however, that other regulatory MLC genes are sufficiently diverged from *mlc-1* and *mlc-2* that they did not hybridize under these conditions. A likely candidate for such genes would be those encoding nonmuscle MLCs. The similarities of cytoplasmic regulatory MLCs to those from muscle are largely unknown. The gene for a chicken cytoplasmic alkali MLC hybridizes to its skeletal muscle counterpart (45), but the generality of this is uncertain.

We analyzed the noncoding regions of *mlc-1* and *mlc-2* for similarities to other *C. elegans* muscle genes. *unc-54* encodes one of two MHC isozymes expressed in body wall muscle (15, 37). Two regions of *mlc-1* are strikingly similar to two regions of the *unc-54* MHC gene. First, a 23-base

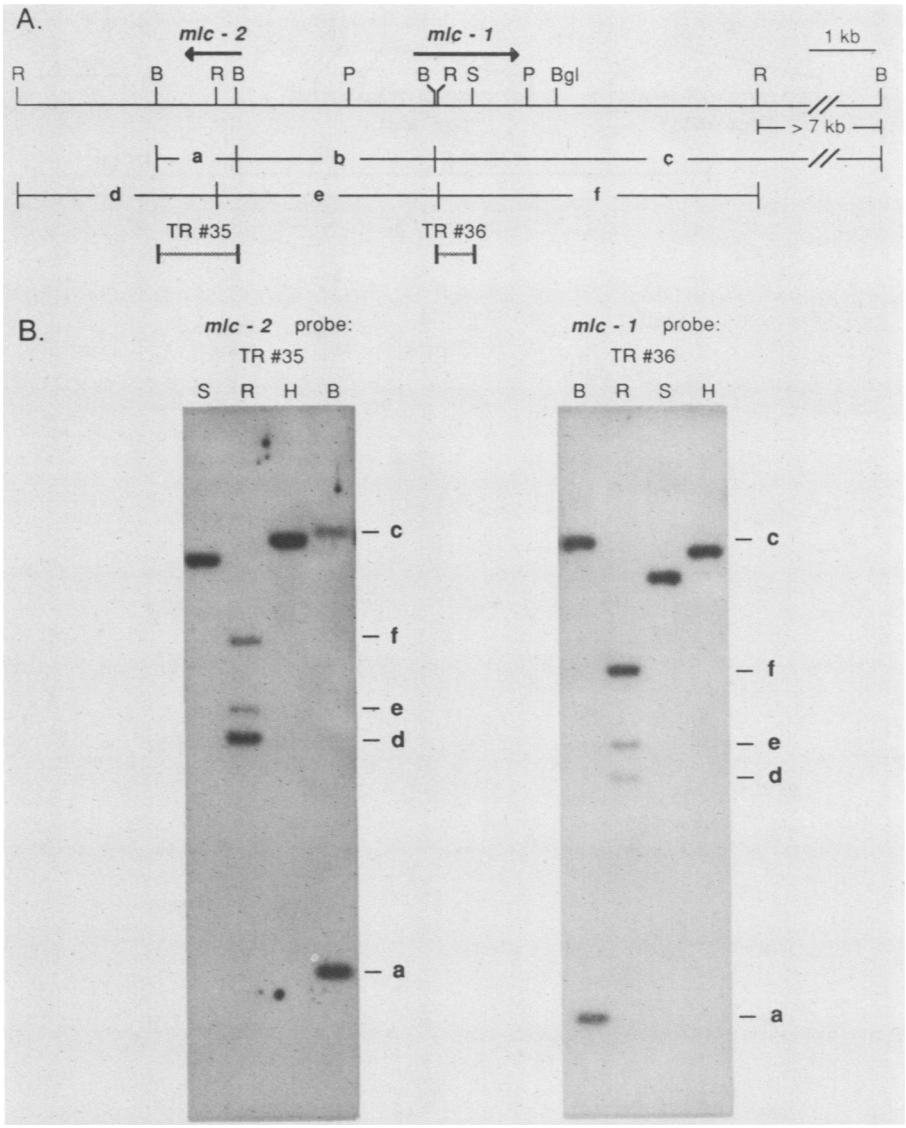

FIG. 5. Copy number of *C. elegans* regulatory MLC genes. (A) Plasmid probes used in Southern hybridizations are shown below a restriction map of *mlc-1* and *mlc-2*. *Bam*HI and *Eco*RI restriction fragments are indicated below the map with lowercase letters. Restriction sites: R, *Eco*RI; B, *Bam*HI; P, *Pvu*II; S, *Sal*I; Bgl, *Bgl*II. Probe TR#35 contains the *Bam*HI restriction fragment "a" cloned into plasmid vector pAT153 (60). Probe TR#36 contains a 560-base-pair *Bam*HI-*Sal*I restriction fragment cloned into pAT153. (B) Genomic Southern blot hybridizations. Each lane contains N2 (wild-type) genomic DNA cut with the indicated restriction enzyme. Restriction enzymes used: S, *Sal*I; R, *Eco*RI; H, *Hin*dIII; B, *Bam*HI. Hybridizations were done using conditions of reduced stringency (see Materials and Methods). Lowercase letters mark the positions of the *Bam*HI and *Eco*RI restriction fragments shown in panel A.

