Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 449–453, January 1985
Developmental Biology

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

(*Drosophila melanogaster*/differential processing/muscle)

SCOTT FALKENTHAL*†, VANN P. PARKER‡§, AND NORMAN DAVIDSON*

*Department of Chemistry and ‡Division of Biology, California Institute of Technology, Pasadena, CA 91125

Contributed by Norman Davidson, August 31, 1984

ABSTRACT **The total sequence of the *Drosophila melanogaster* gene encoding the myosin light chain dissociated by alkali (MLC-ALK) has been determined. By sequence comparisons with an MLC-ALK cDNA clone and by S1 nuclease analyses, the pattern of introns and exons within the gene has been deduced. There are multiple polyadenylylation signals that can account for most of the observed heterogeneity in the lengths of mRNAs. In the 3′ half of the gene, there are two alternative splicing patterns which result in mRNAs that translate to give proteins with two alternative 14 amino acid carboxyl-terminal sequences. There is developmental regulation of the selection of the above splicing sites. One splicing pattern produces an mRNA that translates into a protein used for both larval and adult musculature, whereas the other splicing pattern is used for the latter stage only.**

Drosophila melanogaster is an excellent model system for studies of the structure and function of gene products such as myofibrillar proteins, because one may use both molecular and genetic approaches. To date a number of genes coding for the major myofibrillar proteins have been cloned and characterized. The actin gene family comprises six genes, which are dispersed in the genome (1–3); each gene exhibits a unique tissue and temporal pattern of expression during development (4). On the other hand, *Drosophila* has only one gene encoding myosin heavy chain (MHC) (5, 6) and, as reported by us previously (7), only one gene encoding the myosin light chain that is dissociated by alkali (MLC-ALK). A question of interest then is whether there is only one protein encoded by each gene, so that the several different muscle tissues of the larva and the adult, which use slightly different actin proteins, all use the same MHC and MLC-ALK proteins in their myofibrils.

In the present communication, we report that this is not the case. We find that there is a developmental difference in splicing patterns of the primary transcript of the MLC-ALK gene, resulting in at least two slightly different polypeptides encoded by and expressed by the single gene. A similar situation occurs for the MHC gene (C. Rozek, personal communication).

EXPERIMENTAL PROCEDURES

Isolation and S1 Nuclease Analysis of RNA. Total cellular RNA was prepared by homogenizing developmentally staged whole animals in 4 M guanidine thiocyanate/1 M 2-mercaptoethanol/0.05 M NaOAc/0.001 M EDTA, pH 6.0, and banding in cesium chloride (2, 8). Poly(A)$^+$ RNA was selected by oligo(dT)-cellulose (Collaborative Research, type T3) chromatography as described (9). S1 nuclease mapping of RNA transcripts was done essentially as described (6, 10).

Subcloning, Labeling, and Sequencing of DNA. *Hin*dIII/*Eco*RI fragments from the MLC-ALK λ clone dmpT102 (7) were subcloned in pBR322 by standard methods (11). Additional subclones were generated from one of the resulting plasmids, p102.6 (see Fig. 2), as follows: the insert was isolated from an agarose gel, digested with *Bam*HI and *Hin*dIII or digested with either *Sau*3A1 or *Taq* I, and ligated into *Bam*HI/*Hin*dIII-digested M13mp8 and M13mp9 phage DNA or ligated into *Bam*HI- or *Sal* I-digested M13mp8 replicative form DNA, respectively (11).

Preparation of nick-translated probes and hybridization of [^{32}P]DNA probes to filter-bound DNA were performed essentially as described (12). Nick-translated probes were hybridized at a concentration of 1×10^5 dpm/ml (10^8 dpm/μg).

The sequencing strategy used is shown in Fig. 2*C*. The sequence of nucleotides 1–1705 was determined according to Maxam and Gilbert (13) with modifications (14). Nucleotides 1705–3551 were determined by dideoxy sequencing methods of Sanger *et al.* (15). We had to use the chemical modification techniques of Maxam and Gilbert to determine the sequence of nucleotides 1–1705 due to an inverted repeat from nucleotides 1273–1285 and 1638–1650. Presumably, the indirect repeat formed a stable hairpin that prevented the procession of the Klenow fragment of DNA polymerase through this region in the dideoxy sequencing reactions.

