MOLECULAR AND CELLULAR BIOLOGY, May 1984, p. 956–965
0270-7306/84/050956-10$02.00/0

Vol. 4, No. 5

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

SCOTT FALKENTHAL,[1]†* VANN P. PARKER,[2]‡ WILLIAM W. MATTOX,[2] AND NORMAN DAVIDSON[1]

Department of Chemistry[1] and Division of Biology,[2] California Institute of Technology, Pasadena, California 91125

Received 9 January 1984/Accepted 27 February 1984

A chimeric lambda DNA molecule containing the myosin alkali light-chain gene of *Drosophila melanogaster* was isolated. The encoded amino acid sequence was determined from the nucleic acid sequence of a cDNA homologous to the genomic clone. The identity of the encoded protein was established by two criteria: (i) sequence homology with the chicken alkali light-chain proteins and (ii) comparison of the two-dimensional gel electrophoretic pattern of the peptides synthesized by in vitro translation of hybrid-selected RNA to that of myosin alkali light-chain peptides extracted from *Drosophila* myofibrils. There is only one myosin alkali light-chain gene in *D. melanogaster*; its chromosomal location is region 98B. This gene is abundantly expressed during the development of larval as well as adult muscles. The *Drosophila* protein appears to contain one putative divalent cation-binding domain (an EF hand) as compared with the three EF hands present in chicken alkali light chains.

Myosin light chains are proteins which occur abundantly and in a defined stoichiometry in myofibrils. They are members of an evolutionarily related group of calcium-binding proteins known as the troponin C superfamily, which includes calmodulin, troponin C, and the myosin alkali and 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB) light chains. The primary amino acid sequence has been determined for at least one vertebrate example of each of these polypeptides (2). The principal sequence homology between these proteins resides in the putative Ca^{2+}-binding domains, which are known as EF hands (14). The roles of all of these proteins, except for the myosin alkali light chain, in muscle function have been determined (10, 13, 29; R. A. Murphy, M. O. Askoy, P. F. Dillon, W. T. Gerthoffer, and K. E. Kanim, Fed. Proc. **42**:51–57, 1983).

The skeletal muscle myosin alkali light chains are so named because of the high pH required to dissociate them from the myosin heavy chain (39). For vertebrate muscles, they are sometimes called MLC-1 and MLC-3. The two skeletal muscle alkali light chains of mammals and chickens, which have molecular weights of about 21,000 (MLC-1) and 17,000 (MLC-3), are virtually identical in sequence over their C-terminal 141 residues, but diverge in sequence at the amino terminus. MLC-1, depending upon the tissue from which it is isolated, has an additional alanine-proline- or alanine-lysine-rich sequence of 40 amino acids at its amino terminus. There is evidence that in rats the two proteins are encoded by a single gene (L. Garfinkel, R. Gubits, B. Nadal-Ginard, and N. Davidson, manuscript in preparation). At one time, these peptides were thought to be essential for the actin-activated adenosine triphosphatase activity of myosin (16, 32, 38), but recent in vitro studies have suggested that such is not the case (36). Thus, the function of the alkali light chains is unknown.

Drosophila melanogaster is an advantageous organism for further studies of structure and function of gene products such as the myosin light chains because one may use both molecular and genetic approaches. Here, we describe our initial steps in isolating and identifying the myosin alkali light-chain gene of *D. melanogaster* and in determining from the nucleic acid sequence the primary structure of the protein. We also report on the reiteration frequency and developmental expression of this gene.

* Corresponding author.
† Present address: Department of Genetics, 927 Biological Sciences Building, The Ohio State University, Columbus, OH 43210.
‡ Present address: Amgen, Newberry Park, CA 91320.

MATERIALS AND METHODS

Isolation of RNA. Total cellular RNA was prepared by homogenizing developmental-stage whole animals in 4 M guanidine thiocyanate–1 M 2-mercaptoethanol–0.05 M sodium acetate–0.001 M EDTA (pH 6.0) and banding in cesium chloride (11, 35). Polyadenylate-containing [poly(A)$^+$] RNA was selected by oligodeoxythymidylate-cellulose (type T3; Collaborative Research) chromatography as described by Anderson and Lengyel (1).

Isolation and purification of DNA. Charon phage DNA was isolated as described by Yen and Davidson (41) with the modifications described by Snyder et al. (33). Plasmid DNA was isolated as described by Fyrberg et al. (11). High-molecular-weight pupal DNA was prepared from 40- to 60-h pupae as follows. Two grams of pupae frozen in liquid nitrogen were ground to a fine powder in a mortar at −70°C. The frozen powder was homogenized by 10 to 15 strokes with a B pestle Dounce homogenizer in 30 ml of homogenization buffer (50 mM Tris, pH 7.2, 25 mM KCl, 5 mM $MgCl_2$, 350 mM sucrose, 0.15 mM spermine, 0.15 mM spermidine) and then filtered through Nitex cloth to remove cuticular debris. The nuclei were collected by centrifugation in a Sorvall HB-4 rotor at 4°C for 15 min at 3,000 rpm. The pelleted nuclei were washed with 30 ml of homogenization buffer and centrifuged to reduce the mitochondrial contamination of the nuclear pellet. The washed and pelleted nuclei were then suspended in 1.0 ml of nuclear suspension buffer (60 mM NaCl, 10 mM Tris, pH 7.2, 10 mM EDTA, 0.15 mM spermine, 0.15 mM spermidine). The nuclei were then lysed by the addition of 2.0 ml of lysis buffer (200 mM Tris, pH 8.5, 30 mM EDTA, 2% (wt/vol) Sarkosyl). Proteinase K was added to a concentration of 50 μg/ml, and the solution was gently mixed. After 2 to 4 h at 42°C, CsCl and ethidium

bromide were added to concentrations of 0.925 g/ml and 500 μg/ml, respectively. The solution was spun in a table-top clinical centrifuge at top speed for 20 min to float the denatured protein. The cleared solution was centrifuged for 18 h at 53,000 rpm (20°C) in a Beckman VTi65 rotor. The banded genomic DNA was visualized by UV irradiation, and the band was collected by side puncture. The ethidium bromide was removed by butanol extraction. The sample was diluted to 0.5 ml and dialyzed twice against 5,000 volumes of 10 mM Tris (pH 8.0)–1 mM EDTA. The salt concentration was adjusted to 0.15 M with sodium acetate, and the DNA was precipitated by the addition of 2.0 volumes of absolute ethanol. After 1 h at −20°C, the DNA was collected by centrifugation, rinsed with 70% ethanol, and air dried at 4°C for 24 h. The DNA was resuspended by overlaying the pellet with 10 mM Tris (pH 8.0)–1 mM EDTA and allowing it to hydrate for 48 to 72 h at 4°C.

Screening libraries. A *Drosophila* genomic DNA library prepared by J. Lauer was used. It consisted of randomly sheared Canton S embryonic DNA of 12 to 20 kilobases (kb) inserted into Charon 4 via synthetic *Eco*RI linkers (17). The library was screened by the high-density plaque hybridization technique of Benton and Davis (4).

Subcloning fragments of genomic clones. DNA fragments of chimeric lambda clone λdmpT102 were obtained by digestion with *Hin*dIII and *Eco*RI. These fragments were ligated into plasmid pBR322 as described previously (41).

