MOLECULAR AND CELLULAR BIOLOGY, Feb. 1988, p. 679–694
0270-7306/88/020679-16$02.00/0

Vol. 8, No. 2

The Rat α-Tropomyosin Gene Generates a Minimum of Six Different mRNAs Coding for Striated, Smooth, and Nonmuscle Isoforms by Alternative Splicing

DAVID F. WIECZOREK,† CHRISTOPHER W. J. SMITH, AND BERNARDO NADAL-GINARD*

Laboratory of Molecular and Cellular Cardiology, Howard Hughes Medical Institute, Department of Cardiology, Children's Hospital, and Departments of Pediatrics, Physiology, and Biophysics, Harvard Medical School, Boston, Massachusetts 02115*

Received 7 July 1987/Accepted 23 October 1987

Tropomyosin (TM), a ubiquitous protein, is a component of the contractile apparatus of all cells. In nonmuscle cells, it is found in stress fibers, while in sarcomeric and nonsarcomeric muscle, it is a component of the thin filament. Several different TM isoforms specific for nonmuscle cells and different types of muscle cell have been described. As for other contractile proteins, it was assumed that smooth, striated, and nonmuscle isoforms were each encoded by different sets of genes. Through the use of S1 nuclease mapping, RNA blots, and 5′ extension analyses, we showed that the rat α-TM gene, whose expression was until now considered to be restricted to muscle cells, generates many different tissue-specific isoforms. The promoter of the gene appears to be very similar to other housekeeping promoters in both its pattern of utilization, being active in most cell types, and its lack of any canonical sequence elements. The rat α-TM gene is split into at least 13 exons, 7 of which are alternatively spliced in a tissue-specific manner. This gene arrangement, which also includes two different 3′ ends, generates a minimum of six different mRNAs each with the capacity to code for a different protein. These distinct TM isoforms are expressed specifically in nonmuscle and smooth and striated (cardiac and skeletal) muscle cells. The tissue-specific expression and developmental regulation of these isoforms is, therefore, produced by alternative mRNA processing. Moreover, structural and sequence comparisons among TM genes from different phyla suggest that alternative splicing is evolutionarily a very old event that played an important role in gene evolution and might have appeared concomitantly with or even before constitutive splicing.

Tropomyosin (TM) plays an important role in the regulation of contractility of muscle and nonmuscle cells. This protein forms a coiled-coil dimer and exists in several different forms (α, β, and nonmuscle-cytoplasmic) with apparent molecular weights ranging from 30,000 to 50,000. In striated muscles, the coiled-coil dimers form head-to-tail polymers that lie in the major groove of the thin filament, with each dimer spanning seven actin molecules (63). In striated and possibly in smooth muscle cells, TM plays a fundamental role in regulating the Ca^{2+} activation of the contractile apparatus through its interaction with actin and the Ca^{2+}-regulatory proteins troponin and caldesmon, respectively (see reference 39). However, the precise function of TM in nonmuscle cells remains unknown, although it might play a role in determining the architecture of the cell and modulate its contractile (16) and motile (32) activities.

The different TM isoforms have a tissue- and developmental stage-specific pattern of expression. Skeletal muscle is known to contain the α and β subunits assembled into homo- and heterodimers that are present in variable proportions depending on the fiber type (18–20, 59, 62). These two subunits are 284 amino acids (aa) long and highly conserved in sequence (38, 64, 67). In the adult myocardium of small mammals, only α-TM is detected (18), while fetal and hypertrophic myocardium expresses both the α and β isoforms (S. Jzumo, B. Nadal-Ginard, and V. Mahdavi, Proc. Natl. Acad. Sci. USA, in press). Protein and cDNA sequencing of smooth muscle (28, 31, 57, 58, 60) and some nonmuscle (27, 35, 36, 40) TMs has demonstrated that these isoforms are structurally different from their striated muscle counterparts. Moreover, some nonmuscle cells express up to five different isoforms (40) whose structure and functional characteristics appear to be different (9, 11, 32) but have not been fully elucidated.

Based on the heterogeneity and structural diversity of the TMs summarized above, it was assumed that these isoforms were encoded by a multigene family, with different genes giving origin to the striated α and β, smooth, and nonmuscle isoforms, respectively, as is known to be the case for most other contractile proteins (68). Molecular analyses on TM structure and expression in various organisms spanning several phyla have cast doubt on this assumption. It now appears that the multiplicity of TM isoforms is generated by a limited number of genes, each of which can generate a number of isoforms by alternative mRNA splicing and 3′-end processing. In *Drosophila melanogaster*, muscle-specific isoforms are generated from a single TM gene through alternative splicing of 3′-end exons, in conjunction with selective differential use of polyadenylation sites (2, 30). In mammalian systems, there appear to be at least three to four distinct genes including those that code for fast striated muscle α and β TMs (36, 54, 57; S. Cheley and D. Helfman, Mol. Cell. Biol., in press; Y. Yamawaki-Kataoka and D. Helfman, J. Biol. Chem., in press). Different isoforms are also generated through tissue-specific alternative splicing

* Corresponding author.
† Present address: Department of Microbiology and Molecular Genetics, University of Cincinnati College of Medicine, Cincinnati, OH 45267-0524.

from these genes (24, 24a, 25, 36, 57, 58; Cheley and Helfman, in press) that code for both striated and smooth-nonmuscle isoforms. So far, no gene has been identified that is specific for smooth and/or nonmuscle isoforms.

The isolation and determination of the structure and sequence organization of the rat α-TM gene (57) coupled with the previous isolation of striated and smooth muscle cDNAs (58) permitted a detailed examination of the α-TM gene expression. The results of S1 nuclease and Northern (RNA) blot analyses determined that the rat α-TM gene generates several different mRNAs by alternative splicing. Striated muscle-, smooth muscle-, and nonmuscle-specific isoforms are produced through a process of exon selection at the 3′, middle, and 5′ regions of the gene. This gene produces a minimum of six different mRNAs, including two that appear to be striated muscle specific. Primer extension and S1 nuclease analyses indicate that the gene contains a single promoter region with multiple sites of transcription initiation, none of which demonstrates a regulated preference of usage. The generation of striated (cardiac and skeletal) muscle-, smooth muscle-, and nonmuscle-specific TM isoforms from a promoter that is active in most, if not all, cell types highlights the importance of posttranscriptional processes such as alternative splicing in the generation of cellular phenotypes.

MATERIALS AND METHODS

Preparation of cultures. Rat L_6E_9 myogenic (45) cells were maintained in Dulbecco modified Eagle medium (GIBCO Laboratories, Grand Island, N.Y.) supplemented with 20% fetal calf serum (45). For induction of myogenic differentiation, the medium of exponentially growing cultures was replaced with Dulbecco modified Eagle medium supplemented with 3% horse serum and 3% fetal calf serum. Rat primary fibroblast cell cultures dissociated from skin, skeletal muscle, and cardiac muscle tissues were established on non-collagen-coated dishes (4.5×10^6 cells per 150-mm dish). Rat 3T3 fibroblasts, $NRK_{TS}371$ kidney fibroblasts, and $H_4A_3C_2$ hepatoma cells were passaged and grown in Dulbecco modified Eagle medium supplemented with 10% fetal calf serum.

Hybridization analyses. Total cytoplasmic RNA was isolated from cells grown in culture by the hot-phenol method (23, 65). Total RNA was obtained from rat tissues by modifications of the hot-phenol procedure (65). The procedure of Thomas (71) was used for the blotting and hybridization of RNAs electrophoresed in formaldehyde agarose gels. Southern blot analysis was conducted by established methods (66).

S1 nuclease mapping. RNA-DNA hybridization followed by S1 nuclease mapping analysis was performed by described methods (3) under the conditions used previously (73). The probes used were generated by digestion with restriction endonucleases and labeled at the 5′ end with [γ-^{32}P]ATP (Amersham Corp., Arlington Heights, Ill.) and T4 polynucleotide kinase (Bethesda Research Labortories, Inc., Gaithersburg, Md.) or at the 3′ termini with [α-^{32}P]ddATP (Amersham) and terminal transferase (Bethesda Research Laboratories). The DNA strands were separated, and the strand complementary to the mRNA was purified. Total cellular RNA (25 μg) was hybridized to 2×10^4 cpm of probe in 25 μl of 80% deionized formamide–400 mM NaCl–10 mM PIPES [piperazine-*N*,*N*′-bis(2-ethanesulfonic acid)] (pH 6.4)–0.05% sodium dodecyl sulfate–1 mM EDTA. The hybridization mixture was incubated at 65°C for 1 h, the temperature was adjusted to 60 or 42°C, and the incubation continued for 16 h. S1 nuclease (100 units; New England Nuclear Corp., Boston, Mass.) in 300 μl of 200 mM NaCl–30 mM sodium acetate (pH 4.5)–3 mM $ZnSO_4$ was added to each sample and incubated at 25°C for 1 h. The reaction was terminated with 10 mM EDTA and precipitated with ethanol. Dried pellets were dissolved in 80% formamide and electrophoresed on 6 or 8% polyacrylamide–8 M urea sequencing gels. The gel was dried and exposed for autoradiography on Kodak X-Omat AR film.

DNA sequence analysis. DNA-sequencing reactions were done as described by Maxam and Gilbert (41) or by the dideoxynucleotide chain termination method (61) after subcloning into M13mp18/19 or into the double-stranded sequencing vector pGEM3.

Primer extension analysis. Primer extension analysis was conducted on total RNA by a previously described method (72). Synthetic oligonucleotide primers complementary to the coding sequence were 5′ end labeled with [γ-^{32}P]ATP by using polynucleotide kinase. Total RNA (50 μg) was suspended in 7 μl of primer extension buffer and hybridized with 10^6 cpm of labeled primer at 65°C for 15 min. To initiate primer extension, the incubation mixture was transferred to 42°C and 0.5 μl of 20 mM deoxynucleotide triphosphates and 1 μl of reverse transcriptase (20 U/μl) was added. After 3 h of incubation at 42°C, the samples were ethanol precipitated and electrophoresed on an 8% polyacrylamide–8 M urea gel followed by autoradiography.

