MOLECULAR AND CELLULAR BIOLOGY, July 1989, p. 2957–2974
0270-7306/89/072957-18$02.00/0

Vol. 9, No. 7

Functional Domains of the *Drosophila melanogaster* Muscle Myosin Heavy-Chain Gene Are Encoded by Alternatively Spliced Exons

ELIZABETH L. GEORGE,† MARGARET B. OBER, AND CHARLES P. EMERSON, JR.*
Biology Department, University of Virginia, Charlottesville, Virginia 22901

Received 17 January 1989/Accepted 7 April 1989

The single-copy *Drosophila* muscle myosin heavy-chain (MHC) gene, located at 36B(2L), has a complex exon structure that produces a diversity of larval and adult muscle MHC isoforms through regulated alternative RNA splicing. Genomic and cDNA sequence analyses revealed that this 21-kilobase MHC gene encodes these MHC isoforms in 19 exons. However, five sets of these exons, encoding portions of the S1 head and the hinge domains of the MHC protein, are tandemly repeated as two, three, four, or five divergent copies, which are individually spliced into RNA transcripts. RNA hybridization studies with exon-specific probes showed that at least 10 of the 480 possible MHC isoforms that could arise by alternative RNA splicing of these exons are expressed as MHC transcripts and that the expression of specific members of alternative exon sets is regulated, both in stage and in muscle-type specificity. This regulated expression of specific exons is of particular interest because the alternatively spliced exon sets encode discrete domains of the MHC protein that likely contribute to the specialized contractile activities of different *Drosophila* muscle types. The alternative exon structure of the *Drosophila* MHC gene and the single-copy nature of this gene in the *Drosophila* genome make possible transgenic experiments to test the physiological functions of specific MHC protein domains and genetic and molecular experiments to investigate the mechanisms that regulate alternative exon splicing of MHC and other muscle gene transcripts.

Myosin heavy chain (MHC) is a complex, multifunctional contractile protein with both enzymatic and structural domains. The MHC globular S1 head has ATPase activity essential for contractile activity, and the α-helical tail interacts with the tail of another MHC to form a coiled-coil dimer that associates with other myosins to form the sarcomeric thick filament. Two different myosin light chains are associated with each S1 head and regulate MHC function. Upon activation by actin, ATPase activity in the S1 head generates energy for the production of force during muscle contraction (for recent reviews, see references 26 and 72). Complex organisms have physiologically distinct muscle types that express different MHC isoforms (9, 10, 64, 75), and the divergent protein sequences in specific functional domains of such MHC isoforms determine, at least in part, differences in contractile properties of such physiologically distinct muscle types.

Muscle MHC isoforms are encoded by gene families in mammals (37, 41, 55, 56, 77), in chickens (58), and in nematodes (47). Protein and DNA sequence analyses reveal that these MHC gene families likely arose by gene duplications, followed by divergence of individual gene sequences (see reference 18 for a recent review). *Drosophila melanogaster*, however, is an exceptional organism, having only a single muscle-specific MHC gene, located at subdivision 36B(2L) (6, 59). *D. melanogaster* also has a nonmuscle MHC gene that encodes ubiquitous cytoplasmic MHC (34). Molecular and genetic analyses have established that the 36B(2L) MHC gene is abundantly expressed and is functional in larval and adult muscle types (50). The discovery that *D. melanogaster* has only one muscle-specific MHC gene yet has many functionally distinct and specialized larval and adult muscle types, including the fibrillar indirect flight muscles (IFM), tubular muscles, and body wall muscles (14, 46), raised the question of whether this single MHC gene produces only one MHC isoform that contributes to the distinct contractile activities of all of these different muscle types or, alternatively, whether this gene produces multiple MHC isoforms through posttranscriptional mechanisms such as differential RNA processing. Generation of protein diversity by differential RNA processing occurs in many muscle-specific as well as nonmuscle genes: myosin light chains (19, 20, 53), troponin T (8, 27, 45), tropomyosin (28, 32, 54), calcitonin (1), and P-element transposase (36), to name just a few.

Earlier studies established that the 36B *Drosophila* MHC gene produces two MHC isoforms with different carboxyl termini through regulated alternative RNA splicing involving inclusion or exclusion of the penultimate exon (5, 60). The divergent tailpiece structures of the two MHC isoforms generated by alternative splicing may influence the assembly and organization of myosin into thick filaments in specialized muscle types, but they do not explain functional specializations in muscles such as the IFM, which has an especially high rate of contraction (57). Rate of contraction is determined by sequences in the amino-terminal S1 head region, which has the functional domains for ATP binding, ATPase activity, myosin light-chain binding, and actin interactions that mediate contractile activity.

The objective of this study was to investigate whether the 36B MHC gene produces a diversity of MHC isoforms by alternative RNA splicing of exons encoding protein domains involved in MHC contractile activity. We have determined the complete exon-intron structure of the *Drosophila* 36B muscle MHC gene by DNA sequence analysis of genomic and cDNA clones and characterized the developmental expression of specific MHC exons by RNA hybridization analysis. Our results reveal that the 36B MHC gene has a uniquely complex exon structure. At least 10, and perhaps more, MHC mRNA isoforms are generated by mutually

* Corresponding author.
† Present address: Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139.

exclusive alternative RNA splicing of multiple exons encoding four separate regions in the S1 head and one region in the tail S2 hinge. Alternative RNA splicing of these exons is regulated with both stage and tissue specificity, indicating that these alternatively spliced exon sets may encode functionally significant isoforms of MHC. The alternatively spliced exons in the S1 head encode portions of MHC regions involved in actin and myosin light-chain binding (49) and ATPase activity (72). Alternative exons also encode the S2 hinge region, which has been implicated in force generation during contraction by a helix-coil transition mechanism (69, 70). The regulation of these S2 hinge alternative exons is IFM specific, consistent with a functional role of the S2 hinge in specialized muscle types. Protein regions of unknown function in the S1 head are also encoded by alternatively spliced exons, and these may be involved in generation of force demonstrated by movement of S1 head in vitro (68). These findings provide a basis for transgenic experiments to investigate specific hypotheses concerning the isoform-specific functions of MHC domains in physiologically diverse muscle types in vivo and to investigate the molecular mechanisms of muscle-specific RNA-splicing regulation.

MATERIALS AND METHODS

DNA sequence analysis. Restriction fragments of the MHC gene were isolated from lambda Charon 4 bacteriophage clones (6) from a Canton S genomic DNA library (43) and subcloned into pEMBL vectors 18+ and 19+ (15). DNA sequences were determined on both strands by the dideoxy-chain termination method (62). Nested deletion series of larger subclones were generated by exonuclease III (29).

Isolation of cDNA clones. cDNA was synthesized by the method of Gubler and Hoffman (25) from 2.5 μg of poly(A) RNA isolated from late pupae. Blunt-end cDNA was ligated to *Eco*RI adapters (New England BioLabs, Inc.) and subsequently inserted into the *Eco*RI site of the lambda gt10 vector (30). Nearly full-length clones were selected by plaque hybridization (3), using triplicate nitrocellulose filters and probing with both 5′-end (two filters) and 3′-end (one filter) nick-translated MHC genomic restriction fragments. Two clones, cD301 and cD302, hybridized with both probes. *Eco*RI restriction fragments were subcloned and sequenced as described for the MHC gene.

RNA isolation and Northern (RNA) blots. RNA was isolated by a modification (38) of the guanidinium thiocyanate extraction procedure (12). Poly(A) RNA was selected by oligo(dT)-cellulose chromatography (2). RNA was electrophoresed on a formaldehyde-containing agarose gel (42) and electroblotted to Zetabind. Blots were UV irradiated at 300 μW/cm^2 at 254 nm for 3 min. Hybridizations were performed by the method of Church and Gilbert (13) except that the temperature of hybridization for oligonucleotide probes was T_d − 5°C, where T_d = 4°C per G · C base pair + 2°C per A · T base pair (71, 78). Oligonucleotide probes (20-mer; synthesized at the University of Virginia sequencing facility) were ^{32}P end labeled with polynucleotide kinase (42). Reverse complement sequences of probes correspond to the nucleotide numbers shown in Fig. 1: exon 3a, 2394 through 2414; exon 3b, 2705 through 2725; exon 7a, 4580 through 4600; exon 7b, 5171 through 5191; exon 7c, 5666 through 5686; exon 7d, 6514 through 6534; exon 9a, 8579 through 8598; exon 9b, 9050 through 9069; and exon 9c, 9628 through 9647. Probes for exon 11 were generated by random priming (21) genomic restriction fragments: exon 11a, 265-nucleotide (nt) *Kpn*I-*Sac*I; exon 11b, 495-nt *Sac*I-*Nar*I; exon 11c, 559-nt *Nar*I-*Nco*I; and exon 11d, 775-nt *Nco*I-*Pvu*II. Hybridization and washing with random-primed probes were performed as described for oligonucleotide probes except at 65°C. All blots were exposed to Kodak XAR-5 film (Eastman Kodak Co.) for 0.5 to 2 days at −80°C with an intensifying screen.

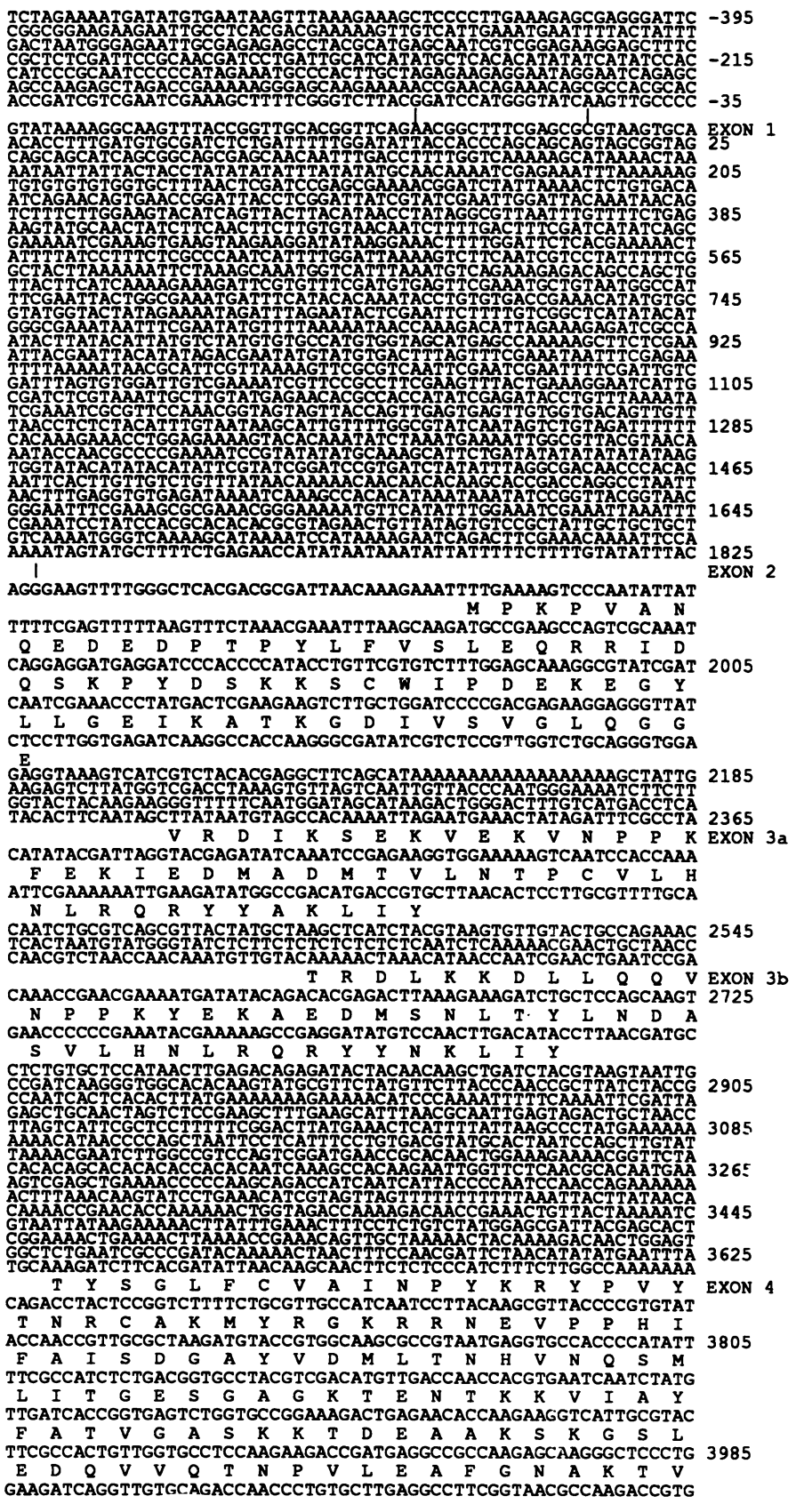
S1 nuclease analysis. The S1 nuclease assay was a modification of the procedure of Berk and Sharp (4). An end-labeled cDNA restriction fragment probe (50,000 cpm) was precipitated with total cellular RNA, washed with 70% ethanol, dried, and suspended in 10 μl of 80% deionized formamide–40 mM piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (PIPES; pH 6.4)–0.4 M NaCl–1 mM EDTA. After denaturation at 65°C for 20 min, RNA-DNA hybrids were allowed to form at 50°C for 16 h (11). S1 nuclease digestion was carried out at 22°C for 1 h with 100 U of S1 (Pharmacia, Inc.) in 300 μl of 0.28 M NaCl–0.05 M sodium acetate (pH 4.6)–4.5 mM $ZnSO_4$–20 μg of salmon sperm DNA. The reactions were stopped by extraction with phenol and chloroform and then precipitated with 5 μg of yeast tRNA and separated by electrophoresis on a denaturing polyacrylamide gel. S1-protected fragments were visualized by exposing the gel to Kodak XAR-5 film with an intensifying screen.