DNA labeling and hybridizations. ^{32}P-labeled pupal cDNA probes were prepared from pupal poly(A)$^+$ RNA by using oligo(dT) primers (P-L Biochemicals) by the method of Mullins et al. (24). Preparation of nick-translated probes and hybridization of ^{32}P-labeled DNA probes to filter-bound DNA were performed essentially as described by Mullins et al. (24). Nick-translated probes were hybridized at a concentration of 10^5 dpm/ml (10^8 dpm/μg), whereas the labeled cDNA probes were hybridized at a concentration of 1×10^6 to 2×10^6 dpm/ml (10^7 dpm/μg).

Modified Okayama and Berg technique for cDNA library preparation. The goal of this approach was to make a cDNA library with efficient utilization of poly(A)$^+$ RNA and with long inserts which contained the entire 3′ untranslated region, the protein-coding region, and most of the 5′ untranslated region, following the general spirit of the method of Okayama and Berg (28). Their procedure can be readily adapted and indeed simplified by using a modern cloning vector with a suitable polylinker. The first version of our procedure is depicted in Fig. 1. In this proposed procedure, intermediate 4 is treated with dGTP and terminal transferase to give intermediate 5 and to provide a G tail for oligodeoxycytidylate priming of second-strand synthesis from intermediate 6. In this procedure, we used Mn^{2+} ion as a catalyst based on the recommendation of Deng and Wu (9). If successful, this procedure would have regenerated the *Eco*RI site upon final ligation. In fact, none of the resulting clones examined at the sequence level had a GC oligonucleotide segment or the *Eco*RI site. However, this procedure has worked well in the hands of other investigators (F. K. Lin, personal communication). We presume that our lack of success in this step was due to a failure of the terminal transferase tailing reaction. We believe that an alternative and equally effective procedure would be to carry out the replacement synthesis directly on intermediate 4. This procedure, which is analogous to that used by Charles Rice (personal communication), relies on RNase H action to generate the primers for second-strand synthesis. Several of our clones for other genes have been shown to include up to ca. 10 nucleotides from the 5′ end of the mRNA. Thus a 5′-terminal primer is hardly necessary. This revised procedure would result in converting the *Bam*HI site into a blunt end. The final ligation would thus preserve the *Eco*RI site.

It should be noted that with the pUC8 vector there is an upstream *lac* promoter and translation start system. Thus, the resulting cDNA clones are useful for expression of the encoded proteins in *E. coli*.

pUC8 DNA (10 μg) was digested with 50 U of *Pst*I in 40 μl containing 100 mM Tris-hydrochloride (pH 7.5), 50 mM KCl, 10 mM $MgCl_2$, and 100 μg of bovine serum albumin (Pentex) per ml. After 60 min the solution was heated to 65°C for 15 min and then quickly cooled. Poly $(T)_{50-80}$ tails were added to the *Pst*I-digested DNA with terminal transferase. The reaction solution (200 μl) contained the following: 10 μg of *Pst*I-cut pUC8 DNA, 0.25 mM unlabeled TTP, 50 μCi of [^{3}H]TTP (77.8 Ci/mmol), 0.2 M cacodylate, 0.05 M Tris (pH 7.0), 1 mM $CoCl_2$. This was incubated for 10 min at 37°C, after which dithiothreitol and terminal transferase were added to concentrations of 0.1 mM and 110 U/ml, respectively. After 20 min the nucleic acid was precipitated by the addition of 2 volumes of absolute ethanol. The precipitated nucleic acid was collected by centrifugation, washed once with 90% ethanol, and dried under vacuum.

At this point, the vector has polythymidylate tails at each end of the molecule. To act as a good primer there must be only one tail per molecule. One tail is eliminated by digesting the molecule with *Bam*HI, whose recognition site is 12 nucleotides 5′ to the T-tailed *Pst*I site. The small resulting oligonucleotide was separated from the tailed vector by gel filtration on Sephadex G-150.

The T-tailed DNA was resuspended in 200 μl of 100 mM Tris (pH 7.5)–50 mM KCl–10 mM $MgCl_2$–100 μg of BSA per ml. *Bam*HI (10 U) was added, and the solution was incubated at 37°C. After 1 h, EDTA was added to a concentration of 50 mM, and the solution was layered onto a 5.0-ml Sephadex G-150 column (0.5 by 28 cm) equilibrated with 50 mM NaCl–10 mM Tris (pH 8.0)–10 mM EDTA. Fractions containing the excluded peak were pooled, and the nucleic acid was recovered by ethanol precipitation and centrifugation.

First-strand cDNA synthesis was performed in a volume of 100 μl containing 2 μg of poly(T)-tailed pUC8 DNA, 10 μg of pupal poly(A)$^+$ RNA, 50 mM Tris (pH 8.3), 50 mM KCl, 8 mM $MgCl_2$, 4 mM sodium PP_i, 2 mM DTT, 2 mM TTP, 2 mM dCTP, 2 mM dATP, 2 mM dGTP, and 5 μCi of [^{32}P]dCTP (410 Ci/mmol). Reverse transcriptase (54 U) was added, and the solution was incubated at 46°C for 45 min. After phenol extraction, the aqueous phase was layered onto a 5.0-ml Sephadex G-150 column (0.5 by 28 cm) equilibrated in 50 mM NaCl–10 mM Tris (pH 8.0)–10 mM EDTA. The excluded peak fractions were pooled, and the nucleic acid was recovered by ethanol precipitation and centrifugation. The second-strand replacement synthesis was done with RNase H and DNA polymerase 1 as described by Okayama and Berg (28). The resulting double-stranded DNA molecules were circularized by blunt end ligation as follows. The DNA was ethanol precipitated after second-strand synthesis, collected by centrifugation, and resuspended in 0.5 ml of a solution containing 50 mM Tris (pH 7.8), 10 mM $MgCl_2$, 20 mM DTT, 1 mM ATP, and 50 μg of BSA per ml. Incubation with 2,000 U of T4 DNA ligase proceeded for 12 h at 12°C. Bacterial cells (*Escherichia coli* MC1061) were transformed, and 5% of the transformation mix was plated onto L plates containing 50 μg of ampicillin per ml, whereas the remainder of the library was amplified for 3 h in liquid culture and then frozen in 50% glycerol at −40°C. Our library contained

200,000 independent recombinants from 2 μg of starting vector DNA.

Purification of *Drosophila* myosin. *Drosophila* indirect flight muscle myofibrils were purified from 24- to 48-h adults essentially as described by Mogami et al. (23). Myosin was prepared by high-salt extraction of isolated myofibrils as described by Whalen et al. (40). EDTA was present at a concentration of 10 mM during the isolation of myofibrils and extraction of actomyosin, which probably resulted in the dissociation of DTNB light-chain peptides from the myosin (V. P. Parker, S. Falkenthal, and N. Davidson, manuscript in preparation).

Electrophoresis of RNA in formaldehyde gels. Electrophoresis of RNA in formaldehyde gels and transfer to nitrocellulose paper was performed as described by Rozek and Davidson (31). Prehybridization, hybridization with 10% dextran sulfate, and the washing of filters after hybridization were done as described by Mullins et al. (25).

