Structure and DNA Sequence of the Tropomyosin I Gene from Drosophila melanogaster*

(Received for publication, March 18, 1985)

Guriqbal S. Basi and Robert V. Storti§

From the Department of Biological Chemistry, University of Illinois Health Sciences Center, Chicago, Illinois 60612

Tropomyosin, along with the troponins, is a key regulatory protein for the process of contraction in skeletal muscle. Tropomyosin is also found in smooth muscle, where it likewise functions in a regulatory capacity in contraction. However, as this tissue is lacking in the troponins, calcium sensitivity is conferred upon this system via the phosphorylation of light chains 1 and 3. In addition to muscle, tropomyosin has also been isolated from a variety of non-muscle sources, e.g., equine platelets (Cote and Smillie, 1981), fibroblasts (Fine et al., 1978; Bautch and Storti, 1983), and smooth muscle tropomyosins (Helfman et al., 1984), which are encoded by multigene families, the members of which are expressed in accordance with the tissue type from which they were isolated, as well as developmental stage of the tissue. The two isoforms each consist of 284 amino acids, with minor variations in sequence, and along with cardiac (Mak et al., 1979; Montarras et al., 1981), and smooth muscle tropomyosins, these isoforms have been observed. Two isoforms, denoted α and β, were originally observed in vertebrate skeletal muscle preparations. The ratio of these isoforms was found to vary according to the tissue type from which they were isolated, as well as developmental stage of the tissue. The two isoforms each consist of 284 amino acids, with minor variations in sequence, and along with cardiac (Mak et al., 1979; Montarras et al., 1981), and smooth muscle tropomyosins (Helfman et al., 1984), have relative mobilities of approximately 34,000–36,000. In contrast, the major form of tropomyosin isolated from non-muscle sources displays a consistently lower relative mobility of approximately 30,000 (Cote et al., 1978; Bautch and Storti, 1983). The lower relative mobility results as a consequence of a shorter chain length of 247 residues in equine platelet tropomyosin (Lewis et al., 1983). More recently, several additional less well-characterized tropomyosins with different relative mobilities have been isolated from additional non-muscle sources (see review by Payne and Rudnick, 1984).

In addition to tropomyosin, isoforms of several other contractile protein genes have now been observed. These isoforms are specific for the development of different muscle types, as well as for the later physiological specialization of different muscles, e.g., into slow twitch and fast twitch fibers (Storti and Rich, 1976; Garrels and Gibson, 1976; Horovitch et al., 1977; Whalen, 1981; Wilkinson and Grand, 1978; Wilkinson, 1980; Roy et al., 1979; Matsuda and Tomino, 1981; Montarras et al., 1981). Different isoforms of many contractile proteins also occur in non-muscle cells.

Recombinant DNA technology has facilitated the isolation of genes encoding the components of the contractile apparatus. With this approach, genes for all the major contractile proteins have been cloned from a broad evolutionary range and characterized to varying extents. There appear to be two basic classes of muscle genes which function during myogenesis and subsequent physiological specialization of muscle to give rise to isoforms of the same protein. On the one hand are proteins which are encoded by multigene families, the members of which are differentially expressed to give rise to the isoforms observed. Examples of this class include the actins and myosin heavy chain genes (Sanchez et al., 1983; Engel et al., 1982a, 1982b; Fyrberg et al., 1986; Hirsch et al., 1982; McKown et al., 1982; Nudel et al., 1982; Umeda et al., 1982). A second class of contractile protein genes have now been described which express multiple isoforms of the same protein by differential splicing of the primary transcript from a single gene. Examples of this class include the Drosophila tropomyosin and myosin light chain genes (Basi et al., 1984; Falken-
thal et al., 1985), the myosin light chain gene from chick, rat, and mouse (Nabeshima et al., 1984; Periasamy et al., 1984; Robert et al., 1984), and the mouse troponin gene (Medford et al., 1984).

We have previously reported the cloning of the tropomyosin genes from Drosophila melanogaster (Bautch et al., 1982). One of these genes encodes a tropomyosin which has a two-dimensional gel electrophoretic mobility similar to that of chicken skeletal muscle α-tropomyosin, and we have designated this gene tropomyosin I (TmI). This gene exhibits a complex pattern of developmental and tissue specific expression. During embryonic and larval stages the gene expresses two transcripts as assayed by RNA blot analysis, a narrow band of 1.6 kb in size and a broad band centering around 1.3 kb in size (Bautch et al., 1982). At pupal stage, and continuing into the adult life of the fly, the gene undergoes a switch in its mRNA processing scheme and expresses two new higher molecular weight transcripts of 1.9 and 1.7 kb in size. In the adult, these transcripts are expressed predominantly in the thorax and leg. Coincident with this processing switch is the appearance of a new isoform of TmI as detected by two-dimensional gel analysis (Rasi et al., 1984). In data presented elsewhere (Basi et al., 1984), we have shown that the processing switch involves alternative splicing of the fourth and fifth exons in the gene. The 5′ ends of these exons encode different carboxy-terminal 27 amino acids in the two isoforms of the TmI protein. Hence the Drosophila tropomyosin I gene is a single copy gene which encodes multiple isoforms of the same protein.