RESULTS

Nucleotide Sequence of the MLC-ALK Gene. Our previous studies suggested that the primary transcript of the MLC-ALK gene lay entirely within a segment of length 3.5 kilobases (kb). There was heterogeneity in molecular length for both the encoded mRNAs and proteins, and there was heterogeneity in isoelectric point for the latter (7). We have sequenced the entire 3.5-kb region, to investigate further the structure and expression of the gene, as well as the nature of the mRNA and protein heterogeneity.

The sequence of length 3552 base pairs is shown in Fig. 1; all of the protein coding sequence was determined from sequencing both strands. Only the 5′ sequence of 150 nucleotides (nt) was determined from reading the sequence of just one strand. We have previously reported the sequence of a cDNA clone (denoted pcMLC.1-10) and interpreted it in terms of a protein sequence (7). By comparing the cDNA sequence with the genomic sequence, we have identified one set of exons of the latter. This interpretation is shown in Figs. 1 and 2. The comparison shows that the gene has a minimum of five introns ranging in length from 59 to 986 nt.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MHC, myosin heavy chain; MLC-ALK, myosin light chain dissociated by alkali; kb, kilobase(s); nt, nucleotide(s).

†Present address: Department of Genetics, 1735 Neil Avenue, Ohio State University, Columbus, OH 43210.
§Present address: Amgen, 1900 Oak Terrace Lane, Thousand Oaks, CA 91320.

```
tgatattcagcactggtccggctgggaatgccacaagcgattattatttgccttcacgaacgaaataaaagaaatattcacagaatagccgaaaaac   100
tcaaaatgcttataaatagctcggctagagagcggacaacaaatggctctggctccataggagagcctcccaatatctggtggtcgttgggatccatggg   200
atcccaaaagagagtggcattcgaaaaagttcccccaagatgaaagtaaattcttgccatcgcacatactacagcatatatttggtgtgtgtatagcct   300
tgcgcaggaaaagagcgagcgtttcgtcatccattcgtattcgtattcgttttgaattcgttttcgaagctatagctgtaaatgtatctgtatctgtgt   400
ctggtgctcggcgatcggctggttggctttcacctcagaattaccttgccgtactgccactgacgaataacgttcagtcgagaccgAGCTCATCCGCGAT   500
ACAGTTGATTTTCGCATCAAACAGACAAAATGgtgagtgcccagtcggccaaggagtcttcacacatcttccaccaatatccagtgatatagatgtacat   600
aaatatatgcagcagctgaaaactgaagatttccacactcgaactaatgccaaactctcaaatcttttcgattccgttgacagGTCGATGTTCCAAAG   700
CGTGAAGTTGAAAgtaagttaaaccccgcacgaaaaggattatctttaacctcaaacgaaatccatcaaagttaaaccaaataaaacatataaaatgaa   800
cgaggtaacggaaaacccagcagaaaggactaaaaacaagtatattaaaatatccataaaaatacagattacataatgtgcgcgtgtgagatacacaaca   900
aaaagtacacaaaagtcggtgaacatatttcgtaatcacttggaatatatcgcatacatatacgcatgcctatatatataacactatgcggaaaatttc  1000
attaaggaatgctgtcgtctatatgcatcgcctgagctatttcatcgatcgaccatccaaaaccagaacggataaaggttgccaacacaattccggtag  1100
agaaaatctagtgcaaagttaagtttgatttgcgttatagaaaagccgcggaaaacaatgaaaatcgtgtgtgtatatgttgcgaacccgaacgcgcgat  1200
cctcgtcttcgagccgaaatgcttctcgtttagagatctaatttggccaagttatttctttagatttatttcgatcaaacgacacagcgaagagatga  1300
gctaactatgcgactttgtgtgttgatatacgcctatatactgtgtatatagcgcagacgagggcgaaggcgttcggttcgggtgtcggattacgtatac  1400
gtactgtgttctcggggctggcaactctcccagatccgcactcgcccagcagccatgtcaagatttcatgatgaccaaaaatgtgcgtaccagaaattt  1500
