Characterization of the Myosin Light-Chain-2 Gene of Drosophila melanogaster

VANN P. PARKER,[1]† SCOTT FALKENTHAL,[2] AND NORMAN DAVIDSON[3]*

Division of Biology[1] and Department of Chemistry and Chemical Engineering,[3] California Institute of Technology, Pasadena, California 91125, and Department of Genetics, Ohio State University, Columbus, Ohio 43210[2]

Recombinant DNA clones encoding the *Drosophila melanogaster* homolog of the vertebrate myosin light-chain-2 (MLC-2) gene have been isolated. This single-copy gene maps to the chromosomal locus 99E. The nucleotide sequence was determined for a 3.4-kilobase genomic fragment containing the gene and for two MLC-2 cDNA clones generated from late pupal mRNA. Comparison of these sequences shows that the gene contains two introns, the positions of which are conserved in the corresponding rat sequence. Extension of a primer homologous to the mRNA reveals two start sites for transcription 12 nucleotides apart. The sequence TATA is not present ahead of the mRNA cap site. There are two major sites of poly(A) addition separated by 356 nucleotides. The protein sequence derived from translation of the cDNA sequence shows a high degree of homology with that for the DTNB myosin light chain (MLC-2) of chicken. A lower degree of sequence homology was seen in comparisons with other evolutionarily related calcium-binding proteins. RNA blots show high levels of expression of several transcripts during the developmental time stages when muscle is being produced. In vitro translation of hybrid-selected RNA produces two polypeptides which comigrate on two-dimensional gels with proteins from *Drosophila* actomyosin, although the cDNA sequence reveals only one 26-kilodalton primary translation product.

Myosin is a hexameric protein which forms the core of the thick filaments of muscle. All myosin filaments share the same architecture of two heavy chains and two pairs of light chains. In vertebrates, different nonallelic forms of these proteins are utilized in different tissues and at different stages of development. The evolution of these proteins can be studied by comparing their sequences both within and between species.

One of the pairs of myosin light chains has been referred to by different investigators discussing the homologous protein in different organisms as the regulatory light chains, the DTNB light chains, the phosphorylatable light chains, EDTA light chains, or myosin light chain (MLC) 2 (MLC-2). For all of these and for the homologous *Drosophila* protein we use the name MLC-2 in this paper, as the term implies no assumption as to function.

Characterization of the MLC-2 gene(s) of *Drosophila melanogaster* is of interest for the following reasons. (i) During myogenesis, profound changes in the expression of genes occur. Knowledge of the structure of MLC-2 may be essential for understanding the mechanisms of coordinate regulation. (ii) The MLCs are members of an evolutionarily related family, termed the troponin C superfamily (1), which includes troponin C, myosin alkali light chain, MLC-2, and the calcium regulatory protein calmodulin. These proteins all share a common structural domain which is requisite for the binding of divalent metal ions, termed the EF hand (29). Comparisons of conserved and diverged features of the amino acid sequences of family members across large evolutionary distances provide insights into molecular evolution and into structure-function relations for these proteins.

In contributing to the study of the questions referred to above, several laboratories have characterized recombinant DNA clones for the major myofibrillar components of *D.*

melanogaster. These components include the *D. melanogaster* actin (17, 49), tropomyosin (2, 25), myosin heavy chain (6, 48) and myosin alkali light-chain genes (13, 14). In this paper we describe the characterization of the MLC-2 gene of *D. melanogaster.* The DNA sequence is given for both the genomic region and cDNA clones generated from poly(A)-containing [poly(A)+] RNA. The protein sequence translated from this region is compared with other MLC-2 protein sequences. We describe the developmental expression of this MLC gene and show that it is present in only one copy in the *Drosophila* genome.

MATERIALS AND METHODS

Enzymes and chemicals. Restriction enzymes and enzymes used for labeling DNA were purchased from New England BioLabs, Inc., Bethesda Research Laboratories, Inc., or Boehringer Mannheim Biochemicals. Nuclease S1 was from Sigma Chemical Co. or Miles Laboratories, Inc. All reactions were performed as recommended by the manufacturer or as described by Maniatis et al. (32) unless otherwise stated.

Generation of plasmid subclones. DNA was subcloned from the lambda clone λdmpT57 (see below) by digestion with *Eco*RI and *Hin*dIII and ligation of the resulting fragments into pBR322 as described elsewhere (7, 32). A subclone, pLP5734, containing a 3.4-kilobase (kb) *Eco*RI-*Hin*dIII fragment which includes the coding region of the gene was used for subsequent analyses. Plasmid subclones used for DNA sequence analysis were generated by digestion of the 1.8-kb *Bam*HI-*Hin*dIII fragment of pLP5734 with *Hpa*II and ligation into the *Acc*I site of pUC8 (54). An *Nde*I subclone was generated by digestion of the *Hpa*II subclone containing the region 3′ to the open reading frame of pLP5734 (see below) with *Nde*I, ligation into the calf intestinal alkaline phosphatase-treated *Nde*I site of pBR322, and selection by colony hybridization (22).

Generation of M13 subclones. M13 subclones were initially generated by digestion of pLP5734 with *Sau*3A or *Hpa*II and

---

* Corresponding author.
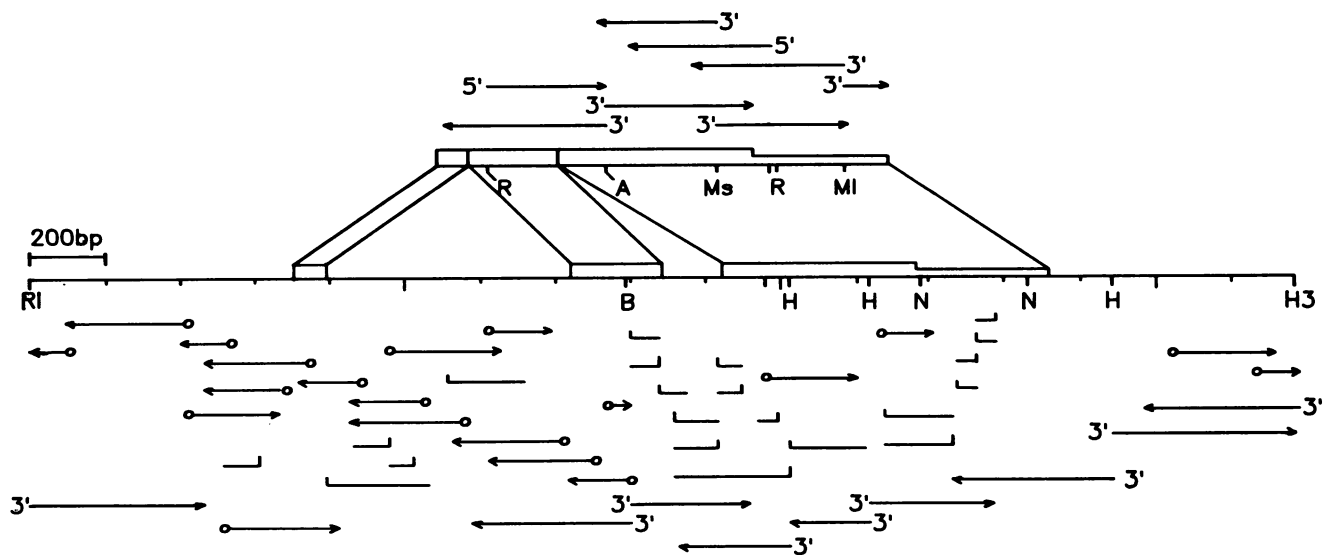† Present address: Amgen, Thousand Oaks, CA 91320.

FIG. 1. Sequencing strategy. The upper part of the figure is the sequencing strategy for the cDNA clones. Symbols: □, transcribed DNA present in mature mRNA; ▭, sequences specific to the longer transcript (see text). The lower part of the figure shows the organization of the exons transcribed from the genomic DNA and the sequencing strategy of the 3.4-kb *Eco*RI-*Hind*III fragment. The direction and extent of sequencing are shown. Symbols: ⊣, M13 clones; ●→, BAL 31-generated M13 clones (see Materials and Methods); 3'→, 3'-end-labeled restriction fragments; and 5'→, 5'-end-labeled fragments. Restriction enzyme sites are A, *Ava*I; B, *Bam*HI; H, *Hpa*II; H3, *Hind*III; Ml, *Mlu*I; Ms, *Mst*II; N, *Nde*I; R, *Rsa*I; and RI, *Eco*RI.

ligation of the resulting fragments into M13mp8 (36), which had been digested with *Bam*HI or *Acc*I, respectively. The 3.4-kb *Drosophila* DNA insert in pLP5734 was isolated by agarose gel electrophoresis and nick translated to probe for clones with desired inserts by plaque filter hybridization (3). The original vector, pBR322, was used as a probe to counterscreen plaques.

The *Eco*RI-*Bam*HI region was underrepresented by this shotgun approach (Fig. 1). To generate molecules which would allow us to rapidly sequence the entire insert region, we have used a procedure based on digestion by the double-strand exonuclease BAL 31 (30). This procedure enabled us to generate a nested set of deletions suitable for sequencing. The method which was independently developed by us (V. P. Parker, Ph.D. thesis, California Institute of Technology, Pasadena, 1985) is generally similar to that described by Poncz et al. (43).