In this report we present a detailed structural and DNA sequence analysis of the TmI gene, the first such analysis of a tropomyosin gene. We show that the gene lacks the TATA consensus at its normal location upstream of the transcription initiation site. Instead the gene contains a series of "TG" sequences. A single major transcriptional initiation site is determined using the method of Maxam and Gilbert (1980). Such short direct repeats have been postulated as playing a role in the generation of deletions in the coding and noncoding regions of β-globin genes during evolution (Efstratiadis et al., 1980).

EXPERIMENTAL PROCEDURES

RESULTS AND DISCUSSION

DNA Sequence of the TmI Gene

The DNA sequence of the TmI gene and flanking regions was determined using the method of Maxam and Gilbert (1980) as described under "Experimental Procedures." The sequencing strategy is presented in Fig. 3B. The gene has been sequenced in its entirety with the exception of approximately 1 kb in the first intron.

Gene Overview—The DNA sequence of the gene and its flanking regions is presented in Fig. 4. Exons 2 and 3, which encode 90% of the TmI molecule are very G + C rich (59.6 and 61.8%, respectively). In contrast to the coding, the noncoding regions are 39.8% G + C for the 5′ noncoding leader sequence, and 41.2 and 38.3% in the fourth and fifth exons in the 3′ noncoding region. Similarly, the 5′ and 3′ flanking regions are much lower in G + C content also, having a value of 33.4 and 45.7%, respectively. The introns are also relatively lower in G + C content with values of 28.6, 39, 43.7, and 45.1% for introns 1–4, respectively. A similar distribution of base composition between coding versus noncoding regions of the small heat shock protein genes of Drosophila has also been reported (Southgate et al., 1983), as well as for the ribosomal protein genes of Drosophila (O’Connell and Roshbash, 1984).

5′ Flanking—We have determined the DNA sequence flanking the 5′ end of the gene up to 1172 base pairs upstream from the site of initiation of transcription. As stated above, the region is very A + T rich. A particularly prominent feature is a 27-base pair stretch of alternating AT residues from -710 to -737 in the sequence. In addition, there are numerous short direct repeats in this region (see Fig. 4 legend for details). Multiple occurrences of such numerous short direct repeats have been noted in the 5′ flanking region of the Drosophila 79B and 88F actin genes (Sanchez et al., 1983). Such short direct repeats have been postulated as playing a role in the generation of deletions in the coding and noncoding regions of β-globin genes during evolution (Efstratiadis et al., 1980).

A conserved sequence feature found in the region -70 to -90 in most eukaryotic genes, known as the CCAAT box (Efstratiadis et al., 1980), is found at position -116 in the TmI gene. This sequence serves to quantitatively effect transcription initiation in eukaryotic genes (Manley, 1983). Most surprisingly, however, the Drosophila TmI gene lacks the very highly conserved and nearly ubiquitous TATA box homology (Goldberg, 1979) found usually between positions -25 and -35 in almost all eukaryotic cellular genes characterized thus far. The first TATA box is located at position -158, and a second is found at -317. In vitro as well as in vivo experiments have verified that the TATA box functions in specifying the site of transcriptional initiation. Indeed, the gene lacks even the minimal component of this conserved sequence, the ATA box (Efstratiadis et al., 1980), as well as the occasionally observed functional variant, TAGA (Grosschedl et al., 1981; Robert et al., 1984). A handful of other eukaryotic cellular genes have now been characterized which are also lacking a TATA box homology in their "promoter" regions. These include the hamster 3-hydroxy-3-methylglutaryl coenzyme A reductase gene (Reynolds et al., 1984), the mouse hypoxanthine phosphoribosyltransferase gene (Melton et al., 1984), and the human adenosine deaminase gene (Valerio et al., 1985). These genes have several features in common. They are referred to by the authors as housekeeping genes, they lack both the "CAAT" and "TATA" box homologies in their promoter regions, and finally, their immediate 5′ flanking sequences are extremely rich in G + C content. The 3′-hydroxy-3-methylglutaryl coenzyme A reductase gene contains the hexanucleotide sequence CCGCCC upstream of its transcription initiation sites. Although not mentioned by the authors, this is also true for the hypoxanthine phosphoribosyltransferase gene. The adenosine deaminase gene, on the other hand, contains the complement of this sequence, GGGCGG, upstream of its transcription initiation site. The Drosophila TmI gene does not contain either version of the hexanucleotide sequence upstream of its initiation site, although it is very G + C rich in its immediate 5′ flanking region. The G + C content of the TmI gene in the region