Positive selection and translation of RNA. RNA was selected by a procedure similar to that described by Ricciardi et al. (30). Bacteriophage DNA (10 μg) was denatured by heating and applied to nitrocellulose filters (0.3 by 0.3 cm) in 10× SSC (1× SSC is 0.15 M NaCl plus 0.015 sodium citrate). Up to 15 filters containing different cloned *Drosophila* DNA sequences were prehybridized in 200 μl of 70% formamide–0.4 M NaCl–0.1 M PIPES [piperazine-*N*,*N*′-bis(2-ethanesulfonic acid)] (pH 6.5) for 2 h at 50°C. Hybridizations were done for 4 h at 50°C in 100 μl of the same buffer containing 100 to 150 μg of pupal poly(A)$^+$ RNA (70 to 75 h postpupariation). After hybridization, the filters were washed en masse 10 times in 5 ml of 1× SSC–0.5% sodium dodecyl sulfate at 65°C, then three times in 5 ml of 0.01 M Tris (pH 7.8)–1 mM EDTA at room temperature. The filters were then placed in individual vials, and the hybridized RNA was eluted in boiling distilled water and recovered by ethanol precipitation. Translation was in a commercial (Bethesda

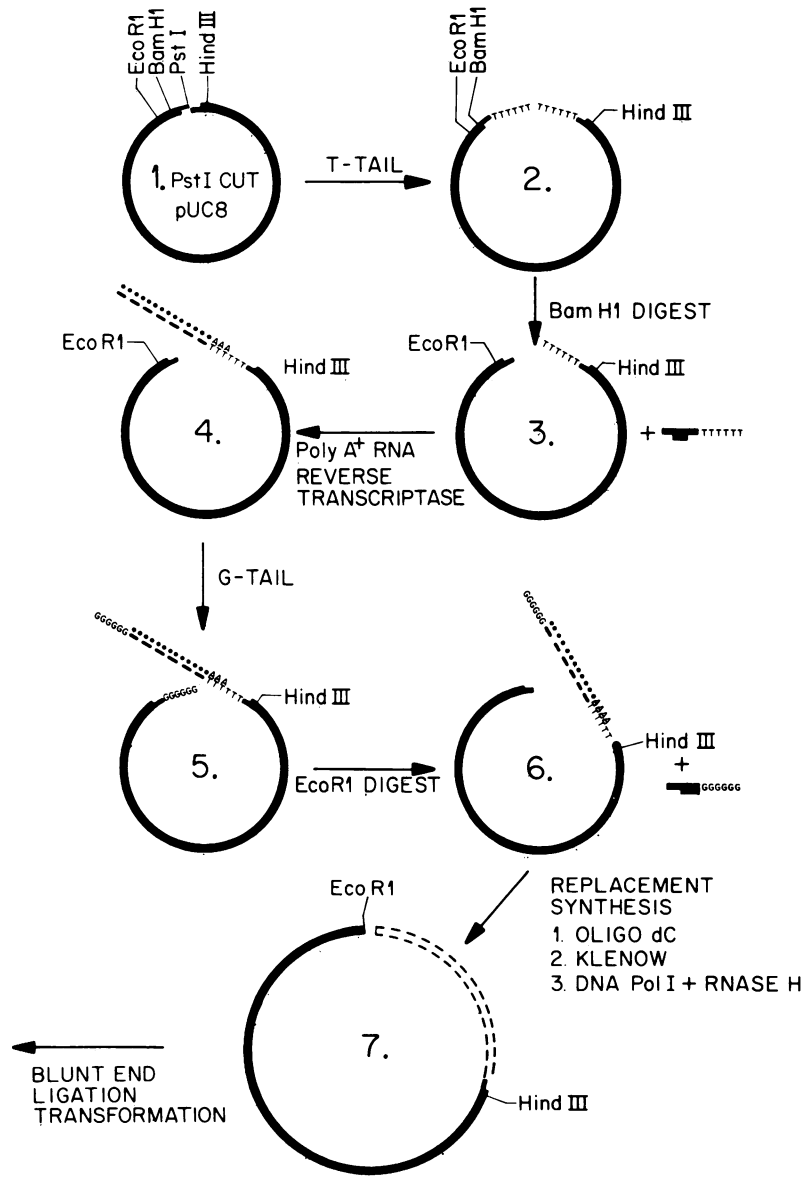

FIG. 1. Diagrammatic representation of the cDNA cloning strategy.

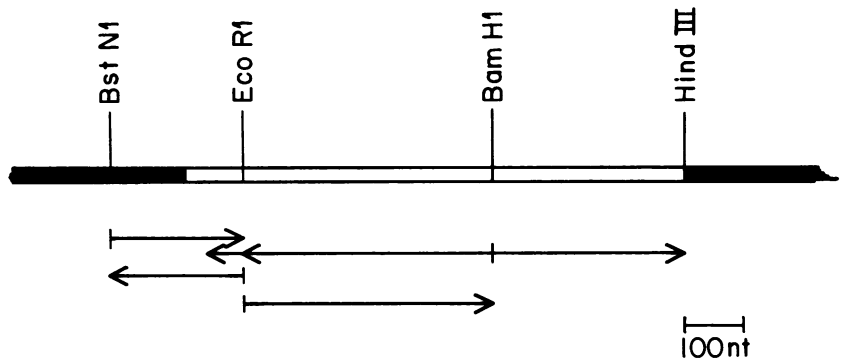

FIG. 2. Sequencing strategy of cDNA clone pcMLC-1.10. The DNA sequencing strategy is such that a vertical line (|) denotes ^{32}P labeling at the 3′ terminus. The arrows indicate the direction and extent of sequencing from the labeled terminus.

Research Laboratories) micrococcal nuclease-treated rabbit reticulocyte translation system. For each 30-μl translation, 50 μCi of 800-Ci/mmol [^{35}S]methionine (New England Nuclear) was used. Translation was terminated after 60 min by the addition of 50 ng each of RNase A and DNase I and subsequent incubation at 4°C for 30 min.

One-third of each translation assay was analyzed on a two-dimensional polyacrylamide gel as described by O'Farrell (27). The lysis buffer was modified to contain 0.1% sodium dodecyl sulfate (6). Molecular weights were determined by electrophoresing ^{14}C-labeled protein standards (Bethesda Research Laboratories) in an adjacent slot. Radioactively labeled proteins were detected by fluorography by the method of Laskey and Mills (15). Dried gels were exposed to preflashed Kodak XAR X-ray film.

DNA sequencing. The DNA sequence of a cDNA clone was determined by the method of Maxam and Gilbert (20) with modifications as described by Snyder et al. (34). The sequencing strategy used is shown in Fig. 2. Regions which were sequenced for one strand only were confirmed by a comparison with the DNA sequence of the chromosomal gene (S. Falkenthal, V. P. Parker, and N. Davidson, manuscript in preparation).

Amino acid homology comparisons. One homology comparison was carried out with a standard dot matrix program that asks for matches of five out of eight contiguous amino acids, but accepts conservative amino acid replacements (Val, Ileu, and Leu; Arg and Lys; Gln and Asn; Glu and Asp) as a match. In practice, this provides a rather stringent comparison of amino acid sequence homology (T. Hunkapiller, personal communication). The second two-dimensional matrix analysis, denoted the "best-fit" program, is useful for comparisons of sequences which are more distantly related. A match of the central amino acid is given a score of one, and matches at increasing distances from the center are given a reduced added score [M. Hunkapiller, S. Kent, M. Carruthers, W. Dreyer, J. Firca, C. Griffin, S. Horvath, T. Hunkapiller, P. Tempst, and L. Hood, Nature (London), in press].