Gel electrophoresis and isolation of fragments. Labeled DNA fragments used for sequencing by the chemical modification method (34) were isolated as described by Garfinkel et al. (18).

RNA was fractionated on a 1% formaldehyde–agarose gel and transferred to nitrocellulose filters as described elsewhere (32).

Two-dimensional gel electrophoresis of proteins followed the technique of O'Farrell (41). ^{14}C-protein standards were from Bethesda Research Laboratories. *Drosophila* proteins were extracted from adult animals as described elsewhere (14).

Sequencing gels were 5, 8, and 20% polyacrylamide (19:1::acrylamide:bisacrylamide)–50% urea in 100 mM TBE (100 mM Tris [pH 8.3], 100 mM boric acid, 2 mM EDTA). Gels were 0.4 mm thick and 40 or 80 cm long.

DNA sequence analysis. DNA sequencing was done by both the chemical modification technique (34) and the dideoxy chain termination method (50) with M13 phage vectors (36). The sequencing strategies for pLP5734 and the cDNA clones pCDM20 and pCDM38 are shown in Fig. 1.

S1 nuclease mapping of transcripts. Transcripts were mapped by S1 nuclease digestion of hybrids of end-labeled DNA with poly(A)[+] RNA (5, 15, 55). In a typical reaction, approximately 6×10^4 cpm of end-labeled probe was hybridized to 7.5 μg of late pupal poly(A)[+] RNA at 56°C in S1 hybridization buffer {80% formamide, 40 mM PIPES [piperazine-*N,N'*-bis(2-ethanesulfonic acid)], 1 mM EDTA [pH 6.8]} for 5 h. Hybridizations were diluted with 300 μl of S1 digestion buffer (280 mM NaCl, 30 mM sodium acetate [pH 4.4], 4.5 mM zinc acetate, 20 μg of denatured salmon sperm DNA per ml) at 0°C and placed immediately on ice. S1 nuclease was added, and the reactions were allowed to incubate for 30 min at the indicated temperature. S1 digestions were terminated by the addition of 75 μl of S1 stop buffer (2.5 mM ammonium acetate, 50 mM EDTA, 400 μg of yeast tRNA per ml) and precipitated with 400 μl of isopropanol. Pelleted samples were suspended in sequencing-gel loading dye (34), boiled, and applied to an appropriate gel.

RESULTS

Isolation of an abundantly expressed myogenic clone. We have previously described the isolation of recombinant DNA clones encoding *Drosophila* genes abundantly expressed during muscle formation (14). We used RNA from animals 70 to 75 h after puparium formation to prepare a cDNA probe for screening a genomic library in the lambda phage vector Charon 4. Pupal RNA at this stage is rich in muscle-specific sequences. After the exclusion of nonmyogenic clones and those encoding actin or myosin heavy-chain gene sequences, we examined 59 clones in further detail. Preliminary characterizations of these clones by restriction enzyme analysis and blotting with a cDNA probe revealed one group of clones which included 11 independent phage isolates. A restriction endonuclease map of these clones is shown in Fig. 2.

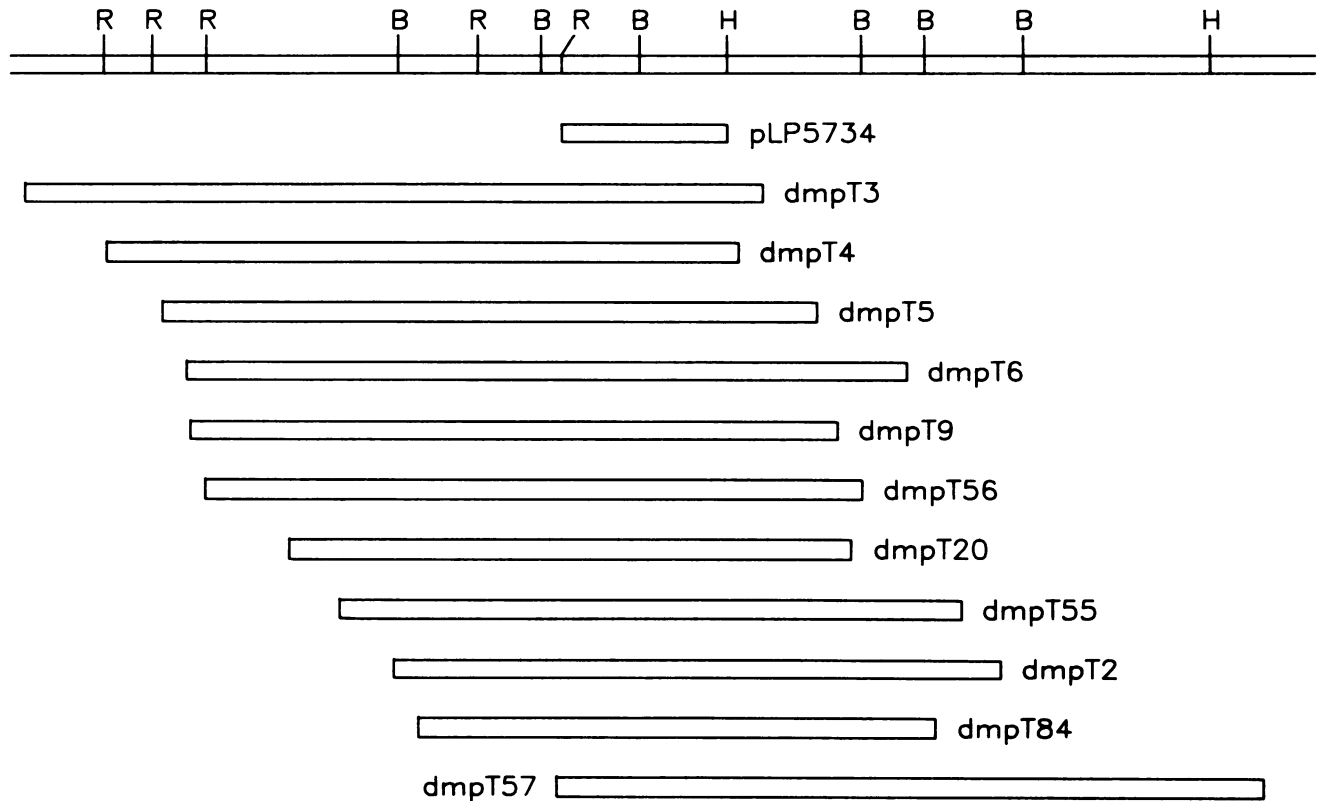The region hybridizing to late pupal cDNA was localized

FIG. 2. Restriction endonuclease map of overlapping λ clones and the subclone pLP5734. The *EcoRI-HindIII* insert of pLP5734 is 3.4 kb. Abbreviations for restriction enzymes used are B, *BamHI*; R, *EcoRI*; and H, *HindIII*.

to a 3.4-kb *EcoRI-HindIII* fragment. This region was subcloned into pBR322 and is designated pLP5734.

Multiple RNA transcripts from one region. Other *Drosophila* muscle-specific genes have been shown to display a specific pattern of expression during different developmental stages (14, 48, 49). We wished to determine whether the gene in pLP5734 conformed to this pattern. RNA blots of poly(A)$^+$ RNA isolated from different developmental stages were probed with pLP5734 (Fig. 3). There are two major transcripts and one minor one with respective molecular lengths of 1.1, 1.4, and 2.7 kb. These transcripts are abundant in late embryonic, second-instar larval, late pupal, and adult RNAs. They are less abundant in first- and third-instar larvae and are not observed in early embryonic and early pupal RNAs. This pattern of expression is the same as that seen with the myosin heavy-chain gene (48) and the myosin alkali light-chain gene (14) when the same protocols for selecting RNA from different stages of development were used. Identical results were obtained when the lambda clone λdmpT56 or λdmpT57 was used (data not shown). Thus, the RNA blots fit the expected pattern for clones encoding a muscle-specific gene.

Copy number and cytogenetic location. Three independent experiments were undertaken to determine the copy number of the sequences within λdmpT57 (Parker, Ph.D. thesis). In the first experiment, *Drosophila* DNA was digested with *BamHI* and *EcoRI* and fractionated on an agarose gel. In adjacent lanes, known equivalents of digested λdmpT57 were loaded. The gel was blotted onto nitrocellulose and probed with λdmpT57 DNA at moderate stringency. By comparison of radiographic intensities, it was determined

that the cloned DNA was single copy within the *Drosophila* genome. In a second set of experiments, cloned DNAs of known copy number were digested, fractionated on an agarose gel, and blotted onto nitrocellulose. Genomic *Drosophila* DNA was nick translated and used to probe the blot. This "reverse Southern blot" also indicated that sequences in λdmpT57 are single copy. A third indication of copy number is provided by in situ hybridization of *Drosophila* polytene chromosomes. Only region 99E hybridized when λdmpT57 or pLP5734 was labeled with either ^{3}H or ^{35}S (14). This result has been independently confirmed (46; J. D. Tofsenetti, D. Mischke, and M. L. Pardue, Genetics 110:S83, 1985).