1 The abbreviations used are: TmI, tropomyosin I; TmII, tropomyosin II; DTT, dithiothreitol; kb, kilobase (pairs); PIPES, pipericline-N,N′-bis(2-ethanesulfonic acid); SDS, sodium dodecyl sulfate.

2 Portions of this paper (including "Experimental Procedures" and Figs. 1–3) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814. Request Document No. 85M-0791, cite the authors, and include a check or money order for $5.60 per set of photocopies. Full size photocopies are also available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814. Request Document No. 85M-0791, cite the authors, and include a check or money order for $5.60 per set of photocopies.
**Structure and DNA Sequence of a Drosophila Tropomyosin Gene**

...the 3' untranslated region are coding regions. Numbering is based on the cap site reading downstream, and **boldface** arrowheads above the sequence denote the CAAT box at approximately -1064 to -1067. The coding region is further subdivided into codon triplets in the protein coding regions. Numbering is +1 beginning at the cap site reading downstream, and -1 at the first nucleotide upstream of the cap site and reading upstream. Representative short direct repeats in the 5' flanking regions are shown as underlined. The coding region is also the CAAT box at -3633. The multiple polyadenylation sites in the 3' untranslated region are shown in boldface and underlined. This is also the CAAT box at -110. The bold arrow beneath the sequence indicates the major cap site predicted by primer extension analysis. The two minor cap sites are determined by primer extension analysis. The multiple polyadenylation sites in the 3' untranslated region are underlined.
bounded by the CAAT homology and the transcription start site (119 base pairs) is 55.1%. The emergence of cellular genes which lack a TATA box may represent a different class of genes, in which the transcriptional initiation site is specified by additional and/or different factor(s), as compared with those genes that contain TATA-like sequences in their promoter regions.

Of possibly greater significance in the context of the TmI promoter are a series of multiple sequences of alternating dTdG (TG elements) residues located in close proximity to the 5' end of the gene. These consist of an imperfect 9-mer at position −105, an 8-mer at −89, and an imperfect 14-mer at −55. In addition, there occurs an imperfect 14-mer of the complement of the TG element at −357. There is also one short TG element (12-mer) in the first intron at +507, as well as three AC elements at +536, 627, and 713 which are a 10-mer, 8-mer, and 16-mer, respectively. These sequences are not observed at any other point in the TmI gene. Hamada et al. (1984a and 1984b) have shown that alternating purine pyrimidine residues of dTdG and its complement dCdA (referred to as the primer extension experiments described above is indicated by a solid arrow below the sequence as deduced from the primer extension experiments de-

promoter regions. Furthermore, no polyadenylation signals are observed in exon 4. In exon 5, 3 different sequences serve as polyadenylation signals as noted earlier. The occurrence and significance of these multiple polyadenylation signals has been discussed in depth elsewhere (Boardman et al., 1985) and hence, will not be discussed further here. The 3’ untranslated portion of the TmI gene comprises a significant proportion of the mRNA. Specifically, it constitutes 26 and 38% of the 1.3- and 1.6-kb embryo mRNAs, respectively, and 40 and 47% of the 1.7- and 1.9-kb thoracic mRNAs, respectively. Such long 3’ untranslated segments are a common feature of a number of other Drosophila genes (Sanchez et al., 1983) as well as numerous other eukaryotic genes (Setzer et al., 1980; Capetanki et al., 1983). To date, no functional significance has been ascribed to these long 3’ untranslated regions save for the polyadenylation signal sequences discussed above.