RESULTS

Isolation of genes abundantly expressed in the indirect flight muscle. Our approach for isolating the myosin light-chain genes was based on the assumption that the concentration of their mRNAs would be high during the developmental stage when maximal synthesis of adult musculature occurs. Protein labeling studies showed that this occurred 70 to 75 h after puparium formation (data not shown). In vitro translation of RNA isolated from dissected thoraces at this developmental stage revealed that greater than 50% of the incorporation of [^{35}S]methionine was into myofibrillar proteins (data not shown). Accordingly, cDNA synthesized from pupal thoracic poly(A)$^+$ RNA was used to screen a Canton S random shear library of *Drosophila* genomic DNA in the vector Charon 4 (17). The 73 positive phages so selected were all plaque purified. These isolates were counterscreened with [^{32}P]cDNA homologous to early embryo RNA, a developmental time in which muscle-specific genes are not expressed, and with actin and myosin heavy-chain probes. The isolates which screened negatively with the above hybridization probes were rescreened with [^{32}P]cDNA synthesized from RNA extracted from the dorsal-lateral indirect flight muscle of late pupae (70 to 75 h). By restriction mapping, the resulting positive clones fell into 24 groups of nonoverlapping DNA inserts.

Initial characterization of these clones involved cytological localization of the DNA inserts by in situ hybridization to salivary gland chromosomes (Table 1). It may be seen that these inserts, representing genes abundantly expressed in the indirect flight muscle, do not show a general pattern of clustering on the *Drosophila* chromosomes. Tight clustering, as has been demonstrated for the cuticle genes on the second

TABLE 1. In situ localization of λdmpT recombinant clones

Clone	Band
31[a]	100B
49	28C
50[a]	97A
57[a]	99E
61[a]	17A
63	36B
73	88F
75	30B
85[a]	64C
101	102EF
102	98B
104	30EF
106	Repeated
115	72DE
116	53F
120	64F
121[a]	66F
123	94E

[a] Clone blots and in vitro translation of hybrid-selected RNA indicated that more than one gene is contained within the *Drosophila* insert (data not shown).

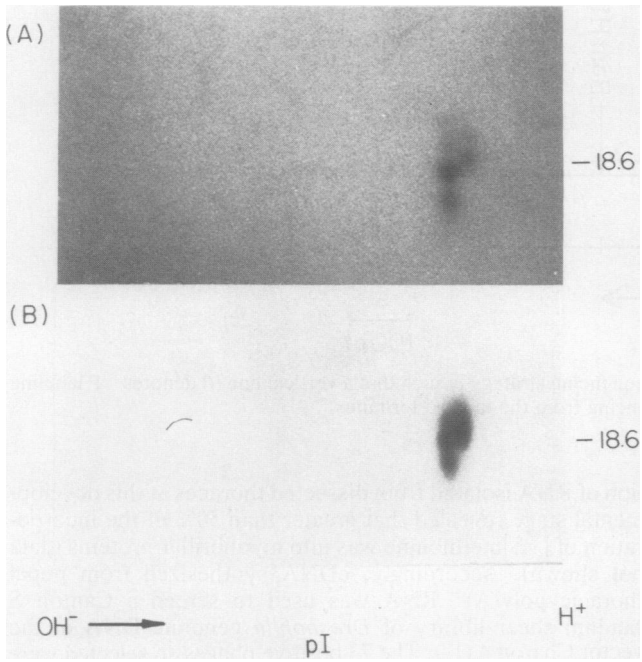

FIG. 3. Translation of myosin light-chain mRNA in a rabbit reticulocyte lysate. Myosin light-chain mRNA was selected by hybridization to filter bound λdmpT102 DNA as described in the text. This RNA was translated in an mRNA-dependent rabbit reticulocyte lysate containing [^{35}S]methionine, and the translation products were coelectrophoresed with 5 μg of *Drosophila* myosin purified from adult myofibrils on a 16% polyacrylamide gel. (A) Coomassie brilliant blue-stained gel. (B) Autoradiographic exposure (36 h) of the gel in A.

chromosome (33, 34) and the glue protein genes on the third chromosome (21), is not excluded by this analysis. Clone blots with labeled pupal cDNA as a hybridization probe and in vitro translation of hybrid-selected RNA revealed that 11 of the inserts contained only one gene which was expressed during pupal myogenesis, whereas the other inserts contained two or possibly three separate but closely linked transcription units (data not shown). Therefore, there is some tight clustering of genes abundantly expressed in the indirect flight muscle. Only one clone, λdmpT73, which probably contains a tropomyosin gene (3) hybridized in situ to the same region, 88F, where an actin gene expressed in the indirect flight muscle maps (12). Mutations which result in dominant flightless behavior have been found at regions 36B, where the myosin heavy-chain gene maps, and 88F (5, 23). None of our other clones mapped in these two regions.

Identification of the clone which contained the myosin light-chain sequence. The DNA insert which most probably encodes a myosin light-chain gene was identified by hybrid selection of RNA and in vitro translation (see above). The resulting ^{35}S-labeled polypeptides were compared to purified *Drosophila* myosin light-chain protein by two-dimensional gel electrophoretic analysis. The in vitro translation products of two of these clones (λdmpT75 and λdmpT102) had molecular weights of 17,000 to 20,000, the molecular weight range expected for the myosin light-chain protein. Whereas the polypeptide encoded by λdmpT75 had a more basic isoelectric point (data not shown), that encoded by λdmpT102 had the identical electrophoretic mobility of the extracted myosin light-chain protein (Fig. 3). The molecular weight of the latter protein(s) ranged from 18,000 to 19,000. Note that there is heterogeneity in both the molecular weight and the isoelectric point of the myosin light-chain isolated from adult muscle as well as for the in vitro-synthesized translation products.

Isolation and sequencing of a myosin light-chain cDNA clone. To achieve a more positive identification of the protein encoded by the insert of λdmpT102, a cDNA clone for which it codes was isolated and sequenced. Initially, the coding region of λdmpT102 was localized onto 3.2 kb of DNA by clone blots with [^{32}P]cDNA from total pupal poly(A)$^+$ RNA as a hybridization probe. This region contains two contiguous restriction fragments, a 1.8-kb *Hin*dIII-*Eco*RI fragment and a 1.35-kb *Eco*RI fragment. These two fragments were subcloned (see Fig. 4 for relevant restriction maps) and used to screen a cDNA library.

The cDNA library was prepared from late pupal poly(A)$^+$ RNA by a simplified procedure modeled on that of Okayama and Berg (28) to increase the probability of obtaining long inserts. Of 10,000 recombinants screened (5% of the total library), 10 clones containing sequences homologous to the insert of p102.6 were obtained. The clone, pcMLC-1.10, which contained the longest insert (880 nucleotides), was

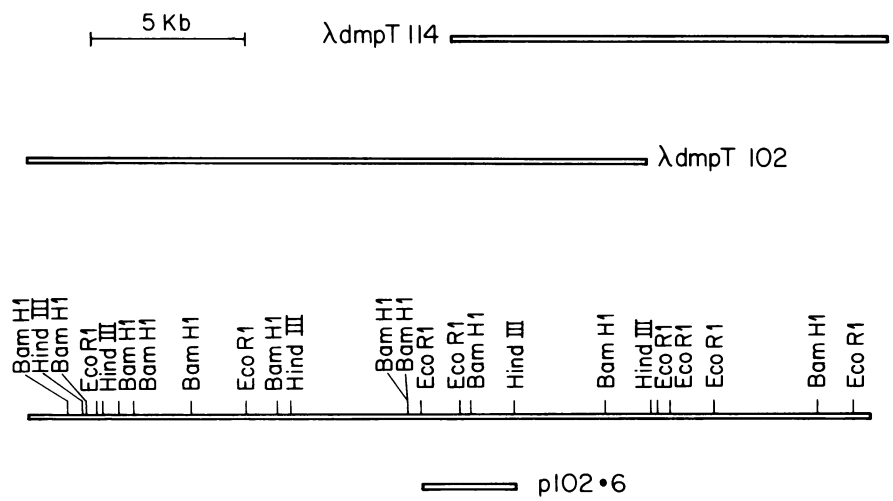

FIG. 4. Restriction endonuclease map of λdmpT102 and λdmpT114. A composite map of the two bacteriophage lambda clone inserts is shown. The subcloned coding region is shown below the map.