RESULTS

Two cDNAs specific for rat striated and smooth muscle α-TM isoforms have been previously isolated and characterized in our laboratory. The sequence of these two clones suggested that these two isoforms were encoded by the same gene (58). We have recently determined the structure of the α-TM gene and demonstrated that this 28-kilobase-pair-long gene produces the smooth and striated isoforms by alternative splicing of five exchangeable isotype-specific exons (57). Analyses of the expression of this gene in liver mRNA and liver-derived cell lines led us to conclude that this gene was not expressed in nonmuscle cells (57, 58). As shown below, however, liver was a poor choice of nonmuscle tissue since it represents the exception rather than the rule for the expression of this gene. Recently, we have identified an additional putative exon in a *Pst*I fragment within the intron 3′ of exon 6 which was not previously sequenced (see below). An update of the organization of the rat α-TM gene and of the previously reported cDNAs is shown in Fig. 1. Probes generated from the two cDNAs and the corresponding genomic clones were used to analyze further the pattern of expression, as well as the structure, of the different mRNAs generated by the α-TM gene in both muscle and nonmuscle cells.

Expression of α-TM mRNAs with the 3′ end of smooth muscle α-TM in all tissues analyzed with the exception of liver. To investigate the pattern of expression of the α-TM gene in muscle and nonmuscle cells and to determine the pattern of exon incorporation into the mature cytoplasmic mRNAs, we prepared a 650-nucleotide (nt)-long 3′-end-labeled single-stranded probe spanning from codon 120 to the middle of the 3′ untranslated (UT) region of the smooth muscle cDNA sequence. This probe was hybridized to total RNA prepared from different tissues and cultured cells, and the hybridization products were digested with S1 nuclease (Fig. 2A). This probe was fully protected from S1 nuclease digestion by all

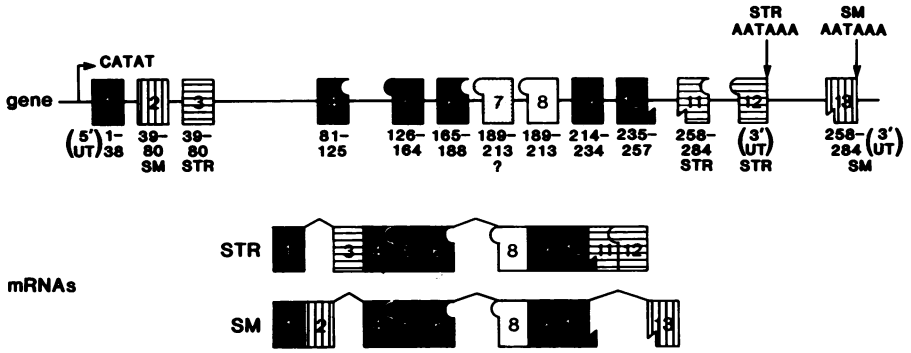


FIG. 1. Schematic representation of α-TM gene intron-exon organization and the striated (STR) and smooth (SM) muscle mRNAs. The gene consists of at least 13 exons with the first encoding the 5′ UT sequences and the amino-terminal 38 aa residues. Black boxes indicate constitutively spliced exons found in all the RNAs so far studied: horizontally striped boxes represent the alternatively spliced exons in the striated muscle cDNA: vertically striped boxes represent alternatively spliced exons in the smooth muscle cDNA. The white box represents an exon that is skipped in one of the mRNAs, while the stippled box is the newly identified putative exon that is thought to replace it in these RNAs (see text). The shape of the boxes indicates the codon triplet structure of each exon. Flush boundaries indicate that the exon begins or ends with an intact codon: concave-convex boundaries indicate that the upstream exon ends in a split codon lacking a single nucleotide contributed by the downstream exon: sawtooth boundaries indicate that the upstream exon ends lacking two nucleotides contributed by the downstream exon in the mature mRNA, as deduced from the genomic sequence.

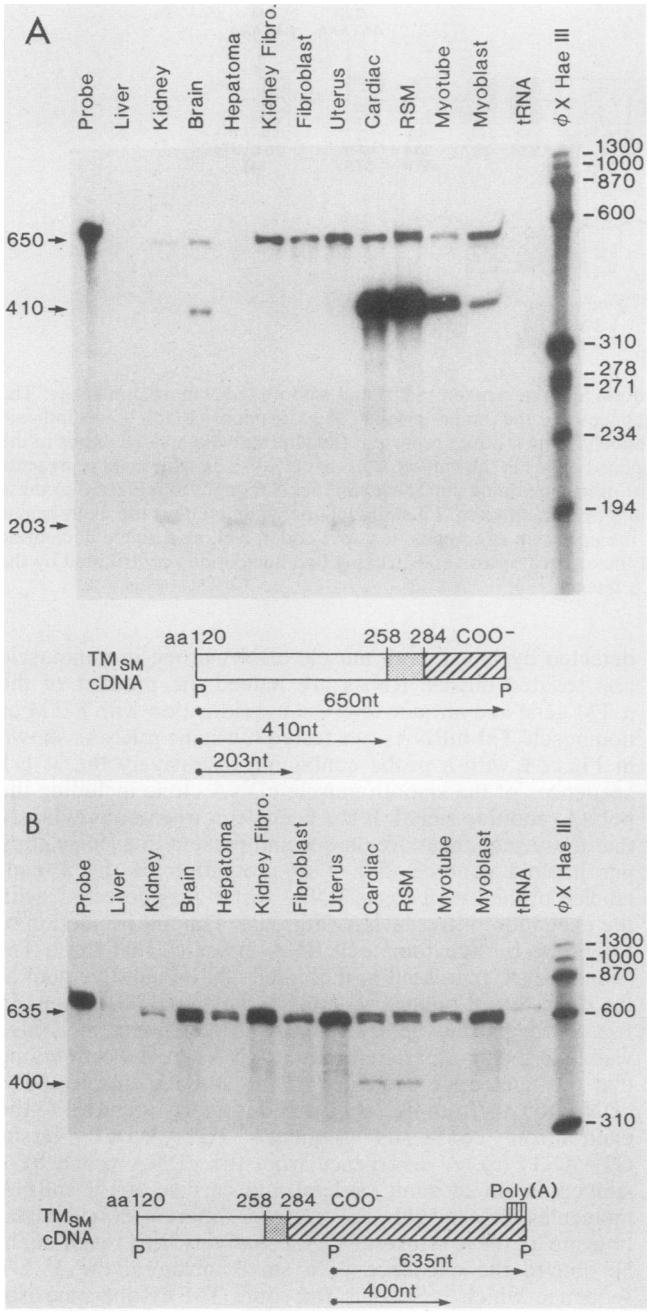
cellular RNAs tested, with the exception of the liver sample, which did not show any protected fragment. Hepatoma cell RNA also failed to give full protection of the probe but produced a 203-nt-long partially protected fragment. This partially protected fragment corresponds to an mRNA that is identical to the probe in the region spanning codons 120 to 189 of the α-TM gene. The divergent point at codon 189 corresponds to the 3′ end of exon 6 (Fig. 1) and indicates that the sequences corresponding to exon 8, present in the probe, are missing from this mRNA. A protected fragment of the same length was also visible in kidney, kidney fibroblast, and uterus RNA lanes. In addition, a 410-nt-long protected fragment was very prominent in all striated muscle and brain RNA lanes. This protected fragment identifies an abundant mRNA that is identical to the probe from codons 120 to 257. This last codon is located at the 3′ end of exon 10, which is the last constitutive exon of the gene. The mRNA represented by this partially protected fragment increases significantly with the progression of muscle differentiation, from myoblast to myotube and adult skeletal or cardiac muscle. This behavior further indicates that it represents the striated muscle isoform, which diverges from the smooth muscle counterpart represented by the probe at the boundary between codons 257 and 258.

The full protection of the smooth muscle probe by all mRNAs (with the exception of RNAs from liver and cells of hepatic origin) shown in Fig. 2A clearly indicates that the α-TM gene is expressed in muscle and nonmuscle cells. Moreover, nonmuscle as well as striated muscle cells accumulate mRNAs that contain the smooth muscle 3′ coding and UT sequences. These mRNAs are indistinguishable from the smooth muscle mRNA at least in the region covered by the probe. The existence of an mRNA that diverges from the smooth muscle cDNA at the boundary of exons 6 and 8 in several nonmuscle tissues indicates that exon 8, until now believed to be constitutive and present in all mRNAs, is spliced out in some tissues. It is also currently unknown whether the faint ~210- and ~180-nt-long bands observed represent additional low-abundance α-TM transcripts, the results of incomplete S1 nuclease digestion, or cross-hybridization with β-TM mRNAs.

To ascertain further that the smooth-cytoplasmic isoforms detected by the smooth muscle cDNA probe in nonmuscle and striated muscle RNAs are indeed the product of the α-TM gene and not due to cross-hybridization with β-TM or nonmuscle TM mRNAs, we tested the same mRNAs shown in Fig. 2A with a probe containing exclusively the 3′ UT sequences of the smooth muscle cDNA clone including the poly(A) addition signal. It has been demonstrated previously that these sequences are unique and present in a single copy per haploid genome (57). Full protection of the 3′-end-labeled probe by all tissue RNAs tested was detected, with the exception of liver RNA (Fig. 2B). The full protection of the probe by hepatoma cell RNA indicates that the α-TM gene mRNA expressed in these cells, although divergent at the exon 6 to 8 boundary, utilizes the same UT region. In Fig. 2B, a partially protected 400-nt-long fragment is also visible in the adult striated muscle RNAs. It is worth noting that the nucleotide sequence of the smooth muscle α-TM cDNA 400 nts from the labeled end is nearly identical to the exon-intron consensus sequence (GTAAGAT versus GTAAGT) (5). A divergence from the cDNA probe by a subpopulation of adult skeletal and cardiac α-TM mRNA molecules may indicate the use of a different polyadenylation site by these transcripts. Alternatively, this result might be due to the existence of a small intron in the 3′ UT sequence which is specific for some TM cytoplasmic isoforms that are expressed in skeletal, cardiac, and nonmuscle tissues. Therefore, although not conclusively proven by the present data, it is possible that mRNA species with heterogeneity only in their 3′ UT sequences are also generated by the α-TM gene.

The results presented above demonstrate that, with the exception of liver, the α-TM gene is expressed in all tissues and cells tested and produces mRNAs that have a 3′-end structure similar to that of the smooth muscle isoform and include exon 13 of the gene (Fig. 1). This alternative 3′ end, therefore, is not unique to the smooth muscle cells and represents a smooth muscle/nonmuscle 3′-end sequence.