cagggttaccacatcgacccgttttcgaaaataaaaagggtggtcaacacccacctctgctaaagtatttttaaaacaattaatgaggcagcataata  1600
aaatgaaacgaataaaaacaaaaatttgatgggatttacgtttgatcgactaacatgaacatgaagatgtaaatgctaatcgagttcctttgcgcagAT  1700
GTCGAATTCGTTTTCGAAGTCATGGGCTCCCCCGGTGAGGCATTGATGCCGTCGACCTAGGTGATGCCCTCCGCGCTCTGAACTTGAACCCCACCTTGG  1800
CGCTGATCGAGAAATTGGGCGGCACCAAGAAGCGCAACGAGAAGAAGATCAAGCTGGATGAGTTCCTGCCCATCTACTCACAGGTCAAGAAGGAGAAGGA  1900
GCAGGGCTGCTACGAGGACTTCATTGAGTGCTTGAAGCTCTACGACAAGGAGGAGAACGGCACCATGCTGCTGGCTGAGCTGCAGCACGCCCTGCTGGCG  2000
CTTGgtgagtaatcaacatgcatatatcctgggctatccctctaaatcattgtgtgccctatagGTGAGAGCTTGGATGACGAGCAGGTCGAGACCCTGT  2100
TCGCTGACTGCATGGATCCCGAGGATGATGAAGGATTTATCCCCTACTCTCgtaagtattcgttcaatggacgagcctcatctaatgtttttgtgcata  2200
tacgatttgtaaactcgctttgtctattcacatgtcactctcagtctgcgcaccaaaccaaaaaaatccccaccattctgcactgtgcactttgtaccac  2300
caactcgacctgaaaaacaccaacaaacccaaaaatatgtgtatgcgtgtatccatgtttgcgtgtccattccttcaacactacactcgaaaaaaaaaa  2400
ccgtccaactcccgaatgtcaaacgtatggtctctgcacggtgattgcgtcttgtttacccgccgtttgtcgccgatccacttgaaattgaatgaaaaca  2500
gAGTTCGTCCAGCGCCTGATGAGCCATCCTGTCGTCTTCGACTAAACAgtaagtagggatggagcccatggcacttcgccgatctctggatgtctacttg  2600
tccaacaaatttgggtaatcccttgctaatatggcccttttcttcgtctactccacagCATTCCTCGCCAGAATGTGTGACAGACCAGATCAGCTAAAAT  2700
AGTTGAAAACAACAGCTTCAACAACAACCAACAAACAACAACAACAACAACACCAACAAACTTTTAGCGACAACCATTCTCCATAGATGTTATATTCATC  2800
GCTAACTCTCTCTATCTCTCTCGTATATTATTTTAATTAAAAATTACTTCTGCTACTACAACTGCAAAATTGAACTTCAAGATAATATTATTTTTTAAAA  2900
AATAAAAAAATTTAAATTATAAAATAACGAAAACAAAAAAACTGAAAAAtattcaaatttgtgcacattatttgttatgaattatttgttggttcag  3000
ccaaaagcaaagcaatagttaatgattatttatatgagacgtgaataaaccaaatagcaaataatataatgtttaggcaaattcaacatccgaaagcgt  3100
ccaaaaagggcagttagcaaaacgccaaatgtgtaaattgtgccttttagccttctcaatttaccaagtgaatgtgtgaaagtgaaataagtcaaatgta  3200
tttaaataaacactaagaataattgtgcggccgatgaggtgttgcacaatcaatataaaagaagatcagggtggtggaaaaatttgggtattatgctata  3300
taggcaggccaatattaatacaagatgtggctgacttgtttaataaactaaaagtaatagagaaagtttaaaaatcttatactgcacattccaccaagaa  3400
atgtctaaatatttcgccgcagactgcaatacttcagctgagacgcaaaccaccgaacgagaagcaggcgaaagactaaaataccgtgtactaaaatgt  3500
ctttttcgctttttctcaacgagagagagctttcgaaatgtttgttaagctt
```

FIG. 1. Sequence of the MLC-ALK gene region. Exon sequences, identified by homology with the MLC-ALK cDNA clone pcMLC.1-10, are uppercase, whereas all other sequences are lowercase. The * denotes the translation initiation codon and ▲ denotes the translation termination codon. The "TATA box," exon–intron junctions, and polyadenylation signals (A-A-T-A-A-A) are underlined.

Exon 2 (28 nt; designated E2) and exon 5 (47 nt; designated E5) are particularly small. The first intron separates the initiator AUG from the amino terminus of the encoded peptide, whereas the most 3′ intron is found 3 nt downstream from the translation termination codon and is thus in the 3′ untranslated region of the gene. E3 contains 66% of the protein-encoding sequences of this gene. Intron 3 interrupts the protein-encoding gene region at the carboxyl terminus of helix F of the EF hand domain, the region of greatest homology to the chicken MLC-ALK protein (7).