sequence within the 3′ untranslated region of *mlc-1* (nucleotides 2457 to 2479) is identical at 19 of 23 positions to a region within the 3′ untranslated region of *unc-54* (nucleotides 8254 to 8275 of Karn et al. [30]; Fig. 2). This sequence similarity occurs approximately 70 bases upstream of the AAUAAA polyadenylation signal in both *mlc-1* and *unc-54*. Second, a 26-base sequence within intron 4 of *mlc-1* (nucleotides 1811 to 1837) is identical at 21 of 26 positions to a region within intron 1 of *unc-54* (nucleotides 1277 to 1302; Fig. 2). The significance of these similarities is unknown.

Transcripts of *mlc-2*, but not *mlc-1*, contain a spliced leader sequence. We sequenced the 5′ termini of both *mlc-1* and *mlc-2* mRNAs and found that the terminal 21 or 22 nucleotides of *mlc-2* mRNA are not encoded within the 4.8 kilobases upstream of *mlc-2* that we have sequenced. Rather, they are identical to nucleotides found at the 5′ termini of three different actin mRNAs (32). Krause and Hirsh (32) propose that these 5′ leader sequences are added posttranscriptionally to actin mRNAs by a mechanism analogous to *trans*-splicing of mRNAs in trypanosomes (43, 58). We believe that *trans*-splicing of the type described by Krause and Hirsh (32) is responsible for generating the 5′ terminus of *mlc-2*. The site in the *mlc-2* genomic sequence to which these nucleotides are attached in the mRNA coincides with a consensus 3′ splice acceptor sequence. The function of the *trans*-spliced leader sequence is unknown. Perhaps it influences the localization, stability, or translational efficiency of mRNAs.

Do *mlc-1* and *mlc-2* serve different functions in *C. elegans*? The inferred amino acid sequences of *mlc-1* and *mlc-2*