Introns

The four introns of the gene vary in size from 62 base pairs for the second intron, to approximately 2.2 kb for the first intron. The sequences at the intron junctions of the TmI gene are shown in Fig. 5. Splicing in this gene obeys the GT/AG rule, and shows close homology with the consensus sequences surrounding the splice sites (Breathnach and Chambon, 1981; Mount, 1982). It has been shown in yeast and other organisms that a sequence found 30–50 base pairs upstream from the 3’ end of introns serves as a recognition signal for acceptor site utilization. This recognition signal has been demonstrated as the site of branch point formation in a lariat structure, a splicing intermediate (for review, see Keller, 1984). For Drosophila, this signal has the consensus sequence CTAAT (Keller and Noon, 1984). We compared the sequences in the TmI introns upstream from the 3’ end of each intron to see if there were any constant features. In addition to the consensus branch point signal (shown boxed in the figure), we found several sequence motifs which were repeated at approximately equal distances from the 3’ ends of each of the four introns (shown underlined in the figure). Most prominent of all is a TACTCA and TATTCA motif found 26 and 25 base pairs upstream of introns 1 and 2, respectively. In the former case the branch point signal overlaps with this motif. The sequence TGATT is also found at positions 17 and 15 base pairs upstream from the 3’ ends of introns 1 and 2, respectively. By contrast, these homologies are not found in introns 3 and 4, the introns which are subject to alternate splicing. Instead, the sequences CAACACAAAC, and CAACACCACCA are found 23 base pairs upstream from the 3’ ends of both introns. As these introns are alternately spliced during development, it would be interesting if this motif were significant in this respect. There is no apparent homology with the alternatively spliced introns of the Drosophila myosin light chain gene (Falkenthal et al., 1985). Finally, we note that the TG and “AC” elements situated downstream of the mini-exon in the first intron have the potential for alternate stem-loop structure formation.

Significant progress has been made regarding the mecha-
**Structure and DNA Sequence of a Drosophila Tropomyosin Gene**

821

**Fig. 5. Intron junction sequences.** The eukaryotic consensus donor and acceptor sites (Mount, 1982) are shown above the DNA sequences at the junctions of the four introns of the Tml gene. The slash denotes the splice point. In the acceptor regions of the four introns, the homologies between the first and second introns, as well as between the third and fourth introns are shown underlined. The putative branch point consensus sequences for lariat formation (Keller and Noon, 1984) are shown in boxes.

**Fig. 6. Drosophila-vertebrate sequence comparison.** Amino acid sequence comparison of rabbit skeletal muscle a- and b-tropomyosin (Rab-aTm, Rab-bTm) (Stone and Smillie, 1978; Mak et al., 1980), chicken skeletal muscle tropomyosin (Chk-sk1) (MacLeod, 1982), chicken smooth muscle tropomyosin (Chk-sm) (Helfman et al., 1984), D melanogaster tropomyosin (Dm-TmI,E, Dm-TmIT denotes carboxyl-terminal 27 residues of thorax isoform) (Basi et al., 1984), and equine platelet tropomyosin (Eq-pltlt). The sequences are presented in the one letter amino acid code. The first 24 amino acids of chick skeletal muscle tropomyosin have not been determined, and are represented by dashes. Identity of the sequences with rabbit skeletal muscle a-tropomyosin are denoted by blank spaces below the residue. Hence, particular residues are only noted when they differ from rabbit skeletal-tropomyosin. Asterisks denote residues which are identical among all the sequences. The equine sequence is diverged from the muscle sequences at its amino terminus, and is presented as a best fit with the rabbit sequence in this region (residues 1–80). The dashes in this case indicate deletions incorporated into the equine sequence to accommodate the best fit (Lewis et al., 1983). Hence, the asterisks indicate identity between muscle sequences for residues 1–80, and between all sequences from residues 81–284. The vertical bars denote every tenth residue in the protein sequence.
Indeed, a great number of the conservative substitutions are Glu-Asp, Asp-Glu, α-Lys-Arg, Arg-Lys substitutions between the two proteins (15 and 10, respectively, or 25 out of the 60 and 64 differences). When the conservative substitutions are taken into account, the level of homology between the embryo and thorax isoforms and rabbit α-tropomyosin increases to 70.1%. The *Drosophila* protein also exhibits a very strong conservation of the pseudo-heptapeptide repeat unit characteristic of vertebrate tropomyosins (McLachlan and Stewart, 1975; Stone and Smillie, 1978).

The *Drosophila* and rabbit tropomyosin proteins show exact identity at their amino termini for the first 9 residues, and are highly homologous over their entire amino termini in general. Indeed, the comparison of all muscle tropomyosins (Fig. 6) shows this region to be highly conserved. The strong conservation in this domain of the molecule can be rationalized in terms of the molecules behavior in dilute salt solutions *in vitro*. Under these conditions, tropomyosin can be induced to aggregate in a head to tail manner, with an overlap of 8–9 residues at the amino and carboxyl termini. The highest degree of divergence between the two proteins occurs between residues 51 and 86, where 21 of the 32 residues consist of non-conservative substitutions. From this point on through to the carboxyl-terminal of the molecule, the two proteins display oligopeptide stretches of homology interrupted by short patches of non-conservative substitutions (see below).