Hybrid selection and in vitro translation. To help identify the gene product(s) encoded by the cloned region, we characterized the polypeptides produced by hybrid selection and in vitro translation (45). Figure 4 shows a two-dimensional gel of the polypeptides synthesized by in vitro translation of hybrid-selected RNA. The apparent molecular weights (M_r) of the two observed protein products are 26,000 and 17,000. Each has a pI of 5. These two products are also seen in standard one-dimensional gels. However, when β-mercaptoethanol was omitted from the loading buffer, only the 26,000 band was seen (data not shown). We occasionally saw an additional spot migrating slightly more slowly than the 17,000 band in the sodium dodecyl sulfate-polyacrylamide dimension. Hybrid selection using the 1.8-kb *BamHI-HindIII* fragment of pLP5734 gave the same pattern.

DNA sequence analysis of pLP5734. The sequence of the cloned insert is presented in Fig. 5. All regions were sequenced on both strands except bases 2723 through 2885 and

FIG. 3. Developmental expression of RNAs from pLP5734. Poly(A)$^+$ RNA (1 μg) isolated from different developmental stages by synchronized populations of *D. melanogaster* was separated by electrophoresis on a 1.5% agarose–2.2 M formaldehyde gel. After transfer to nitrocellulose, the RNA was hybridized to a nick-translated probe of pLP5734 DNA. Time (in hours) is from egg deposition. Pupae were resynchronized at 5 h postpupariation by flotation (37) and were assumed to be 125 h old. RNA size standards are *Escherichia coli* 16S (1,547-nucleotide) and 23S (2,904-nucleotide) rRNAs, and HeLa cell 18S (1,760-nucleotide) and 28S (5,090-nucleotide) rRNAs (9, 10).

2886 through 3041, which were sequenced in the same direction from two different clones, and nucleotides 1 through 392, which were sequenced from several clones in both directions. For further comments on difficult regions, see the legend to figure 5.

Isolation and characterization of cDNA clones. We isolated cDNA clones corresponding to the chromosomal gene in pLP5734 to analyze the exon-intron structure and to deduce an encoded protein sequence. The late pupal cDNA library described previously (14) was screened with a probe from the 1.8-kb *Hind*III-*Bam*HI fragment of pLP5734. Two clones with long inserts, pCDM20 and pCDM38, were chosen for sequence analysis.

These cDNA sequences were in complete agreement with the genomic sequence. As discussed below, a translated amino acid sequence can be derived from the cDNA sequence. Comparison of the genomic and cDNA sequences shows that the gene contains two introns and three exons. The introns occur between amino acids 1 and 2 and amino acids 82 and 83 (Fig. 5). The lengths of the two introns are 669 and 169 nucleotides, respectively. The junctions of the intervening sequences (IVSs; underlined in Fig. 6) conform to the consensus rules for eucaryotic genes (8) and for those of *D. melanogaster* (51). Both of the cDNA clones have a poly(A) sequence at their respective 3' ends. The poly(A) tail from pCDM20 is approximately 200 nucleotides long and

begins at nucleotide 2361 of the genomic sequence, 21 nucleotides 3' to the first A in the recognition sequence AATAAA (44). The insert of pCDM20 is 918 nucleotides long [excluding the poly(A) sequence].

The insert of pCDM38 contains 1,111 base pairs of genomic sequence and a poly(A) segment of more than 150 nucleotides. It is colinear with pCDM20 starting at nucleotide 1443 (Fig. 5). It has the same intron spliced out at position 1619 through 1840 but extends in the 3' direction for an additional 356 nucleotides beyond the site of poly(A) addition of pCDM20. This clone does not include the 5' exon of pCDM20.

Homology to chicken DTNB MLC. Examination of the cDNA sequence revealed an open reading frame following an ATG codon 666 nucleotides long. The predicted protein molecular weight is 24,000, corresponding to the larger in vitro translation product of the gene. Dot matrix sequence comparisons (23) were used to compare the amino acid sequence derived from the *Drosophila* DNA sequence with the protein sequence of the chicken MLCs and other myofibrillar components (data not shown). Figure 6 shows a histogram of the homology of the *Drosophila* sequence with the chicken MLC-2 polypeptide. In this comparison, stretches of 10 amino acids are compared. A match in all 10 positions would be 100% homology, a match in 5 positions would be 50% homology, etc. Comparisons were made either requiring identical amino acids for a match or allowing for conservative amino acid replacements. In the more stringent I comparison, two regions are especially similar in sequence: amino acids 65 through 101 and 115 through 134 of *D. melanogaster* with amino acids 11 through 48 (65% homology with a 20-amino-acid stretch of 80% homology) and 63 through 82 (80% homology) of chicken, respectively. The overall similarities (Fig. 6; see Fig. 9), are the primary basis for the identification of the encoded *Drosophila* protein as MLC-2.

S1 nuclease analyses. To confirm the 3' endpoints of transcripts within pLP5734, an S1 nuclease experiment was performed. Hybrids formed between late pupal RNA and end-labeled DNA were digested with S1, and the products were separated by electrophoresis on a denaturing acrylamide gel (Fig. 7). With a probe which had been labeled at an *Hpa*II site (position 2227), protected fragments of 146 and 496 nucleotides were obtained. These correspond to the position of the poly(A) tails in the cDNA clones pCDM20 and pCDM38, respectively. This 350-nucleotide difference is approximately the same as the size difference observed in RNA blots (Fig. 3). A subclone containing the 282-base-pair *Nde*I fragment (nucleotides 2383 to 2664 in Fig. 5) hybridizes to the 1.4-kb but not to the 1.1-kb RNA in an RNA blot, in agreement with the interpretation that the poly(A) addition sites for these two RNAs lie at nucleotide positions 2722 and 2372, respectively (data not shown).

We observed heterogeneity of 3 to 5 nucleotides in the length of the fragments protected by nuclease S1. In addition, enzymes which were purchased from different commercial sources digested the hybrids to end products which differed from each other by 2 nucleotides. We found that varying the temperature used in the S1 nuclease digestion from 0 to 20°C did not alleviate the observed heterogeneity. We suggest that the different digestion products result from digestion of A-T base pairs formed at the end of the duplex.

The transcript of the myosin alkali light-chain gene of *D. melanogaster* undergoes differential processing to generate mRNAs that encode different protein products (13). To determine whether differential splicing of the MLC-2 pri-
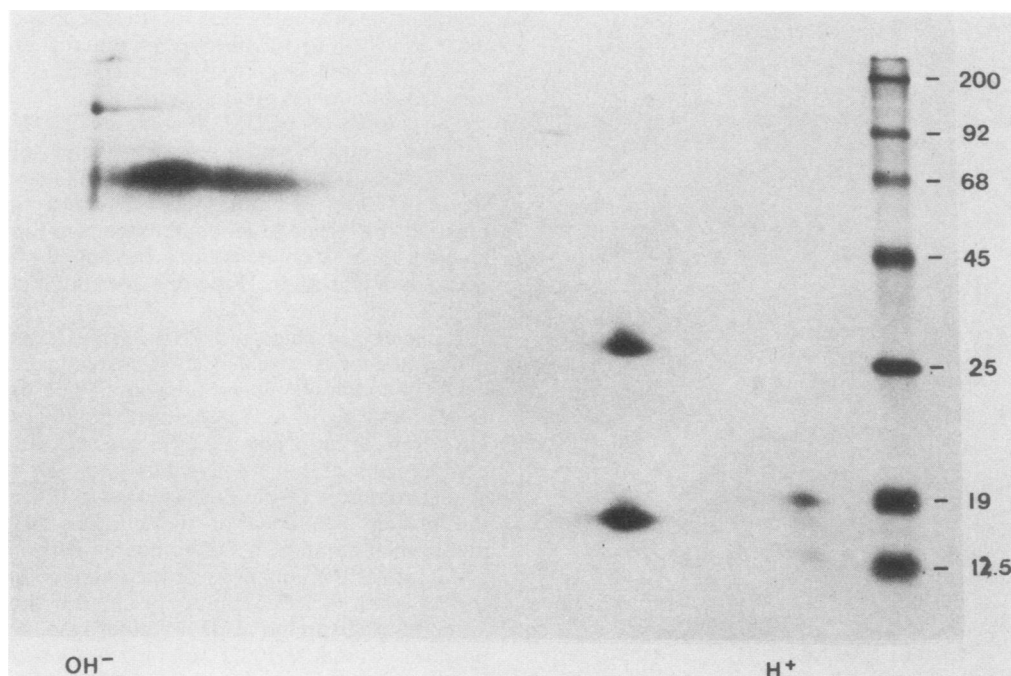
FIG. 4. In vitro translation of hybrid-selected RNA. Linearized DNA (10 μg) from the clone λdmpT55 was spotted onto a nitrocellulose filter. Poly(A)$^+$ RNA from late pupae was hybridized to select the mRNAs transcribed from this cloned region during muscle formation. Following elution from the filter, the mRNA was translated in an mRNA-dependent rabbit reticulocyte system with [^{35}S]methionine. Products were analyzed by two-dimensional gel electrophoresis. Ampholines of pH range 3.5 to 10 were used for the first dimension, and a 12% polyacrylamide gel was used for the second dimension.

mary transcript generates two mRNAs with different coding potentials, S1 nuclease mapping experiments were done with the cDNA clone pCDM20. This cDNA clone contains an open reading frame capable of generating the larger in vitro translation product. Two hybridization probes were generated. One was 3' end labeled at the NcoI site which is at the initiation of translation (CCATGG, nucleotides 765 through 770) in exon 1, and the other was 5' end labeled at the MstII site at the termination of translation (CCTAAGG, nucleotides 2259 through 2265) within exon 3. If MLC-2 transcripts were present at significant concentrations in the pupal mRNA population and differed in their exon organization from that revealed by DNA sequence analysis of pCDM20, S1-protected fragments would appear which were shorter than that predicted from the pCDM20 sequence. No evidence of alternative splicing patterns downstream from exon 1 to the poly(A) site of pCDM20 and upstream from exon 3 to the 5' end of pCDM20 was observed in these experiments (Fig. 8).