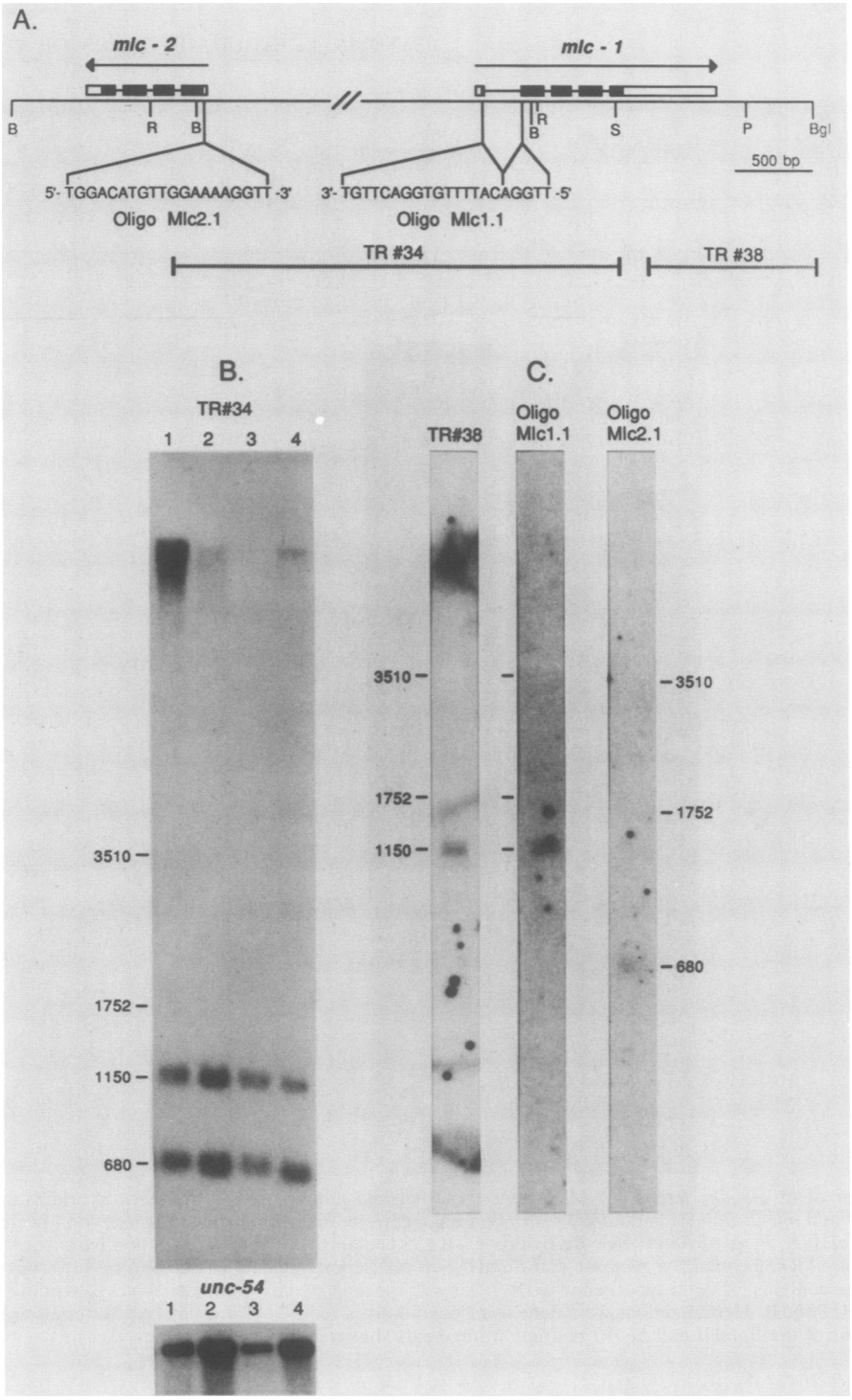

FIG. 6. Analysis of *mlc-1* and *mlc-2* transcripts. (A) Hybridization probes used in Northern blots are shown relative to the genomic organization of *mlc-1* and *mlc-2*. Plasmid clone TR#34 hybridizes to both *mlc-1* and *mlc-2*; it contains a 3.7-kilobase *Sau*3A-*Sal*I restriction fragment cloned into pAT153. Plasmid clone TR#38, an M13 clone, is specific for *mlc-1* transcripts; it contains an insert of approximately 1,100 base pairs that includes the terminal 500 nucleotides of the *mlc-1* 3′ untranslated region. Oligonucleotide Mlc1.1 is specific for *mlc-1* transcripts. Its sequence is discontinuous with respect to the genomic sequence of *mlc-1*. The 5′-terminal five nucleotides of Mlc1.1 are complementary to the first five nucleotides of exon 2 (coordinates, 1290 to 1294; Fig. 2). The remaining 15 nucleotides of Mlc1.1 are complementary to the last 15 nucleotides of exon 1, which include the ATG initiator codon and 12 nucleotides of 5′ untranslated region (coordinates 1016 to 1030; Fig. 2). Oligo-Mlc1.1, therefore, is fully complementary to mature *mlc-1* mRNA. Oligonucleotide Mlc2.1 is specific for *mlc-2* transcripts; it contains seven nucleotides that are complementary to the *mlc-2* protein-coding region and 13 nucleotides that are complementary to the 5′ untranslated region of *mlc-2* (coordinates 1629 to 1648; Fig. 3). We used the 18S (1,752 bases) and 26S (3,510 bases) rRNAs (14) as size markers in panels B and C. (B) Developmental Northern blot. Total RNA was prepared from synchronized populations of wild-type worms. Lanes: 1, L1 and L2 larvae; 2, L2 and L3 larvae; 3, L4 larvae and young adults; 4, gravid adults. The hybridization probe was plasmid clone TR#34. As a control to indicate the relative amounts of RNA loaded in each lane, we performed duplicate hybridizations using a plasmid probe specific for the *unc-54* MHC mRNA. These hybridizations are shown below those of TR#34. (C) Northern blots using gene-specific hybridization probes. The gene-specific probes are indicated above each lane. We calculated the sizes of hybridizing RNAs by using rRNA size markers. We confirmed these assignments by comparing the autoradiograms to control hybridizations that were done in parallel by using a probe which hybridizes to both *mlc-1* and *mlc-2* mRNAs (not shown).