RESULTS

MHC gene structure. DNA sequence analyses of genomic and cDNA clones revealed that the 36B muscle MHC gene in *D. melanogaster* spans 21,867 nt. The complete sequence is presented in Fig. 1. Primer extension and S1 nuclease analyses have shown that the MHC gene transcripts initiate at a single site in all muscles (73). Two sites of polyadenylation were localized by S1 nuclease protection assay and cDNA sequence analysis. Use of either polyadenylation site does not appear to be regulated (5, 60).

To define the exon structure of the MHC gene, we compared the genomic nucleotide sequence with the nucleotide sequence of an essentially full-length cDNA clone, cD301. Two nearly full-length cDNA clones, cD301 and

FIG. 1. Complete nucleotide and derived amino acid sequences of the *Drosophila* muscle MHC gene. The nucleotide sequence of the coding strand only is shown, and nucleotide numbers are at the right of every fourth 60-character line. Nucleotide 1 is the first nucleotide of untranslated exon 1 (73). Positive numbers indicate nucleotides within the gene; negative numbers indicate 5′-flanking sequence of the gene. Vertical lines above nucleotides indicate nontranslated exon borders. Encoded amino acids (aa) are in one-letter code above the second nucleotide of each codon and align with exons as follows: exon 2, aa 1 through 67 (excluding initiator methionine); exon 3, aa 68 through 115; exon 4, aa 116 through 242; exon 5, aa 243 through 263; exon 6, aa 264 through 296; exon 7, aa 297 through 331; exon 8, aa 332 through 467; exon 9, aa 468 through 524; exon 10, aa 525 through 721; exon 11, aa 722 through 760; exon 12, aa 761 through 848; exon 13, aa 849 through 918; exon 14, aa 919 through 1214; exon 15, aa 1215 through 1240; exon 16, aa 1241 through 1632; exon 17, aa 1633 through 1934; exon 18, aa 1935; and exon 19, aa 1935 through 1961. Amino acids encoded by split codons are in parentheses. Duplicated exons except exon 11e are named in alphabetical order according to the direction of transcription. *, TAA termination codon. The nucleotides underlined in exon 19 are consensus polyadenylation signals; known polyadenylation sites are indicated by arrowheads above the sequence (5). Nucleotides −335 through 780 are from reference 73; nucleotides 21306 through 22208 are from reference 5.