The region around cysteine 190 and extending carboxyl-terminal in rabbit tropomyosin has been implicated as one of the two binding sites for troponin-T on the tropomyosin molecule. McLachlan and Stewart (1976) have proposed residues 297–217 to be the localized region of troponin binding. In this regard, although Cys-190 of the rabbit is substituted by Val-190 in the *Drosophila* protein, the two proteins, nevertheless, exhibit significant stretches of homology immediately carboxyl-terminal to residue 190. Only 3 non-conservative substitutions are observed between residues 191 and 217, suggesting that this locally higher conservation most likely reflects the role of this region as the site of interaction with the chymotryptic T2 subfragment (residues 159–258) of troponin-T (Chong and Hodges, 1982; Ohtuski, 1979; Stewart and McLachlan, 1976; Morris and Lehrer, 1984). Interestingly, this region of the tropomyosin molecule (residues 190–217) is one of two regions that displays the highest divergence in chicken smooth muscle tropomyosin from *Drosophila* and skeletal muscle tropomyosin. This can be rationalized by the fact that smooth muscle lacks a troponin-mediated regulatory system for contraction. It may be noteworthy in this regard that the *Drosophila* gene is split by an intron in this vicinity of the protein molecule, *i.e.* specifically between amino acid residues 198/199. There is no information regarding the troponin-T gene number in *Drosophila*, nor its mode of expression at the present time. However, a recent report on the rat troponin-T gene (Medford et al., 1984) indicates that it is also a single gene encoding multiple isoforms by alternative splicing. Furthermore, the alternately spliced region encodes different amino acids between residues 229 and 242, a region within the tropomyosin interacting T2 subfragment mentioned above.

In addition to the region extending carboxyl-terminal from Cys-190, residues 258–284, encompassing Tyr-261 and Tyr-267 of α-tropomyosin, have also been implicated as a second binding site for troponin-T (Mak and Smillie, 1981; Pearlstone and Smillie, 1982). This region, therefore, serves a dual function as both a troponin binding site, and in the head to tail polymerization of the molecule. This region of the molecule shows the highest amino acid divergence between all the tropomyosins (see below). Furthermore, this is the second region of the protein which is encoded by separate exons in the *Drosophila* gene. Indeed, these exons are the ones which undergo alternative splicing in different *Drosophila* muscle.

We have discussed previously the possible functional significance of this region of the protein, and isoform switching, as it relates to the function of the different *Drosophila* muscle types (supercontracting in embryos, and indirect flight muscle in adult thorax) in which it is expressed (Basi et al., 1984). We have evidence that other tropomyosin mRNAs and isoforms in *Drosophila* arise by splicing at the same position in the molecule. Hence, two important functional domains of the molecule correlate closely with amino acid differences among the tropomyosins and splice points in the gene.

**Tropomyosin Function and Evolution**—The developmentally regulated alternative splicing strongly suggests a functional basis for tropomyosin isoform diversity, particularly with regard to the carboxyl-terminal domain of the molecule. This raises the question as to whether other organisms generate tropomyosin diversity in a similar manner. For that reason, as well as to gain insights into evolutionary relationships among the different tropomyosins, we have compared the *Drosophila* tropomyosin sequence with the known tropomyosin protein sequences from other organisms. The asterisks in Fig. 6 indicate identity at amino acid positions 1–80 for the muscle proteins, and identity for all tropomyosins from positions 81 to 284. The platelet molecule possesses a 5-residue extension at its amino terminus, and two deletions between residues 1 and 80, as compared with muscle tropomyosins (Lewis et al., 1983). Therefore, it is not included in the comparison of the first 81 amino acids.

The amino acid sequences of all tropomyosins up to amino acid 257 are very highly conserved. However, if the differences among the molecules are compared, it can be seen that the differences in this region of the molecule fall into two groups. This grouping, however, is different than that found for the carboxyl-terminal domain of the molecule (see below). For instance, the rabbit β differs from the rabbit α sequence in 34 of the first 257 positions. However, of the 34 differences, 28 of the amino acids of β-tropomyosin are identical with the chick smooth muscle sequence. Most of the differences between the rabbit β and chicken smooth muscle tropomyosin occur in the region which binds the T2 fragment of troponin-T, and probably represents the different functions of the two proteins. Thus at the level of amino acid sequence, the β skeletal muscle is more closely related to the smooth muscle tropomyosin than it is to a skeletal muscle tropomyosin. The platelet sequence, while being highly diverged from the muscle tropomyosins in the amino-terminal domain, also shows much stronger homology to the rabbit β and chick smooth muscle than to the rabbit α and chick skeletal muscle sequence.