Analysis of transcription initiation by primer extension. The 5' ends of the mRNA transcripts were mapped by primer extension (19). A single-stranded 51-nucleotide fragment from the middle exon (bases 1439 through 1489) was gel isolated and used as a primer for reverse transcription of a late pupal mRNA template. Products of 109 and 121 nucleotides were observed. To determine the relationship of the two products, they were gel purified and sequenced. The sequences of these two fragments were colinear, with the larger fragment extending an additional 12 nucleotides 5' on the genomic sequence (data not shown). To rule out the possibility that this result was an artifact of reverse transcription, an S1 nuclease protection experiment was done with a probe labeled at the NcoI site of pLP5734 (nucleotide

766). Two products of the expected length were observed (data not shown). Primer extension experiments with different primers also confirmed the positions of the two start sites (data not shown). These two start sites, at nucleotides 700 and 712, are marked in Fig. 5.

DISCUSSION

We have previously described the isolation of recombinant lambda phage clones encoding proteins that are in high concentration in the muscles of D. melanogaster (14). One of these recombinant clones has been shown to encode the Drosophila myosin alkali light chain (13, 14). This protein has a similar molecular weight and isoelectric point to one of the proteins encoded by λdmpT55, the clone studied here. These results suggested that λdmpT55 encoded another member of the troponin C superfamily. Comparison of the translated sequence of a cDNA clone homologous to the gene region of λdmpT55 with that of chicken MLC-2 decisively identifies the Drosophila protein as an MLC-2. The gene is present at 1 copy per haploid genome and is located at band 99E on Drosophila polytene chromosomes.

MLC-2 proteins are members of the troponin C superfamily. This group of evolutionarily related proteins all contain one or several domains designated as EF hands (1). In some cases, this domain provides a structure capable of binding a calcium ion by chelation. Members of the troponin C superfamily are believed to have evolved from an ancestral Ca^{2+}-binding protein with a single EF hand by two gene duplication events to give a prototype protein with four EF hands. Further gene duplication events and subsequent divergent evolution gave rise to the different modern members of the troponin C superfamily. During this evolution, some family members have lost one or several EF-hand

```
1    GAATTCTCGGCGCTCTTCATATTGCACAATGTGCGGGCATTTCGAAAACAGAAAAAGGAACGTAGGCGAGCAAAACATCGAGTATATAGGCGTGAATAGAATACATACATATGCTATATG   120
121  TTGGCATATATGTATGCTGTCCGCCGCCAAGGAAAATGTAGGTGATTTCAGCGCAAGAGGACACCTTTCGGAGGTGGCCTCTGGCAGAGAGACGCACCGAGTAGTGAAGCAGAGCAAAAT   240
241  AAAAGAAAACAAAACAGAATCCGAGGATGGGGAAATGTATCCACAACAGAGAGCCGCTATATGAACCAATCCGAATCGCCCCGAAATAAATTCCGAACATATGTATAGGCTTCTGCTGCC   360
361  CAAAAGCCAAAGCCAAAGCAATTTCTCTGCCTTTAAAATTAGATTTTTGATTAGAGCAGCGCCCGATGGATGAGTAAACACCTCGGATACAGCTACACCATAGGCTATAGCTATAGCCAT   480
481  GGCTATATCTATATAGTTGTAGCTGCCCGCCCCCGCCGATCTCCTGGGCGGACCTCGCCTTGCAGATACTAAAAAAGGCAGCACGAATATTGCAGATACGGATACAGCGAATGGCGCCCG   600
601  AAGATCCCCGAAGGTGGCCGAACCTCCACGAGCGGCCCCAATGCTTGGAAGAAGAAGAGCTCCTCCGAGCTCTTAAAACATCATTGTCTTGGCGATAAAATCAGCCAGAGATTCAGTGTG   720

                                                         Met                                                                          1
721  CTCTGACGGCTCGACTCGATTAGACAACACATCAAAGTTGAGCACCATGGTAAGGACTCGTCCCGGAATATCACTACAATCCAACTGAACCGAACTGAACCAGAACCGAATCGAACTGGA   840
841  CCCCAGTGCCTGGATATATGCAGCAAGGATAATGCCGATCAGTGCAGTGAATCGTTTGCATTCGAGAATGACCCCGAGGATATGAATATGAAGTGAACCATAACTCTAAATCATGGTGAT   960
961  CCCGATTCACCTTCTTCACATACGCAGTGAGAATAGTGCAGAAAGTTCTATAAATGGATCGATCCGATCCGAGTGTTACGCGTCAGTAATGTAATCTGATATCCGTGCCGGGGAGCGAGC   1080
1081 GAGACAGCGAGTGGCCACGGCGGAGAGCTTACGTAATATGCGTCCGGTATATTTTTAGAATGAGTGGAAAACTTTCGCCTGCAGATACATATTTTCTGGCCTAGTCGGGTAGGCCTTTTG   1200
1201 AAAAGCATACCTTACATCCCTTAAATCCCTATTAAGGGATAACTTTCTACCCATATAATTCCAATTGAAGATCCTTCAACGCGGATAACTTACTAAGATCCCTAAAACGAATAGTATAAC   1320
1321 TTCCGGAATCGGAACTTTCAAGTGACAGCTAAAGACGGACTATAACGACTAATGACCGAGTAGATACCCCCTTTCTGCCACGAGTCCCTACTAATAAATCCCTCTTTTCTCTCCTCAG    1438

2    AlaAspGluLysLysLysValLysLysLysLysThrLysGluGluGlyGlyThrSerGluThrAlaSerGluAlaAlaSerGluAlaAlaThrProAlaProAlaAlaThrProAlaPro   41
1439 GCCGATGAGAAGAAGAAGGTTAAGAAGAAGAAGACCAAGGAAGAGGGTGGTACTTCCGAAACCGCTTCTGAGGCCGCATCCGAGGCAGCAACCCCGGCACCAGCTGCCACTCCTGCCCCG   1558

42   AlaAlaSerAlaThrGlySerLysArgAlaSerGlyGlySerArgGlySerArgLysSerLysArgAlaGlySerSerValPheSerValPheSerGlnLysGlnIleAlaGluPheLys   81
1559 GCCGCATCTGCCACTGGTTCGAAGAGAGCGTCGGGCGGATCCCGTGGCTCCAGGAAGTCGAAGCGCGCTGGCTCCTCGGTCTTCTCTGTGTTCTCCCAGAAGCAGATCGCCGAGTTCAAG   1678

82   Glu                                                                                                                            82
1679 GAGGTGAGTTTTGGGCGTTACTCATATATCGCATCATATGTCGGATCACTCGAGACCCATCCCATCATCCTGTCATCCTACCATCTATCGCACTTGTCAAGGACCAAGTCACCAACACAC   1798

83                                    AlaPheGlnLeuMetAspAlaAspLysAspGlyIleIleGlyLysAsnAspLeuArgAlaAlaPheAspSerValGly            108
1799 TCTTTCTTTCTATCTCTCTGTCTCTGTGCGGGATCGATAGGCCTTCCAACTCATGGATGCCGACAAGGACGGTATTATTGGCAAGAACGATCTGCGCGCTGCCTTCGACTCCGTCGGC    1918

109  LysIleAlaAsnAspLysGluLeuAspAlaMetLeuGlyGluAlaSerGlyProIleAsnPheThrGlnLeuLeuThrLeuPheAlaAsnArgMetAlaThrSerGlyAlaAsnAspGlu   148
1919 AAGATCGCCAACGACAAGGAGTTGGACGCCATGCTGGGCGAGGCCTCGGGTCCGATCAACTTCACCCAGTTGCTGACCCTGTTCGCCAACCGCATGGCCACCTCCGGTGCCAACGATGAA   2038

149  AspGluValValIleAlaAlaPheLysThrPheAspAsnAspGlyLeuIleAspGlyAspLysPheArgGluMetLeuMetAsnPheGlyAspLysPheThrMetLysGluValAspAsp   188
2039 GACGAAGTTGTTATTGCTGCCTTCAAAACATTCGATAACGATGGTCTCATCGACGGTGACAAATTCCGCGAAATGCTCATGAACTTCGGTGACAAGTTCACCATGAAGGAGGTTGATGAT   2158

189  AlaTyrAspGlnMetValIleAspAspLysAsnGlnIleIleAspThrAlaAlaLeuIleGluMetLeuThrGlyLysGlyGluGluGluGluGluGluGluGluAlaAla            222
2159 GCCTACGATCAGATGGTGATCGACGACAAGAACCAGATCGATACCGCCGCCCTGATCGAGATGCTCACCGGCAAGGGTGAAGAGGAGGAGGAGGAGGCCGCCTAAGGCACATCCGATCAC   2278
2279 ATCCTAACTTCTAACTCAACATCCATTTTACAGCCAGAATCAGAACGTTAAATCAATCAGTCAAATCAGTAATAATAAACGTAATTAAAAGAAACGCTCGTTTACATATGTAGATTCTCA   2398
2399 GTATCTTGTAATCTCTTTTGTTAACCATCACCAAGTACCTTGTACCAAGAACAACACCACCAACCAGATCAGCAACATCCGATCAAACTGAACCTCAAAAGAACTTCTTCAGCAACATTT   2518
2519 GGTAGCGATCAGAACCATTTACCATGTATCTGTCCATCATAGTTAGGGATACTTACGATCACGTTCACTGTCACGGCAAGAACTGACGCGTAGTTTTCTAGTTGATTTACTGTTAAATTT   2638
2639 TACACGTAAAATTTATAAAACATATGCTATCTATTTTATTTAAAACACCTATAAAATCAACAAAAAAAACTTAAAAACAACAAACAAGCCGTGTCTATTTTTATATCGCTGTCCGTTGAA   2758
2759 GAACTAAAGAGAAGAACCCCCAGAGAGAGATGATTACCAGATCGAGATCGAGATCGGAATCGAAATGAGGATGTATGATGTATGATTTATGATGTAGGGAGGACGAAGACGATGTTGATG   2878
2879 GGTAGCCCGGCTGTATTCGTTAGCCCTGCACCCGCGAGTTCAATTTGATTATGAATATGAATTTTAGCCAGTTTTGTGAGTTACAATGTTTTTCGTTCAAGAAATCTCGCAGGACCATGA   2998
2999 GAGAGAGAGGTAAACCCATAACAAACTATTGTTAGGGATCATTTAACTGAGAGCCAAATATATATCCTACTATCACAAAATATTTATCGTTATACATATCTATCGTTCGTTACTTACTTA   3118
3119 TTTATCAGTTATGAGTTATCAATATTATTGCAAGTTATCGAGAGTTAGGCTAATAACAGAGAAACATTAAGTTCATTCACTCAAATTGTTTAATGCGAATTTATATAACTTAACCATCGC   3238
3239 CACCACCACAACTTTCTGCTTATGAACTATCAGCGATATAAACCACCCAAACGAACAAACAACCCAAAATGATATCTGAAATATGAAGAAACATTCAAATAACAAACCCAGTGTTTCTAG   3358
3359 TTAATGCGTTTTGAAGCTT                                                                                                         3377
```