The cDNA sequence extends at its 5′ end up to nucleotide 501 from the *Hin*dIII site. The structure of the transcripts further upstream has not yet been determined. The best candidates for a Goldberg–Hogness box that would lie 20 to 30 nt upstream of the origin of transcription are the sequences T-A-T-A-A-A-T and A-T-A-T-A-T-T located around position 115 and 276, respectively. Only the sequence located around position 276 has an additional sequence located 30 nt downstream, C-G-C-A-G-G-A, which shows homology to other *Drosophila* transcription start sequences (14). The hypothesis that this is the transcription start is consistent with the observed mRNA lengths and the polyadenylation sites that have been identified, but it has not been confirmed by experiment.

The genomic sequence shows five possible polyadenylation signals (A-A-T-A-A-A) within or downstream of E6. These are marked a, b, c, d, and e in Fig. 2*B*. S1 nuclease analyses of the 3′ ends of transcripts present in pupae show

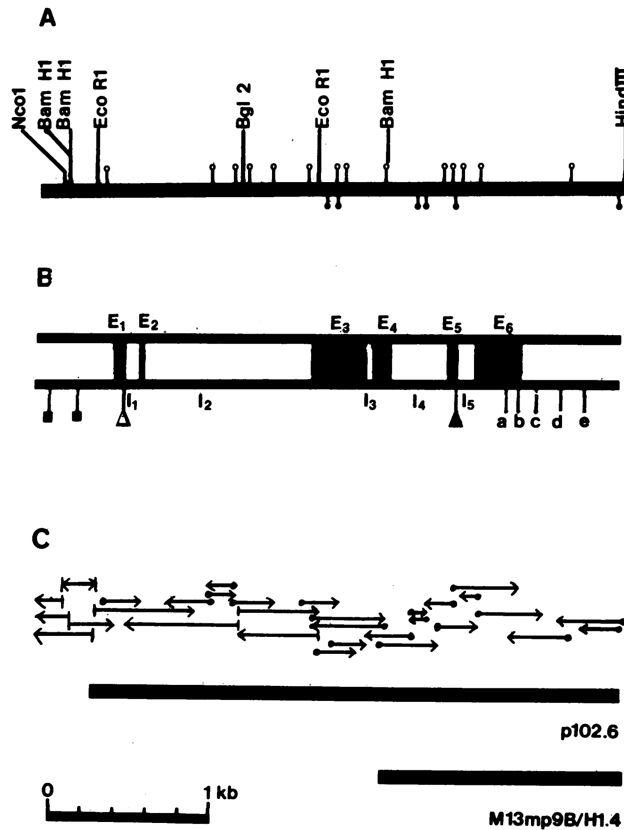

FIG. 2. Summary of MLC-ALK gene structure. (*A*) Restriction map of the sequenced region of genomic DNA that encodes the MLC-ALK gene (●, *Taq* I; ○, *Sau*3A1). (*B*) Arrangement of introns and exons that were identified by homology with the MLC-ALK cDNA clone pcMLC.1-10. E refers to exon and I refers to intron; a, b, c, d, and e are polyadenylylation sequences (A-A-T-A-A-A); ■, TATA box; △, translation initiation codon; ▲, translation termination codon. (*C*) Sequencing strategy of the MLC-ALK genomic DNA. The DNA sequencing strategy is such that a vertical line (|) denotes ^{32}P labeling at the 3′ terminus and a circle (●) denotes those subregions sequenced by using the dideoxy sequencing methods. The arrows indicate the direction and extent of sequencing for a particular run.