A comparison of amino acids 258–284 shows that they also fall into two groups. This grouping, however, is different from that for the first 257 amino acids. In this comparison, the rabbit α, β, and chicken skeletal sequences (all skeletal muscle tropomyosins) show only 5 differences in the last 27 amino acids. The chicken smooth muscle sequence, on the one hand, differs from the skeletal muscle tropomyosins at 21 of 27 positions. The smooth muscle sequence is identical with the platelet sequence at all but 8 positions. The similarity of the sequences in the two groups indicates that the members of each group probably have a common evolutionary origin for their last 27 amino acids, and that both groups have diverged from a common ancestor. The *Drosophila* sequence, while...
having diverged from both groups, is more similar to the skeletal muscle sequences than the cytoplasmic/smooth muscle sequences. Thus a curious dichotomy emerges for rabbit skeletal muscle β-tropomyosin wherein, the first 257 amino acids of the rabbit β sequence are more closely related to the chick smooth muscle and equine platelet sequences, while the carboxy-terminal domain is more closely related to the skeletal muscle sequences.

A final point worth noting in regards to the amino and carboxyl termini of tropomyosin is based on the following observations. First is the fact that the muscle tropomyosins show strikingly higher homology at the amino terminus as compared with the remainder of the molecule. The fact that such strong conservation in primary sequence is observed in this region of the molecule over such a broad evolutionary range, whereas the remainder of the molecule shows a considerably higher level of tolerance to evolutionary divergence, argues strongly in favor of a very strict functional requirement for this region of the molecule. Namely, the head to tail overlap of the molecule when it polymerizes. The carboxy-terminal 9–11 amino acids, on the other hand, have diverged relatively more between the sequences. Therefore, the amino terminus of the molecule must play a constant role in the polymerization of the molecule, and the functional specialization of the molecule to meet different physiological needs most likely comes about by modulation at the carboxy terminus, as reflected by the differences observed in the carboxy-terminal 27 amino acids of the tropomyosins.

The above analysis of the published tropomyosin protein sequences taken in the light of the *Drosophila* gene structure reveals a striking pattern of differences between the first 257 residues and the final 27 residues in the molecule. This analysis strongly suggests a different evolutionary history for these two domains. Given the above, it is reasonable to speculate that they have probably arisen by independent duplication and separate divergence from common ancestors. The structure and expression of the *Drosophila* tropomyosin gene provides a basis for the mechanism of exon duplication and/or rearrangement which could account for the groupings of tropomyosin domains described above.

The finding that single genes can generate multiple isoforms of the same protein by alternative splicing of a primary transcript has important implications concerning the functional significance of multiple isoforms via a vis multiple gene fusion. The model by Caplan et al. (1983) holds that isoform switching during myogenesis is functionally significant in that successive isoforms provide functional specialization to the muscle type during development. The alternative hypothesis (Lee et al., 1984) holds that isoforms arose as a consequence of duplication of an ancestral gene, followed by divergence of the genes. The duplication allowed for different genes to be controlled by different developmental programs, and the differences in the protein isoforms arose as a secondary consequence of the duplication event. While the issue of isoform specialization in this context for proteins encoded by multi-gene families (e.g. the actins) remains an open one, the finding of single genes encoding multiple isoforms whose expression is developmentally regulated argues in favor of a functional significance for the isoforms. Finally, the demonstration of isoforms which are encoded by a single gene, as well as by multi-gene families, indicates that the two hypotheses are no longer necessarily mutually exclusive, and in addition, provides evidence that each hypothesis, in its own context, may be correct.

Finally, an analysis of actin protein sequence deduced from the DNA sequence of the genes in *Drosophila* have revealed that the muscle actins of *Drosophila* are more similar in sequence to vertebrate cytoplasmic actins than they are to vertebrate muscle actins (Fyrberg et al., 1981). Our analysis reveals that tropomyosin is one contractile protein in *Drosophila* which does not follow this pattern, in that it is indeed more similar to other vertebrate muscle tropomyosins rather than vertebrate non-muscle tropomyosins. Hence, the muscle and non-muscle tropomyosins must have diverged from one another prior to the time of arthropod-vertebrate radiation.