Specificity of the 3′-end sequence encoded by exons 11 and 12 for striated muscle α-TM mRNAs. The finding of nonmuscle and striated muscle mRNAs with the 3′ end of the smooth muscle cDNA clone raised questions about the specificity of the 3′ end corresponding to the striated muscle



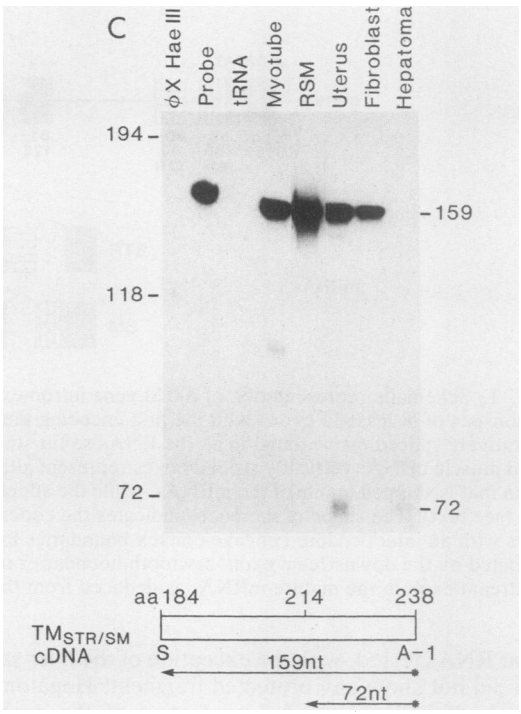


FIG. 2. Detection of α-TM gene mRNAs in muscle and nonmuscle cells and tissues by using common and smooth muscle sequences. (A) The smooth muscle 3′ end is expressed in muscle and nonmuscle cells. A single-stranded probe derived from the smooth muscle cDNA clone (TM_{SM}) extending from codon 120 to the middle of the 3′ UT sequence was labeled at the *Pst*I (P) restriction site (*). Hybridization and S1 nuclease digestion were as described in Materials and Methods. The length and sequence organization of the probe, as well as the partially protected fragments detected, are indicated schematically at the bottom of the figure. It is clear that the partially protected fragments are cleaved at the boundaries between exons (see text). Myoblast and myotube RNAs are from L_6E_9 cells; RSM is adult rat skeletal muscle; cardiac, uterus, brain, kidney, and liver represent RNAs from the respective tissues obtained from adult animals; fibroblast is RNA from rat 3T3 cells; kidney fibro. is RNA from $NRK_{TS}371$ cells; and hepatoma is from $H_4A_3C_2$ cells. *Hae*III-digested φX nucleotide size markers are indicated on the right. (B) Detection of smooth muscle (SM) and cytoplasmic-nonmuscle α-TM mRNA probing with the 3′ UT region extended through the poly(A) tail. Full protection of the 635 nts corresponding to the 3′ UT sequences is detected in all samples except liver and the tRNA used as the control. An additional partial protection of 400 nts is visible in the adult striated muscle RNA samples. Markers are as described for panel A. (C) Analysis of the α-TM mRNA by using an internal region common to the striated muscle (STR) and smooth muscle (SM)-nonmuscle cDNAs. Full protection of the 159-nt probe by muscle and nonmuscle RNAs (with the exception of liver) is detected. A partial protection of 72 nts representing a protection of codons 214 to 238 is detected in hepatoma and uterine RNA. The identity of the band at ~107 nt is unknown.

isoform. For this reason, a probe generated from the striated muscle α-TM cDNA clone was used to determine the pattern of expression of the presumed striated muscle-specific 3′-terminal sequences encoded by exons 11 and 12 (Fig. 1). The probe was 3′ end labeled at codon 184 and contained sequences extending to the poly(A) addition site.

The results (Fig. 3) demonstrate that complete protection of the probe from S1 nuclease digestion is only obtained with striated muscle RNAs in which the abundance of the homologous mRNA increases with the differentiated state of the muscle. No fully protected probe could be detected in the nonstriated muscle samples even after longer exposures of the autoradiograms. This result indicates that, in contrast to the findings with the smooth muscle probe, the α-TM 3′ end specified by exons 11 and 12 is highly restricted to striated muscle.

As expected, all samples except hepatoma RNA protected a 219-nt-long fragment which corresponds to the region shared between striated muscle and smooth muscle-nonmuscle mRNAs. The absence of a full protection of the probe by brain RNA is noteworthy given the results shown in Fig. 2A. This result demonstrates that the 410-nt-long fragment protected with brain RNA (Fig. 2A) does not contain a sequence identical to that of the fragment of the same length protected by striated muscle RNA. It is possible that the brain isoform omits the sequences corresponding to

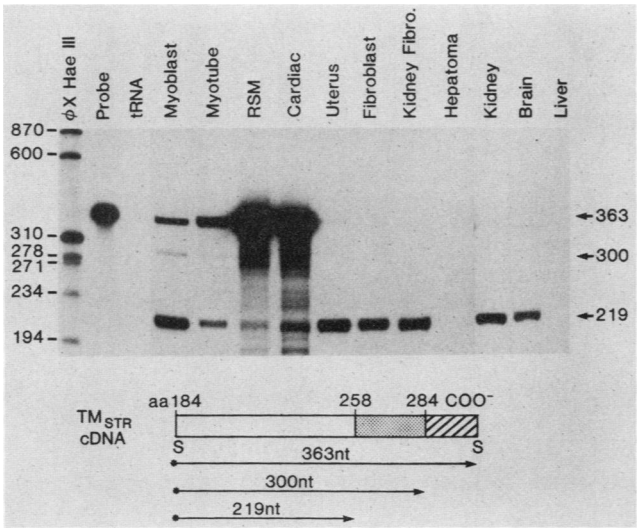


FIG. 3. Detection of α-TM mRNAs probing with common and striated muscle-specific sequences. A 3'-end-labeled single-stranded striated muscle (STR) cDNA probe derived from the *Suc*I (S) restriction fragment depicted in the diagram at the bottom was hybridized to the same RNA samples used for Fig. 1. Full protection of the probe is detected only in striated muscle samples. Partial protection of 219 nts corresponding to the 3'-end boundary of exon 9 is seen in all RNA samples except liver. Another partial protection of 300 nts is seen in the striated muscle samples. This band represents an mRNA that has the sequences corresponding to exon 10 (codons 258 to 284) but lacks the sequences corresponding to exon 11, containing the striated muscle-specific 3' UT sequences.

exon 11 (coding for aa 258 to 284) and splices exon 10 directly to exon 12, which contains the stop codon. An example of this splicing pattern has been demonstrated in the avian system by the isolation of a corresponding cDNA clone (24a). Attempts to demonstrate this mRNA structure in brain RNA by probing with the corresponding synthetic oligonucleotides were unsuccessful. Alternatively, it is possible that there is an as yet unidentified exon(s) downstream of exon 10. This possibility has gained credence from the isolation of a brain cDNA clone with sequences 3' of exon 10 that are not present in either smooth or striated muscle α-TM cDNAs (D. Helfman, personal communication).

In Fig. 3, the partial protection of a 300-nt fragment in striated muscle RNAs corresponds to a minor isoform which diverges in nucleotide sequence after aa 284. Rather than splicing exon 11 to exon 12, which contains the striated muscle UT sequences, this isoform may splice exon 11 to exon 13, the nonmuscle-smooth muscle 3'-terminal exon. To test this possibility, probes end labeled within the 3' UT region of the smooth muscle cDNA and extending up to a region corresponding to codon 213 demonstrated that striated muscle RNAs protected a fragment extending up to codon 258 (the 5' end of exon 13) of the smooth muscle-nonmuscle cDNA probe (data not shown). This protection pattern can only be obtained with striated muscle mRNAs which do not splice exon 10, the last common exon coding for aa 235 to 257, to exon 13, which contains codon 258 and the 3' UT region of the nonmuscle isoform, since in that case a fully protected fragment would be expected with this probe. These combined results indicate that, in this particular isoform, exons 10, 11, and 13 are spliced together (see Fig. 7).

The results shown in this section demonstrate that exons 11 and 12 are striated muscle specific but that in addition to these major forms, most splicing combinations involving exons 11, 12, and 13 can be found in mRNA from muscle tissues, albeit in low abundance.

Mutually exclusive splicing of exons coding for aa 189 to 213 in nonmuscle cells. The TM mRNA isoform expressed in hepatoma cells as well as in other cell types shown in Fig. 2 and 3 diverges from the smooth muscle cDNA sequence at codon 189 and employs the same 3' UT region as the other nonmuscle tissues. Yet, this mRNA only partially protects the internal exons from S1 nuclease digestion. To map the 3' end of the divergence of this mRNA, we prepared a probe to examine the exons spanning codons 184 to 238 (Fig. 2C). This probe was 5' end labeled at a nucleotide corresponding to codon 238 and used for S1 nuclease analyses. Hepatoma cell RNA protected only 72 nts of the probe, which maps the divergence between codons 213 and 214 (Fig. 2C), that is, at the boundary between exons 8 and 9. The same band was also visible in uterus RNA, but not in fibroblast, myotube, or skeletal muscle RNAs, in agreement with the data shown in Fig. 2A. This result demonstrates that the TM mRNA isoform produced by the α-TM gene and expressed in hepatoma, uterus, kidney, and kidney fibroblast RNAs does not use exon 8, which codes for aa 189 to 213 in the smooth muscle cDNA clone. It is very likely that this exon is skipped in the generation of the RNA transcript. The identity of the partially protected fragment of ~107 nts observed in low abundance in myotube RNA is presently unknown.

The splicing of exon 6 directly to exon 9 would result in the generation of a TM protein of 206 aa, which is smaller than any TM so far identified. Yet, the previously reported rat α-TM genomic sequence (57) did not appear to contain any alternative exons other than exon 8 between exons 6 and 9. However, a reexamination of the relevant genomic clone (57) revealed that a 277-base-pair *Pst*I fragment had been missed from the sequence between exons 6 and 8. This fragment had presumably been missed owing to its small size and our focus on the known sequences from the cDNAs. We therefore cloned this *Pst*I fragment and an overlapping 1.3-kilobase *Sac*I-*Hin*dIII fragment and sequenced across the *Pst*I sites. The sequence of this fragment is shown in Fig. 6b. With the inclusion of this sequence, it became clear that between exons 6 and 8 there is an alternative open reading frame, flanked by consensus slice donor and acceptor sites. The open reading frame contained the same number of nucleotides and had the same 5' split codon structure as exon 8. Moreover, the predicted amino acid sequence had the expected 3-4-3-4 periodicity of hydrophobic residues required for the coiled-coil packing of TM (58). We tentatively concluded that this putative exon is part of a pair of mutually exclusive exons, along with exon 8. We therefore refer to this open reading frame as exon 7 (Fig. 1). Although its inclusion into mRNA has not been directly demonstrated, this exon is most likely included in those RNAs which lack the sequences corresponding to exon 8.