FIG. 5. DNA sequence of pLP5734. The DNA sequence was determined as described in the text and shown in Fig. 1. The sites mapped for initiation and termination of transcription are denoted by an arrow (◀). Intron sequences at the intron-exon junctions are underlined. The predicted amino acid sequence is depicted over the DNA sequence. C20 and C38 are the 5' ends of the cDNA clones pCDM20 and pCDM38, respectively. We occasionally experienced difficulty reading nucleotides adjacent to inverted repeats with M13-based dideoxy sequencing. An example of this is the sequence from 1215 to 1241. The last base of the sequence CATCCCTTAAAATCCCTATTA AGGGATA follows a sequence which is an inverted repeat (underlined) of two adjacent nucleotide stretches and gave a signal in all four lanes. Raising the temperature of polymerization to 30°C did not alleviate this problem entirely. Nucleotides which gave an ambiguous signal were determined by a combination of the following: sequencing from the opposite direction, use of the chemical modification technique, and use of an M13 clone generated with BAL 31, which interrupted the inverted repeat.

domains. In some cases, apparent EF-hand domains have been retained but have lost the ability to bind Ca^{2+}. For example, troponin C and calmodulin have retained all four sites, myosin alkali light chain has lost three sites (it retains one site which is incapable of binding calcium), and MLC-2 has lost three sites (1). Figure 9 illustrates the alignment of the primary structure of the MLC-2 of chicken skeletal muscle, chicken cardiac muscle, chicken smooth muscle, rabbit skeletal muscle, and scallop with that of the *Drosophila* MLC-2. First note that after alignment, the *Drosophila* MLC-2 contains an amino-terminal segment of 50 to 55 residues that is not present in the vertebrate MLC-2

sequences. The identified calcium-binding domain of the vertebrate proteins corresponds to amino acid residues 70 to 101 of the *Drosophila* protein. In Fig. 9 these residues are designated by the rules of Kretsinger (29), where n denotes amino acid residues in the hydrophobic core of the calcium binding domain and X, Y, Z, -X, -Y, and -Z denote the residues involved in coordination with the calcium ion. To bind calcium, there must be at least four acidic amino acid residues in the calcium loop, but others may be serine, threonine, asparagine, and glutamine. The *Drosophila* MLC-2 contains four acidic residues in the correct positions. In addition, there is a glycine residue at the -Y position.
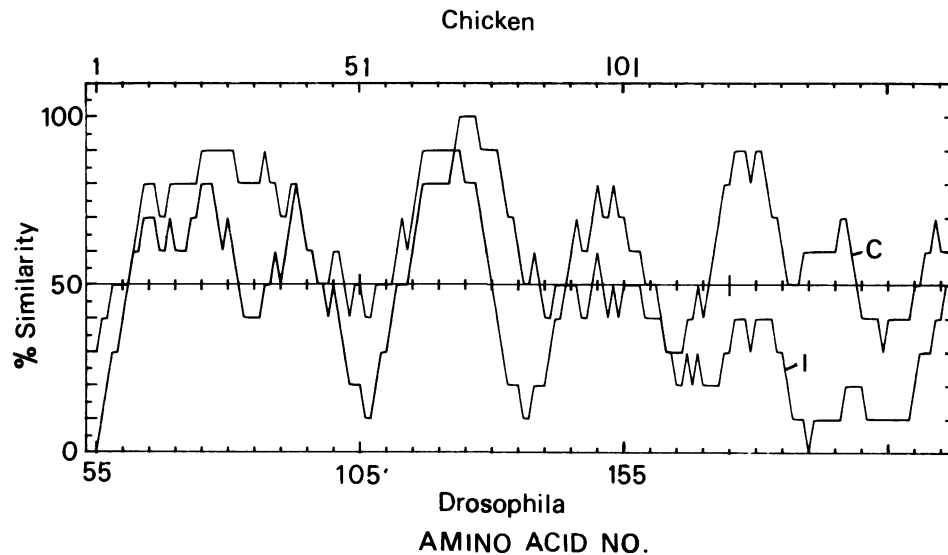
FIG. 6. Similarity comparison of the predicted *Drosophila* protein sequence with the chicken DTNB MLC. The protein sequences for the *Drosophila* polypeptide and the chicken DTNB light chain were aligned by dot matrix comparisons (data not shown). These histograms plot the percent homology of the two sequences over stretches of 10 amino acids. The comparisons were done by requiring an exact match of the two sequences (trace I) and allowing for conservative amino acid replacements (trace C). Conservative replacements are any amino acid within the same functional group. Since all matches in trace I would also be scored in trace C, the percent similarity in trace C will always be equal to or greater than that of trace I.

Glycine at this position is thought to augment the binding of calcium by means of a hydrogen-bonded water molecule (29). From these structural comparisons we predict that the *Drosophila* MLC-2 protein binds calcium with high affinity.

A second highly conserved region is in the ancestral

calcium-binding domain 2 of the aligned MLCs, where 13 of 16 amino acids in *Drosophila*, rabbit, and chicken sequences are identical (*Drosophila* amino acids 115 through 130; Fig. 9). Goodman et al. (20) have suggested that this striking conservation in sequence for rabbit, chicken, and scallop reflects the conserved function of myosin heavy-chain binding.

The site of phosphorylation has been identified in the rabbit MLC-2 as serine residue 15 (24) and in chicken smooth muscle as serine residue 19 (33). It has been proposed that phosphorylation is a recent evolutionary event, as other members of the troponin C family are not phosphorylated at this site, nor is the MLC-2 of scallop (which functions as a regulatory light chain). The *Drosophila* protein has a serine residue (residue 67) which aligns with the rabbit serine residue 15 and shows conservation of the preceding two amino acids as well. It has been reported that the *D. melanogaster* MLC-2 is phosphorylated and that this phosphorylation is necessary for assembly of the myosin filament (Y. Hayashi and Y. Hotta, Dev. Growth Differ. **24**:417, 1982).