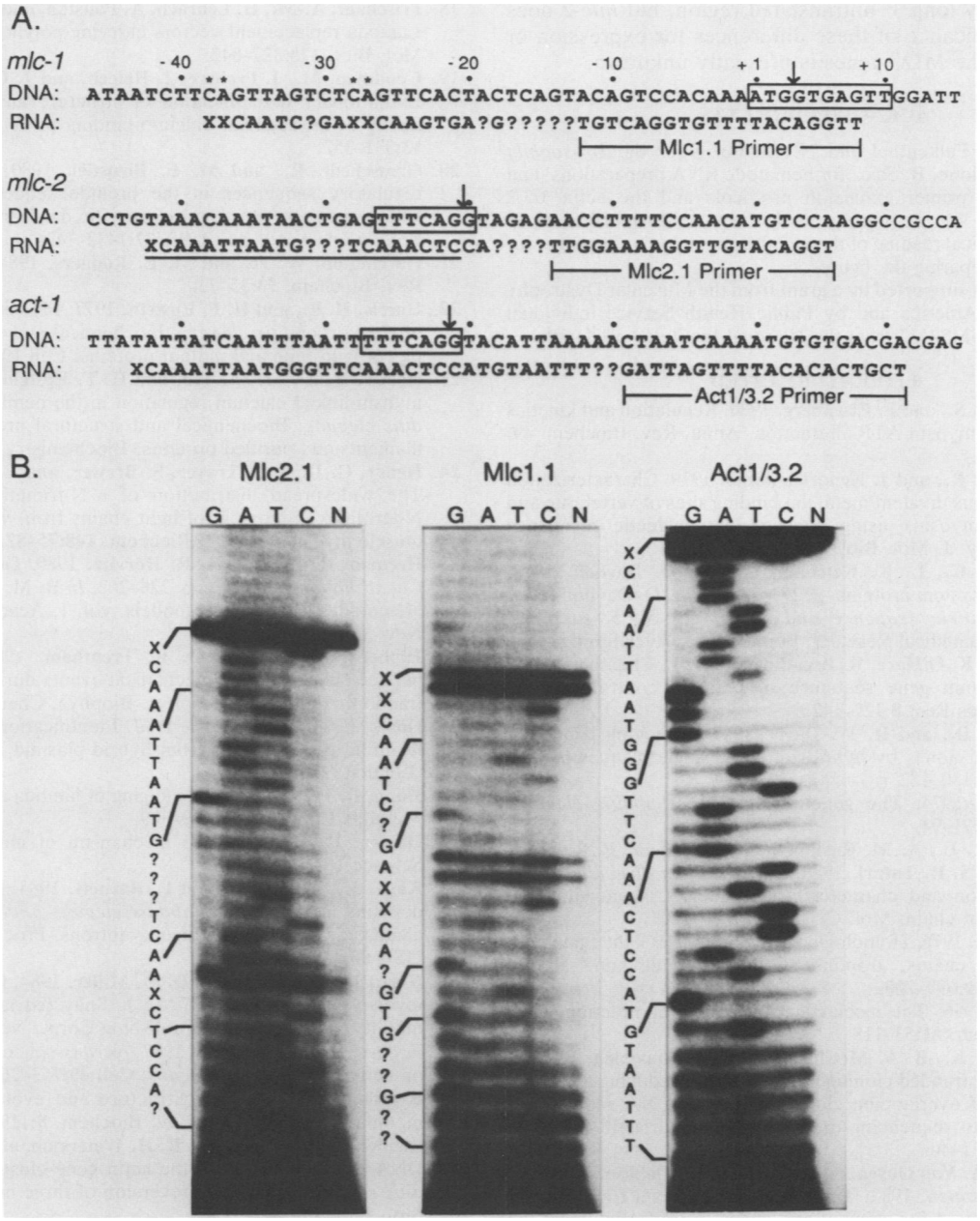

FIG. 7. Primer extension sequencing analysis of *mlc-1* and *mlc-2* mRNAs. (A) Summary of 5′-terminal sequences. For each gene, the sequence that we obtained by using an mRNA template is shown below the sequence of the DNA coding strand. The RNA sequences shown are the antisense strands (3′ → 5′, left to right) as read directly from autoradiograms. The sequencing primers used for each gene are indicated. Oligonucleotide primer Mlc1.1 spans the first intron of *mlc-1*. Thus, the 5′-terminal 5 nucleotides of Mlc1.1 are not complementary to the DNA strand shown above it; Mlc1.1 is, however, fully complementary to *mlc-1* mRNA (Fig. 6A). The 5′ and 3′ splice sites are shown in boxes with arrows indicating the splice junctions. Symbols: ?, positions where our sequences are ambiguous; X, positions of strong termination in all sequencing lanes, including reactions performed in the absence of dideoxynucleoside triphosphates (ddNTPs); *, sites of transcription initiation that we deduce from these results. We included analysis of the *act-1* 5′ terminus as a control. The spliced leader sequence common to *mlc-2* and *act-1* mRNAs is underlined. The sequence summary shown in panel A is compiled from several experiments for each oligonucleotide primer. A representative sequencing experiment for each primer is shown in panel B. (B) Primer extension sequencing results. G, A, T, and C indicate the ddNTPs added to a given reaction. Each reaction contained all four dNTPs. For lanes labeled N, ddNTPs were omitted. Symbols: ?, uncertainties in the RNA sequences; X, termination sites in the absence of ddNTPs.