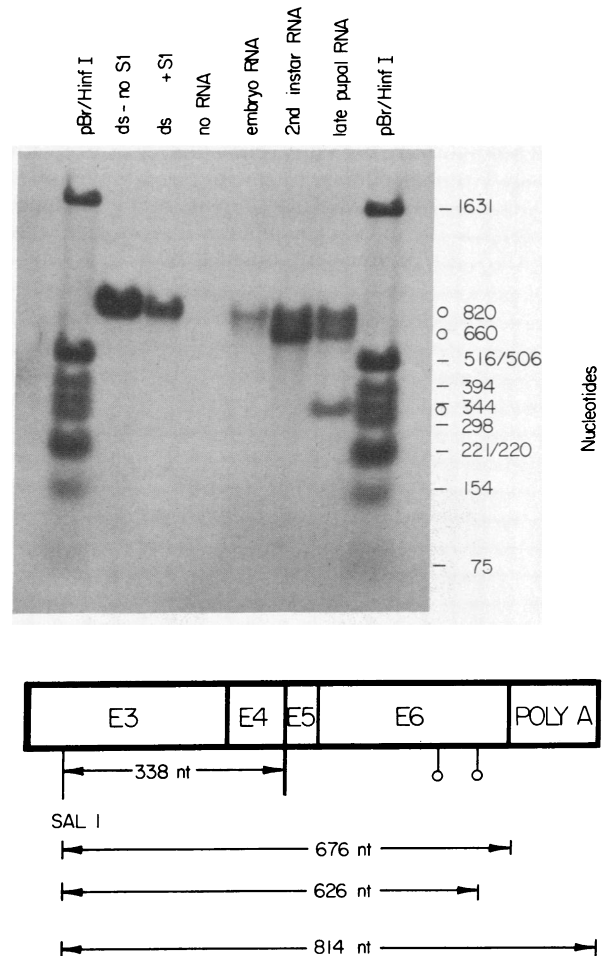

FIG. 3. S1 nuclease analysis of MLC-ALK mRNAs. Poly(A)$^+$ RNA (1 μg) from different developmental stages of synchronized populations of *D. melanogaster* (embryo, 0–4 hr; 2nd instar larvae, 55 hr; late pupae, 195 hr) was hybridized with a 3′-end-labeled *Sal* I/*Hin*dIII fragment isolated from the cDNA clone pcMLC.1-10 and digested with S1 nuclease. The S1 nuclease-resistant fragments were electrophoresed on a 1.5% alkaline agarose gel. After electrophoresis, the protected fragments were visualized by autoradiography. Size markers are *Hin*fI-digested pBR322; controls are the double-stranded (ds) *Sal* I/*Hin*dIII fragment incubated with or without S1 nuclease and the *Sal* I/*Hin*dIII fragment hybridized without RNA.

that sites b and c are used for relatively abundant RNAs (data not shown).

Developmental Regulation of Transcript Splicing. Having determined the structure of the gene as it is expressed as one species of pupal mRNA, we asked whether there are other splicing patterns besides E1-E2-E3-E4-E5-E6 that generate an MLC-ALK-encoding mRNA. Examination of the sequences showed that a splicing pattern of E1-E2-E3-E4-E6 would maintain a reading frame generating a polypeptide with a different carboxyl terminus from that encoded by the cDNA clone pcMLC.1-10. We therefore carried out S1 nuclease analyses of MLC-ALK mRNAs from larval and pupal mRNAs to search for alternative splicing patterns besides that of the contiguous E3-E4-E5-E6.

A *Sal* I 3′-end-labeled *Sal* I/*Hin*dIII fragment of the cDNA clone pcMLC.1-10 was hybridized to early embryo, larval, and pupal poly(A)$^+$ RNA, digested with S1 nuclease, and electrophoresed on an alkaline agarose gel. The logic of the experiment and the autoradiograms are shown in Fig. 3. Note that this fragment was labeled at the *Sal* I site within E3 and extends to the *Hin*dIII site in the poly-linker of the pUC8 cloning vector. This *Hin*dIII site is 8 nt 3′ to the poly(A) sequence of the cDNA clone pcMLC.1-10 (7). Note further that the cDNA clone pcMLC.1-10 uses polyadenylylation site b (Fig. 2*B*). The poly(A) stretch of length approximately 140 nt in the pcMLC.1-10 cDNA clone begins 35 nt downstream from this site. Therefore, hybrids of the DNA with this RNA should protect a fragment of maximum length 814 nt. The renatured DNA band of length 820 nt is seen as a weak signal in the autoradiogram in the embryo lane (0–4 hr), where by RNA blots no detectable muscle mRNA is present. The much more intense band of about this length in the larval and pupal lanes is therefore attributed to the RNA that uses polyadenylylation site b. RNAs using polyadenylylation site c diverge in sequence from pcMLC.1-10 at the junction of the poly(A) stretch with the 3′ untranslated region of the cDNA clone, resulting in the band of length approximately 660 nt in these lanes.

The most significant result is the additional prominent band of length approximately 340 nt found only in the late pupal lane. That is, there is an abundant RNA in pupae that includes E3 (at least downstream from the *Sal* I site) and E4 but differs from the mRNA sequence of clone pcMLC.1-10 at the E4–E5 junction.