We wish to note two unexpected similarities of amino acid sequence of the *Drosophila* MLC-2 protein with vertebrate myosin alkali light chains. The myosin alkali light chains of
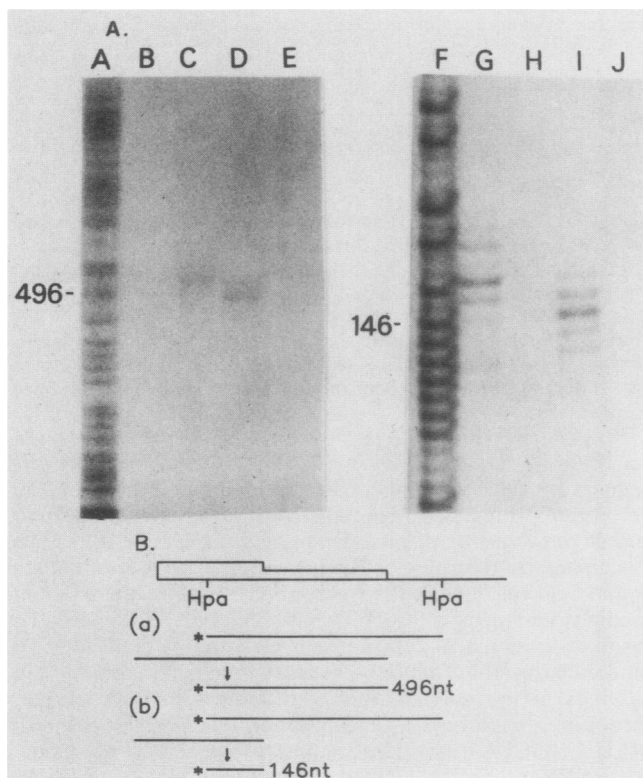
FIG. 7. S1 nuclease mapping of the 3' ends of MLC mRNAs. (A) Late pupal RNA hybridized to a probe labeled at the *Hpa*II site at nucleotide 2227. Following S1 nuclease digestion, the products were run on a 5% polyacrylamide–50% urea gel. To size the products accurately, two loadings (lanes A through E and F through J) were applied. Lanes A and F, Purine-specific sequence ladder (gift of L. Yu); lanes B and H, tRNA plus Miles S1; lanes C and G, 1 µg of late pupal RNA plus Miles S1; lanes D and I, 1 µg of late pupal RNA plus Sigma S1; lanes E and J, no S1. (B) Diagrammatic representation of expected fragment size predicted by cDNA clones pCDM38 (a) and pCDM20 (b).
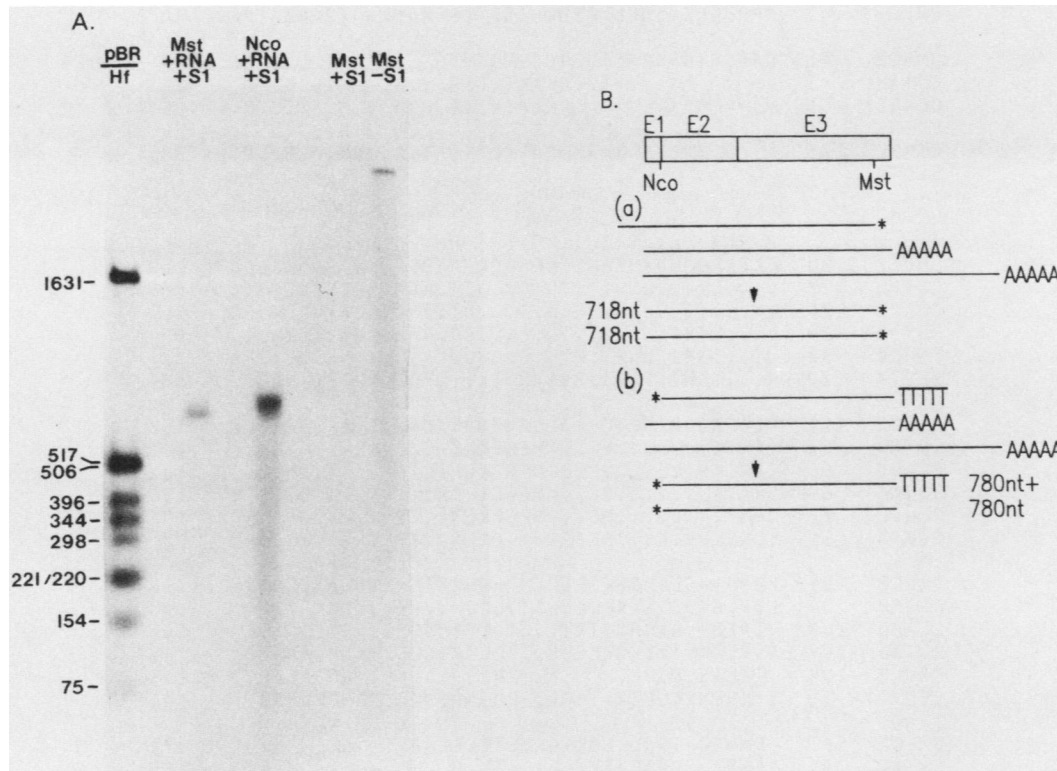
FIG. 8. (A) Late pupal RNA was hybridized to probes generated by 3' end labeling the *Nco*I site or 5' end labeling the *Mst*II site of pCDM20. Fragments protected from S1 nuclease digestion were analyzed on a 2% alkaline agarose gel. pBR/Hf, pBR322 DNA digested with *Hin*fI and end labeled with ³²P; Mst + RNA + S1, 5'-end-labeled pCDM20 digested with *Mst*II plus late pupal RNA plus S1; Nco + RNA + S1, 3'-end-labeled pCDM20 digested with *Nco*I plus late pupal RNA plus S1; Mst + S1, the probe in lane 2 with tRNA and S1; and Mst − S1, the probe in lane 2 with tRNA but no S1. (B) Diagrammatic representation of expected fragment sizes predicted by cDNA clones. E1, E2, and E3 refer to exons 1, 2, and 3. The products shown are those expected for RNAs homologous to pCDM20 and pCDM38 for the *Mst*II end-labeled probe (a) and the *Nco*I end-labeled probe (b).

vertebrate striated muscle are composed of two classes of proteins, LC1 and LC3. In chicken, rat, and mouse, these proteins are identical over their carboxy-terminal 141 amino acids (39, 42, 47). As a result of differential promoter utilization and RNA splicing, LC1 has 42 amino-terminal residues not present in LC3, and LC3 has 8 amino-terminal residues not present in LC1 (39, 42, 47). Amino acids 11 through 23 of the LC1 protein are AAAPAPAPAPAPA. The *Drosophila* MLC-2 sequence has some striking similarities to this pattern. As noted above, it has 53 amino-terminal residues which extend beyond the amino terminus of the chicken MLC-2 sequence. Secondly, the sequence AATPAPAATPAPA (residues 30 through 42) is identical to the chicken skeletal muscle LC1 in 10 of 13 residues (underlined). The function, if any, of this region is unclear, but we find it noteworthy that a region in the alkali light chain of vertebrates is found in the MLC-2 of *D. melanogaster*. We find no other significant region of homology between these proteins.

There are several conserved features between the structure of *Drosophila* and vertebrate MLC genes. In all of the MLC genes which have been studied, i.e., the *Drosophila* MLC-2 gene structure reported here, the rat MLC-2 gene (40), the *Drosophila* myosin alkali light-chain gene (13), and the MLC-3 genes of chicken (39), rat (42), and mouse (47), the initiation codon ATG occurs precisely at the 3' end of the first exon. Thus, the intriguing possibility exists that this structural characteristic may apply to MLC genes for higher eucaryotes over a large evolutionary distance.

The nucleotide sequence of the rat skeletal muscle MLC-2 gene has been determined (40). The translated amino acid sequence for this protein is identical with the rabbit protein presented in Fig. 9 except that it lacks one of two glycines at positions 12 and 13 of the rabbit sequence. As noted above, the first exon of the rat gene ends precisely at the initiation ATG codon, just as for the other *Drosophila* and vertebrate MLC genes. Furthermore, the second intron occurs between the amino acid residues EFKE and AFTV (rabbit residues 28 through 31 and 32 through 35; Fig. 9). This position aligns precisely with the position of the second intron of the *Drosophila* sequence which lies between residues 82 and 83 (Fig. 5 and 9). There are additional introns in the rat gene, but none in the *Drosophila* gene. It may be recalled that, in general, there is no conservation whatsoever of intron positions between any of the six *Drosophila* actin genes and any deuterostome, despite considerable conservation of intron positions within deuterostomes (for details, see references 4, 16, 53, 56). The sole known exception is an identical intron position for the *Drosophila* cytoskeletal actin 5C (16) and a newly discovered chicken cytoskeletal actin (4), each of which has an intron 8 nucleotides 5' to the initiator ATG codon. Despite this exception for one actin, we regard the conservation of two intron positions for the *Drosophila* and rat MLC-2 genes as striking.