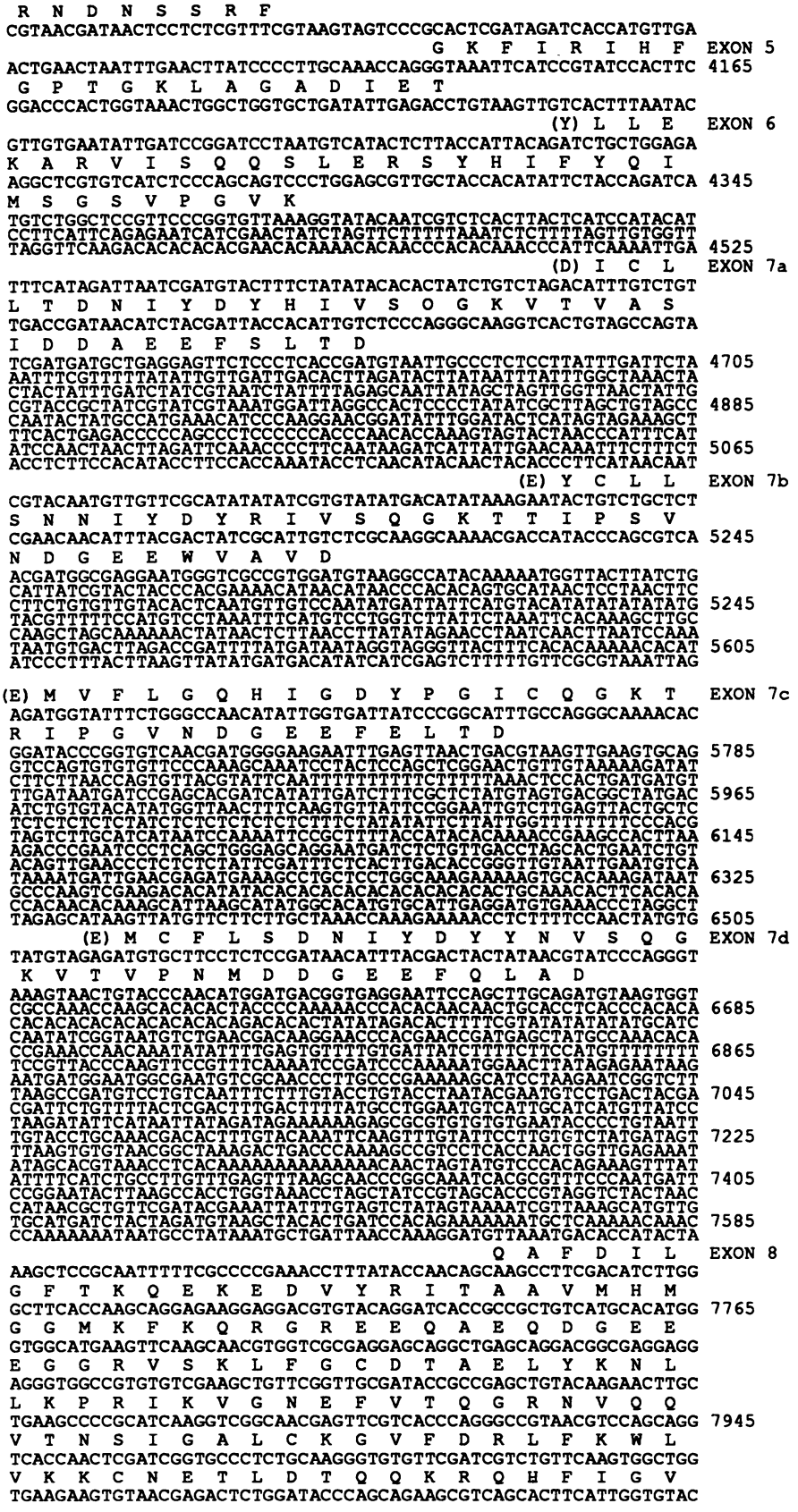


FIG. 1—*Continued*

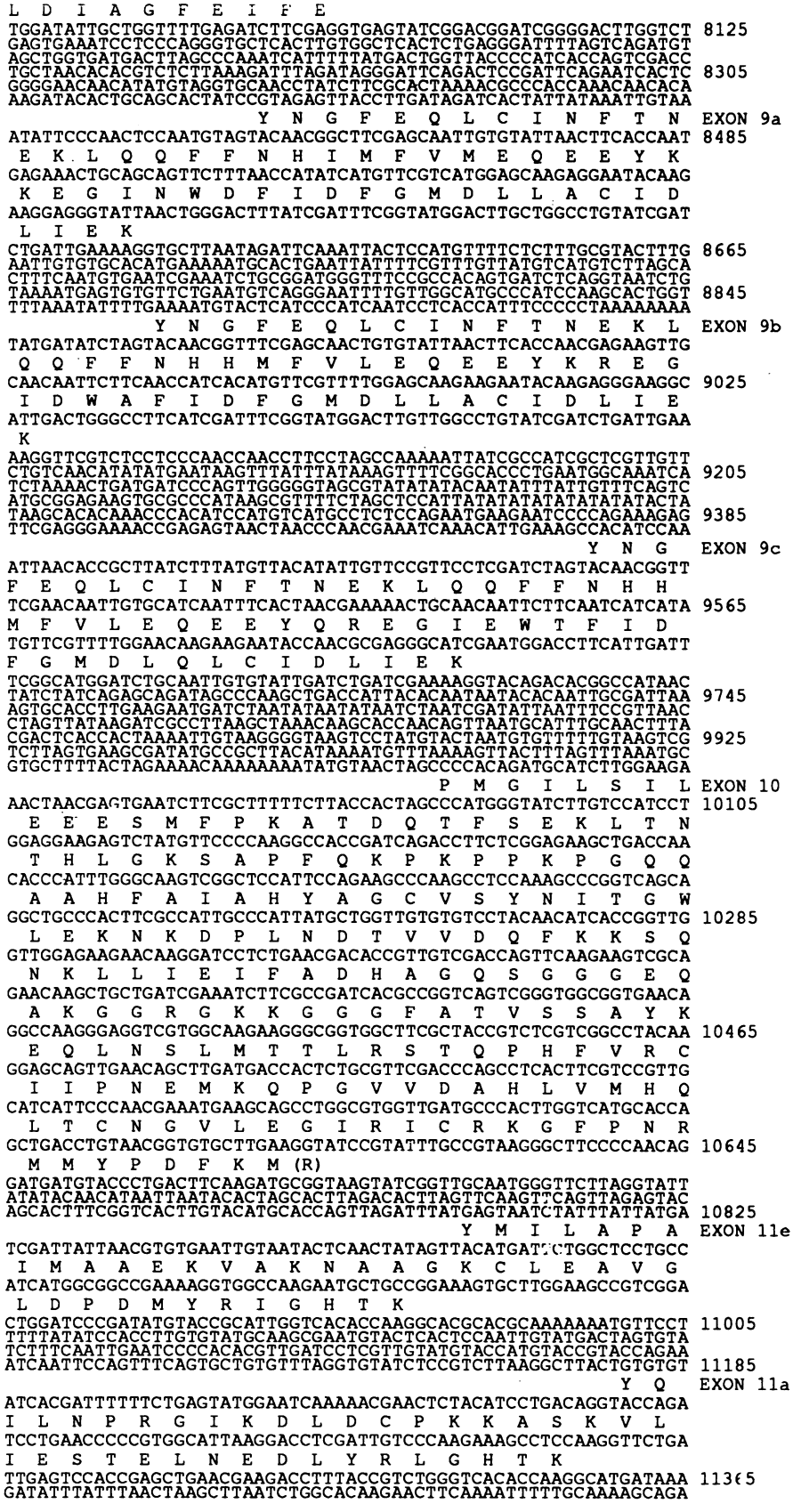


FIG. 1—*Continued*

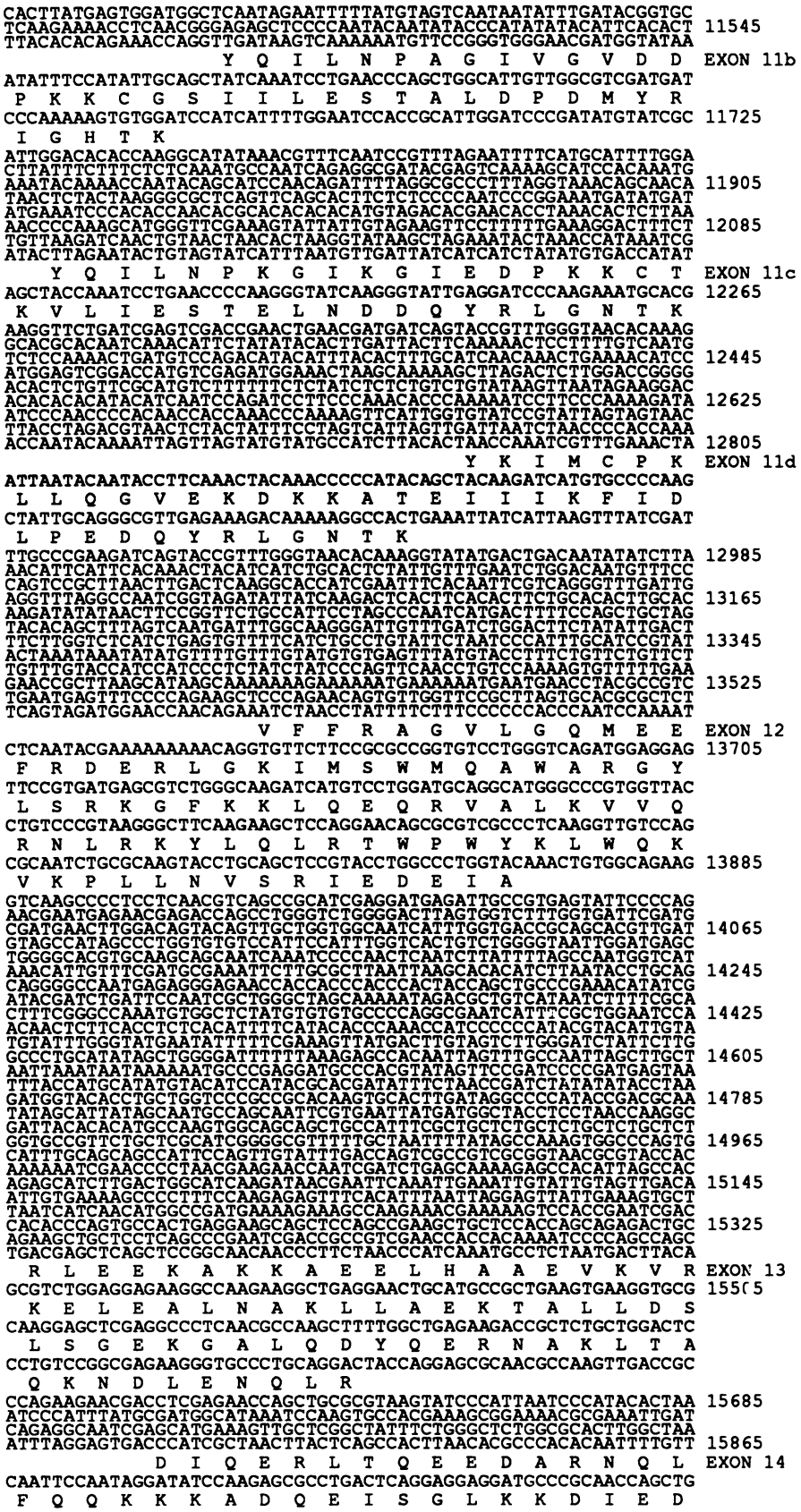


FIG. 1—*Continued*

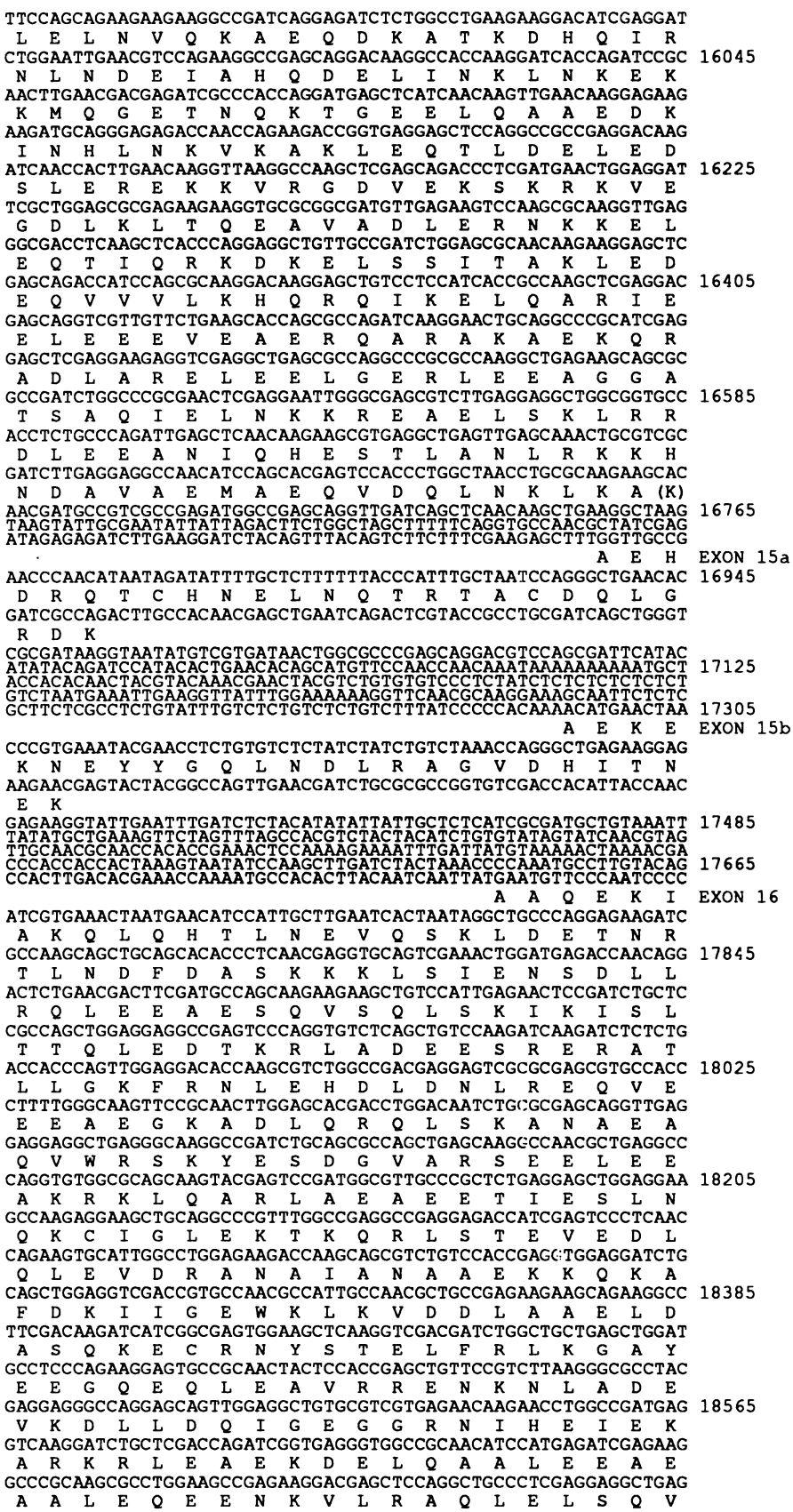


FIG. 1—*Continued*

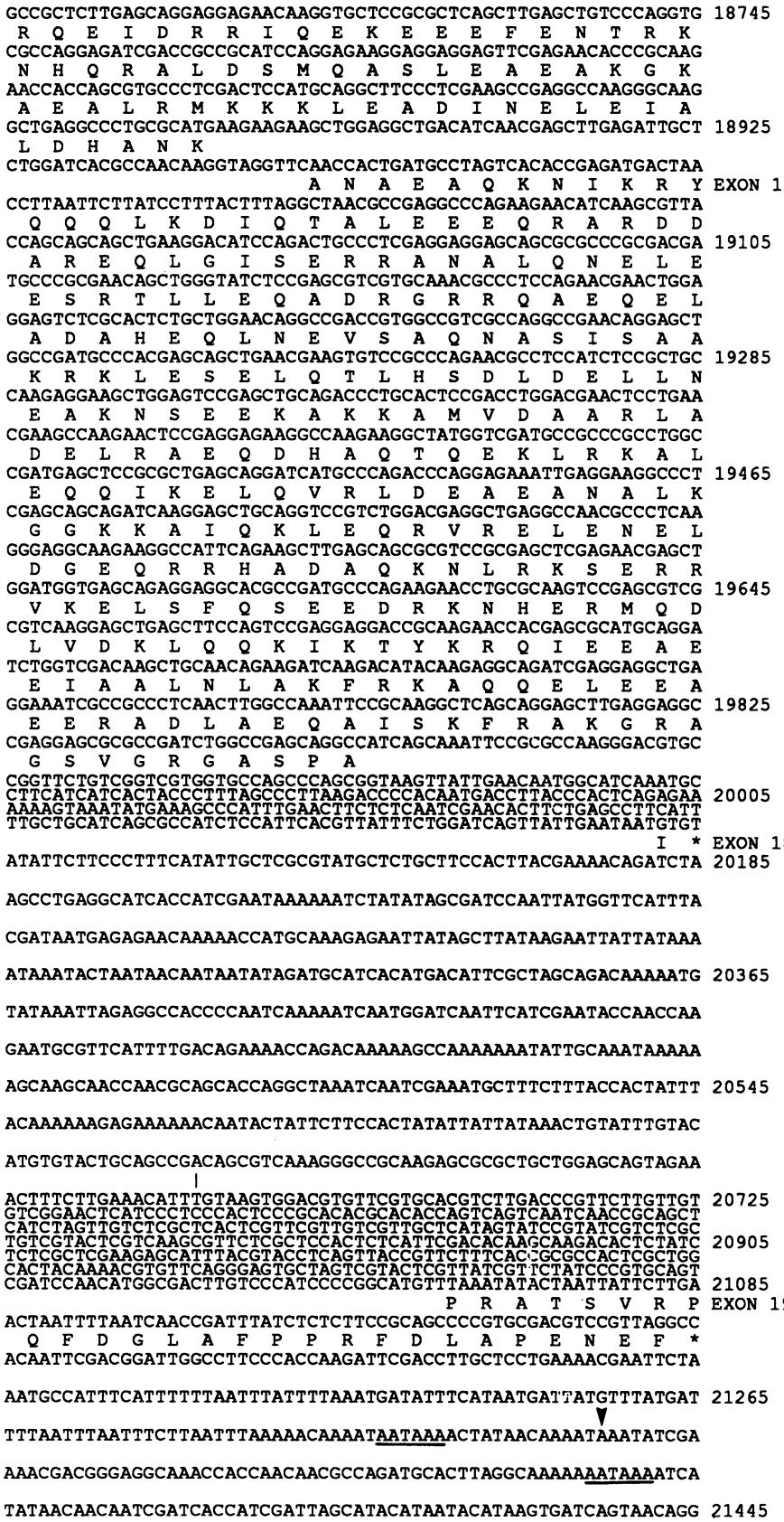


FIG. 1—*Continued*

```
CTATTGTGTAAGCGCGACATTTGCCATGGGGACACTGCTCGCTGAACCTGTATTTGTATT

TGTATTTATATATTTTTATACAACTTTCGAAATGCATTCAACCAACTAACTATGAAAGAA

CCATCGACTCTCAACCGAACAACATGAGAATAATCGCGAAACGTGAGCCCCGTCAAAATG 21625

AAGGCCCCGCGGAGGTGTCCGCGGTAAATGTCGCCCGCAGTTCGAATATTTTGCTCTATA

TGTACTCCCAATATTGTTTGTCATGCAGACCAGACAAGACTTACGGACGACACGAGCTAC

ACACTGCGCTTATGTTTATGTTTAACTCTGACTCTATCTAACAATGATAGGGAGATAAGA 21805

TCATCACTTGACACTAAATGAAAAGTATACAAAAAAGGAACAAAATAAATGTAAATCAAT

TTAACAAATGTGTATTATTTTCCGTGATCAAATCATTCGACTGGGACATATTCATTCATA
TTTACTGGCCATACGTGTTAAGAAATTGTCACAGGTCTTTTGTAAACTAATAAGATGTTG 21985
TGGATATATTTATTTACTTTTACTTAAGCCCCATACAAATTAAGTCCTACTTTCCCGCCC
AGGAGCCCTGGGATTGGGCTCGATGCGCAATAGTTGCTTGTGCTGGCCAATCATGTGCAC
GATGTGGTTGCACTCGCGGAGATGACCCTTGAACTCATCGAGAGTCTGGTAAAGCACCGG 22165
ATAGAGGGCGGGCAGGAGGCCATGATCCAGGTCGGGAAGAGTC                   22208
```

FIG. 1—*Continued*

cD302, were recovered from a late-pupal lambda gt10 cDNA library constructed in our laboratory. A complete DNA sequence analysis indicated that cD301 contains 5,940 nt and spans MHC exons 2 to 19. The nucleotide sequence of cD301 matched exon sequences in the *Drosophila* MHC gene except for three nucleotides, at positions 17833, 18205, and 18481 (Fig. 1). All three differences were in the third nucleotide of a codon and did not change the encoded amino acid. cD302 spans exons 4 to 19, and representative regions were sequenced. Figure 2 displays a comparison of cD301-encoded amino acid sequence with two other complete sequences of muscle MHC: *Caenorhabditis elegans* body wall isoform of MHC (33) and rat embryonic skeletal isoform of MHC (65). cD301 contains all exons required for a structurally complete MHC protein. Interspecies comparison of the MHC head (S1) and tail (S2 plus light meromyosin) showed the high degree of homology that this *Drosophila* MHC shares with the MHCs of these distantly related organisms. Percent identity of cD301-encoded amino acids with nematode body wall MHC was as follows: S1, 60%; S2, 53%; and light meromyosin, 58%. Comparison with rat embryonic MHC yielded the following values: S1, 53%; S2, 49%; and light meromyosin, 58%.