To determine the sequences that are continguous with E4 in this new set of pupal MLC-ALK mRNAs, we have isolated and sequenced a corresponding cDNA clone. We sur-

mised that the arrangement of exons would be E3-E4-E6. Therefore, an end-labeled *Rsa* I/*Nco* I fragment (Fig. 1, nucleotides 2294–2565) containing E5 sequences was isolated from the genomic plasmid p102.6 and used to rescreen 10 isolated MLC-ALK cDNA clones (7). Three cDNA clones, which did not hybridize with this probe, did hybridize with a more amino-terminal coding fragment from E3 (data not shown). One of these, pcMLC.1-1, was partially sequenced (Fig. 4 *A* and *B*). The relevant sequences (Fig. 4*C*) show that in this mRNA, E4 is precisely joined to E6. This junction

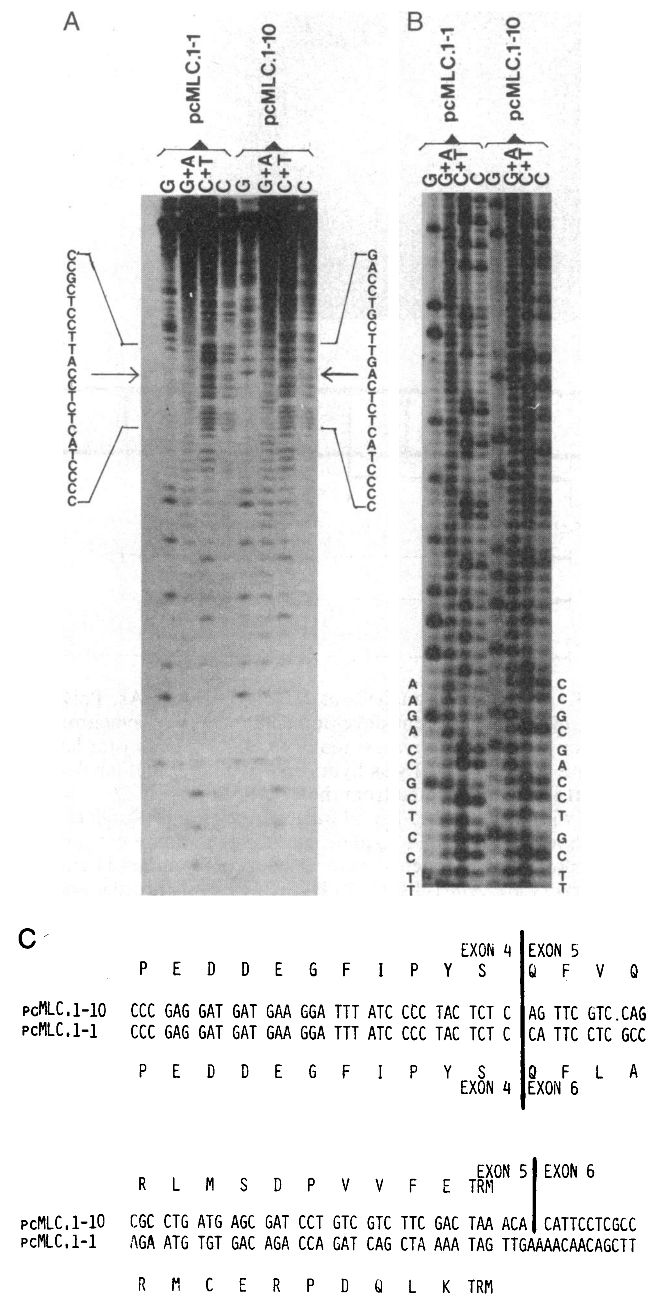

FIG. 4. Sequence comparisons of cDNA clones pcMLC.1-1 and pcMLC.1-10. The *Bam*HI sites of cDNA clones pcMLC.1-1 and pcMLC.1-10 were 5′-end-labeled with [γ-^{32}P]ATP and polynucleotide kinase and subsequently digested with *Hin*dIII. The labeled *Bam*HI/*Hin*dIII fragments were isolated by gel electrophoresis and subjected to the chemical sequencing methods of Maxam and Gilbert (13). The products of the sequencing reactions were electrophoresed on 20% acrylamide/urea gels (*A*) and 8% acrylamide/urea gels (*B*). (*C*) Comparison of the nucleotide sequences of pcMLC.1-1 and pcMLC.1-10 around the E4–E5 and the E4–E6 junction.

generates a new reading frame in E6 leading to a carboxyl-terminal fragment 14 amino acids long. Note that the mRNA containing E4-E5-E6 translates to give a carboxyl-terminal fragment in E5 of quite different amino acid sequence but again 14 amino acids long (see Fig. 4*C*).