**Acknowledgments**—We gratefully acknowledge the assistance of Dr. R. Kaul in operation of the DNA synthesizer, and Dr. P. J. Jagadeeshwaran for helpful discussions of Maxam and Gilbert sequencing methodology. G. S. B. acknowledges the support of J. K. Basi during the course of this work.

**REFERENCES**


Chirgwin, J. M., Pryzbyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) *Biochemistry* 18, 5294–5299


were sequenced min. The reactions were termmated by rapid dilution with 20 Chirgwin et
of ice cold S1
nuclease.

5' and labeling of DNA

5' and labeling was carried out using (-32P)dATP (sp. act. >7000 CPM/umol) and 74 polymerase kinase (Pharmacia-L
Biochemical) as described in Malecski et al. (1982). 3' end
labeling was carried out using (-32P)dCTP (sp. act. >7000 CPM/umol, American) and either T4 DNA polymerase, or Elenkow fragment of DNA pol.

5' Nucleic Acid Preparation

Total RNA isolated from e-7 Roman stocks of 12-24 h embryos, or 1-3 week old adult flies isolated according to the method of Chirgwin et al. (1979) for double stranded RNA, was used for hybridization and data analysis. Total RNA was extracted from embryos and adults according to the method of Malecski et al. (1982). RNA was prepared by phenol-extraction and precipitation as described in the legend to figure 1. RNase was omitted from the final washes to permit detection of residual RNase contamina
tion. In addition, RNA was prepared by a modification of the method of Chirgwin et al. (1979) in which RNase was present in the final washes to deplete the RNA of RNase.

RNA Sequencing

DNA sequencing was carried out essentially according to the method of Maxam and Gilbert (1980). The sequencing reactions were performed in a 10-20 umol volume of 0.1X TBE (pH 8.3) at 60°C. The sequencing reactions contained 0.1X TBE, 20 umol EDTA, and 1 umol each dGTP, dCTP, dATP, and dTTP. The sequencing reactions were allowed to proceed for 3-5 hours at 60°C. In each case, 2 umol of each of the 4 labeled nucleotides was added to the reaction mixture. The sequencing reactions contained 10 umol of purified DNA fragment. The sequencing reactions were terminated with 0.5X TBE rigor mortis solution (40 umol EDTA, 10 umol MgCl2, 100 umol NaCl, and 10 umol Tris-HCl, pH 7.4) and were separated into 10 volume ice cold 2X digestion buffer containing 500 to 6000 u of SI nuclease.

DNA Sequencing

DNA sequencing was carried out essentially according to the method of Chirgwin et al. (1982) for double stranded RNA. Hybridization and sequencing reactions were generally run overnight at 60°C in 10-20 umol volumes of 0.1X TBE. After hybridization, the DNA sequencing gels, followed by autoradiography of the gel at

Primer Extension

Primer extension was carried out essentially as described by Ghosh et al. (1987) with modifications described below. A 36 nt
single stranded DNA fragment was prepared as described in Malecski et al. (1982). The fragment was end labeled with

Transcript Mapping Results

Transcript map of the Fos gene (Kitamura et al. 1982) was in fig 1A. a. single stranded DNA fragment was prepared as described in Malecski et al. (1982). The fragment was end labeled with

Location of Initiation of Transcription

The 5' end of exon 2 mapped upstream of the Hind III site in fig 1C. lanes 1 and 3, and contains a 3' untranslated leader. The 5' end of exon 1 was at least 540 base pairs upstream of the Hind III site in fig 1C. lane 2, and contains a 3' untranslated leader. The 5' end of exon 1 was determined by primer extension analysis with the upstream primer used in fig 1A. Lane 2.

For high resolution mapping we employed SI nuclease analysis using 5' or 3' end labeled probes as described below. The SI digestion ladder was prepared by cleaving Sal I fragment 1. 2, 3, 4, 6, 8, 10, 12, 16, 20, 24, 30, 40, 60, 80, 100, and 120 base pairs upstream of the Hind III site with SI nuclease. The results from hybridization of this probe with adults 12-24 h embryos total RNA and adult total RNA we presented in fig 1, lanes 1, 2, 3, and 4. Both RNAs showed protection against nuclease digestion of a series of bands which migrate between approximately 16-200 nucleotides (nts.) in size. A more precise analysis of the 5' end of this exon is shown in fig 1, lanes 5-8, 9 and 10. 5' end. In lane 5 of fig 1, the probe, hybridized to 3300 CPM of S1 digested adult embryo RNA, was diluted with 3300 CPM of S1 end digested total RNA. Under these conditions, a single predominant band 88 nts. in size is observed from the adult RNA. This result is in agreement with the data shown in lanes 1, 2, 3, and 4 above. In lane 6 of fig 1, hybridized to 3300 CPM of S1 digested 12-24 h embryo RNA, a weaker band of 350 nts. is observed. These results are consistent with the results shown in lanes 1, 2, 3, and 4 above.