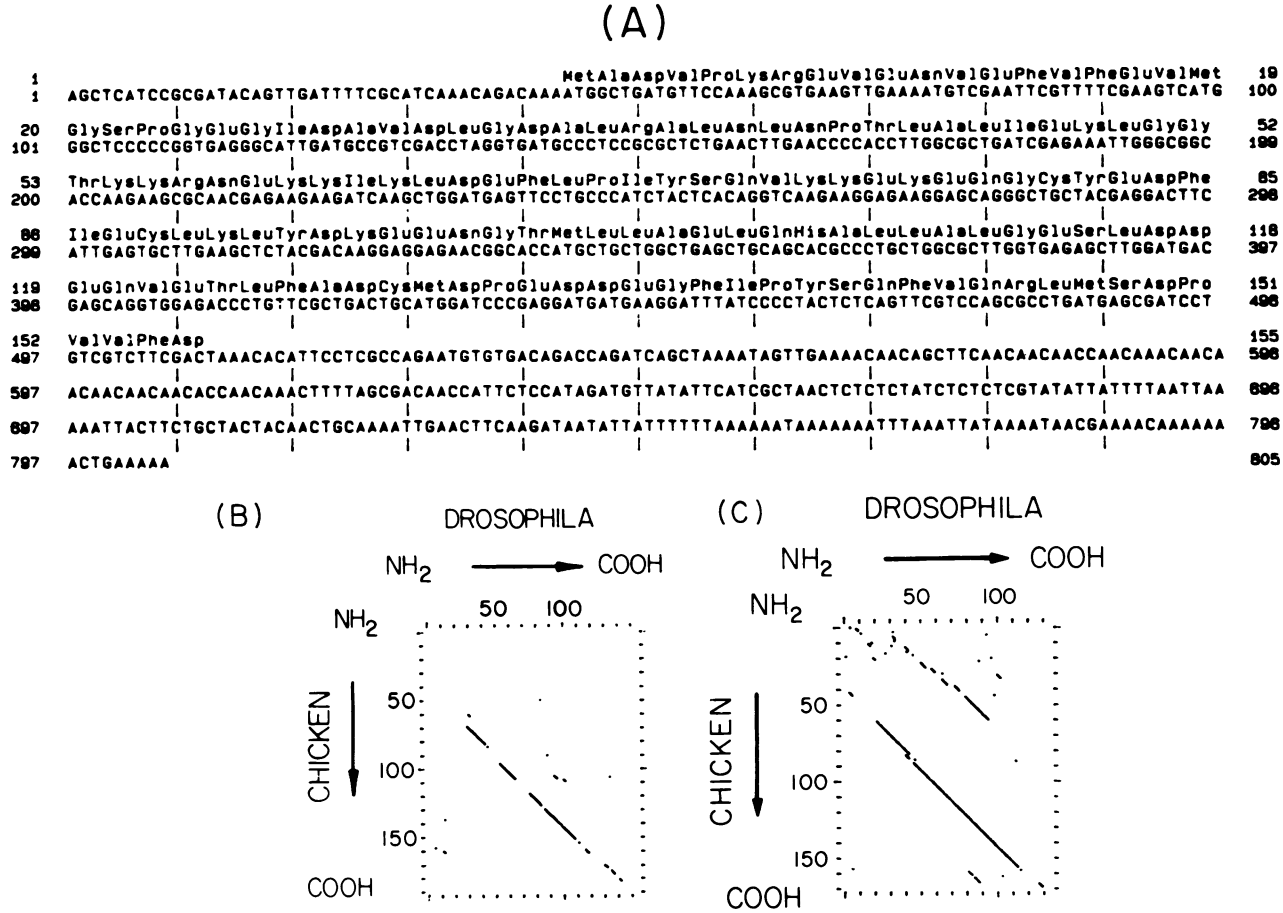

FIG. 5. Predicted protein coding sequence of the insert of pcMLC-1.10. (A) The amino acid sequence of the putative myosin light chain is depicted over the encoding DNA sequence of the 880-nucleotide insert of pcMLC-1.10. The DNA sequence was determined as described in the text. (B) Homology search between the predicted amino acid sequence of the *Drosophila* protein and the amino acid sequence of chicken skeletal muscle myosin light chain-1 protein. The comparison is such that every eight amino acids of the *Drosophila* protein are compared with the chicken protein (a positive score is marked if five out of eight amino acids match). A diagonal line indicates homology. (C) A best-fit comparison between the predicted amino acid sequence of the *Drosophila* protein and the chicken skeletal muscle myosin light chain-1 protein calculated by the method of Hunkapiller et al. (in press). The comparison is such that 10 amino acids of the Drosophila protein are compared with the chicken protein, allowing conservative amino acid replacements. The confidence limit for this fit was set at 99%.

sequenced (Fig. 5A). An AUG codon at nucleotide 44 begins a reading frame 456 nucleotides in length, which terminates with a TAA at nucleotide 509. Downstream from this termination codon are two additional in-frame termination signals (TAA and TGA at nucleotides 552 and 561, respectively).

The amino acid sequence of this reading frame was deduced and compared with the amino acid sequence of bovine brain calmodulin (37), rabbit troponin C (7), chicken skeletal muscle myosin light chain-1 (MLC-1) (19), and chicken skeletal muscle myosin light chain-2 (MLC-2) (18). These evolutionarily related proteins all possess the divalent metal binding (Ca^{2+}, Mg^{2+}) structure referred to as the EF hand (14).

By the stringent dot matrix amino acid homology comparison described above, only the chicken skeletal muscle myosin alkali light chains (MLC-1 and MLC-3) gave a significant positive score with the *Drosophila* protein (Fig. 5B). This fact and the in vitro translation study described above are the primary bases on which we identified the gene as that for the *D. melanogaster* alkali light chain.

In all regions where the dot matrix method showed homology between the *Drosophila* protein and the chicken MLC-1 protein, the latter is identical in sequence with the chicken MLC-3 protein (see Fig. 8 and below). The *Drosophila* protein has approximately the same molecular length as does the chicken MLC-3 protein, without the amino-terminal tail of the MLC-1 protein. A more detailed comparison of these sequences is presented later. We note here simply that overall the region of homology extends from amino acids 32 to 144 of the *Drosophila* protein with amino acids 72 to 167 and 31 to 143 of chicken MLC-1 and MLC-3, respectively (see Fig. 8). On this basis, we identified this gene as the *Drosophila* myosin alkali light chain (MLC-ALK).

The *Drosophila* MLC-ALK and chicken MLC-1 sequences were compared by the best-fit program which is designed to identify more distant sequence relationships (Fig. 5C). This comparison confirmed the major segment of homology noted above. In addition, it revealed that the sequence from amino acids 90 to 105 of *D. melanogaster* is related to the amino terminus proximal sequence from amino acids 56 to 71 of chicken MLC-1.