The exons encoded by cD301 were used as test sequences in a computer homology analysis to search for the presence of exon duplications in the *Drosophila* MHC gene. This analysis revealed an extraordinary complexity of intragenic tandem duplications of exons 3, 7, 9, 11, and 15 in the *Drosophila* MHC gene. Putative MHC intron sequences were compared in all translation reading frames with the open reading frames of all exons in cD301, using the FASTP amino acid sequence homology program (39). Sequences detected at the amino acid level as related exons were further confirmed as exons by the identification of flanking consensus intron-exon splice junction donor and acceptor sequences (7). Computer comparisons involved the sequential removal of homologous exon sequences from the intron sequence, followed by repeated comparison with the cD301 exon test sequence. All newly discovered putative exons were also used as exon test sequences against the remaining intron sequence. Exon 15b was discovered by computer-assisted comparison with the second MHC cDNA clone, cD302, and not by homology with exon 15a, which is highly divergent. The complete structures of the MHC gene and the two alternatively spliced cDNA clones are illustrated in Fig. 3.

The duplicated exons in the MHC gene were found to be alternatively spliced in a mutually exclusive manner. In cD301 and cD302 cDNA clones, only one form of each duplicated exon was present and was flanked by common exons. The inclusion of only one exon from each set by alternative splicing produced an MHC protein structure predicted from other MHCs. Furthermore, exon sets 7, 11, and 15 had split codons at their 5′ splice junctions. Combinatorial splicing of these exons, resulting in more than one version of a particular exon per mRNA, would produce aberrant MHCs with altered translational reading frames. cD301 and cD302 illustrate two different combinations of MHC alternative splicing. They shared exon 9b and exclusion of exon 18, but alternatives for exons 7, 11, and 15 were different. Exon 18 is differentially spliced by selective inclusion or exclusion in pre-mRNA (5, 60) rather than by mutually exclusive exon choice that we report here for alternative exons 3, 7, 9, 11, and 15. Thus, *Drosophila* MHC gene expression is accomplished by two distinct modes of differential RNA splicing.

We compared the MHC intron sequences and found no obvious sequence correlations in introns flanking alternatively spliced exons that we know are spliced together by their presence in the same cDNA clone. All introns had consensus dinucleotide borders except those downstream of alternative exons 11a, 11b, 11c, and 11e, which had a GC dinucleotide at the 5′ donor site rather than the consensus GT. A donor site GC dinucleotide at the 5′ donor site is also present in the alternatively spliced αA-crystallin gene (35) and in an α-globin gene (17). MHC intron sequences immediately upstream of the 3′ splice acceptor of exons 3a, 3b, 4, 7b, 7c, 9b, 11c, 12, and 16 were purine rich, as previously reported for the region preceding the 3′ splice acceptor of the differentially spliced exon 18 (5), rather than similar to the consensus pyrimidine-rich sequence (52).

We compared exon borders of the *Drosophila* MHC gene with those of other muscle MHC genes to determine whether the alternative exons and the exons of other MHC genes are ancestral or whether these alternative exons are unique to the *Drosophila* gene. Conserved exon borders in *Drosophila*, nematode body wall (33), and rat embryonic skeletal (65) MHC genes are demarcated by lines drawn between exon borders in Fig. 4. This comparison illustrates the conservation of exon borders in the MHC head and the lack of conservation of exon borders in the MHC tail, which begins in *Drosophila* exon 12. The *Drosophila* MHC tail shares a single intron location with the rat gene and follows alternatively spliced exon 15a/15b. The nematode and *Drosophila* genes share few exon borders, but those that are shared are in the head region. *Drosophila* exons 3a and 3b

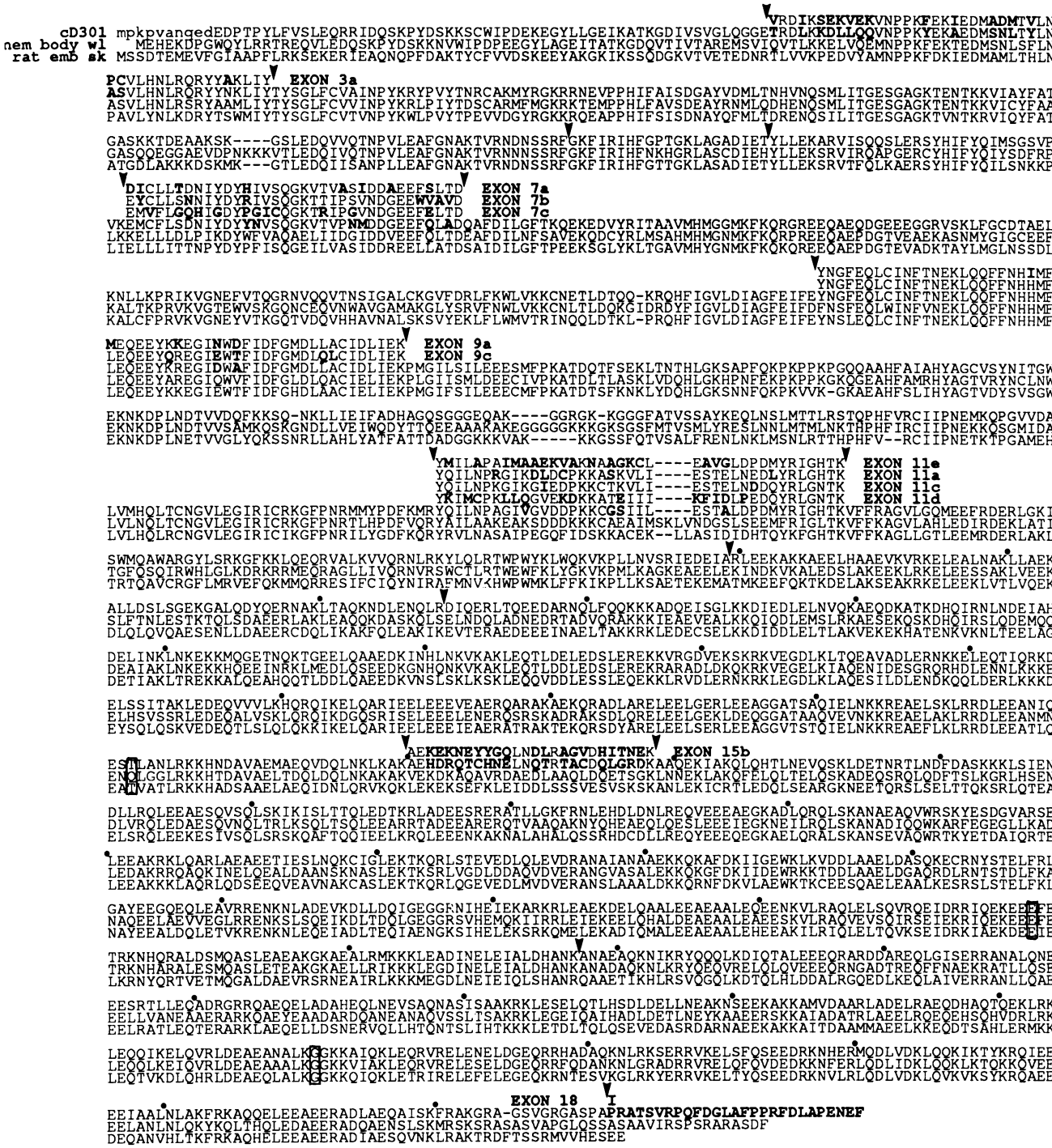


FIG. 2. MHC amino acid sequence comparisons between organisms and *Drosophila* alternative exons. The encoded amino acid sequence of one *Drosophila* isoform, cD301, is compared with those of nematode body wall MHC (33) and rat embryonic skeletal MHC (65) sequences. The cD301 cDNA does not contain the N-terminal 10 residues (in lowercase). The sequences are aligned and gaps are introduced in order to maximize identity of residues. *Drosophila* alternative exons not encoded by cD301 are aligned above the cD301 sequence and labeled. Note that cD301 contains alternative exons 3b, 7d, 9b, 11b, and 15a. Boldface characters highlight amino acid differences between *Drosophila* alternative exons. Residues in boldface differ from any other *Drosophila* residue at that position. Symbols: ▼, exon border; ●, 28-residue repeat characteristic of MHC coiled-coil tail sequences. The conserved locations and identities of the skip residues between the 28-residue repeats (44) are boxed.

and rat exon 4 also share approximate borders with nematode exon 3. Nematode exon 3 has two additional amino acids at the 5′ border, but the 3′ border is precise. All four forms of alternative exon 7 share precise borders with rat exon 11, including a 2-nt split codon at the 5′ border. Alternative exons 9a/9b/9c share a precise 3′ border with nematode exon 5 and both borders with rat exon 15. Alternative exons 11a/11b/11c/11d/11e and rat exon 20 share