Gene Express. In addition, the 5' end of exon 1 was mapped by primer extension analysis. The results shown in lanes 1, 2, 3, and 4 above are consistent with the results shown in lanes 5, 6, 7, 8, 9, and 10 above. In lane 5 of fig 1, the probe was hybridized to 3300 CPM of S1 digested 12-24 h embryo RNA, was diluted with 3300 CPM of S1 end digested total RNA. Under these conditions, a single predominant band 88 nts. in size is observed. This result is in agreement with the data shown in lanes 1, 2, 3, and 4 above.
The complete structure, and DNA sequence of the Tml gene is shown in fig 3 A. The embryonic Tml mRNA consists of exons 1, 2, 3, and 5, and is spliced together in embryonic tissue as shown in the figure.
G. Nuclease protection experiment used to confirm the presence and map location of the 5' mini-exon, and mRNA cap site, deduced from experiments shown in part F and fig. 2 B. A single stranded 66 nt, Dde I fragment, 5' end labeled at the Dde I site indicated in B (probe G), was hybridized with equal amounts of yeast tRNA (lane 1), embryo RNA (lane 2), and adult RNA (lane 3) as described in materials and methods. The hybrids were treated with 5,000 U of enzyme at 37°C for 1 h. The DNA sequence ladder of the probe is shown alongside the S1 results. Again the sequence read from the ladder is of the non-coding strand. The arrows next to the sequence denote the nucleotides which comigrate with the 3 major nuclease resistant bands observed in lanes 2 and 3. Note: the products resulting from SI. digestion of DNA are retarded by 1 1/2 nts. in mobility with respect to the bands resulting from chemical cleavage of DNA (Weaver and Weissman 1979).

Exposure time in F was 10 days at -80°C with an intensifying screen.

Figure 3 Confirmation of primer extension results

A. Primer extension melting experiment. The 36 nt, single stranded Hco I – Hind III fragment, 5' end labeled at the Hind III site (probe F in fig. 2 B) was hybridized with embryo RNA at temperatures (C) indicated above the lanes. The annealed primer was subjected to reverse transcriptase catalyzed extension as described in materials and methods. P indicates the unhybridized primer. Lanes 1 and 2 contain end labeled Hae III and Hind I fragments, respectively, of phi174RF DNA. The sizes of the marker (in nts.) fragments are indicated alongside the lanes. Note the absence of the 127 nt, product when the primer is hybridized with RNA at 75°C, whereas the 139 nt, product persists. Both extension products are absent at hybridizations above 80°C.

B. Chemical cleavage DNA sequence ladder of the 139 nt, extension product from a preparative primer extension experiment. The sequence written in alongside the ladder begins at the Hae III site at the 3' end of the primer (also shown in fig. 1 F), and continues until the region of ambiguity in the middle of the ladder. The sequence continues above this ambiguous region. Note that the sequence here does not match with the corresponding sequence from the genomic fragment shown in fig. 1 F.

C. Chemical cleavage DNA sequence ladder of a 340 base pair genomic Sau 3A1 fragment to which the 5' mini-exon was mapped. The sequence shown is of the coding strand, and hence is complementary to the sequence above the ambiguous region in 2 B. The complementary nucleotides between the two ladders are shown underlined. The sequence alongside the ladder begins at the deduced mRNA cap site (see text), ind going up the ladder, reads downstream through the full length of the 50 nt, mini-exon. Note exact complementarity of this sequence with that in figs. 2B and 1G.

Figure 4 Gene structure and sequencing strategy

A. Transcript map of the Drosophila melanogaster Tmi gene. The solid boxes represent exons, and thin line represents introns and flanking DNA in the Tmi genomic DNA. Thin lines denote the splicing pathways used in embryonic and adult thoracic muscle to produce the mRNAs and tropomyosin isoforms expressed by the gene. The direction of transcription in A and B is from left to right.

B. Sequencing strategy used to determine the DNA sequence of the gene. This map is collinear with the transcript map in A. The arrows represent individual gel readings of DNA fragments 5' end labeled at the site above the base of the arrow. The arrows with solid circles at their base denote sequences which were determined by both the chemical method, as well as dideoxy sequencing of M13 subcloned fragments. Only the appropriate sites used in determining the sequence are shown. The length of the arrows is directly proportional to the length of sequence read from the gel.