The splice donor sequences are all very similar and conform to *Drosophila* and eukaryotic consensus sequences (14, 16, 17). The splice acceptor sequences also conform to the consensus, with the notable exception of that preceding E5. In this case, the consensus C-A-G at the junction is preceded by a purine-rich sequence rather than the consensus pyrimidine-rich sequence (18) (T-G-A-A-A-T-T-G-A-A-T-G-A-A-A-A-*C*-*A*-*G*-A-G-A-T).

DISCUSSION

The sequence of a 3.5-kb region containing the entire MLC-ALK gene has been determined. The precise 5′ end of the gene has not yet been mapped. However, by consideration of a cDNA sequence that extends into the 5′ untranslated region and of the genomic sequence, a plausible start site for transcripts such as that which gave rise to the cDNA clone pcMLC-1.10 has been identified.

We reported previously that the most abundant mRNA species have molecular lengths in the range of 950 to 1050 nt, but species exist with lengths about 1300 nt (7). There are several consensus polyadenylylation sites at the 3′ end of the gene that are used (see Fig. 2*B*, data not shown). These sites generate mRNAs that are greater in molecular length than the shortest species by 220 nt. The observed lengths of the putative primary transcript (3.2 kb) and of the mature mRNAs (0.9–1.3 kb) (7) are consistent with the hypothesis that transcription starts at the site proposed for pcMLC.1-10 and ends at the mapped poly(A) site.

The nucleotide sequence of one MLC-ALK cDNA clone corresponds to an mRNA that is expressed during both larval and pupal muscle development and that contains six exons from the genomic sequence. This mRNA has the sequence E1-E2-E3-E4-E5-E6. A surprising result is that another MLC-ALK mRNA, present in pupae only, includes the sequence E3-E4-E6; that is, it is missing E5 (Fig. 3). The translation of the two mRNAs results in different carboxyl-terminal amino acid sequences, each of length 14 amino acids, within E6 and E5, respectively.

The cause for the observed diversity of about 10% in molecular weight from sodium dodecyl sulfate gel electrophoretic velocity of the MLC-ALK polypeptides (7) is not readily explained by the observed differences in carboxyl termini. Possibilities for the heterogeneity include sequence effects on electrophoretic velocity, post-translational modifications, protease degradation, or some as-yet-undiscovered differences in transcript initiation sites, splicing patterns, and translation starts at the 5′ region of the gene.

Nevertheless, the important point is that there is a developmental difference in splicing that provides two different carboxyl terminal MLC-ALK polypeptide sequences for the pupal and adult musculature, but only one of these in the larval musculature. Thus, in this case, as for the MHC gene of *Drosophila* (C. Rozek, personal communication), developmentally regulated RNA processing (specifically splicing) of transcripts from a single gene is used to provide different proteins for the myofibrils of larvae and pupae. The functional significance of these sequence differences is of course unknown; this ignorance is aggravated by the fact that the function of the MLC-ALK polypeptide in the myofibril is unknown (19).

It is not yet known whether the two predicted MLC-ALK polypeptides in pupae are made in different or the same muscle cells. We note that the intron 4 3′ splice site used for E4-E5 splicing is nonconsensus. It is used in both larval and

pupal RNA processing. The competing E4-E6 splice, which occurs in pupae only, involves consensus sequences. Speculative models as to whether there are positive or negative regulatory factors that act specifically on the nonconsensus sequence can be more fruitfully formulated when it is known whether or not the two patterns of pupal splicing occur in the same or different muscle tissues.

There are now a number of cases known in which a single transcribed region of DNA gives rise to a diverse but related set of polypeptides by a combination of alternative transcription initiation sites, splicing choices, or polyadenylylation sites (20–26).

In the present context, it is particularly interesting that the two chicken skeletal muscle alkali light chains (LC1 and LC3) are encoded by one gene. The polypeptide sequence differences between chicken LC1 and LC3 occur exclusively at the amino termini and arise through differential use of transcription initiation sites and splice sites in the 5′ region of the gene. However, identical use is made of exons and introns in the 3′ region of the single gene, in marked contrast to that seen for the *Drosophila* MLC-ALK gene.