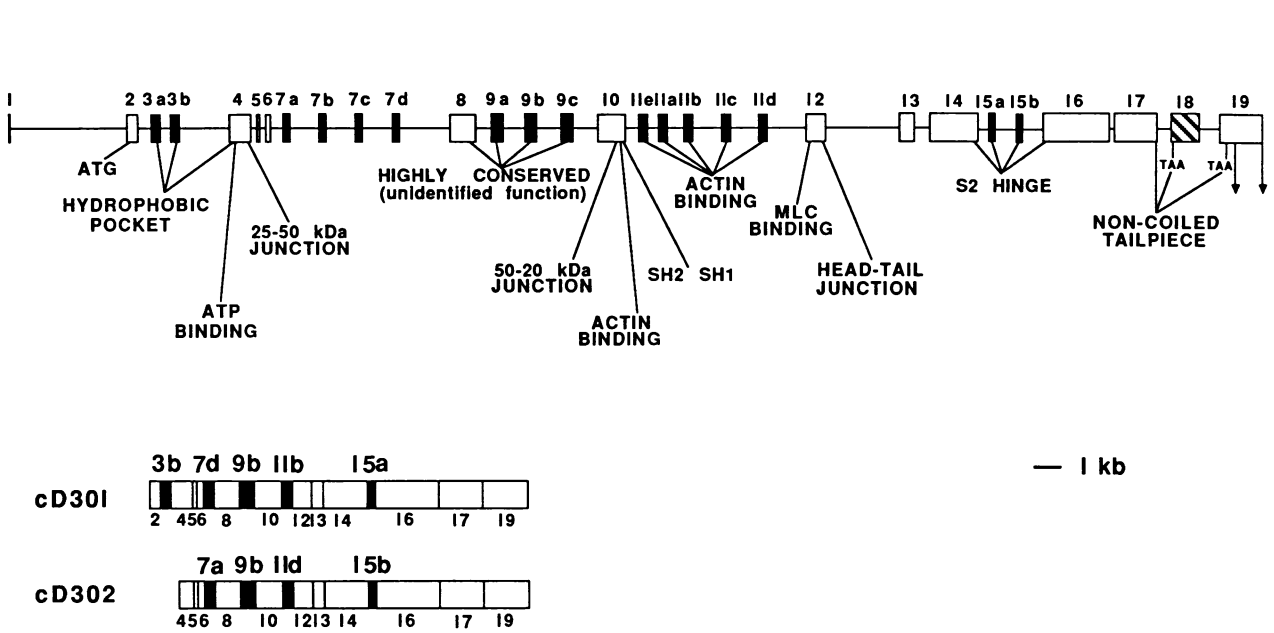


FIG. 3. Diagram of *Drosophila* muscle MHC gene structure and two alternatively spliced cDNAs. (A) Symbols: ——, intron sequences; □, common exons; ■, mutually exclusive alternative exons; ↓, polyadenylation sites within and at the 3′ end of common exon 19 (5, 60). Exon 18 is striped to distinguish it as a differentially included exon. Exons and introns are drawn to the same scale. Protein domain landmarks determined from MHC amino acid sequence comparison are indicated relative to the encoding exon(s) and are discussed in the text. If a protein domain is encoded by more than one exon, then all exons are indicated; if a protein domain is encoded by a duplicated exon, then all versions of the exon are indicated. Head-tail junction refers to the conserved proline near that junction which is residue 836 in *Drosophila* MHC. ATG, Initiator codon; TAA, termination codon. (B) cDNA structures illustrated by the symbols described above. The clones represent two different MHC mRNAs generated by alternative splicing. Note that both mRNAs exclude exon 18. Both cDNAs were isolated from a lambda gt10 library generated from polyadenylated late-pupal RNA. cD301 extends to MHC codon 11 in exon 2, and cD302 extends to MHC exon 4. Both clones contain 3′ nontranslated sequence but no poly(A) tail sequence. MLC, Myosin light chain.

precise borders, including a 1-nt split codon at the 5′ border. In contrast to the alternative exons in the head, alternative exons 15a and 15b in the tail share only the 3′ border with rat exon 27. Alternative exon borders are also precisely conserved with those of a chicken embryonic MHC gene (51). This conservation of the boundaries of the alternative exons of the *Drosophila* MHC gene with those of the exons of other MHC genes reveals that these exons are ancient and were probably established before the split of the invertebrate and vertebrate lineages.

The alternative exons encode functional MHC domains. The protein sequences encoded by the *Drosophila* alternative exons and common exons were compared with the putative functional domains of MHC, as ascribed to sequences of MHCs of other organisms (Fig. 3).

Alternative exons 3a/3b encode amino acids 68 to 115, which are part of a conserved hydrophobic region neighboring the ATP-binding domain (72). The C-terminal 32 codons of alternative exons 3a/3b encode a conserved hydrophobic pocket sequence required for MHC function (16). The remainder of this hydrophobic pocket extends through residue 128 into common exon 4. Thus, alternative exon 3 overlaps but does not include the entire hydrophobic pocket domain. Most divergent sequences in exons 3a and 3b lie in a stretch of 11 amino acids N terminal to this hydrophobic domain. Common exon 4 encodes the entire, highly conserved ATP-binding domain (67), followed by the highly divergent proteolytic digestion site that defines the 25- to 50-kilodalton (kDa) junction within the MHC head.

The 50-kDa fragment contains two regions of remarkable sequence conservation. A function for these conserved regions has not been determined, but their degree of evolu-

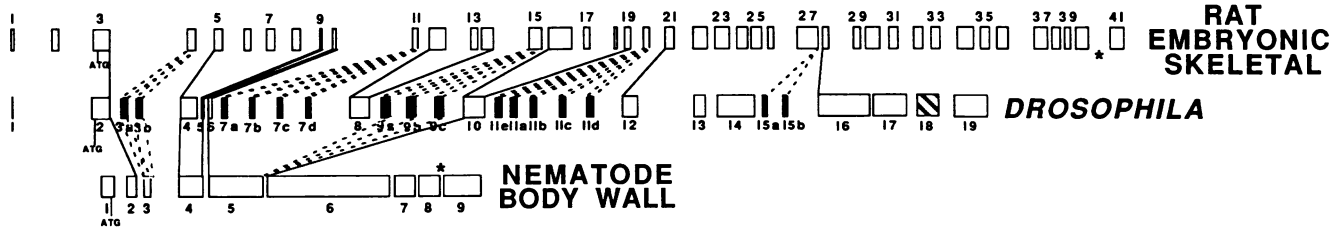


FIG. 4. Comparison of three MHC gene structures. Conserved exon borders between three MHC genes are indicated by connecting lines: ——, common exons in *D. melanogaster*; – – –, exons that are alternatively spliced in *D. melanogaster*. Genes are drawn to scale as for Fig. 3 except that introns are spaces rather than lines. *, The only intron position that is conserved between the rat embryonic gene and the nematode body wall gene but not in the *Drosophila* gene. Exons are numbered below the diagram for the *Drosophila* and nematode genes and above the diagram for the rat gene.

tionary conservation and unusual amino acid composition suggest that they have important functions (72). The N-terminal conserved region is encoded by common exons 4, 5, and 6 at residues 225 through 270. In contrast, the C-terminal conserved region begins in common exon 8 and crosses the N-terminal border of alternative exons 9a/9b/9c. This conserved region includes residues 455 through 510, and alternative exons 9a/9b/9c encode the C-terminal 43 residues. As with alternative exons 3a/3b, alternative exons 9a/9b/9c extend beyond the highly conserved region. A poorly conserved region, encoded by alternative exons 7a/7b/7c/7d and common exon 8, lies between the two highly conserved sequence regions. Like the conserved regions, this region has not been assigned a function.

Common exon 10 encodes the remainder of the 50-kDa domain and a region of divergent sequence at the 50- to 20-kDa junction, followed by two regions of high sequence conservation. The conserved primary actin-binding site (66) and reactive thiols SH2 and SH1 (74) are near the C terminus of common exon 10, and a likely secondary actin-binding site is encoded by alternative exons 11a/11b/11c/11d/11e (49). Binding of both myosin light chains has been demonstrated in sequences encoded by common exon 12, near the swivel or neck region of the MHC head (49).

The MHC tail is composed almost exclusively of common exons. However, alternative exons 15a/15b and differentially included exon 18 encode the only two regions of the tail that deviate from the repeating coiled-coil inflexible rod structure (44). Alternative exons 15a/15b encode 26 residues of the S2 hinge, at residues 1215 through 1240. The entire S2 hinge region includes residues 1125 through 1276. Thus, only an internal segment of the entire S2 hinge structural domain is encoded by the alternative exons 15a/15b. The sequence encoded by the alternative exons is the major helix-breaking sequence of the hinge, as revealed by computer-assisted secondary-structure predictions (24). The C-terminal noncoiled tailpiece is encoded by the last two residues of common exon 17 and differentially included exon 18 or common exon 19. Thus, alternative exon 18 encodes part of the noncoiled tailpiece by substituting a 3-residue form for the exon 19-encoded 29-residue form (5, 60).

This sequence analysis, therefore, shows that alternative exons of the *Drosophila* MHC gene encode sequences close to or within regions thought to have important enzymatic or structural functions, consistent with the view that these alternative exons have functional significance. Putative functional domains and regions of exceptional sequence conservation, however, are most often distributed in both alternative and common exons, and in some cases the boundaries of the alternative exons only partially overlap or include such domains.

Accumulation of the alternative exons in specific mRNAs. Alternatively spliced exon expression in MHC transcripts was assayed during the larval and pupal stages of muscle synthesis. Larval muscle types are synthesized at the late embryo and early larval stages, and adult muscle types are synthesized at the late pupal stage. To assay the stage specificity of expression of each alternatively spliced exon, exon-specific probes were either synthesized or isolated from regions of exon nucleotide sequence that are more than 30% divergent and do not cross-hybridize with related exons in genomic DNA. Synthetic oligonucleotide probes were used under appropriate conditions to prevent annealing of all but perfectly matched sequences (71). These probes were hybridized to Northern blots of larval and late-pupal RNAs to assay expression of alternative exons in MHC transcripts. These Northern blots also allowed us to assay the coexpression of specific exon sequences in transcripts that include the thorax-specific, differentially spliced exon 18 in pupae. Exon 18 accumulates only in late-pupal and adult MHC mRNAs, which are represented by three MHC mRNA size classes of 6.1, 6.6, and 7.1 kilobases (kb) (59). Exon 18 is expressed only in the thoraxes of late pupae and adults (5), in the 6.6- and 7.1-kb size classes of MHC mRNAs. These two size classes differ only in their 3′ ends because of selection of different polyadenylation sites (Fig. 3). In contrast, larvae expressed 6.1- and 6.6-kb MHC mRNAs, neither of which include exon 18. The late pupal 6.6-kb transcript that include exon 18 used the upstream polyadenylation site, whereas the larval 6.6-kb transcript uses the downstream polyadenylation site. These 6.6-kb transcripts could not be distinguished by migration on Northern gels, but since the 7.1-kb thorax-specific transcript has a distinct size, detection of hybridization of exon probes to 7.1-kb transcripts provides a marker for inclusion of this exon in transcripts containing the thorax-specific exon 18. Those MHC mRNAs that do not include exon 18 are also present in the thorax.

Figure 5 displays autoradiograms of Northern blots that were hybridized under stringent conditions with exon-specific probes. Probes were generated either by end labeling oligonucleotides or by uniformly labeling genomic restriction fragments. Alternative exons 3a and 3b are present in mRNAs from both larvae and late pupae (Fig. 5, row A). Exon 3a hybridized only with the 6.1- and 6.6-kb MHC mRNAs, whereas exon 3b hybridized with all MHC mRNA size classes, including pupal 7.1-kb transcripts. Thus, exons 3a and 18 are not spliced into the same mRNA. Exon 3b, in contrast, is present in mRNAs that either include or exclude exon 18.

Figure 5, row B, shows developmental-stage specificity and restriction of alternative splicing of the four alternative forms of exon 7. Exons 7a and 7b hybridized with the 6.1- and 6.6-kb mRNAs only. Thus, both exons 7a and 7b are not present in mRNAs that contain exon 18 but are expressed in both larval and late pupal RNAs. Exon 7b appears to be the least abundant form of exon 7 at both larval and pupal stages. Both 7c and 7d hybridized with 6.1-, 6.6- and 7.1-kb MHC mRNAs, indicating that they are expressed in mRNAs that include and exclude exon 18. Exon 7c and 7d autoradiograms (Fig. 5, row B) represent sequential hybridization of exon-specific probes to a single RNA blot. Exon 7d accumulated almost exclusively in late-pupal RNA, whereas exon 7c was not stage specific. Exon 7d differed from exon 18, which was also pupal specific, because exon 7d was present in all three MHC mRNA size classes, including the 6.1-kb mRNA, which was detectable upon fivefold-longer exposure (data not shown).

Alternative exons 9a, 9b, and 9c were present in both larval and pupal RNAs (Fig. 5, row C). The autoradiograms represent sequential hybridization of exon-specific probes to a single RNA blot. Exon 9c was the most abundant form at both larval and pupal stages of muscle synthesis (the film shown is a 10-fold-shorter exposure than those shown for exons 9a and 9b). All three alternatives of exon 9 hybridized with all size classes of MHC mRNA and therefore are expressed in mRNAs that either include or exclude exon 18.