differ at a single position (an Ile-to-Val substitution). We think it unlikely that this conservative substitution is functionally significant, but definitive proof will likely result from in vitro manipulation of the genes. The amounts of *mlc-1* and *mlc-2* mRNAs are relatively constant throughout postembryonic development (Fig. 6), but we have not examined the spatial distribution of the mRNAs or their expression during early development. *mlc-1* and *mlc-2* could serve different functions if they are expressed in different tissues or at different times during development. Outside of their protein-coding regions, *mlc-1* and *mlc-2* are strikingly different. For example, *mlc-2* is *trans*-spliced, but *mlc-1* is not. *mlc-1*

contains a very long 3′ untranslated region, but *mlc-2* does not. The significance of these differences for expression or regulation of the MLC genes is presently unknown.

ACKNOWLEDGMENTS

We thank S. Falkenthal and N. Davidson for the *Drosophila* MLC-2 cDNA clone, B. Saari for nematode RNA preparations, and M. Krause for primer extension protocols and the actin 1/3.2 oligonucleotide. We are grateful to B. Carr, A. Rushforth, and E. Spangler for critical reading of the manuscript and to S. Sprunger for assistance in preparing the figures.

This work was supported by a grant from the Muscular Dystrophy Association of America and by Public Health Service individual research grant GM30132 from the National Institutes of Health.