We also note that Karlick *et al.* (26) report additional examples for the generation of muscle protein polymorphisms in *Drosophila* by alternative patterns of RNA processing from a single DNA segment. Whether this general strategy for the generation of protein diversity is more frequently used in muscle rather than in other tissues of diverse species or in *Drosophila* rather than in species with larger genome sizes remains to be determined.

Note Added in Proof. Primer extension experiments have mapped the 5′ end of embryo, larval, and pupal MLC-ALK mRNAs to nucleotide 473 (Fig. 1), indicating that there is only one promoter used throughout development.

We thank E. Vanin and P. Perlman for valuable discussions during the preparation of this manuscript. S.F. was supported by a Muscular Dystrophy Postdoctoral Fellowship and National Research Service Award 1 F32 GM08895-01. The work was supported by grants from the National Institutes of Health to N.D. and S.F. (GM33270) and from the American Cancer Society (IN-16V) to S.F.

1. Tobin, S. L., Zulauf, E., Sanchez, F., Craig, E. A. & McCarthy, B. J. (1980) *Cell* **19**, 121–131.
2. Fyrberg, E. A., Kindle, K. L., Davidson, N. & Sodja, A. (1980) *Cell* **19**, 365–378.
3. Fyrberg, E. A., Bond, B. J., Hershey, N. D., Mixter, K. S. & Davidson, N. (1981) *Cell* **24**, 107–116.
4. Fyrberg, E. A., Mahaffey, J. W., Bond, B. J. & Davidson, N. (1983) *Cell* **33**, 115–123.
5. Bernstein, S. I., Mogami, K., Donady, J. J. & Emerson, C. P. (1983) *Nature (London)* **302**, 393–397.
6. Rozek, C. E. & Davidson, N. (1983) *Cell* **32**, 23–34.
7. Falkenthal, S., Parker, V. P., Mattox, W. W. & Davidson, N. (1984) *Mol. Cell. Biol.* **4**, 956–965.
8. Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W. J. & Goodman, H. M. (1977) *Science* **196**, 1313–1319.
9. Anderson, K. V. & Lengyel, J. A. (1979) *Dev. Biol.* **70**, 217–231.
10. Favaloro, J., Treisman, R. & Kamen, R. (1980) *Methods Enzymol.* **65**, 718–749.
11. Yen, P. H. & Davidson, N. (1980) *Cell* **22**, 137–148.
12. Mullins, J. I., Casey, J. W., Nicolson, M. O., Burck, K. B. & Davidson, N. (1981) *J. Virol.* **38**, 688–703.
13. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
14. Snyder, M., Hunkapiller, M., Yuen, D., Silvert, D., Fristrom, J. & Davidson, N. (1982) *Cell* **29**, 1027–1040.
15. Sanger, F., Nicklen, S. & Coulsen, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
16. Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459–472.
17. Sharp, P. A. (1981) *Cell* **23**, 643–646.
18. Weiringa, B., Hofer, E. & Weissmann, C. (1984) *Cell* **37**, 915–925.
19. Wagner, P. D. & Giniger, E. (1981) *Nature (London)* **292**, 560–562.
20. Cheng, H.-L., Blattner, F. R., Fitzmaurice, L., Mushinski, J. F. & Tucker, P. W. (1982) *Nature (London)* **296**, 410–415.
21. Fitzmaurice, L., Owens, J., Blattner, F. R., Cheng, H.-L., Tucker, P. W. & Mushinski, J. F. (1982) *Nature (London)* **296**, 459–462.
22. Rogers, J. & Wall, R. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1877–1879.
23. Amara, S. G., Jonas, V., Rosenfeld, M. G., Ong, E. S. & Evans, R. M. (1982) *Nature (London)* **298**, 240–244.
24. Inana, G., Piatigorsky, J., Norman, B., Slingsby, C. & Blundell, T. (1983) *Nature (London)* **302**, 310–315.
25. Nabeshima, Y., Fujii-Kuriyama, Y., Muramatsu, M. & Ogata, K. (1984) *Nature (London)* **308**, 333–338.
26. Karlik, C. C., Mahaffey, J. W., Coutu, M. D. & Fyrberg, E. A. (1984) *Cell* **37**, 469–481.