Expression of alternative exons 11a, 11b, 11c, and 11d was found to be regulated (Fig. 5, row D). Exons 11a and 11d probes hybridized with both larval and late-pupal RNAs and with all three size classes of MHC mRNA. Exon 11c, in contrast, hybridized only with the 6.1- and 6.6-kb MHC mRNAs and therefore is not present in 7.1-kb mRNAs that

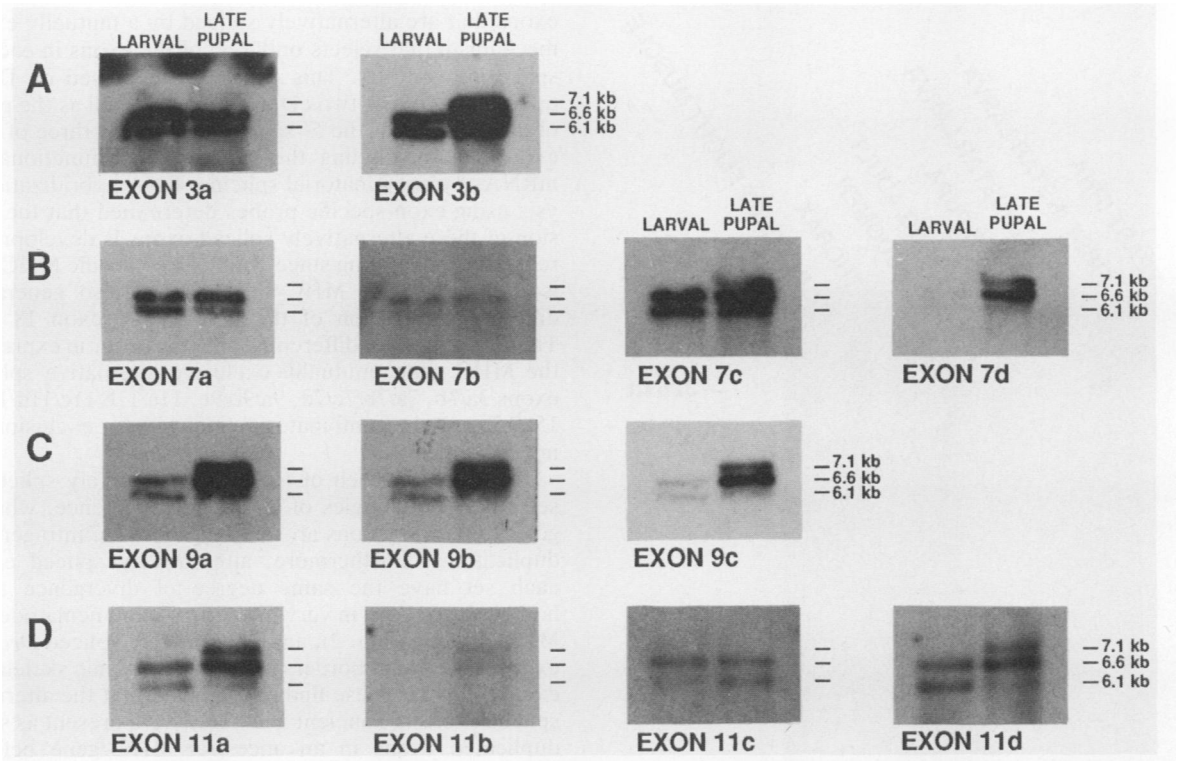


FIG. 5. Developmental accumulation and restriction of MHC S1 head alternative exons. In rows A to C, 2-μg samples of polyadenylated RNAs from first-instar larvae and late pupae were separated on a 12-cm-long formaldehyde-agarose gel and electroblotted to Zetabind; in row D, 0.5-μg samples of the same RNAs were separated and blotted in the same manner except that the gel was 15 cm long. Blots in rows A, B, and C were hybridized with exon-specific end-labeled oligonucleotide probes. High-stringency conditions were established by temperature of hybridization as calculated from individual probe sequences (71, 78). Oligonucleotide probes were tested for cross-hybridization with related exons by primer extension DNA sequence analysis of genomic DNA (data not shown). Blots in row D were hybridized with random-primed exon-specific genomic restriction fragments under stringent conditions. The blots within each row were exposed to Kodak XAR-5 film with an intensifying screen for equivalent times except for row C, in which a 10-fold-shorter exposure of exon 9c is shown. RNA sizes were determined by comparison with known RNAs. The 7.1-, 6.6-, and 6.1-kb bands correspond, respectively, to the 8.6-, 8.0-, and 7.2-kb bands of references 5 and 60.

include exon 18. Exon 11b was stage specific, accumulating only in late-pupal RNA. Exon 11b was expressed predominantly in the 6.6- and 7.1-kb size classes of MHC mRNA, but very weak expression could be detected in 6.1-kb RNA after fivefold-longer exposure of the autoradiogram (data not shown). Therefore, exon 11b expression is specific to the late pupal stage of muscle synthesis and is predominant in RNAs that also contain exon 18. Expression of exon 11e has not been assayed, but conserved amino acid sequence and splice junction placement suggest that it is a fifth alternative exon 11.

Exons 15a and 15b are the only alternative exons within the coiled-coil MHC tail. To assess their regulation, we assayed both their stage specificity and anatomical distribution of expression. Figure 6 shows results of an S1 nuclease protection assay of RNA isolated from developmentally staged whole animals and dissected anatomical regions from young adult flies. These results show that larval, total adult, and adult abdominal RNAs included exons 15a and 15b. In contrast, thorax RNA contained almost exclusively exon 15a, as seen as a single protected band of 326 nt. A faint signal at 84 nt, indicative of use of exon 15b, was detectable upon 10-fold-longer exposure (data not shown). Dissected IFM RNA, which represents approximately 90% of total thorax RNA, expressed only exon 15a. This abundance relationship suggests that the small amount of exon 15b detected in the whole thorax is contained in muscles other than the IFM. In situ hybridization to young adult thoraxes with exon-specific probes confirmed these results. Exon 15a mRNA is present in IFM and jump muscles, and exon 15b mRNA is present only in the direct flight muscle (G. Hastings and C. P. Emerson, Jr., unpublished results). Absence of protected fragments after yeast tRNA hybridization indicates that the probe does not reanneal under our hybridization conditions. Relative abundance levels of exons 15a and 15b can be compared in total animal RNAs because equal amounts by weight of first- and third-instar-larval, late-pupal, and young-adult RNAs were hybridized with probe. Whereas expression of exons 15a and 15b is not specific to developmental stage, 15b predominates in larval RNA and 15a predominates in adult RNA. Thus, while exon 15a is not uniquely expressed in the IFM, it is the only form expressed in IFM and the major form in the whole thorax. The results of these studies of exon hybridization, therefore, show that alternative exon expression in abundant MHC transcripts is precisely regulated in stage specificity, in combination with exon 18, and in muscle-type specificity. Exons 7d, 11b, and 18 are pupal specific. Exons 3a, 7a, 7b, and 11c are not present in transcripts containing the thorax-specific exon 18, whereas exons 7d and 11b are predominantly expressed in pupal RNAs including exon 18. Exons 15a and 15b are both expressed in the thorax, but exon 15a is the exclusive form expressed in the IFM.

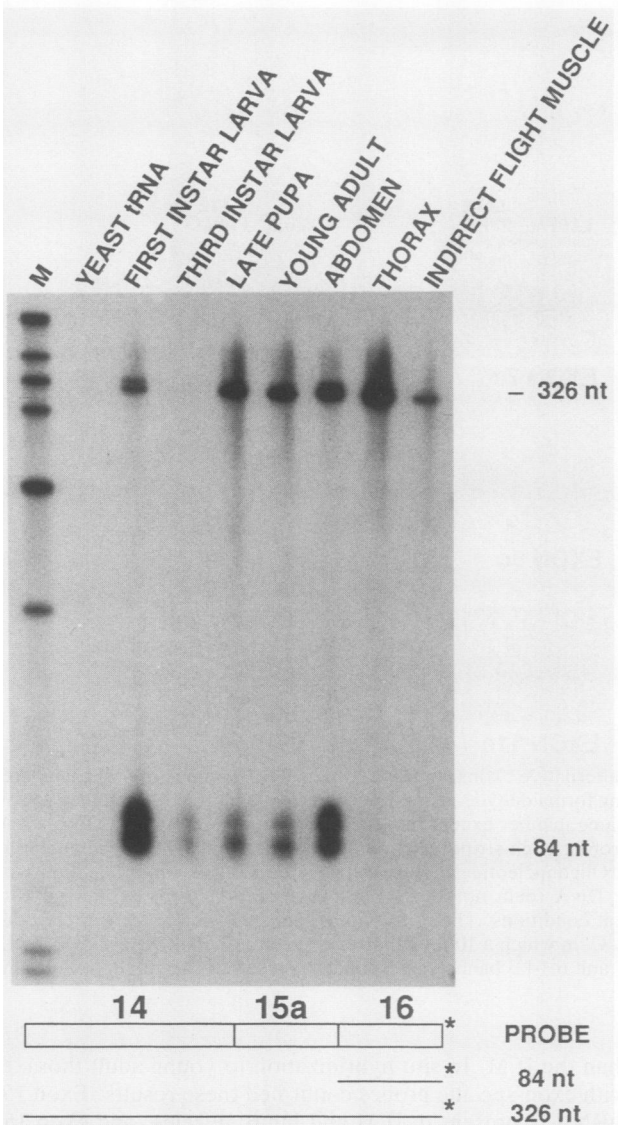


FIG. 6. S1 nuclease protection assay of exon 15 alternative splicing. The probe was generated by 5′ end labeling a 326-nt *Xmn*I-*Sst*I fragment of cD301. Probe (50,000 cpm) was hybridized with each of the following total cellular RNAs: 35 μg of yeast tRNA, 35 μg of first-instar larva, 35 μg of third-instar larva, 35 μg of late pupa, 35 μg of 2-day adult, 20 abdomens, 20 thoraxes, and 10 dissected IFM. Hybridization was at 50°C for 16 h. S1 nuclease-protected fragments were separated on a denaturing 8% polyacrylamide gel. The gel was exposed to Kodak XAR-5 film for 12 h with an intensifying screen. Two S1 nuclease-protected fragments were detected with this probe. The fragments are indicated on the right and in the diagram showing their relationship to cD301. Lane M (marker) is end-labeled *Hin*fI-digested pBR322 DNA.

DISCUSSION

We have determined the complete exon-intron organization of the single muscle MHC gene of *D. melanogaster*. This MHC gene contains 4 sets of tandemly duplicated exons coding for domains in the S1 head, 1 duplicated exon coding for part of the S2 hinge in the coiled-coil tail, and 14 unique exons. Exons were predicted by computer search of genomic DNA sequence, and their inclusion in mRNAs was confirmed by Northern analysis. One alternatively spliced exon is represented by as many as five tandemly duplicated exons that are alternatively spliced by a mutually exclusive mechanism that selects only one of the exons in each set to splice into mRNA. This conclusion is based on DNA sequence analysis of two cDNA clones as well as the presence of split codons at the 5′ splice junction for three of the five exon sets, precluding the generation of functional MHC mRNAs by combinatorial splicing. RNA hybridization analysis using exon-specific probes determined that the expression of these alternatively spliced exons is developmentally regulated, generating stage- and tissue-specific MHC mRNA isoforms. Distinct MHC mRNAs are also generated by differential inclusion of the penultimate exon 18 (5, 60). Thus, two types of differential splicing occur in expression of the MHC gene: mutually exclusive alternative splicing of exons 3a/3b, 7a/7b/7c/7d, 9a/9b/9c, 11a/11b/11c/11d/11e, and 15a/15b and the combinatorial (inclusion or exclusion) splicing of exon 18.

The exons in each of the five alternatively spliced exon sets share homologies of amino acid sequence, which suggests that these exons arose by a process of intragenic exon duplications. Furthermore, alternatively spliced exons in each set have the same degree of divergence as their homologous exons in very distantly related nematode and rat MHC proteins (Fig. 2), and alternatively spliced *Drosophila* exons share exact borders with rat embryonic skeletal MHC exons (Fig. 4). These findings indicate that the alternatively spliced exons are ancient and likely were present as single or duplicated exons in an ancestral MHC gene before the divergence of vertebrates and insects. The evolutionary maintenance of the protein-coding capacity of these apparently ancient exons in *D. melanogaster* and their regulated developmental expression suggest that these alternatively spliced exons have been subject to selection for sequence-specific protein functions.