We wish to note several other conserved features of the *Drosophila* MLC-2 gene and other eucaryotic genes. The nucleotide sequence CCAAT is often found 65 base pairs 5' of the mRNA start site and is thought to be responsible for

```
DMLC2    1    MADEKKKVKKKKTKEEGGTSETASEAASEAATPAPAATPA    40

DMLC2   41    PAASATGSKRASGGSRGSRKSKRAGSS-VFSVFSQKQIAE    79
CDTNB    1                   PKKAKRRAAE...N...M.D.T..Q.    26
CCARD    1                   PKKA.K.VE.G.N ..M.E.T..Q.    25
RDTNB    1                   PKKAKR.AAAEG...N...M.D.T..Q.    28
CSMLC    1           SSKRKRPQRAKAKTTKAT.N..AM.D.S..Q.    32
SEDTA    1                         ADKAA.G.LTKLP....Q.    19

                          -Y -X
              n   nn   nX Y ZG I   -Zn  nn   n
DMLC2   80    FKEAFQLMDAUKDGIIGKNDLRAAFDSVGK-IANDKELDA    118
CDTNB   27    .....TVI.QNR....D.D...ET.AAN.RLNVKNE....    66
CCARD   26    .....TI..QNR..F.D.A...DT.AAL.RLNVKNE..ED    65
RDTNB   29    .....TVI.QNRN...D.E...DT.AAM.RLNVKEED...    68
CSMLC   33    .....NMI.QNR..F.D.E..HDMLA.M..-NPT.EY.EG    71
SEDTA   20    M....SMI.V.R..FVS.D.IK.ISEQL.R-TPD....T.    58

DMLC2  119    MLGEASGPINFTQLLTLGANRMATSGANDEDEVVIAAFKT    158
CDTNB   67    .IK.........VF..-MFGEKLK-..DP..-.IMG...V    103
CCARD   66    .VK..P......VF..-MFGEKLK-.TDP.E-TILN...I    102
RDTNB   69    .-K.........VF..MMFGEKLKK...P..-.ITG...V    106
CSMLC   72    .MS..P......MF..-MFGEKLN-.TDP..-.IRN..AC    108
SEDTA   59    ..K..P..L...MF.S----DKLS-.TDS.E-TIRN..AM    92

DMLC2  159    FDND--GLIDGDKFREMLMNFGDKFTMKEVDDAYDQMVID    196
CDTNB  104    L.P.GK.S.KKSFLE.L.TTQC.R..PE.IKNMWAAFPP.    143
CCARD  103    ..PEGK.H.KA.YIK....TQEGR.SQE.INQMFAAFPP.    142
RDTNB  107    L.PEGK.T.KKQFLE.L.TTQC.R.SQE.IKNMWAAFPP.    146
CSMLC  109    ..EEAS.F.HE.HL..L.TTM..R..DE...EM.REAP..    148
SEDTA   93    ..EQENKKLNIEYIKDL.EDM.NN.NKD.MRMTFKEAPV-    131

DMLC2  197    DKNQID-TAALIEMLTGKGEEEEEEAA    222
CDTNB  144    VAGNV.YKNICY-VI.H-..DK.G.    166
CCARD  143    VSGNL.YKNLCY-VI.H-...KD    163
RDTNB  147    VGGNV.YKNICY-VI.H-.DAKD.Q    169
CSMLC  149    K.GNFNYVEFTR-I.KH-.AKDKDD    171
SEDTA  132    EGGKF.YVKFTAMIK-.S..D.A    153
```

FIG. 9. Comparison of the *Drosophila* protein sequence with MLC sequences from other organisms. The *Drosophila* sequence (DMLC2) is aligned with the chicken skeletal muscle DTNB MLC (CDTNB) (52), the chicken cardiac DTNB MLC (CCARD) (33), the rabbit skeletal muscle DTNB MLC (RDTNB) (11), the chicken smooth muscle MLC (CSMLC) (31), and the scallop EDTA regulatory light chain (SEDTA) (27). Identity to the *Drosophila* sequence is denoted by a period, and a hyphen denotes the absence of a residue at that position. Above amino acids 81 to 108 of the *Drosophila* sequence, the letters X, Y, Z, -Z, -X, and -Y denote the positions of amino acids which may be involved in calcium binding, n refers to the position of hydrophobic residues which form the adjacent α helices, and G and I refer to glycine and isoleucine residues often found in Ca^{2+}-binding domains (29).

high levels of expression of genes (12). This sequence is present at position 638 through 642 (Fig. 5), 62 and 74 nucleotides upstream of the mapped mRNA cap sites. The sequence TATA is generally 20 to 30 nucleotides prior to the cap site and is hypothesized to be responsible for precise positioning of the mRNA start (21). The *Drosophila* MLC-2 gene is not preceded by the TATA sequence, a finding which may be related to the multiple start sites of transcription.

The nucleotide sequence ATCAGTC is present at the beginning of each of the four *D. melanogaster* larval cuticle genes (51). A similar sequence is present at the 5' terminus of other *Drosophila* mRNAs (51). The mapped initiation sites of the MLC-2 gene are ATCAGCC and TTCAGTG (marked in Fig. 5), which are in good agreement with this observation.

Several conserved sequences have been observed in the IVSs of eucaryotic genes. The sequences at the 5' and 3' boundaries of IVSs have been well documented (8, 51). The junction sequences in the two IVSs of the *D. melanogaster* MLC-2 conform to these rules. The consensus sequence CTAAT has been observed near the 3' boundary of *Drosophila* IVSs and has been proposed to be a "3' splice signal" (26). Furthermore, the A in the fourth position of this sequence has been proposed to be the nucleotide involved in lariat formation of splicing intermediates (26). In the first

IVS, the sequence CTAAT occurs twice, at positions −70 and −28 relative to the 3' splice junction. In the second IVS, the sequence CCAAC, which we presume is the 3' splice signal for this IVS, occurs 51 nucleotides before the 3' boundary.

Preceding the site of polyadenylation, most eucaryotic genes have the hexanucleotide AATAAA (44). Recently, the conserved sequence YGTGTTYY (where Y is a pyrimidine) has been shown to be proximal to the poly(A) site of approximately two-thirds of the genes examined (35). In the *Drosophila* MLC-2 gene, the polyadenylation site at position 2372 is preceded by the sequence AATAAA but is not flanked by a close homolog of the YGTGTTYY signal. The polyadenylation site at position 2722 is preceded by the sequence AAAAAA (but not AATAAA) and has the sequence CGTGTCTA 6 nucleotides downstream. This latter sequence may explain why polyadenylation is observed at this site.

A question left unresolved in the present study is the relation of the DNA sequence of this gene to the two observed polypeptides. The open reading frame deduced from the cDNA sequence predicts a protein 222 amino acids long which would have a molecular weight of 24,000. Hybrid selection and in vitro translation produce two protein products with apparent molecular weights of 17,000 and 26,000

which comigrate on a two-dimensional gel with abundant proteins extracted from *Drosophila* indirect flight muscle actomyosin. However, when the in vitro translation product is run on a one-dimensional gel in the absence of β-mercaptoethanol, only one band of 26,000 is seen. There are no cysteine residues in the protein as shown by nucleotide sequencing of the cDNA clone to account for this unexplained result. Since this is a single-copy gene and we find no evidence for alternative RNA splicing patterns, we see no obvious explanation for the two translation products. Apparently, identical protein spots were observed by Mogami et al. (38) in two-dimensional gels of adult *Drosophila* myofibrils (their spots 181 and 184), but no functional identifications were done by them. There are several mechanisms not addressed in the current study by which the shorter polypeptide could be generated. These mechanisms include protease action or other specific modification which is present in both the in vitro translation system and in vivo in *D. melanogaster*, initiation of translation at the internal methionine at amino acid position 87 (Fig. 5), or initiation elsewhere within the gene. We note, however that the ATG codon at position 767 (amino acid 1 in our protein sequence) has a good consensus sequence for initiation of translation as defined by Kozak (28). Another possibility is that the shorter peptide is an artifact resulting from β-mercaptoethanol-induced cleavage of the 26,000 protein during the preparation for gel electrophoresis. This result would be analogous to the observation that nonenzymatic cleavage catalyzed by dithiothreitol or β-mercaptoethanol occurs with yeast glutamine synthetase (S. Rhee, personal communication). The resolution of this problem and the determination of the effect of β-mercaptoethanol on the electrophoretic behavior of the in vitro translation products remain for future work.