Developmental expression of the MLC-ALK gene. The MLC-ALK gene was selected by using a probe isolated from the pupal stage of development. The question arose as to whether the MLC-ALK gene was transcribed at other times in development, particularly at other times of muscle biosynthesis. To answer this question, total poly(A)$^+$ RNA was

isolated from animals at various developmental stages, electrophoresed on denaturing formaldehyde agarose gels, transferred to nitrocellulose paper, and hybridized with nick-translated MLC-ALK-coding DNA (p102.6). The results (Fig. 6) demonstrated that the MLC-ALK gene is transcribed during late embryogenesis (the time of formation of the larval musculature), during the larval instars (a time of rapid tissue, and therefore muscle, growth), and during the late stages of pupariation (the time of rapid synthesis of the indirect flight muscle and other adult muscles). There was no accumulation of MLC-ALK transcripts during early pupal development, the time of histolysis of larval musculature (8). The most abundant RNA showed a broad distribution with a mean molecular length of around 0.95 kb. There is a less intense band at 1.3 kb, which was most evident in Fig. 6 in the late pupal indirect flight muscle RNA lane, because of the greater fraction of muscle-specific RNA in this preparation. A 3.2-kb band which we believe is an unspliced nuclear transcript (unpublished data) is faintly visible in several lanes.

Reiteration frequency of the MLC-ALK gene. There are indications of multiple polypeptides in the in vitro translation products of RNA selected by the MLC-ALK gene we isolated (Fig. 3). Additionally, cross-hybridizing RNAs are expressed in several stages of development (Fig. 6). Similar results were observed for the actin genes of *D. melanogaster*, which comprise a small multigene family (11), and for the myosin heavy-chain gene which is single copy (5, 31). If the MLC-ALK gene is reiterated, all of the copies must reside at the chromosomal region 98B (Table 1), because this is the only site labeled by in situ hybridization.

We carried out genome blot studies to resolve this question. The gel blots were hybridized to the MLC-ALK probes at a moderate stringency. Quantitative comparisons of intensity were made by genome reconstruction experiments. All these data (Fig. 7) show that at the moderate stringency of the hybridization experiments there is only one MLC-ALK gene per haploid genome.

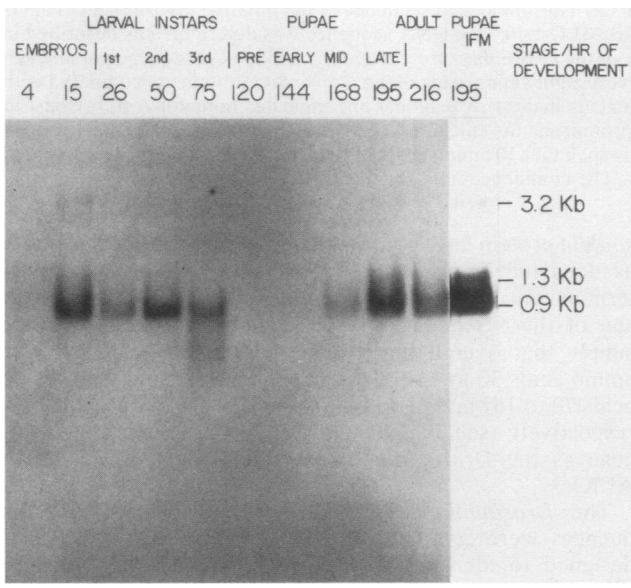

FIG. 6. Developmental expression of myosin alkali light chain mRNAs. Poly(A)$^+$ RNA (1 μg) from different developmental stages of synchronized populations of *D. melanogaster* was subjected to electrophoresis on 1.5% agarose gels containing 2.2 M formaldehyde. The RNA was transferred to nitrocellulose and hybridized with a nick-translated probe synthesized from p102.6 DNA. Time (in hours) was calculated from the time of egg deposition. Pupae were resynchronized by floatation on water at 5 h post-pupariation (22). *E. coli* 16S and 23S rRNAs as well as HeLa 28S and 18S rRNAs were used as length standards.

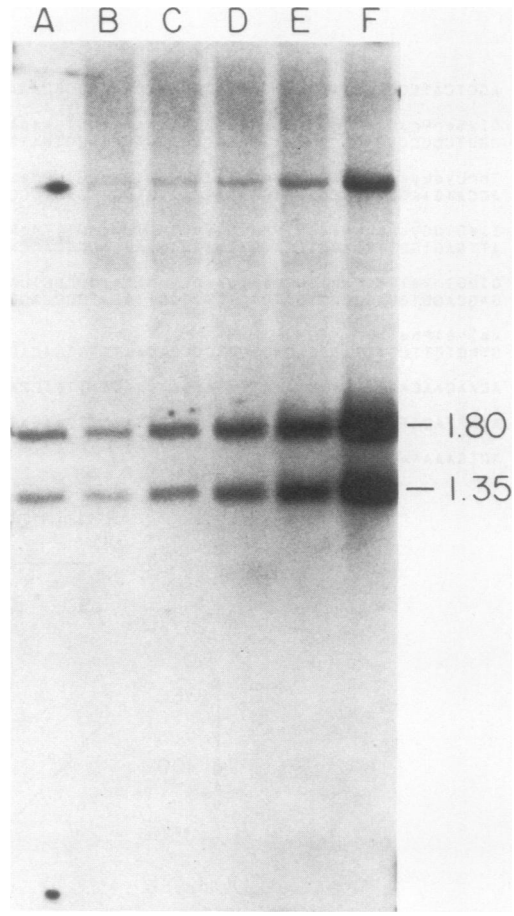

FIG. 7. Genomic representation of the coding region of λdmpT102. (A) Genomic *Drosophila* Canton S DNA (4 μg) was digested with *Eco*RI and *Hin*dIII, and the resulting fragments were separated on 1% agarose gels. In adjacent lanes of the same gels, amounts of *Eco*RI-*Hin*dIII-restricted λdmpT102 DNA equivalent to (B) 0.5, (C) 1.0, (D) 3.0, (E) 5.0, and (F) 25.0 copies per genome (a haploid genome size of 1.6×10^8 base pairs was assumed) plus 4 μg of *Eco*RI-*Hin*dIII-restricted *E. coli* DNA were subjected to electrophoresis. After electrophoresis, the fragments were transferred to nitrocellulose and hybridized with ^{32}P-labeled nick-translated p102.6 DNA.

DISCUSSION

We cloned a *D. melanogaster* gene coding for a myosin alkali light-chain protein. The gene is expressed to provide a relatively high concentration of RNA during those specific times in development when extensive myofibrillar assembly occurs, and it is not expressed before larval myogenesis or during early pupariation when larval muscles are undergoing histolysis (8). We showed by in vitro translation of hybrid selected RNA that the gene encodes a polypeptide which comigrates in two-dimensional gels with myosin alkali light-chain protein extracted from adult indirect flight muscle. The strongest evidence that the gene encodes the MLC-ALK protein was obtained from the nucleotide sequence of a cDNA clone derived from the mRNA transcript of the gene.