Mutually exclusive splicing of 5' exons: exon 2 is smooth muscle specific, while exon 3 is used in striated muscle and nonmuscle cells. We have previously shown that exons 2 and 3 of the α-TM gene, coding for aa 39 to 81 of the smooth and striated muscle mRNAs, respectively, are alternatively spliced in a mutually exclusive fashion (57). Only one of the two exons is incorporated in each muscle mRNA, and no mRNA species was identified in which both exons were spliced out in muscle cells. However, in light of the complexity of slicing patterns detected at the 3' end of the α-TM transcripts, it became appropriate both to reexamine the 5' splicing patterns in muscle cells and to determine the 5'

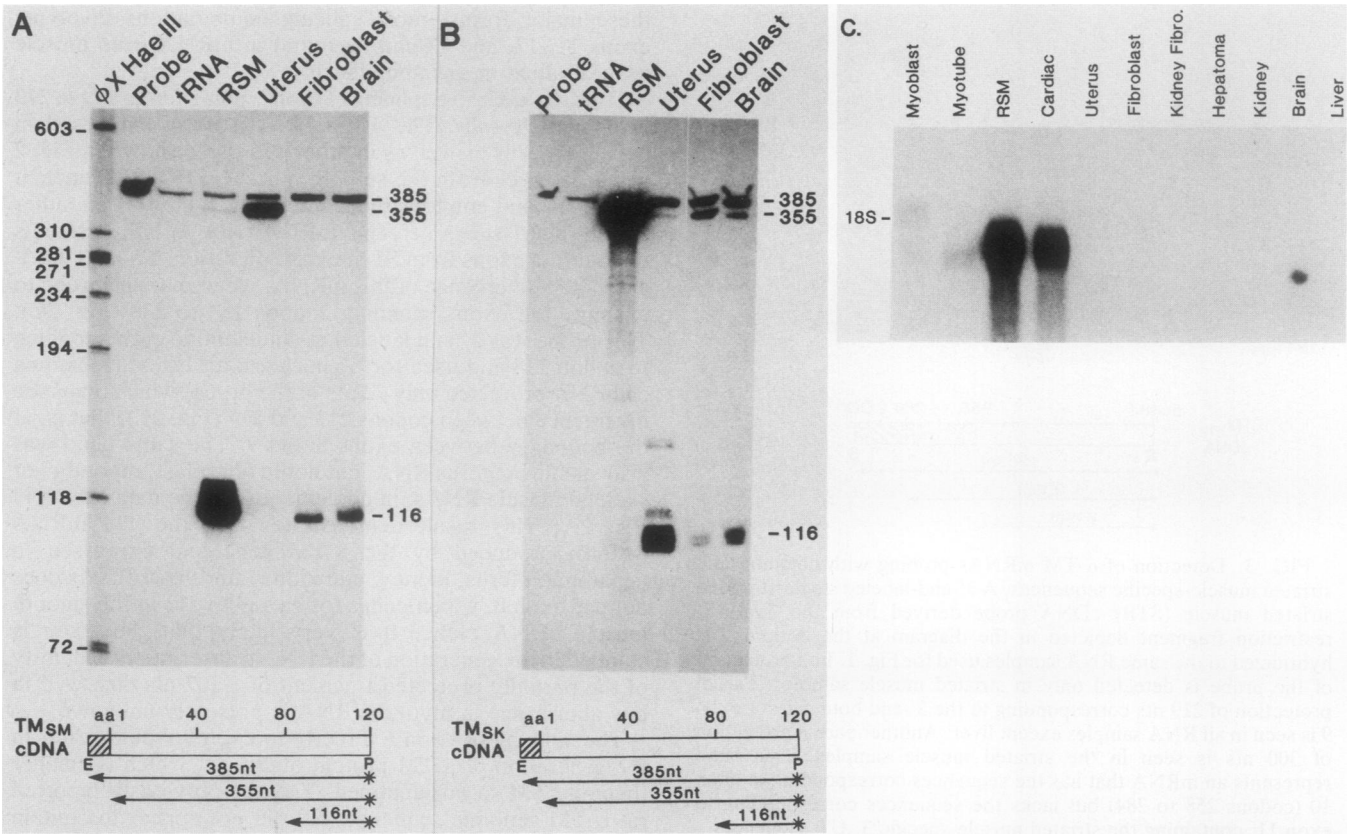


FIG. 4. Structure of the 5′ region of the α-TM mRNAs as detected by S1 nuclease analysis (A, B) and Northern blots (C). (A) The smooth muscle (SM) cDNA *Eco*RI (E)-*Pst*I (P) restriction fragment with a polylinker tail (hatched) depicted at the bottom of the figure was 5′ end labeled at a nucleotide corresponding to codon 120. Hybridization was done at 60°C. Significant full protection of this probe (355 nts) was visible only with the smooth muscle (uterus) sample. A partial protection (116 nts) corresponding to the 5′-end boundary of exon 4 (codon 80) is visible in striated muscle and nonmuscle samples. (B) The striated muscle cDNA restriction fragment with a polylinker tail (hatched) (SK) depicted at the bottom of the figure was 5′ end labeled at a nucleotide corresponding to codon 120. S1 nuclease mapping was done as detailed in the Materials and Methods except that the probe was double stranded and hybridization was at 60°C. Full protection (355 nts) was observed with fibroblast RNA, while uterus RNA gave mainly partial protection (116 nts). (C) Northern blot of various RNAs hybridized to striated muscle-specific sequences corresponding to exon 3. Total RNA from various sources was size separated in a formaldehyde agarose gel. The blotted RNA was hybridized to a synthetic oligonucleotide corresponding to codons 51 to 67 of exon 3, the striated muscle-specific exon (57). Hybridization of the probe was detected only with the striated muscle RNA samples even after very long exposure of the autoradiograms.

structure of mRNAs derived from the α-TM gene in nonmuscle cells. To address these issues, we generated a 5′-end-labeled probe spanning from codon 120 up to codon 1 of the smooth muscle cDNA clone with a 30-base-pair tail of plasmid vector sequence. Substantial full protection of this probe (355 nts) was only observed with smooth muscle RNA (Fig. 4A). Striated muscle and nonmuscle RNAs protected mainly a 116-nt fragment, the region of the probe corresponding to exon 4. A small amount of 355-nt fragment was also observed in these RNAs, indicative of a minor fraction of mRNA containing exon 2. These results clearly indicate that although the splicing of the α-TM transcripts in smooth muscle and nonmuscle cells is identical at the 3′ end, it is different at the 5′ end. In contrast to smooth muscle, nonmuscle cells do not use exon 2 to generate their major α-TM isoform.

To determine the usage of exon 3 in the various mRNAs, we carried out both Northern blots and S1 nuclease analysis. For the Northern blot, a synthetic oligonucleotide probe complementary to codons 51 to 67 of the striated muscle isoform was hybridized to a panel of RNAs. Figure 4C demonstrates that the probe hybridized to RNAs of striated muscle origin including L_6E_9 myoblasts and myotubes. In the myoblasts, both the 1.8- and 1.25-kilobase α-TM mRNAs, corresponding to the smooth muscle/nonmuscle and striated muscle 3′ ends (58), respectively, were detected with this probe. The lack of hybridization in nonmuscle cells could be due either to the lower levels of α-TM transcripts or to the possibility that in these cells both exons 2 and 3 are spliced out. These possibilities were tested by generating a 5′-end-labeled probe from a new full-length striated muscle α-TM cDNA (F. Schachat and B. Nadal-Ginard, unpublished data). The probe extended from the 5′ label at codon 120 to codon 1 and had a 30-base-pair tail of polylinker sequence. Full protection of this probe (355 nts) was observed with striated muscle and nonmuscle RNA (Fig. 4B), indicating that the major α-TM mRNA in these tissues contains exon 3. The substantial partial protection of the probe (116 nts) by uterus RNA corresponds to smooth muscle α-TM mRNA in which exon 2 is spliced in. The partial protection by brain and fibroblast RNA could be due to minor RNA species in which exon 2 is spliced in or in which both exons 2 and 3 are spliced out. We have no direct evidence for the latter species of mRNA, although it is worth

noting that a direct splice from exon 1 to 4 would maintain the correct reading frame of the transcript. The substantial partial protection of both probes by brain RNA is particularly suggestive of the presence of such an RNA species.

These results therefore show that α-TM transcripts preferentially include exon 3 in all cells except smooth muscle, in which exon 2 is specifically incorporated.

The promoter region of α-TM gene has multiple start sites of transcription and a sequence characteristic of housekeeping genes. The fact that α-TM transcripts are found in all tissues and cells analyzed with the exception of liver, but with splicing patterns that are characteristic for each tissue, raises the question as to whether these alternatively spliced transcripts originate from a single or multiple promoters. For this reason, the start sites of transcription of this gene were mapped by S1 nuclease and primer extension analyses.

A genomic DNA fragment 5′ end labeled at a nucleotide corresponding to codon 1 and extending 168 nts upstream was hybridized to the same battery of muscle and nonmuscle RNAs used in the previous experiments and digested with S1 nuclease (Fig. 5A). Four distinct size fragments were protected, with the most prominent being 78 nts upstream from the AUG. This site corresponds to the cap site mapped previously (57). The same protection pattern was observed with RNAs from skeletal, cardiac, and smooth muscle and nonmuscle cells, although the intensity of the bands correlated with the relative abundance of the α-TM gene transcripts. Although this result demonstrated that muscle and nonmuscle cells utilize the same 5′ exon and probably have identical initiation of transcription sites, the existence of several different promoters separated by intervening sequences cannot be ruled out by these data, especially given the fact that most contractile protein genes have their 5′ UT sequence interrupted by one or two introns (69). For this reason, a primer extension analysis was performed with striated and smooth muscle and nonmuscle RNAs. A 5′-end-labeled 21mer synthetic oligonucleotide extending from −20 up to +1 of the translated sequence was used as a primer. The extension products generated were identical for all mRNAs tested (Fig. 5B). Furthermore, four of the five extension products (A, B, D, E) had sizes identical to those of the S1 nuclease-protected bands shown in Fig. 5A. This pattern of primer extension was confirmed by using another synthetic oligomer corresponding to the nucleotides +22 to +42 (data not shown). The nucleotide sequence of the main extension product had been obtained previously (57) and shown to be continuous with the sequences of exon 1, without introns. Therefore, the close correspondence between the results of the S1 nuclease and primer extension analyses indicated that all the transcription start sites are closely spaced and obviated the need for sequencing these low-abundance extension products. There does not appear to be a tissue-specific utilization of any of the start sites, although in some of the experiments band E is proportionally more prominent in the skeletal muscle sample.