If the *Drosophila* alternative exons share an ancestral relationship with exons in nematode, chicken, and rat genes, one would predict that duplicated and alternatively spliced exons might exist in other MHC genes. For instance, the rat embryonic skeletal MHC gene contains fragmented, apparently nonfunctional copies of known exons (65). These exon fragments correlate with both alternative and common *Drosophila* exons, which suggests that some exon duplications of alternatively spliced exons may have occurred before the separation of the vertebrate lineage. Continued characterization of MHC cDNA clones from the different muscle types of other organisms may reveal alternatively spliced exons in other MHC genes. Comparison of *Drosophila* alternatively spliced exons with sequences of rat and chicken MHC genes encoding the skeletal, cardiac, smooth, and nonmuscle isoforms may reveal relationships between individual alternatively spliced *Drosophila* exons and corresponding MHC gene family members.

At this time, *D. melanogaster* is unique in that alternative splicing of the 36B(2L) MHC gene, rather than a family of MHC genes, generates a complexity of MHC isoforms. The processes that have led to selection of an alternative RNA-splicing mechanism for generation of MHC and other contractile protein isoforms (20, 32) during *Drosophila* evolution are of interest, and comparative studies of the structures of muscle MHC genes in other insects and invertebrates may be revealing in this regard. One speculation for the adaptiveness of alternative splicing as opposed to a multigene-family mechanism for generating functional diversity of MHC proteins is related to the sensitivity of myofilament assembly to MHC gene dosage in the development of specialized muscles such as the IFM (50). An alternative splicing mechanism

allows for the generation of isoform diversity without perturbing the levels of protein expression, whereas gene duplications that lead to gene families alter dosage relationships of expression until establishment of a complexity of muscle-specific gene regulatory systems. This idea can be tested in vertebrates and insects by experimentally changing MHC expression levels transgenically.

Alternatively spliced MHC exons in the *Drosophila* MHC gene can generate functionally divergent MHC isoforms that may have important roles in the functional specialization of different muscle types. This expectation is based on the observation that these alternatively spliced exons encode protein sequences in regions of putative functional and structural MHC domains (Fig. 3). The protein sequences of the various alternative exon sets are related but, in most cases, divergent. Furthermore, the protein sequences contributed by alternatively spliced exon sets are embedded in larger MHC domains or regions of sequence conservation, which suggests that the specific sequence regions generated by these exons have local significance within the context of the function of the larger domain. These considerations assume that these alternatively spliced exons in MHC mRNAs are translated into MHC protein. A functional role of these exon sequences, however, seems probable, since all of the alternative exons are expressed as abundant, polyadenylated mRNAs, all have open translation reading frames that encode related protein sequences, and the amino acid changes in alternative exon sequences are largely conservative, indicating that these regions of sequence are structurally constrained.

The extraordinary complexity of the MHC gene exon structure creates a potential for this single gene to produce as many as 480 combinations of MHC isoforms, although as few as five combinations are sufficient to include all alternative exons in transcripts. It seems unlikely that all potential combinations are produced, a conclusion supported by our finding that three of the five sets of alternative exons have at least one version not present in mRNAs that include exon 18. Such restrictions favor the idea that the gene produces as few as 10 MHC isoforms, a number similar to that of MHC genes encoding mammalian MHC isoforms (76).

The complexity of *Drosophila* MHC isoforms expressed in individual muscle types is of considerable interest from the viewpoint of understanding MHC protein and isoform function in contractile activities of specialized muscles such as the IFM. Other contractile proteins such as actin (23) and myosin light chains (19) have isoforms expressed in muscles such as the IFM, and these isoforms likely interact with alternatively spliced exon domains of MHC. *Drosophila* actins are encoded by a family of six genes (22), and the actin gene at 88F is specifically expressed in the IFM (23). The MHC region encoded by alternative exon 11 has been tentatively defined as the secondary site of actin binding (49). Actin 88F may interact with MHC isoforms that include pupal-specific exon 11b, generating contractile properties specific to the IFM. Exons 11a and 11d are also expressed in the IFM and thus likely interact with actin 88F. The myosin alkali light-chain gene produces an IFM-specific isoform by regulated alternative splicing involving inclusion or exclusion of the penultimate exon, resulting in substitution of the C-terminal 12 amino acids (19). Both the alkali and the regulatory myosin light chains appear to bind MHC in the region encoded by common exon 12 (49). These binding models do not take into account the three-dimensional nature of the MHC head, so binding regions may actually be quite different. Alternative exon 3 encodes part of a hydrophobic region near the ATP-binding site of the MHC head. A single amino acid change in this region results in nonfunctional body wall muscle in *C. elegans*, indicating that the hydrophobic region is required for MHC function (16). The nonconservative amino acid differences in exon 3a and 3b sequences may result in MHC isoforms with specific properties. Exon 3a is not found in MHC 7.1-kb mRNAs that contain thorax-specific exon 18. This result suggests that mRNAs which contain exon 3a may encode MHCs that are specific to larval and adult intersegmental muscles rather than to the IFM. The IFM contains only one of the two forms of alternative exon 15, which encodes part of the S2 hinge. This flexible hinge region located within the rigid helical tail has been proposed as the site of motive force generation by helix-coil transitions (69, 70). Recent in vitro assays of MHC movement along actin fibers show that the S1 head alone is capable of movement and therefore can contribute to the generation of motive force (68). However, the hinge cannot be ruled out yet as a second site of force generation (31). The precisely regulated, IFM-specific expression of the alternatively spliced exon 15 strengthens the argument that the S2 hinge plays an important role in MHC function. Analysis of MHC isoform expression at the single-fiber level, using exon-specific RNA and antibody probes, will be necessary to define the cellular, subcellular, and thick-filament localization (48) of the different alternatively spliced exons in the various MHC protein isoforms in different muscle types.

We have determined that alternative splicing generates diversity in six separate regions of the multifunctional MHC protein and, as discussed above, that these alternatively spliced exons encode functionally significant enzymatic and structural domains in muscle-specific MHC isoforms. These alternatively spliced exons do not encode entire domains but rather encode segments of domains. Therefore, these exons may distinguish subregions within functional domains, which may determine specific contractile properties of MHC. *D. melanogaster* is an excellent organism for investigations of specific functions of such specific sequences in MHC isoforms. The presence of a single muscle MHC gene in the *Drosophila* genome makes feasible the use of both genetic and molecular genetic analyses to examine MHC function in vivo and in vitro. Since this single MHC gene, through alternative exon splicing, encodes all muscle MHC isoforms of the fly, genetic manipulations that affect specific protein functions can be undertaken without the complications of coexpression and complementation by other MHC genes in a multigene family. Alternative exons can be substituted in vitro and then reintroduced into the fly by P-mediated germ line transformation (61) to test function in vivo. Biochemical analysis of purified MHC isoforms also can be accomplished by expression of specific alternatively spliced cDNA clones in vitro. Thus, the functional significance of alternatively spliced MHC isoforms can be tested both in vitro and in the whole fly.

Genetic analysis of MHC function is also possible, and MHC gene mutations likely already exist with lesions that cause isoform-specific defects in IFM-specific exon splicing or protein function. The IFM has MHC requirements distinct from those of other muscle types (50). Two classes of dominant flightless mutations map to the MHC locus. The first class is dominant flightless and homozygous lethal. The second class is dominant flightless but homozygous viable. Several types of genetic lesions can account for these IFM-specific phenotypes, including mutations in IFM-specific transcriptional control elements, in IFM-specific splicing, or in amino acid sequences of alternative exons such as

15a, which is the only form present in the MHC mRNA of the IFM. The structure of the 36B MHC gene described here, therefore, provides a basis for analysis of the structure and expression of these IFM-specific MHC mutations.

The complex organization of the *Drosophila* muscle MHC gene described here raises several questions concerning the mechanisms involved in alternative splicing of MHC pre-mRNAs. Our analysis of restricted alternative exon combinations suggests that only a few regulated splice choices occur, which implies that alternative splice choice must be precise. This regulation includes selection of one exon in a set of as many as five possible alternatives. These mechanisms must involve *cis*-acting recognition sequences contained within the primary transcript and *trans*-acting factors, such as proteins and small nuclear RNAs, that recognize these sequences or their resultant secondary structures and either repress or direct splicing (see reference 63 for a recent review). We have examined and have not found any sequence correlations at the nucleotide level between introns flanking common versus alternative exons. We have also not found any homologous sequence elements near the 5′ and 3′ splice donors and acceptors of alternatively spliced exons that our analysis suggests are coregulated. To date, the only other example of a gene with more than two alternative splice choices is *Drosophila* tropomyosin gene 1, exons 9A through 9D (32). However, the splicing of these tropomyosin gene exons differs from that of the MHC gene, since the tropomyosin exon 9 is the most 3′ exon. The tropomyosin gene alternative RNA splicing involves a single alternative splice choice and differential utilization of polyadenylation sites, whereas the alternatively spliced exons of MHC are internal and thus alternative splice choices must be made at both the 5′ and 3′ ends of each alternative exon. Molecular genetic studies should reveal the mechanisms that regulate the complex alternative splicing of MHC gene transcripts and that coordinate the regulation of MHC splicing with expression and splicing of transcripts expressed from other muscle genes.

Transgenic experiments are now also possible to test specific hypotheses concerning MHC domains and their interactions in the diverse muscle types of the fly. Muscle-specific transcription of alternatively spliced MHC cDNA clones or modified gene segments can now be achieved by fusion with the 5′ end of the MHC gene, which directs marker gene expression in all larval and adult muscle types (E. L. George and C. P. Emerson, Jr., manuscript in preparation). Further analysis of transcription by the MHC gene 5′ end may reveal separable muscle type-specific transcriptional control elements, which will make expression of alternatively spliced isoforms in individual muscle types possible.

ACKNOWLEDGMENTS

We thank Anne Glenn for excellent technical assistance in construction and screening of the cDNA library and Deborah Pinney and Sonia Pearson-White for critically reviewing the manuscript. We thank Sanford Bernstein for communication of DNA sequence results before publication and William Kronert, Kevin Edwards, and S. Bernstein for informing us of potential exon 11e.

This work was supported by a Public Health Service Grant to C.P.E. from the National Institutes of Health. E.L.G. was supported by a Public Health Service predoctoral training grant from the National Institutes of Health.