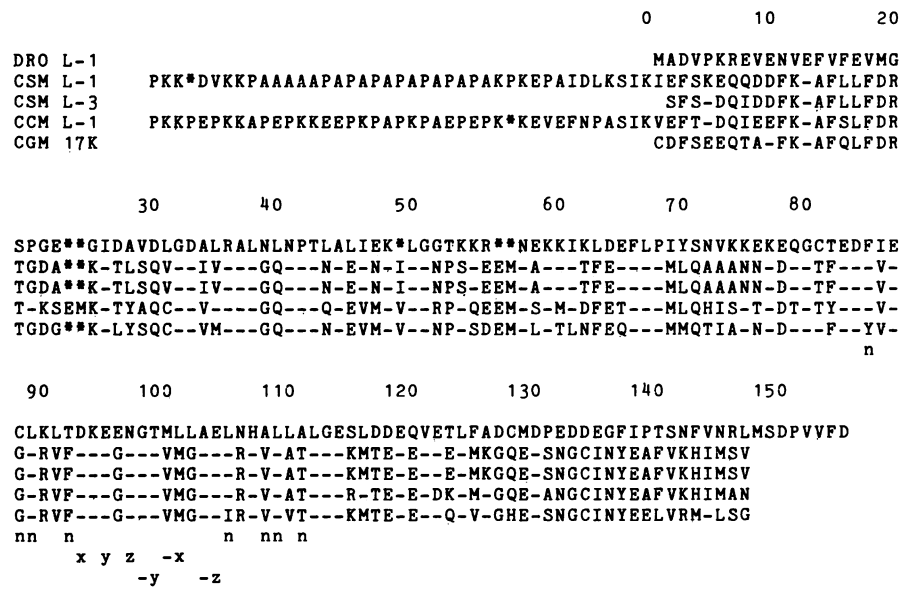

FIG. 8. Comparison of the amino acid sequence of *Drosophila* myosin light chain with the sequence of chicken myosin light chains (7, 19). Abbreviations: DRO L-1, *Drosophila* myosin alkali light chain; CSM L-1, chicken skeletal muscle myosin light chain-1; CSM L-3, chicken skeletal muscle myosin light chain-3; CCM L-1, chicken cardiac muscle myosin light chain-1; CGM 17K, chicken gizzard muscle 17,000-dalton myosin light chain. *Drosophila* amino acids identical to those of the chicken myosin light chains are indicated by a hyphen. The absence of a particular amino acid in the sequence is denoted by an asterisk; n indicates those residues which form the core of the E and F α helices. X, Y, Z, −X, −Y, and −Z are residues which might be involved in the binding of divalent metal ions in EF hand domain III.

The corresponding amino acid sequence showed a high degree of sequence homology with that of the chicken alkali light chains and much less homology with any other of the evolutionarily related calcium ion-binding proteins.

The consensus EF hand structural domain of the calcium-binding proteins consists of an α helix of 9 amino acids (the E helix), a loop of 12 amino acids (6 of which have side chains containing oxygen atoms which are capable of octahedral coordination to a Ca^{2+} ion), and an α helix of eight amino acids (the F helix) (14). This structure is denoted n(2), n(5), n(6), n(9), X(10), Y(12), Z(14), −Y(16), −X(18), −Z(21), n(22), n(25), n(26), n(28), n(29). The numbers denote the relative positions of residues, where n represents the hydrophobic or inner core of the α helix and X, Y, Z, −X, −Y, and −Z represent the calcium ligands. The amino acids at these positions are variable from protein to protein. To bind Ca^{2+}, there must be at least four aspartic and glutamic acid residues in the calcium loop, but others may be serine, threonine, asparagine, and glutamine. In some calcium-binding proteins, one of the six positions is occupied by glycine which is capable of coordinating Ca^{2+} through the oxygen of a hydrogen-bonded water molecule (14). The vertebrate myosin alkali light chains do not, in fact, bind Ca^{2+} in vitro (39). However, because of the very similar amino acid composition of their EF hand domains with those of other calcium-binding proteins, they are included in this group.

A comparison of the amino acid sequence of the *Drosophila* MLC-ALK gene with those of the highly homologous chicken skeletal, cardiac, and gizzard alkali light chains is shown in Fig. 8. The region of greatest homology between the chicken proteins and the *Drosophila* protein extends from amino acids 31 to 124 of the latter. There is an EF hand (denoted EF hand III) of the chicken proteins in a region of very high sequence homology with the *Drosophila* protein, amino acids 85 to 112. The chicken MLC-ALK proteins have four domains which are identified as resembling the EF hand domains of other calcium-binding proteins. However, due to an insertion of two amino acids in the Ca^{2+} loop of EF hand domain II, it is believed that this structure has been disrupted. Weeds et al. (39) note that only the EF hand in domain III has four acidic residues among its six ligating groups.

Sequence comparisons of aligned amino acids for chicken MLC-3 and the *Drosophila* protein are shown in Table 2. The highest degree of homology is in domain III, with a moderate degree of homology in domain II and very low homology in domains I and IV. Analysis of domain II did not show a good EF hand structure. Therefore, we believe that the only potential calcium-binding site of the *Drosophila* protein is that found in the region homologous to domain III of the chicken proteins.

The best-fit matrix analysis showed a definite but more distant relationship between domain III of the *Drosophila* protein and domain I of the chicken protein (Fig. 5C), calmodulin, and troponin C (data not shown). This is reasonable because the four Ca^{2+}-binding domains of calcium-binding proteins are thought to have arisen from an ancestral Ca^{2+} binding domain that underwent two gene duplication events followed by sequence divergence (14). Similar comparisons show sequence homology between this domain of the *Drosophila* protein with the EF hand III domains of bovine brain calmodulin and of rabbit skeletal muscle troponin C (data not shown). This EF hand domain contains the high-affinity Ca^{2+}–Mg^{2+}-binding site.

We conclude that the *Drosophila* alkali light chain contains one EF hand domain which is related in sequence to those of the family of calcium-binding proteins.

The *Drosophila* MLC-ALK gene is present at only one copy per haploid genome and is transcribed during larval and pupal myogenesis. In this respect, it is analogous to the myosin heavy-chain gene (5, 31). However, the latter shows developmental differences in transcript lengths and splicing patterns (31) which lead to developmental differences in amino acid sequences (C. Rozek, personal communication).

TABLE 2. Homology by domain of the chicken and *Drosophila* proteins

Domain (amino acids)	% Homology		% Protein homology allowing conservative amino acid changes	% DNA homology discounting third-base changes	% Nonhomology due only to third-base changes	Protein homology/DNA homology
	Protein[a]	DNA[b]				
I (1–32)	15	34	21	39	8	0.44
II (33–76)	46	52	50	59	15	0.88
III (77–115)	58	57	70	68	33	1.02
IV (116–157)	15	27	18	30	4	0.55

[a] Homology between the chicken skeletal muscle myosin light chain-3 and *Drosophila* myosin alkali light chain (Fig. 4).
[b] Homology between the *Drosophila* myosin light-chain sequence and that of chicken skeletal muscle myosin light chain-3 cDNA clone (26).

There are indications of heterogeneity in molecular weight and isoelectric point for the *Drosophila* MLC-ALK in vitro translation product and for this protein extracted from adult indirect flight muscle. This heterogeneity could be due to post-translational modification, such as phosphorylation. Alternatively, or in addition, there could be subtle differences in transcripts, thereby encoding proteins which differ slightly in amino acid composition, which are not revealed by the RNA gel blots. We have evidence that some of this heterogeneity is due to differential splicing generating transcripts encoding two proteins which differ at their carboxy-terminal ends. Studies of these and related questions dealing with the fine structure of the gene will be reported in a later communication (Falkenthal et al., in preparation).

ACKNOWLEDGMENTS

We thank E. Fyrberg and C. Rozek for their generous gifts of actin and myosin heavy-chain-coding DNAs and T. Hunkapiller for help in the execution of the protein homology comparisons. We also thank H. Mitchell and N. Peterson for guidance and direction in pupal dissection and sharing their results of pupal indirect flight muscle development.