To further localize the transcription initiation sites, sequencing reactions (41) were performed on the genomic fragment used as a probe in Fig. 5A and run alongside an S1 nuclease protection reaction (data not shown). In addition, the nucleotide sequence of the region upstream from the start sites was also determined to search for canonical upstream regulatory sequences, e.g., TATA and CCAAT boxes (5). The multiple initiation sites were located 20 to 35 nts apart and within 100 nts of each other (Fig. 6). No typical TATA- or CCAAT-like boxes were found. The only TATA-like sequence was ~100 nts upstream from the 5′-most initiation site, at position −275 from the AUG. It is remarkable that multiple initiation sites together with an absence of canonical TATA and CCAAT boxes are a common feature of housekeeping genes (44, 55) that are constitutively expressed in most cell types.

Thus, the results shown in Fig. 5 and 6 indicate that the α-TM gene is constitutively expressed in most cell types and that striated and smooth muscle and nonmuscle cell α-TM transcripts originate from a single promoter region that has several transcription initiation sites. The cell-specific pattern of expression is determined, however, by regulated alternative splicing and poly(A) addition site selection. The results of the S1 nuclease data presented above demonstrate that the rat α-TM gene generates a minimum of six different mRNAs in a tissue-specific manner. The structures of these mRNAs, as determined from cDNA and genomic sequences (57, 58), as well as from the S1 data presented here, are summarized in Fig. 7. In this figure, the different mRNAs have been labeled according to the tissue or cell type in which they are predominantly expressed. Splices between exons that have been definitively demonstrated are joined by a solid line, while splicing events that have been inferred but not directly confirmed are shown by broken lines. With one exception (see below), all six major isoforms code for a protein which is 284 aa long. The putative minor nonmuscle isoforms which appear to lack the exon coding for aa 39 to 80 would maintain the correct reading frame and produce a protein of 244 aa if no other sequences replace the missing exon.

Phylogenetic comparisons of TM gene structural organization suggest that alternative splicing arose early in the evolution of this gene. Striated and smooth muscle TM binds to seven actin monomers, with the binding sites spaced approximately every 40 aa (67). There is also a 14-fold pattern of apolar and acidic residues that have been interpreted as 14 quasi-equivalent zones that may act as two alternative sets, on and off, of binding sites to seven actin monomers (42, 48). The shorter nonmuscle TMs display a similar pattern, but the molecule spans only six actin monomers. Recently, Phillips (51) has demonstrated that the on sites (termed α) are more regular than the alternative sites (termed β) and postulated that the α sites are the true actin-binding sites in both the on and off state (53, 54).

On the basis of the structural organization of the α-TM gene and the nonrandom location of the exon-intron boundaries with respect to the heptamer repeat, we have suggested that the TM genes originated by repeated duplications of an ancestral 21-aa-long sequence (57). The interruption of the heptamer repeat by introns at specific locations (57) as well as the alternating distribution (42) and higher degree of regularity of the seven α zones compared with the seven β zones (51) suggests the evolutionary sequence depicted in the top portion of Fig. 8. Duplication of the primordial exon encoding 21 aa would have allowed for the divergence between α and β zones; a further duplication with the loss of the intron between the 5′ exons would have established the different structure of the 5′ and 3′ portions of the gene characterized by exons coding for 42 and 21 aa on the average, respectively. This point of view is reinforced by the amino acids and cDNA-derived sequence of a nonmuscle TM that has two internal deletions of 21 aa each (35; Yamawaki-Kataoka and Helfman, in press).

This putative evolutionary origin, however, does not give any indication as to whether the complex pattern of alternative splicing of the TM genes, such as described here, is a recently acquired or an old evolutionary feature. The splic-

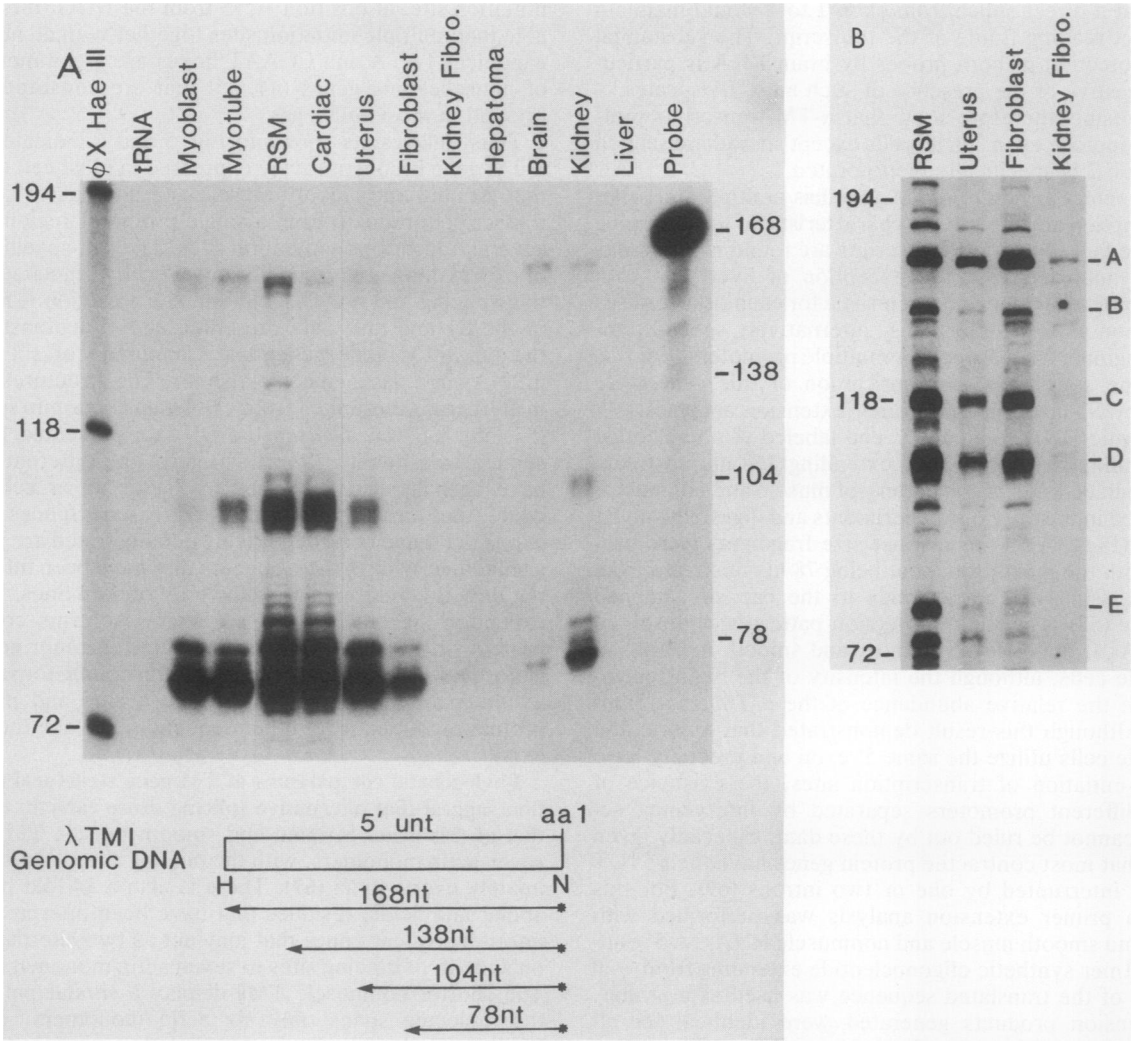


FIG. 5. Mapping of the sites of transcription initiation by S1 nuclease mapping (A) and 5′ primer extension analyses (B). (A) A genomic DNA fragment corresponding to the 5′ UT region of the α-TM gene was 5′ end labeled at the *Nco*I site corresponding to codon 1. The protection pattern is similar in all RNA samples except liver and shows four main bands of 78, 104, 138, and 168 nts, respectively. The intensity of the bands closely correlates with the abundance of α-TM mRNAs in the sample. (B) The primer and the primer-extended products were analyzed as described in the text. Five major extension products are visible in both muscle and nonmuscle samples. The lengths of four of these products, labeled A, B, D, and E, correspond to the S1 nuclease-protected fragments shown in panel A.

ing patterns of present-day TM genes could be the result of recent exon duplications or could have arisen early in TM gene evolution with the specific patterns of exon organization and splicing characteristic of each gene having evolved subsequently. In the first case, alternatively spliced exons, being the result of recent duplications, should exhibit a higher degree of intraspecies rather than interspecies conservation. In the second case, if all or most of the alternatively spliced exons were present in the ancestral gene, the opposite result would be expected. This hypothesis is based on the assumption that the different exons are under similar evolutionary pressures. The high degree of structural and sequence conservation of the TM genes from *D. melanogaster* to humans (57) strongly suggests that this is the case.

To test which of these two hypotheses is correct, we compared the organization, nucleotide and amino acid sequence, and alternative splicing behavior of six TM genes from different phyla spanning more than 600 million years of evolutionary time and ranging from *D. melanogaster* to human (Fig. 8). From this comparison, several interesting facts became evident. First, most TM exons are alternatively spliced in a mutually exclusive manner in one or other species, and each of the TM genes produces more than one TM mRNA isoform. Only exons 1, 3–5, and 8 are constitutively spliced in all species so far analyzed. Second, from insects to mammals, the exon-intron boundaries of the existing exons are conserved to the precise nucleotide. Third, in all species, the 5′ half of the gene has exon lengths clustered around 42 codons, while in the 3′ portion they are clustered around 21 codons. Fourth, there is a very high degree of sequence conservation among constitutive exons, especially at the amino acid level. Fifth, although in all pairs of mutually exclusive and alternatively spliced exons the two members of a pair have diverged considerably from each other, there remains enough sequence conservation to support the conclusion that they represent exon duplications.