LITERATURE CITED

1. **Adelstein, R. S., and E. Eisenberg.** 1980. Regulation and kinetics of the actin-myosin-ATP interaction. Annu. Rev. Biochem. **49:** 921–956.
2. **Bagshaw, C. R., and J. Kendrick-Jones.** 1979. Characterization of homologous divalent metal ion binding sites of vertebrate and molluscan myosins using electron paramagnetic resonance spectroscopy. J. Mol. Biol. **130:**317–336.
3. **Barker, W. C., L. K. Ketcham, and M. O. Dayhoff.** 1978. Contractile system proteins, p. 273–284. *In* M. O. Dayhoff (ed.), *Atlas of protein sequence and structure,* vol. 5, suppl. 3. National Biomedical Research Foundation, Silver Spring, Md.
4. **Benoist, C., K. O'Hare, R. Breathnach, and P. Chambon.** 1980. The ovalbumin gene sequence of putative control regions. Nucleic Acids Res. **8:**127–142.
5. **Benton, W. D., and R. W. Davis.** 1977. Screening lambda-gt recombinant clones by hybridization to single plaques *in situ*. Science **196:**180–182.
6. **Brenner, S.** 1974. The genetics of *Caenorhabditis elegans*. Genetics **77:**71–94.
7. **Chisholm, R. L., A. M. Rushforth, R. S. Pollenz, E. R. Kuczmarski, and S. R. Tafuri.** 1988. *Dictyostelium discoideum* myosin: isolation and characterization of cDNAs encoding the essential light chain. Mol. Cell. Biol. **8:**794–801.
8. **Collins, J. H.** 1976. Homology of myosin DTNB light chain with alkali light chains, troponin C and parvalbumin. Nature (London) **259:**699–700.
9. **Cooke, R.** 1986. The mechanism of muscle contraction. Crit. Rev. Biochem. **21:**53–118.
10. **Dale, R. M. K., B. A. McClure, and J. P. Houchins.** 1985. A rapid single stranded cloning procedure for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18 S rDNA. Plasmid **13:**31–40.
11. **Dierks, P., A. Van Ooyen, N. D. Cochran, C. Dobkin, J. Reiser, and C. Weissman.** 1983. Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell **32:**695–706.
12. **Eide, D., and P. Anderson.** 1985. The gene structures of spontaneous mutations affecting a *Caenorhabditis elegans* myosin heavy chain gene. Genetics **109:**67–79.
13. **Eisenberg, E., and T. L. Hill.** 1985. Muscle contraction and free energy transduction in biological systems. Science **227:**999–1006.
14. **Ellis, R. E., J. E. Sulston, and A. R. Coulson.** 1986. The rDNA of *C. elegans*: sequence and structure. Nucleic Acids Res. **14:** 2345–2364.
15. **Epstein, H. F., R. H. Waterston, and S. Brenner.** 1974. A mutant affecting the heavy chain of myosin in *Caenorhabditis elegans*. J. Mol. Biol. **90:**291–300.
16. **Falkenthal, S., V. P. Parker, W. W. Mattox, and N. Davidson.** 1984. *Drosophila melanogaster* has only one myosin alkali light-chain gene which encodes a protein with considerable amino acid sequence homology to chicken myosin alkali light chains. Mol. Cell. Biol. **4:**956–965.
17. **Fire, A.** 1986. Integrative transformation of *Caenorhabditis elegans*. EMBO J. **5:**2673–2680.
18. **Frischauf, A.-M., H. Lehrach, A. Poustka, and N. Murray.** 1983. Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. **170:**827–842.
19. **Goodman, M., J. Pechere, J. Haiech, and J. G. Demaille.** 1979. Evolutionary diversification of structure and function in the family of intracellular calcium-binding proteins. J. Mol. Evol. **13:**331–352.
20. **Grosschedl, R., and M. L. Birnstiel.** 1980. Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants *in vivo*. Proc. Natl. Acad. Sci. USA **77:**1432–1436.