ACKNOWLEDGMENTS

LITERATURE CITED

1. **Barker, W. C., L. K. Ketcham, and M. O. Dayhoff.** 1978. The troponin C superfamily, p. 273–284. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure, vol. 5. National Biomedical Research Foundation, Silver Springs, Md.
2. **Basi, G. S., M. Boardman, and R. V. Storti.** 1984. Alternative splicing of a *Drosophila* tropomyosin gene generates muscle tropomyosin isoforms with different carboxy-terminal ends. Mol. Cell. Biol. **4**:2828–2836.
3. **Benton, W. D., and R. W. Davis.** 1978. Screening of λgt recombinant clones by hybridization to single plaques *in situ*. Science **196**:180–182.
4. **Bergsma, D. J., K. S. Chang, and R. J. Schwartz.** 1985. Novel chicken actin gene: third cytoplasmic isoform. Mol. Cell. Biol. **5**:1151–1162.
5. **Berk, A., and P. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids. Cell **12**:721–732.
6. **Bernstein, S. I., K. Mogami, J. J. Donady, and C. P. Emerson.** 1983. *Drosophila* muscle myosin heavy chain is encoded by a single gene located in a cluster of muscle mutations. Nature (London) **302**:23–34.
7. **Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Crosa, and S. Falkow.** 1977.

Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene **2**:95–113.
8. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
9. **Brosius, J., T. J. Dull, and H. F. Noller.** 1980. Complete nucleotide sequence of a 23S ribosomal RNA gene from *Escherichia coli*. Proc. Natl. Acad. Sci. USA **77**:201–204.
10. **Brosius, J., M. L. Palmer, P. J. Kennedy, and H. F. Noller.** 1978. Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. Proc. Natl. Acad. Sci. USA **75**:4801–4805.
11. **Collins, J. H.** 1976. Homology of myosin DTNB light chain with alkali light chains, troponin C and parvalbumin. Nature (London) **259**:699–700.
12. **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.
13. **Falkenthal, S., V. P. Parker, and N. Davidson.** 1985. Developmental variations in the splicing pattern of transcripts from the *Drosophila* gene encoding myosin alkali light chain result in different carboxyl-terminal amino acid sequences. Proc. Natl. Acad. Sci. USA **82**:449–453.
14. **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.
15. **Favaloro, J., R. Treisman, and R. Kamen.** 1980. Transcription maps of polyoma virus-specific RNA: analysis by two-dimensional nuclease S1 gel mapping. Methods Enzymol. **65**:718–749.
16. **Fyrberg, E. A., B. D. Bond, N. D. Hershey, K. S. Mixter, and N. Davidson.** 1981. The actin genes of *Drosophila*: protein coding regions are highly conserved but intron positions are not. Cell **24**:107–116.
17. **Fyrberg, E. A., K. L. Kindle, N. Davidson, and A. Sodja.** 1980. The actin genes of *Drosophila*: a dispersed multigene family. Cell **19**:365–378.
18. **Garfinkel, M. D., R. E. Pruitt, and E. M. Meyerowitz.** 1983. DNA sequences, gene regulation and modular protein evolution in the *Drosophila* 68C glue gene cluster. J. Mol. Biol. **168**:765–789.
19. **Ghosh, P. K., V. B. Reddy, M. Paitak, P. Lebowitz, and S. M. Weissman.** 1980. Determination of RNA sequences by primer directed synthesis and sequencing of their cDNA transcripts. Methods Enzymol. **65**:580–595.
20. **Goodman, M., J.-F. 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.
21. **Grosschedl, R., and M. 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.
22. **Grunstein, M., and D. Hogness.** 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. USA **72**:3961–3965.
23. **Hunkapiller, M., S. Kent, M. Caruthers, W. Dreyer, J. Firca, C. Griffin, S. Horvath, T. Hunkapiller, and L. Hood.** 1984. A microchemical facility for the analysis and synthesis of genes and proteins. Nature (London) **310**:105–111.
24. **Jakes, R., F. Northrup, and J. Kendrick-Jones.** 1976. Calcium binding regions of myosin "regulatory" light chains. FEBS Lett. **70**:229–234.
25. **Karlik, C. C., J. W. Mahaffey, M. D. Coutu, and E. A. Fyrberg.** 1984. Organization of contractile protein genes within the 88F subdivision of the *Drosophila melanogaster* third chromosome. Cell **37**:469–481.
26. **Keller, E. B., and W. A. Noon.** 1984. Intron splicing: a conserved internal signal in introns of animal pre-mRNAs. Proc. Natl. Acad. Sci. USA **81**:7417–7420.

27. Kendrick-Jones, J., and R. Jakes. 1977. Myosin-linked regulation: a chemical approach, p. 28–40. In G. Riecker, A. Weber, and J. Goodwin (ed.), International Symposium on Myocardial Failure. Springer-Verlag, New York.

28. Kozak, M. 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucleic Acids Res. 9:5233–5252.

29. Kretsinger, R. H. 1980. Structure and evolution of calcium-modulated proteins. Crit. Rev. Biochem. 8:119–174.

30. Legerski, R. J., J. L. Hodnett, and H. B. Gray, Jr. 1978. Extracellular nucleases of pseudomonas Bal31 III. Use of the double-strand deoxyribonuclease activity as the basis of a convenient method for the mapping of fragments of DNA produced by cleavage with restriction enzymes. Nucleic Acids Res. 5:1445–1464.

31. Maita, T., J. Chen, and G. Matsuda. 1981. Amino-acid sequence of the 20,000-molecular-weight light chain of chicken gizzard-muscle myosin. Eur. J. Biochem. 117:417–424.

32. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1980. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

33. Matsuda, G., T. Maita, Y. Kato, J. Chen, and T. Umegane. 1981. Amino acid sequence of the cardiac L-2A, L-2B, and gizzard 17,000 M light chains of chicken muscle myosin. FEBS Lett. 135:232–236.

34. Maxam, A., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. 65:499–560.

35. McLauchlan, J., D. Gaffney, J. L. Whitton, and J. B. Clements. 1985. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. Nucleic Acids Res. 13:1347–1368.

36. Messing, J., and J. Vieira. 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene 19:269–276.

37. Mitchell, M. K., and A. Mitchell. 1964. Mass culture and age selection in Drosophila. Dros. Inf. Ser. 39:135–137.

38. Mogami, K., S. C. Fujita, and Y. Hotta. 1982. Identification of Drosophila indirect flight muscle myofibrillar proteins by means of two-dimensional electrophoresis. J. Biochem. 91:643–650.

39. Nabeshima, Y., Y. Fujii-Kwriyama, M. Murimatsu, 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.

40. 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.

41. O'Farrell, P. H. 1975. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. 250:4007–4021.

42. 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. 21:13595–13604.

43. Poncz, M., D. Solowiejczyk, M. Ballantine, E. Schwartz, and S. Surrey. 1982. "Nonrandom" DNA sequence analysis in bacteriophage M13 by the dideoxy chain-termination method. Proc. Natl. Acad. Sci. USA 79:4298–4302.

44. Proudfoot, N. J., and G. G. Brownlee. 1976. 3' Non-coding region sequences in eukaryotic messenger RNA. Nature (London) 263:211–214.

45. Ricciardi, R. P., J. S. Miller, and B. E. Roberts. 1979. Purification and mapping of specific RNAs by hybridization-selection and cell-free translation. Proc. Natl. Acad. Sci. USA 76:4927–4931.

46. Roark, M., P. A. Mahoney, M. L. Graham, and J. A. Lengyel. 1985. Blastoderm-differential and blastoderm-specific genes of Drosophila melanogaster. Dev. Biol. 109:476–488.

47. Robert, B., P. Daubas, M.-A. Akimenko, A. Cohen, I. Garner, J.-L. Guenet, and M. Buckingham. 1984. A single locus in the mouse encodes both myosin light chain 1 and 3, a second locus corresponds to a related pseudogene. Cell 39:129–140.

48. Rozek, C. E., and N. Davidson. 1983. Drosophila has one myosin heavy chain gene with three developmentally regulated transcripts. Cell 32:23–34.

49. Sanchez, F., S. L. Towbin, U. Rdest, E. Zulauf, and B. J. McCarthy. 1983. Two Drosophila actin genes in detail: gene structure, protein structure and transcription during development. J. Mol. Biol. 163:533–551.

50. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467.

51. Snyder, M., M. Hunkapiller, D. Yuen, D. Silvert, J. Fristrom, and N. Davidson. 1982. Cuticle protein genes of Drosophila: structure, organization, and evolution of four clustered genes. Cell 29:1027–1040.

52. Suzuyama, Y., T. Umegane, T. Maita, and G. Matsuda. 1980. The amino acid sequence of the L-2 light chain of chicken skeletal muscle myosin. Hoppe-Seyler's Z. Physiol. Chem. 361:119–127.

53. Ueyama, H., H. Hamada, N. Battula, and T. Kakunaga. 1984. Structure of a human smooth muscle actin gene (aortic type) with a unique intron site. Mol. Cell. Biol. 4:1073–1078.

54. Vieira, J., and J. Messing. 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259–268.

55. Weaver, R. F., and C. Weissmann. 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5' termini of 15S β-globin mRNA precursor and mature 10S β-globin mRNA have identical map coordinates. Nucleic Acids Res. 7:1175–1193.

56. Zakut, R., M. Shani, D. Givol, S. Neuman, D. Yaffe, and U. Nudel. 1982. Nucleotide sequence of the rat skeletal muscle actin gene. Nature (London) 298:857–859.