Sixth, despite this conservation, it is striking that there is a significantly higher degree of intraspecies (between α- and β-TM genes in the same species) and interspecies conservation between the equivalent alternatively spliced exons of different genes than between two members of a pair of alternatively spliced exons in the same gene. The sequence divergence between the members of any pair of alternatively spliced exons in any TM gene analyzed so far is as high or higher than the divergence between any constitutive exon from a mammalian TM gene and its counterpart in the *Drosophila* TM genes. This is exemplified in Table 1 in which the percent identity between the exons coding for aa 39 to 80 in the rat (exons 2 and 3 corresponding to smooth muscle and striated muscle-nonmuscle, respectively) and between each one of them and the constitutive exon coding for aa 81 to 125 has been determined and compared with the corresponding exons in human TM, rat β-TM, and *Drosophila* TM genes. The rat α-TM exons 2 and 3 have a sequence identity of 33%, while exon 3 has a sequence identity ranging from 90 to 36% when compared with the human and *Drosophila* counterparts. A similar result is obtained when the other pairs of mutually exclusive exons are compared in the same manner (data not shown). The fact that there is a higher degree of conservation between the corresponding rat and human exons and their *Drosophila* homolog than between the members of a given pair in the same gene strongly suggests that all pairs of mutually exclusive exons found so far in TM genes (Fig. 8) were already present in the TM gene before the radiation of the insects, more than 600 million years ago. This conclusion is reinforced by the fact that the interspecies conservation among homologous alternatively spliced exons is still high. This conservation indicates that these exons are not subject to rapid evolutionary sequence drift, but rather are under remarkable selective pressure to maintain their sequence. In light of these facts, the alternative explanation, namely, that the mutually exclusive pair are the result of a recent exon duplication followed by rapid sequence divergence, is difficult to support.

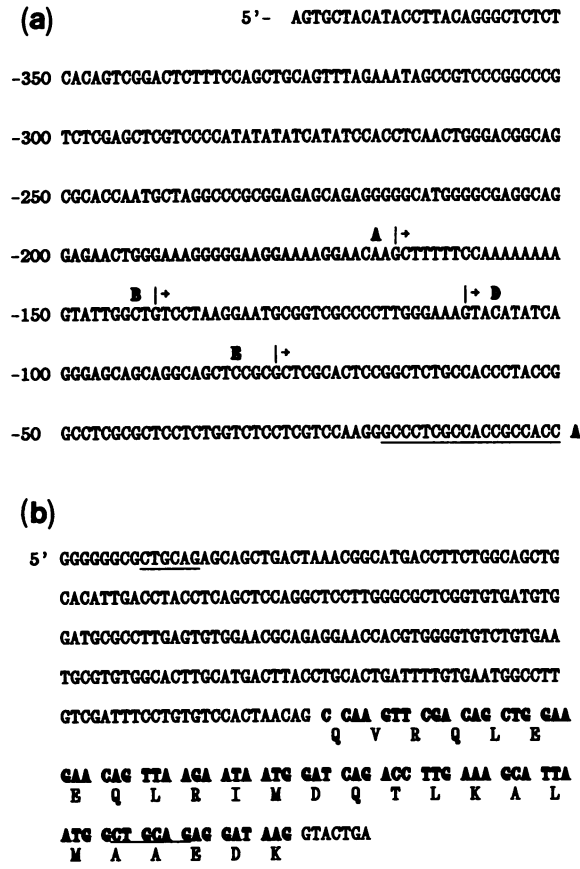


FIG. 6. (a) Nucleotide sequence of the 5′ UT and flanking regions of the α-TM gene. The starts of transcription sites identified in Fig. 5 are marked ⊢→ and are labeled A, B, D, E as in Fig. 5. The sequence of the primer used for the 5′ extension experiment is underlined. (b) Nucleotide sequence of the genomic region encompassing exon 7. A 1.3-kilobase *Hin*dIII-*Sac*I fragment of the rat α-TM gene from within exon 6 to 3′ of exon 9 (57) was subcloned into pGEM3 and sequenced across the *Pst*I sites (underlined). The open reading frame identified as exon 7 is shown in bold type. The translated amino acid sequence is shown in the single-letter code beneath the codons. Like exon 8, this exon begins with the third base of a split codon, codes for 25 aa with the 3-4-3-4 hydrophobic repeat characteristic of coiled-coil proteins such as TM, and ends between codons at the 5′ end.

The above conclusion, if correct, implies that the ancestral TM gene, present at the time of the radiation of the insects, had a more complex structure than current TM alleles, but was very similar to the rat α-TM gene, with the major differences being at the 3′ end. At a minimum, the structural organization of this ancestral TM gene had to be as complex as that shown in Fig. 8. Given this organization, the pattern of alternative splicing of this ancestral gene must have been at least as complex as that of the present-day rat α-TM gene. The possibility that the exons were constitutively spliced to generate a longer TM molecule seems very unlikely given the number, nature, and remarkable evolutionary conservation of split codons at the ends of alternatively spliced exons. Constitutive splicing of current alternatively spliced exons would generate nonsense codons and frameshifts resulting in either truncated TMs or proteins of unrelated primary and secondary structure.

These data strongly suggest that, at least for the TM gene family, alternative mutually exclusive splicing is not an evolutionary recent acquisition but, on the contrary, was already present early in the evolution of multicellular organisms. The data are compelling that the ancestral TM gene had a structure and pattern of alternative splicing more complex than that of any of the present-day genes. Its pattern seems to have been simplified in the course of speciation. Selective loss of specific exons in different alleles and different species would have generated the species-specific TM genes of current organisms.

DISCUSSION

The α-TM gene produces a minimum of six mRNAs, each with the potential to code for a different protein. The results presented here demonstrate that the rat α-TM gene has the capability to generate a minimum of six, and possibly eight, different mRNAs in a tissue-specific manner (Fig. 7). All these isoforms are produced by three groups of exons that are spliced in a mutually exclusive manner. These are located near the 5′ end, in the middle, and at the 3′ end of the gene. The usage pattern of these three groups of exons is noteworthy. In each group, one exon is spliced into mature mRNA in all but a restricted subset of cell types in which the alternative exon is used. At the 5′ end, exon 3 is used predominantly in all cells and tissues tested with the exception of smooth muscle, in which exon 2 is specifically

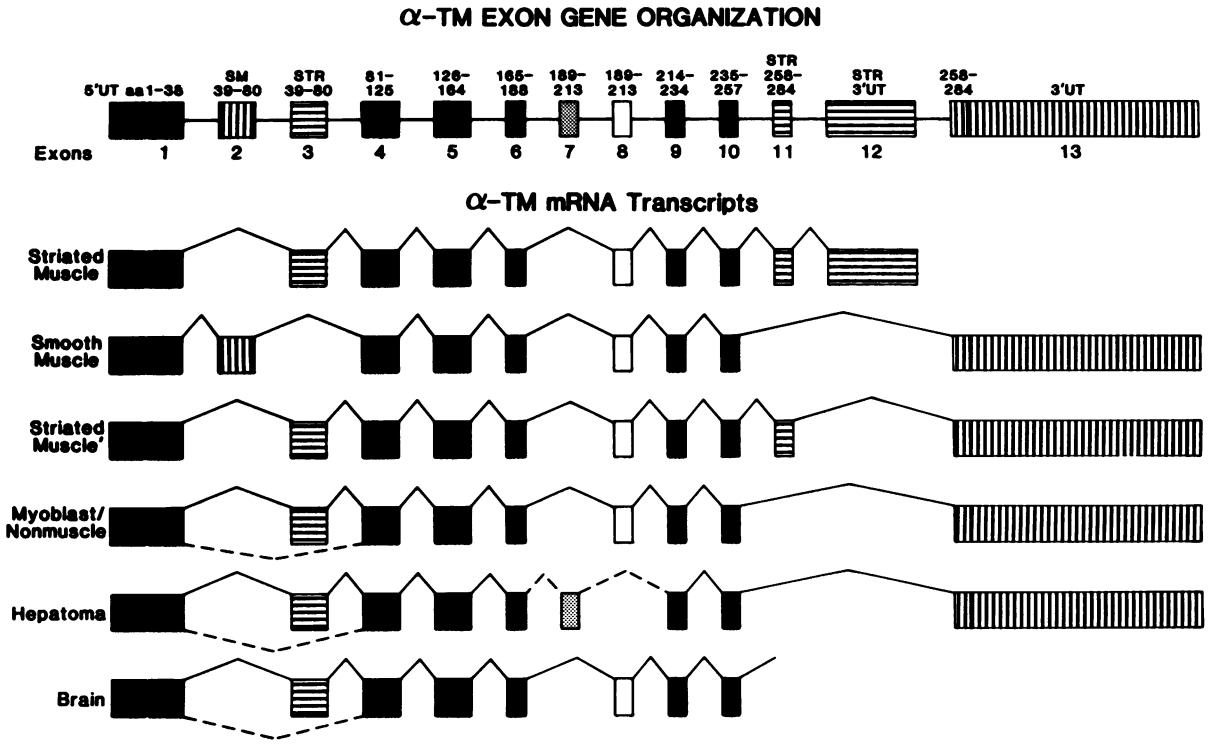


FIG. 7. Diagram of the α-TM gene organization and its associated mRNA transcripts. Boxes represent exons as in Fig. 1, i.e., black, constitutive; horizontal stripes, striated muscle and striated muscle-nonmuscle; vertical stripes, smooth muscle-nonmuscle; white and stippled, alternative exons for aa 189 to 213. The numbers above the exons indicate the codons in each exon. The exons have been numbered below in consecutive sequence from the 5′ end. Solid lines between boxes indicate confirmed splicing patterns. Broken lines indicate probable splice patterns but where the actual joining of the two exons indicated has not been directly proven and only indirect evidence is available. See text for full explanation.

incorporated. In the middle of the gene, exon 8 is used in most cells and tissues with the exception of hepatoma cells, uterus, and kidney, in which exon 7 is apparently incorporated instead to a variable extent. At the 3′ end, exon 13 is used in all cells, while exons 11 and 12 are incorporated only in differentiated striated muscle. The picture that emerges from these data is that one exon from each pair of mutually exclusive exons constitutes the ground-state splicing pattern and is incorporated constitutively unless specifically replaced by the alternative exon(s) in response to the environment of specific cell types. The main nonmuscle isoform therefore appears to be the product of this ground-state splicing apparatus incorporating the default exons 3, 8, and 13. The striated muscle, smooth muscle, and hepatoma isoforms are produced by the alternative utilization of their specific exons (exons 11 and 12 for striated muscle, exon 2 for smooth muscle, and exon 7 for hepatoma) in one part of the transcript while maintaining the nonmuscle splicing pattern elsewhere in the transcript. All these mRNAs code for proteins of 284 aa. The minor striated muscle isoform in which exon 11 is spliced to exon 13 codes for a 294-aa protein owing to a readthrough of part of exon 13 before a termination codon is encountered. The extra 10 aa in this isoform do not have the hydrophobic repeat structure required for coiled-coil packing, but since the ends of TM dimers are not thought to be in a tightly packed coil, this sequence should not be incompatible with a functional TM protein.