S.F. was supported by a Muscular Dystrophy Postdoctoral Fellowship and NRSA Fellowship 1 F32 GM08895-01. The research was supported by Public Health Service grants from the National Institutes of Health.

LITERATURE CITED

1. **Anderson, K. V., and J. A. Lengyel.** 1979. Rates of synthesis of major classes of RNA in Drosophila embryos. Dev. Biol. **70**:217–231.
2. **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 Spring, Md.
3. **Bautch, V. L., R. V. Storti, D. Mischke, and M. L. Pardue.** 1982. Organization and expression of Drosophila tropomyosin genes. J. Mol. Biol. **162**:231–250.
4. **Benton, W. D., and R. W. Davis.** 1978. Screening λgt recombinant clones by hybridization to single plaques in situ. Science **196**:180–182.
5. **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**:393–397.
6. **Brandhorst, B. P.** 1976. Two-dimensional gel patterns of protein synthesis before and after fertilization of sea urchin eggs. Dev. Biol. **52**:310–317.
7. **Collins, J. H., J. D. Potter, M. J. Horn, G. Wilshire, and N. Jackman.** 1973. The amino acid sequence of rabbit skeletal muscle troponin C: gene replication and homology with calcium binding proteins from carp and hake muscle. FEBS Lett. **36**:268–272.
8. **Crossley, A. C.** 1978. The morphology and development of the Drosophila muscular system, p. 499–560. *In* M. Ashburner and T. R. F. Wright (ed.), The genetics and biology of Drosophila, vol. 2b. Academic Press, Inc., New York.
9. **Deng, G., and R. Wu.** 1981. An improved procedure for utilizing terminal transferase to add homopolymers to 3′ termini of DNA. Nucleic Acids Res. **9**:4173–4188.
10. **Ebashi, S.** 1963. Third component participating in the superprecipitation of "natural actomyosin." Nature (London) **200**:1010–1012.
11. **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.
12. **Fyrberg, E. A., J. W. Mahaffey, B. J. Bond, and N. Davidson.** 1983. Transcripts of the six Drosophila actin genes accumulate in a stage and tissue specific manner. Cell **33**:115–123.
13. **Klee, C. B., T. H. Crouch, and P. G. Richman.** 1980. Calmodulin. Annu. Rev. Biochem. **49**:489–515.
14. **Kretsinger, R. H.** 1980. Structure and evolution of calcium modulated proteins. Crit. Rev. Biochem. **8**:119–174.
15. **Laskey, R. A., and A. D. Mills.** 1975. Quantitative fluorography of acrylamide gels. Eur. J. Biochem. **56**:335–341.
16. **Lowey, S., and D. Risby.** 1971. Light chain from fast and slow muscle myosins. Nature (London) **234**:81–85.
17. **Maniatis, T., R. C. Hardison, E. Lacy, Jr., J. Lauer, C. O'Connell, D. Quon, G. K. Sim, and A. Efstratiadis.** 1978. The isolation of structural genes from libraries of eukaryotic DNA. Cell **15**:687–702.
18. **Matsuda, G., T. Maita, Y. Kato, J. Chen, and T. Umegane.** 1981. Amino acid sequences of the cardiac L-2A, L-2B, and gizzard 17,000 M light chains of chicken muscle myosin. FEBS Lett. **135**:232–236.
19. **Matsuda, G., T. Maita, and T. Umegane.** 1981. The primary structure of L-1 light chain of chicken fast skeletal muscle myosin and its genetic implication. FEBS Lett. **126**:111–113.
20. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
21. **Meyerowitz, E. M., and D. S. Hogness.** 1982. Molecular organization of a Drosophila puff site that responds to ecdysone. Cell **28**:165–176.
22. **Mitchell, M. K., and A. Mitchell.** 1964. Mass culture and age selection in Drosophila. Dros. Inf. Ser. **39**:135–137.
23. **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.
24. **Mullins, J. I., J. W. Casey, M. O. Nicolson, K. B. Burck, and N. Davidson.** 1981. Sequence arrangement and biological activity of cloned feline leukemia virus proviruses from a virus-productive human cell line. J. Virol. **38**:688–703.
25. **Mullins, J. I., J. Casey, M. Nicolson, and N. Davidson.** 1980. Sequence organization of feline leukemia virus DNA in infected cells. Nucleic Acids Res. **8**:3287–3305.
26. **Nabeshima, Y., Y. Fuju-Kuriyama, M. Muramatsu, and K. Ogata.** 1982. Molecular cloning and nucleotide sequences of the complementary DNAs to chicken skeletal muscle myosin two alkali light chain mRNAs. Nucleic Acids Res. **10**:6099–6110.
27. **O'Farrell, P. H.** 1975. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. **250**:4007–4021.
28. **Okayama, H., and P. Berg.** 1982. High efficiency cloning of full length cDNAs. Mol. Cell. Biol. **2**:161–170.
29. **Pemrick, S. M.** 1980. The phosphorylated L2 light chain of

skeletal myosin is a modifier of the actomyosin ATPase. J. Biol. Chem. **255**:8836–8841.
30. **Ricciardi, R. P., J. S. Miller, and B. E. Roberts.** 1979. Purification and mapping of specific mRNAs by hybridization-selection and cell-free translation. Proc. Natl. Acad. Sci. U.S.A. **76**:4927–4931.
31. **Rozek, C. E., and N. Davidson.** 1983. Drosophila has one myosin heavy chain gene with three developmentally regulated transcripts. Cell **32**:23–34.
32. **Sarkar, S., F. A. Streter, and J. Gergely.** 1971. Light chains of myosins from white, red, and cardiac muscles. Proc. Natl. Acad. Sci. U.S.A. **68**:946–950.
33. **Snyder, M., J. Hirsh, and N. Davidson.** 1981. The cuticle genes of Drosophila: a developmentally regulated gene cluster. Cell **25**:165–177.
34. **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.
35. **Ullrich, A., J. Shine, J. Chirguin, R. Pictet, E. Tischer, W. J. Rutter, and H. M. Goodman.** 1977. Rat insulin genes: construction of plasmids containing the coding sequences. Science **196**:1313–1319.
36. **Wagner, P. D., and E. Giniger.** 1981. Hydrolysis of ATP and reversible binding to F-actin by myosin heavy chains free of all light chains. Nature (London) **292**:560–562.
37. **Watterson, D. M., F. Sharief, and T. Vanaman.** 1980. The complete amino acid sequence of the calcium dependent modulator protein (calmodulin) of bovine brain. J. Biol. Chem. **255**:962–975.
38. **Weeds, A. G., and S. Lowey.** 1971. Substructure of the myosin molecule II. The light chains of myosin. J. Mol. Biol. **61**:701–725.
39. **Weeds, A., P. Wagner, R. Jakes, and J. Kendrick-Jones.** 1977. Structure and function of myosin light chains, p. 222. *In* R. H. Wasserman et al. (ed.), Calcium binding proteins and calcium function. Elsevier/North-Holland Publishing Co., Amsterdam.
40. **Whalen, R. G., G. S. Butler-Browne, and F. Gros.** 1978. Identification of a novel form of myosin light chain present in embryonic muscle tissue and cultured muscle cells. J. Mol. Biol. **126**:415–431.
41. **Yen, P. H., and N. Davidson.** 1980. The gross anatomy of a tRNA gene cluster at region 42A of the Drosophila melanogaster chromosome. Cell **22**:137–148.