21. **Harrington, W. F., and M. E. Rodgers.** 1984. Myosin. Annu. Rev. Biochem. **53:**35–73.
22. **Harris, H. E., and H. F. Epstein.** 1977. Myosin and paramyosin of *Caenorhabditis elegans*: biochemical and structural properties of wild-type and mutant proteins. Cell **10:**709–719.
23. **Harris, H. E., M. W. Tso, and H. F. Epstein.** 1977. Actin and myosin-linked calcium regulation in the nematode *Caenorhabditis elegans*. Biochemical and structural properties of native filaments and purified proteins. Biochemistry **16:**859–865.
24. **Henry, G. D., I. P. Trayer, S. Brewer, and B. A. Levine.** 1985. The widespread distribution of α-N-trimethylalanine as the N-terminal amino acid of light chains from vertebrate striated muscle myosins. Eur. J. Biochem. **148:**75–82.
25. **Herman, R. K., and H. R. Horvitz.** 1980. Genetic analysis of *Caenorhabditis elegans*, p. 228–262. *In* B. M. Zuckerman (ed.), Nematodes as biological models, vol. 1. Academic Press, Inc., New York.
26. **Hibberd, M. G., and D. R. Trentham.** 1986. Relationships between chemical and mechanical events during muscular contraction. Annu. Rev. Biophys. Biophys. Chem. **15:**119–161.
27. **Hicks, J., and G. R. Fink.** 1977. Identification of chromosomal location of yeast DNA from hybrid plasmid pYeleu10. Nature (London) **269:**265–267.
28. **Hohn, B.** 1979. In-vitro packaging of lambda and cosmid DNA. Methods Enzymol. **68:**299–309.
29. **Huxley, H. E.** 1969. The mechanism of muscle contraction. Science **164:**1359–1366.
30. **Karn, J., S. Brenner, and L. Barnett.** 1983. Protein structural domains in the *Caenorhabditis elegans unc-54* myosin heavy chain gene are not separated by introns. Proc. Natl. Acad. Sci. USA **80:**4253–4257.
31. **Karn, J., N. J. Dibb, and D. M. Miller.** 1984. Cloning nematode myosin genes, p. 185–237. *In* J. Shay (ed.), Cell and muscle motility, vol. 6. Plenum Publishing Corp., New York.
32. **Krause, M., and D. Hirsh.** 1987. A *trans*-spliced leader sequence on actin mRNA in *C. elegans*. Cell **49:**753–761.
33. **Kretsinger, R. H.** 1980. Structure and evolution of calcium-modulatd proteins. Crit. Rev. Biochem. **8:**119–174.
34. **Landel, C. P., M. Krause, R. H. Waterston, and D. Hirsh.** 1984. DNA rearrangements of the actin gene cluster in *Caenorhabditis elegans* accompany reversion of three muscle mutants. J. Mol. Biol. **180:**497–513.
35. **Lehman, W.** 1978. Thick-filament-linked calcium regulation in vertebrate striated muscle. Nature (London) **274:**80–81.
36. **Lehman, W., and A. G. Szent-Gyorgyi.** 1975. Regulation of muscular contraction: distribution of actin control and myosin control in the animal kingdom. J. Gen. Physiol. **66:**1–30.
37. **Mackenzie, J. M., Jr., F. Schachat, and H. F. Epstein.** 1978. Immunocytochemical localization of two myosins within the same muscle cells in *Caenorhabditis elegans*. Cell **15:**413–419.
38. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
39. **Matsuda, G., Y. Suzuyama, T. Maita, and T. Umegane.** 1977. The L-2 light chain of chicken skeletal muscle myosin. FEBS Lett. **84:**53–56.
40. **Meinkoth, J., and G. Wahl.** 1984. Hybridization of nucleic acids immobilized on solid supports. Anal. Biochem. **138:**267–284.
41. **Miller, D. M., and I. Maruyama.** 1986. The *sup-3* locus is closely linked to a myosin heavy chain gene in *C. elegans*. UCLA Symp. Mol. Cell. Biol. **29:**629–638.
42. **Moerman, D. G., G. M. Benian, R. J. Barstead, L. A. Schriefer,**