The existence of a number of different nonmuscle TM isoforms has been documented by criteria of electrophoretic mobility, partial peptide mapping, and sequence analyses. These isoforms have been grouped according to size or electrophoretic mobility into a high-M_r class (M_r = 36,000 to 46,000), with a size of 284 aa for the isoforms for which data are available (28, 31, 36, 38, 57, 58, 60, 74), and a low-M_r class (M_r = 30,000 to 36,000), with a size of about 247 aa (35, 60; Yamawaki-Kataoka and Helfman, in press). Both human and rat fibroblasts, for instance, contain five TM isoforms, three of the high-M_r class and two of the low-M_r class (40). It is likely that one or possibly two of the high-M_r forms are coded for by the nonmuscle and hepatoma transcripts described here. The two fibroblast low-M_r forms are coded for by separate genes (54; Yamawaki-Kataoka and Helfman, in press), each of which, however, can also give rise to a muscle-specific transcript. For the human TM30nm gene, a nonmuscle 248-aa protein and a 285-aa muscle-specific protein are produced from the same gene by alternative RNA processing (37). The rat α-TM gene may give rise to up to three nonmuscle TM isoforms lacking either of the identified exons for aa 39 to 80. It is possible that a thus far unidentified exon replaces exons 2 and 3 in these isoforms. However, direct splicing of exon 1 to 4 would not generate a frame shift and could result in the production of proteins of 244 aa. If such proteins are made, they represent distinct isoforms from those previously characterized. In the low-M_r TMs for which sequence data are available, the missing sequences correspond to aa 23 to 43 and 60 to 80 (35; Yamawaki-Kataoka and Helfman, in press) or to aa 1 to 37 (54) when compared with the sequence of the 284-aa isoforms.

The production of a number of TM isoforms from a single gene is not entirely unprecedented. Both the rat β-TM and the human slow α-TM genes have been shown to generate an

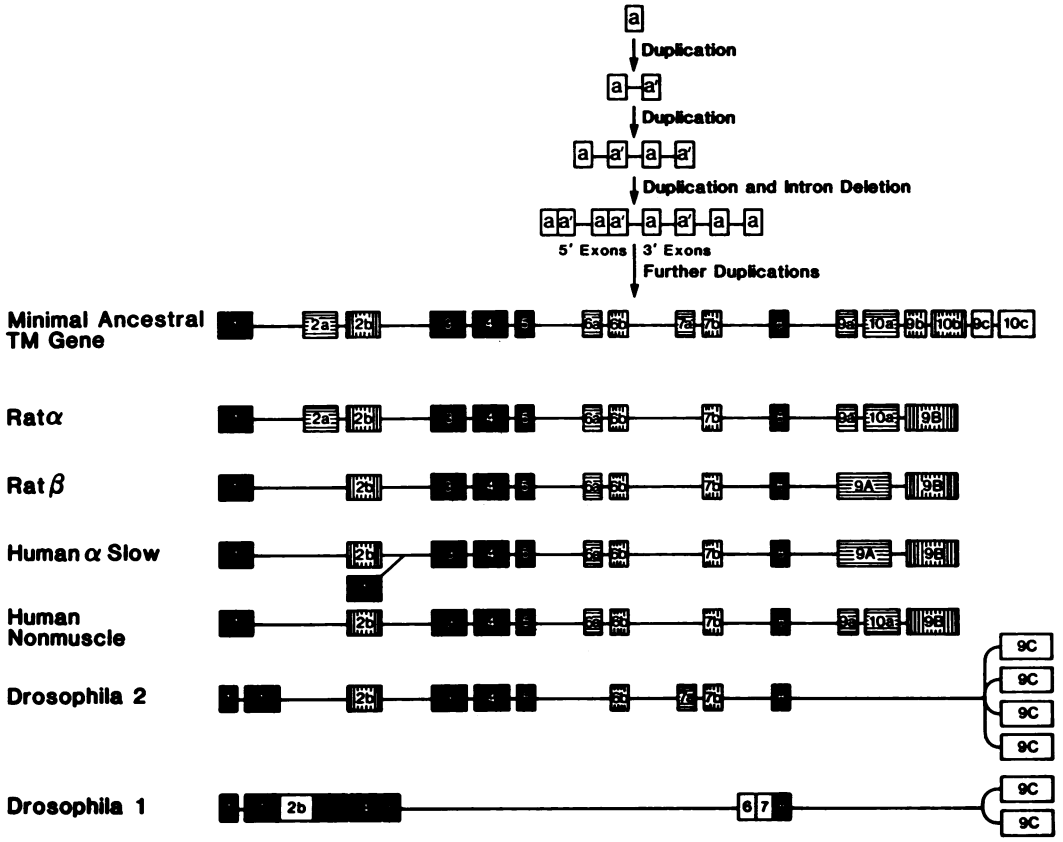


FIG. 8. Phylogenetic comparisons and evolutionary origin of the α-TM gene. The blocks labeled a and a' in the upper portion of the figure represent the ancestral exons coding for the α and β sites, respectively. The structural information for the genes represented was obtained from the following publications: rat α (57); rat β (27); human slow α (54); human nonmuscle (36); *Drosophila* 2 (30); *Drosophila* 1 (2). Alternative exons are labeled with the same number followed by a, b, or c. Black exons are constitutive in all the phyla surveyed. Homologous exons that are alternatively spliced in one or other species are marked identically: striped vertically or horizontally. The size of the boxes indicates whether the corresponding exon codes for ~21 or 42 aa. The exons marked 9C contain an extended open reading frame at the 3' end not related to the α-TM sequence (27, 57). Note that the exon numbers do not correspond to the numbering system used for the rat α-TM exons in Fig. 1 and 7.

isoform common to smooth muscle and nonmuscle cells by a combined process of alternative splicing and polyadenylation site selection (27, 36). In these cases, however, only two splicing patterns were characterized, compared with the six described here for the rat α-TM gene. It is interesting to note, nevertheless, that in all three instances the striated muscle splicing pattern involved the use of a specific 3'-end exon not used in the smooth muscle-nonmuscle isoforms.

The functional significance of the alternatively spliced TM domains and the physiological role of the TM isoforms produced by this and other TM genes remain to be determined. The isoforms reported here contain the sequences involved in head-to-tail polymerization and all the sequences involved in the interaction with troponin-T, despite the fact that troponin does not occur in smooth muscle and nonmuscle cells (39, 52). Yet, the alternatively spliced regions are involved in head-to-tail polymerization and troponin-T binding. It remains to be determined whether any of the different primary sequences documented here are responsible for the different assembly properties, interactions with Ca^{2+}-regulatory proteins, or actin-binding properties of smooth muscle and nonmuscle TMs (9, 17, 39, 60).

The choice of striated muscle-, smooth muscle-, and nonmuscle-specific isoforms produced by the α-TM gene is regulated through alternative splicing and poly(A) addition site selection. Most components of the contractile apparatus in striated and smooth muscle and nonmuscle cells are encoded by different sets of genes that are specific for each contractile system. The degree of divergence between sarcomeric and nonsarcomeric isoforms is very high. Most antibodies and cDNA clones developed against a given sarcomeric contractile protein do not recognize the nonsarcomeric isoforms and vice versa (68). This feature indicates that the isoforms expressed in each of the three cell types are encoded by different genes that, at least for the striated muscle isoforms, are grouped in multigene families (68). Until recently, the only known exception to this type of arrangement was the actin genes that exhibit a high degree of nucleotide and amino acid sequence conservation among muscle and nonmuscle types (10, 26). Yet, despite this high degree of conservation among the six known mammalian actins, each is encoded by a different gene that exhibits a characteristic developmental and tissue-specific pattern of expression (56). Therefore, the main selection mechanism that determines whether a striated muscle, smooth muscle, or nonmuscle contractile protein isoform is going to be expressed in a given cell type is dependent on which particular promoter is activated (56). Once the gene to be expressed has been selected, the range of isoform diversity available to a particular cell type can be further increased by the availability of

TABLE 1. Homology comparisons among two mutually exclusive and a constitutive exon from different species[a]

Amino acids	% Homology with rat α-TM		
	Exon 2 (39–80 SM)	Exon 3 (39–80 STR)	Exon 4 (81–125)
RAT α			
Exon 2 (39–80 SM)	**100**	33	29
Exon 3 (39–80 STR)	33	**100**	38
Exon 4 (81–125)	29	38	**100**
RAT β			
39–80	36	**79**	21
81–125	29	38	**98**
Human α slow			
39–80	38	**90**	36
81–125	26	40	**98**
Human NM			
39–80	33	**75**	31
81–125	29	36	**98**
Drosophila 2			
39–80	21	**40**	45
81–125	19	21	**60**
Drosophila 1			
39–80	24	**36**	33
81–125	19	31	**47**

[a] The relatedness among the sequences is shown as percentage of amino acid homology. Similar results are obtained when the nucleotide sequences are compared. All the sequences were compared with exon 2 of the rat α-TM gene that encodes aa 39 to 80 of the smooth muscle α-TM isoform (39–80 SM), to exon 3 that encodes the corresponding region of the striated muscle and nonmuscle α-TM isoform (39–80 STR), and to exon 4 that encodes aa 81 to 125 and is present in all the isoforms. The homology between analogous sequences is presented in boldface characters. The sequence data used for the comparisons were obtained from the following sources: rat α-TM (57); rat β-TM (Cheley and Helfman, in press); human α slow TM (54); human nonmuscle (NM) (36); *Drosophila* 2 (30); and *Drosophila* 1 (2).

different alternative splicing patterns to the transcripts from some of these genes (1, 6). However, with the exception of TM, the different contractile protein isoforms so far described that are produced by a single gene through alternative splicing are restricted to the striated muscles. None of them is expressed in smooth muscle and nonmuscle cells.