LITERATURE CITED

1. **Amara, S. G., V. Jonas, M. G. Rosenfeld, E. S. Ong, and R. M. Evans.** 1982. Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products. Nature (London) **298**:240–244.
2. **Aviv, H., and P. Leder.** 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA **69**:1408–1412.
3. **Benton, W. D., and R. W. Davis.** 1977. Screening lambda gt recombinant clones by hybridization to single plaques *in situ*. Science **196**:180–182.
4. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids. Cell **12**:721–732.
5. **Bernstein, S. I., C. J. Hansen, K. D. Becker, D. R. Wassenberg II, E. S. Roche, J. J. Donady, and C. P. Emerson, Jr.** 1986. Alternative RNA splicing generates transcripts encoding a thorax-specific isoform of *Drosophila melanogaster* myosin heavy chain. Mol. Cell. Biol. **6**:2511–2519.
6. **Bernstein, S. I., K. Mogami, J. J. Donady, and C. P. Emerson, Jr.** 1983. *Drosophila* muscle myosin heavy chain encoded by a single gene in a cluster of muscle mutations. Nature (London) **302**:393–397.
7. **Breathnach, R., and R. Chambon.** 1981. Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
8. **Breitbart, R. E., H. T. Nguyen, R. M. Medford, A. T. Destree, V. Mahdavi, and B. Nadal-Ginard.** 1985. Intricate combinatorial patterns of exon splicing generate multiple regulated troponin T isoforms from a single gene. Cell **41**:67–82.
9. **Butler-Browne, G. S., and R. G. Whalen.** 1984. Myosin isozyme transitions occurring during the postnatal development of the rat soleus muscle. Dev. Biol. **102**:324–334.
10. **Caplan, A. I., M. Y. Fiszman, and H. M. Eppenberger.** 1983. Molecular and cell isoforms during development. Science **221**: 921–927.
11. **Casey, J., and N. Davidson.** 1977. Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. Nucleic Acids Res. **4**:1539–1552.
12. **Chirgwin, J. M., A. E. Przybla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
13. **Church, G. M., and W. Gilbert.** 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA **81**:1991–1995.
14. **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.
15. **Dente, L., G. Cesarini, and R. Cortese.** 1983. pEMBL: a new family of single stranded plasmids. Nucleic Acids Res. **11**: 1645–1655.
16. **Dibb, N. J., D. M. Brown, J. Karn, D. G. Moerman, S. L. Bolten, and R. H. Waterston.** 1985. Sequence analysis of mutations that affect the synthesis, assembly and enzymatic activity of the *unc-54* myosin heavy chain of *Caenorhabditis elegans*. J. Mol. Biol. **183**:543–551.
17. **Dodgson, J. B., and J. D. Engel.** 1983. The nucleotide sequence of the adult chicken α-globin genes. J. Biol. Chem. **258**:4623–4629.
18. **Emerson, C. P., Jr., and S. I. Bernstein.** 1987. Molecular genetics of myosin. Annu. Rev. Biochem. **56**:695–726.
19. **Falkenthal, S., M. Graham, and J. Wilkinson.** 1987. The indirect flight muscle of *Drosophila* accumulates a unique myosin alkali light chain isoform. Dev. Biol. **121**:263–272.
20. **Falkenthal, S., V. P. Parker, and N. Davidson.** 1985. Developmental variations in the splicing pattern of transcripts from the *Drosophila* myosin alkali light chain gene result in different carboxyl-terminal amino acid sequences. Proc. Natl. Acad. Sci. USA **82**:449–453.
21. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.
22. **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.
23. **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.
24. **Garnier, J., D. J. Osguthorpe, and B. Robson.** 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120**:97–120.
25. **Gubler, U., and B. J. Hoffman.** 1983. A simple and very efficient method for generating cDNA libraries. Gene **25**:263–269.
26. **Harrington, W. F., and M. Rogers.** 1984. Myosin. Annu. Rev. Biochem. **53**:35–73.
27. **Hastings, K. E. M., E. A. Bucher, and C. P. Emerson, Jr.** 1985. Generation of troponin T isoforms by alternative RNA splicing in avian skeletal muscle. J. Biol. Chem. **260**:13699–13703.
28. **Helfman, D. M., S. Cheley, E. Kuismanen, L. A. Finn, and Y. Yamawaki-Kataoka.** 1986. Nonmuscle and muscle tropomyosin isoforms are expressed from a single gene by alternative RNA splicing and polyadenylation. Mol. Cell. Biol. **6**:3582–3595.
29. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**: 351–359.
30. **Huynh, T. V., R. A. Young, and R. W. Davis.** 1985. Constructing and screening cDNA libraries in lambda gt10 and lambda gt11. p. 49–78. *In* D. M. Glover (ed.), DNA cloning, vol. 1. IRL Press, Washington, D.C.
31. **Hynes, T. R., S. M. Block, B. T. White, and J. A. Spudich.** 1987. Movement of myosin fragments *in vitro*: domains involved in force production. Cell **48**:953–963.
32. **Karlik, C. C., and E. A. Fyrberg.** 1986. Two *Drosophila melanogaster* tropomyosin genes: structural and functional aspects. Mol. Cell. Biol. **6**:1965–1973.
33. **Karn, J., S. Brenner, and L. Barnett.** 1983. Protein structural domains in the *Caenorhabditis elegans unc-54* myosin heavy chain gene are not separated by introns. Proc. Natl. Acad. Sci. USA **80**:4253–4257.
34. **Kiehart, D. P., M. S. Lutz, D. Chan, A. S. Ketchum, R. A. Laymon, B. Nguyen, and L. S. B. Goldstein.** 1989. Identification of the gene for fly non-muscle myosin heavy chain: *Drosophila* myosin heavy chains are encoded by a gene family. EMBO J. **8**:913–922.
35. **King, C. R., and J. Piatigorsky.** 1983. Alternative RNA splicing of the murine αA-crystallin gene: protein-coding information within an intron. Cell **32**:707–712.
36. **Laski, F. A., D. C. Rio, and G. M. Rubin.** 1986. Tissue specificity of *Drosophila* P element transposition is regulated at the level of mRNA splicing. Cell **44**:7–19.
37. **Leinwand, L. A., L. Saez, E. McNally, and B. Nadal-Ginard.** 1983. Isolation and characterization of human myosin heavy chain genes. Proc. Natl. Acad. Sci. USA **80**:3716–3720.
38. **Linzer, D. I. H., and D. Nathans.** 1983. Growth-related changes in specific mRNAs of cultured mouse cells. Proc. Natl. Acad. Sci. USA **80**:4271–4275.
39. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
40. **MacLeod, A. R., J. Karn, and S. Brenner.** 1981. Molecular analysis of the *unc-54* myosin heavy-chain gene of *Caenorhabditis elegans*. Nature (London) **291**:386–390.
41. **Mahdavi, V., M. Periasamy, and B. Nadal-Ginard.** 1982. Molecular characterization of two myosin heavy chain genes expressed in the adult heart. Nature (London) **297**:659–664.
42. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
43. **Maniatis, T., R. C. Hardison, E. Lacy, 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**: 687–701.
44. **McLachlan, A. D.** 1984. Structural implications of the myosin amino acid sequence. Annu. Rev. Biophys. Bioeng. **13**:167–189.
45. **Medford, R. M., H. T. Nguyen, A. T. Destree, E. Summers, and B. Nadal-Ginard.** 1984. A novel mechanism of alternative RNA splicing for the developmentally regulated generation of troponin T isoforms from a single gene. Cell **38**:409–421.
46. **Miller, A.** 1950. The internal anatomy and histology of the imago of *Drosophila melanogaster*, p. 468–481. *In* M. Demerec (ed.), Biology of *Drosophila*. John Wiley & Sons, Inc., New York.
47. **Miller, D. M., F. E. Stockdale, and J. Karn.** 1986. Immunological identification of the genes encoding the four myosin heavy chain isoforms of *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA **83**:2305–2309.
48. **Miller, D. M., III, I. Ortiz, G. C. Berliner, and H. F. Epstein.** 1983. Differential localization of two myosins within nematode thick filaments. Cell **34**:477–490.
49. **Mitchell, E. J., R. Jakes, and J. Kendrick-Jones.** 1986. Localisation of light chain and actin binding sites on myosin. Eur. J. Biochem. **161**:25–35.
50. **Mogami, K., P. T. O'Donnell, S. I. Bernstein, T. R. F. Wright, and C. P. Emerson, Jr.** 1986. Mutations of the *Drosophila* myosin heavy chain gene: effects on transcription, myosin accumulation, and muscle function. Proc. Natl. Acad. Sci. USA **83**:1393–1397.
51. **Molina, M. I., K. E. Kropp, J. Gulick, and J. Robbins.** 1987. The sequence of an embryonic myosin heavy chain gene and isolation of its corresponding cDNA. J. Biol. Chem. **262**:6478–6488.
52. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–472.
53. **Nabeshima, Y., Y. Kurijama-Fujii, M. Muramatsu, and K. Ogata.** 1984. Alternative transcription and two modes of splicing result in two myosin light chains from one gene. Nature (London) **308**:333–338.
54. **Pearson-White, S. H., and C. P. Emerson, Jr.** 1987. A novel hybrid α-tropomyosin in fibroblasts is produced by alternative splicing of transcripts from the skeletal muscle α-tropomyosin gene. J. Biol. Chem. **202**:15998–16010.
55. **Periasamy, M., D. F. Wieczorek, and B. Nadal-Ginard.** 1984. Characterization of a developmentally regulated perinatal myosin heavy-chain gene expressed in skeletal muscle. J. Biol. Chem. **259**:13573–13578.
56. **Periasamy, M., R. M. Wydro, M.-A. Strehler-Page, E. E. Strehler, and B. Nadal-Ginard.** 1985. Characterization of cDNA and genomic sequences corresponding to an embryonic myosin heavy chain. J. Biol. Chem. **260**:15856–15862.
57. **Pringle, J. W. S.** 1967. The contractile mechanism of insect fibrillar muscle. Prog. Biophys. Mol. Biol. **17**:1–60.
58. **Robbins, J., G. A. Freiger, D. Chisholm, and T. C. Gilliam.** 1982. Isolation of multiple genomic sequences coding for chicken myosin heavy chain protein. J. Biol. Chem. **257**: 549–556.
59. **Rozek, C. E., and N. Davidson.** 1983. *Drosophila* has one myosin heavy chain gene with three developmentally regulated transcripts. Cell **32**:23–34.
60. **Rozek, C. E., and N. Davidson.** 1986. Differential processing of RNA transcribed from the single-copy *Drosophila* myosin heavy chain gene produces four mRNAs that encode two polypeptides. Proc. Natl. Acad. Sci. USA **83**:2128–2132.
61. **Rubin, G. M., and A. C. Spradling.** 1982. Genetic transformation of *Drosophila* with transposable element vectors. Science **218**:348–353.
62. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
63. **Sharp, P. A.** 1987. Splicing of messenger RNA precursors. Science **235**:766–771.
64. **Stockdale, F. E., and J. B. Miller.** 1987. The cellular basis of myosin heavy chain isoform expression during development of avian skeletal muscles. Dev. Biol. **123**:1–91.
65. **Strehler, E. E., M.-A. Strehler-Page, J.-C. Perriard, M. Periasamy, and B. Nadal-Ginard.** 1981. Complete nucleotide and encoded amino acid sequence of a mammalian myosin heavy chain gene. J. Mol. Biol. **190**:291–317.
66. **Sutoh, K.** 1983. Mapping of actin binding sites on the heavy chain of myosin subfragment 1. Biochemistry **22**:1579–1585.
67. **Tong, S., and M. Elzinga.** 1983. The sequence of the NH_2-terminal 204-residue fragment of the heavy chain of rabbit skeletal muscle myosin. J. Biol. Chem. **258**:13100–13110.

68. **Toyoshima, Y. Y., S. J. Kron, E. M. McNally, K. R. Niebling, C. Toyoshima, and J. A. Spudich.** 1987. Myosin subfragment-1 is sufficient to move actin filaments *in vitro*. Nature (London) **328**:536–539.
69. **Ueno, H., and W. F. Harrington.** 1986. The temperature-dependence of local melting in the myosin subfragment-2 region of the rigor cross-bridge. J. Mol. Biol. **190**:59–68.
70. **Ueno, H., and W. F. Harrington.** 1986. Local melting in the subfragment-2 region of myosin in activated muscle and its correlation with contractile force. J. Mol. Biol. **190**:69–82.
71. **Wallace, R. B., M. Schold, M. J. Johnson, P. Dembek, and K. Itakura.** 1981. Oligonucleotide directed mutagenesis of the human β-globin gene: a general method for producing specific point mutations in cloned DNA. Nucleic Acids Res. **9**:3647–3656.
72. **Warrick, H. M., and J. A. Spudich.** 1987. Myosin structure and function in cell motility. Annu. Rev. Cell Biol. **3**:379–421.
73. **Wassenberg, D. R., II, W. A. Kronert, P. T. O'Donnell, and S. I. Bernstein.** 1987. Analysis of the 5′-end of the *Drosophila* muscle myosin heavy chain gene. J. Biol. Chem. **262**:10741–10747.
74. **Wells, J. A., and R. G. Yount.** 1982. Chemical modification of myosin by active-site trapping of metal nucleotides with thiol crosslinking reagents. Methods Enzymol. **85**:93–115.
75. **Weydert, A., P. Daubas, M. Caravatti, A. Minty, G. Bugaisky, A. Cohen, B. Robert, and M. Buckingham.** 1983. Sequential accumulation of mRNAs encoding different myosin heavy chain isoforms during skeletal muscle development in vivo detected with a recombinant plasmid identified as coding for an adult fast myosin heavy chain from mouse skeletal muscle. J. Biol. Chem. **258**:13867–13874.
76. **Wieczorek, D. F., M. Periasamy, G. S. Butler-Browne, R. G. Whalen, and B. Nadal-Ginard.** 1985. Co-expression of multiple myosin heavy chain genes, in addition to a tissue-specific one, in extraocular musculature. J. Cell Biol. **101**:618–629.
77. **Wydro, R. M., H. T. Nguyen, R. M. Gubits, and B. Nadal-Ginard.** 1983. Characterization of sarcomeric myosin heavy chain genes. J. Biol. Chem. **258**:670–678.
78. **Zoller, M. J., and M. Smith.** 1982. Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any fragment of DNA. Nucleic Acids Res. **10**:6487–6500.