and R. H. Waterston. 1988. Identification and intracellular localization of the *unc-22* gene product of *Caenorhabditis elegans*. Genes Dev. **2**:93–105.

43. **Murphy, W. J., K. P. Watkins, and N. Agabian.** 1986. Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for *trans* splicing. Cell **47**:517–525.
44. **Nabeshima, Y., Y. Fujii-Kurijama, M. Muramatsu, and K. Ogata.** 1984. Alternative transcription and two modes of splicing result in two myosin light chains from one gene. Nature (London) **308**:333–338.
45. **Nabeshima, Y., Y. Nabeshima, Y. Nonomura, and Y. Fujii-Kuriyama.** 1987. Nonmuscle and smooth muscle myosin light chain mRNAs are generated from a single gene by the tissue-specific alternative RNA splicing. J. Biol. Chem. **262**:10608–10612.
46. **Nudel, U., J. M. Calvo, M. Shani, and Z. Levy.** 1984. The nucleotide sequence of a rat myosin light chain 2 gene. Nucleic Acids Res. **12**:7175–7186.
47. **Padgett, R. A., P. J. Grabowski, M. M. Konarska, S. Seiler, and P. A. Sharp.** 1986. Splicing of messenger RNA precursors. Annu. Rev. Biochem. **55**:1119–1150.
48. **Parker, V. P., S. Falkenthal, and N. Davidson.** 1985. Characterization of the myosin light-chain-2 gene of *Drosophila melanogaster*. Mol. Cell. Biol. **5**:3058–3068.
49. **Periasamy, M., E. E. Strehler, L. I. Garfinkel, R. M. Gubits, N. Ruiz-Opazo, and B. Nadal-Ginard.** 1984. Fast skeletal muscle myosin light chains 1 and 3 are produced from a single gene by a combined process of differential RNA transcription and splicing. J. Biol. Chem. **259**:13595–13604.
50. **Pulliam, D. L., V. Sawyna, and R. J. C. Levine.** 1983. Calcium sensitivity of vertebrate skeletal muscle myosin. Biochemistry **22**:2324–2331.
51. **Reinach, F. C., K. Nagai, and J. Kendrick-Jones.** 1986. Site-directed mutagenesis of the regulatory light-chain Ca^{2+}/Mg^{2+} binding site and its role in hybrid myosins. Nature (London) **322**:80–83.
52. **Robert, B., P. Daubas, M. Akimenko, A. Cohen, I. Garner, J. Guenet, and M. Buckingham.** 1984. A single locus in the mouse encodes both myosin light chains 1 and 3, a second locus corresponds to a related pseudogene. Cell **39**:129–140.
53. **Ross, J.** 1976. A precursor of globin messenger RNA. J. Mol. Biol. **106**:403–420.
54. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
55. **Slightom, J. L., A. E. Blechl, and O. Smithies.** 1980. Human fetal $^{G}\gamma$- and $^{A}\gamma$-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell **21**:627–638.
56. **Spieth, J., K. Denison, E. Zucker, and T. Blumenthal.** 1985. The nucleotide sequence of a nematode vitellogenin gene. Nucleic Acids Res. **13**:7129–7138.
57. **Sulston, J. E., and S. Brenner.** 1974. The DNA of *C. elegans*. Genetics **77**:95–104.
58. **Sutton, R. E., and J. C. Boothroyd.** 1986. Evidence for *trans* splicing in trypanosomes. Cell **47**:527–535.
59. **Toffenetti, J., D. Mischke, and M. L. Pardue.** 1987. Isolation and characterization of the gene for myosin light chain two of *Drosophila melanogaster*. J. Cell Biol. **104**:19–28.
60. **Twigg, A. J., and D. Sherratt.** 1980. Trans-complementable copy-number mutants of plasmid ColE1. Nature (London) **283**:216–218.
61. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp17-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
62. **Waterston, R. H., R. M. Fishpool, and S. Brenner.** 1977. Mutations affecting paramyosin in *Caenorhabditis elegans*. J. Mol. Biol. **117**:679–697.
63. **Waterston, R. H., and G. R. Francis.** 1985. Genetic analysis of muscle development in *C. elegans*. Trends Neurosci. **8**:270–276.
64. **Waterston, R. H., D. Hirsh, and T. R. Lane.** 1984. Dominant mutations affecting muscle structure in *Caenorhabditis elegans* that map near the actin gene cluster. J. Mol. Biol. **180**:473–496.
65. **Winter, B., H. Klapthor, K. Wiebauer, H. Delius, and H. H. Arnold.** 1985. Isolation and characterization of the chicken cardiac myosin light chain (L-2A) gene. J. Biol. Chem. **260**:4478–4483.
66. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.
67. **Zaug, A. J., J. R. Kent, and T. R. Cech.** 1984. A labile phosphodiester bond at the ligation junction in a circular intervening sequence RNA. Science **224**:574–578.