The existence of well-defined size and/or sequence differences among striated and smooth muscle and nonmuscle TMs strongly suggested that, like the other contractile proteins, they would also be encoded by separate multigene families. Recent data obtained with α- and β-TM cDNA and genomic clones (27, 36, 54, 57; Cheley and Helfman, in press), as well as the results presented in this report, clearly demonstrate that this is not the case. It is now established that the genes encoding the sarcomeric α- and β-TMs also encode isoforms expressed in smooth muscle (27, 57) and nonmuscle cells. It is surprising, however, that to date not a single TM gene has been identified that exclusively codes for striated muscle, smooth muscle, or nonmuscle isoforms. From the data presently available, it seems likely that all the TM isoforms are encoded by genes that are also expressed in striated muscles. The production of several tissue- and developmental stage-specific isoforms with unique primary sequences by a gene that is transcribed from a single "housekeeping" promoter region is a feature that so far has not been described for any other gene. It is likely, however, that not all the TM isoforms present in vertebrate cells are generated by the α and β genes so far described. The lack of expression in liver of a bona fide mRNA from the α-TM gene, peptide maps, mRNA translation, and cDNA sequence data (40; Yamawaki-Kataoka and Helfman, in press) all suggest the existence of at least another gene with a high degree of nucleotide sequence divergence.

The structure and sequence organization of the α-TM gene promoter region support the conclusion that this is in fact a housekeeping promoter. Distinctive features of this type of gene, such as the genes for hydroxymethyl-glutaryl coenzyme A reductase (55) or mammalian dihydrofolate reductase (44), include the existence of multiple, closely spaced transcription start sites within a single 5′ exon and the absence of TATA and CCAAT boxes in their characteristic locations. Furthermore, with very few exceptions (69), the 5′ UT regions of the sarcomeric contractile protein genes are interrupted by one or two introns. This feature is also absent from the α- and β-TM genes. The conclusion that the α-TM gene is not a muscle-specific gene but a housekeeping one generating muscle-specific isoforms is further reinforced by the finding that the translational ability of the α-TM mRNA also sets it apart from the other muscle-specific genes. Conditions that specifically inhibited the translation of muscle-specific but not of constitutive mRNAs failed to inhibit the translation of the TM transcripts (21), thus indicating that α-TM is a housekeeping, rather than a muscle-specific, transcript.

The generation of a minimum of six different isoforms described here, each one specifically expressed in striated muscle, and smooth muscle, or nonmuscle cells, is, therefore, not regulated at the transcriptional level by activating a particular member of a multigene family. On the contrary, all six isoforms are produced by an α-TM gene which appears to be transcribed in all cell types with the exception of liver. Interestingly, we have recently documented that the smooth muscle myosin light-chain 2 gene is also transcribed in all cell types but not in liver cells (70). For the α-TM, isoform selection is exclusively accounted for by cell-specific alternative splice site selection between pairs of alternative exons in combination with the alternative use of two alternative poly(A) addition sites.

The splicing pattern of the α-TM gene transcripts implies the existence of cell- and developmental stage-specific splicing factors. Alternative splicing is now known to operate in more than 60 different genes (1, 6, 34, 47). This mode of generating protein diversity is particularly prevalent in contractile protein genes in which alternative splicing has been demonstrated for myosin alkali light chain (49), myosin heavy chain (4), troponin-T (8, 43), and tropomyosins (25, 27, 36, 57). Despite this apparent ubiquity, however, little is known about the detailed mechanisms that permit certain exons but not others to be variably incorporated into the mature mRNA.

From the limited number of experiments designed to address the mechanisms involved in alternative splicing (see reference 1, 6, 7, and 46), it is clear that in some cases the existence of primary transcripts with different structures plays a fundamental role in determining the splicing pattern. In these cases, different transcripts originate from the existence of more than one transcription unit in a gene, either from multiple promoters (6, 34) or through the utilization of different poly(A) addition sites (6, 34). For the myosin light-chain 1/3 gene, it is clear that promoter choice determines which of two alternative exons will be utilized (1, 46, 49, 69), while in the immunoglobulin gene, poly(A) addition site utilization may dictate whether the secreted or membrane-bound form will be produced (50). Yet, when several different mRNAs are generated from an identical transcript in a regulated manner, it is clear that *cis*-acting elements are not sufficient to determine the splicing pattern and that the

existence of specific *trans*-acting elements must be invoked (8). The existence of these specific factors has recently been documented in brain and muscle cells for the calcitonin (33) and troponin-T (7) genes, respectively.

From the structures of the six α-TM mRNAs documented here and shown in Fig. 7, it is evident that in this gene poly(A) addition site selection is tightly regulated. Fully differentiated muscle cells utilize the proximal site, while all other cells use the distal one. However, at a minimum, four of these mRNAs are produced from identical primary transcripts and have characteristic cell-specific patterns of expression. This can only be accomplished if each cell type provides a different splicing environment that specifically determines the relative affinities of different splicing sites for each other. It is unavoidable that in certain cells a specific donor and acceptor site exhibit a strong functional affinity for each other and are efficiently spliced together, while in other cell types or developmental stages the same pair is completely overlooked by the splicing machinery. In this respect, it appears that exons 1, 3 to 6, 8 to 10, and 13 are spliced together in the absence of any specific factors and represent the prototypical nondifferentiated isoform produced by this gene. The differential incorporation of other exons requires a specific splicing environment. Only smooth muscle cells incorporate exon 2 significantly into mature mRNA, exons 10 and 11 are present only in striated muscle cells, and exon 7 is expressed in a limited set of cells and tissues. The change from the production of the nonmuscle to the striated muscle isoform of α-TM can be readily observed during myogenic differentiation. This suggests the existence of developmental factors which quantitatively and qualitatively regulate TM isoform production via alternative splicing. Whether the differences among different cell types and developmental states are due to qualitative or quantitative differences in the components of the splicing pathway that affect the efficiency of specific splice sites is not known at this time. The data obtained with the troponin-T gene suggest that specific splicing factors are induced during myogenesis (7). Whether the same factors involved in the correct splicing of troponin-T are also involved in the splicing of α-TM transcripts remained to be determined.

Whatever the precise mechanism(s) involved in alternative splicing, the phenotype of the α-TM transcripts reinforces a notion that is becoming more apparent as the number of genes known to be alternatively spliced increases (see references 1 and 6). That is, the splicing sites in a gene transcript (*cis*-elements) even when conforming to the canonical consensus sequences are not necessarily equivalent in a functional sense. In addition, their behavior is dramatically influenced by the cell environment (*trans*-elements) in which the transcript is expressed. The splicing apparatus of different cells is not equivalent either, since as shown here for the α-TM transcript and previously for the troponin-T gene (7), different cells process an identical transcript in significantly different ways. This behavior, however, must be reconciled with the fact that transcripts of most genes involved in general metabolic pathways must be correctly and uniformly spliced in every cell type. This could be accomplished by the existence of a wholly nonspecific splicing apparatus common to all cells, used to splice the universally constitutive transcripts, over which a more specialized one involved in cell-specific splicing is superimposed. Alternatively, constitutively spliced housekeeping genes might have evolved intrinsically strong splice sites that are consistently recognized by the various splicing components of different cell types. These strong splice sites, however, need not all be equivalent. There might be a hierarchy of affinities among pairs of splice sites in these genes as well. As shown for the dihydrofolate reductase gene (44), mutation of one of these sites can disrupt this hierarchy and alternative splicing can emerge in an otherwise constitutively spliced gene.

The generation of different cell phenotypes is already known to be complex process that includes transcriptional (46) and, at least in some cases, translational (21, 70) regulation of the genes involved in producing the characteristic phenotype. This complexity is compounded by the alternative processing of the same transcript into different mRNAs in distinct cell types, as shown for the α-TM and other genes (1, 6). As clearly exemplified here, transcriptional induction of a gene producing a cell-specific protein is not sufficient and, as for the striated muscle and smooth muscle α-TM, might not even be involved in the production of the differentiated phenotype. The production of the appropriate splicing factors able to splice the transcript in a tissue-specific manner is an essential component of the cell differentiation process.

Alternative splicing was prevalent in the ancestral α-TM gene and appears to be a most convenient mechanism to favor gene evolution. RNA splicing is a very ancient process since it is present in archaebacteria (29) and T-even bacteriophages (14, 15) in addition to eucaryotic cells. It evolved, perhaps, from autocatalyzed RNA ligation, as suggested by the evidence of self-splicing introns (12, 13). However, the role of introns in gene evolution, as well as the selective advantage that maintained them in eucaryotes, remain an unsolved puzzle. Alternative splicing, with its power to increase the generation of protein diversity, is made possible only by the existence of introns and provides a rationale for their possible selective advantage. The outstanding question, however, is whether alternative splicing represents a predecessor or a refinement of constitutive splicing. As shown previously, at least for the troponin-T gene (1, 7) and the dihydrofolate reductase gene (44), the distinction between constitutive and alternative exons is a relative one that is not exclusively due to the intrinsic properties of the exon and its flanking splice sites. The behavior of an exon is greatly influenced by the characteristics of the neighboring splice sites (22, 53), as well as by the availability of particular *trans*-acting factors in the cell in which the gene is expressed (7, 33), as shown here for the α-TM transcripts.

From the data presently available (1, 7, 57), it is clear that most alternative splicing uncovered so far occurs among exons which originated through duplication events. The question is whether these duplications are recent or whether they were present in the ancestral gene. For the TM gene at least, the evidence shown here strongly suggests that the duplication occurred early in evolution. Only five exons in this gene are not alternatively spliced in some species. The degree of divergence between pairs of alternatively spliced exons within any given TM gene is greater than that among the same isoform-specific exons in isogenes across the evolutionary ladder, ranging from *D. melanogaster* to humans. This feature suggests that all the alternatively spliced exons so far detected were already present at the time of the radiation of insects more than 600 million years ago (57). As pointed out in the Results, this conclusion implies that the primitive TM gene was more complex than its contemporary counterparts and had the capacity to code for a larger molecule, a larger number of isoforms, or both. The fact that all the TM genes in existence are differentially spliced strongly suggests that the TM gene ancestor was also alter-

natively spliced and that some of the exons have been selectively lost during evolution.

From the phylogenetic data available for the TM genes, it is clear that analysis of the patterns of alternative splicing offers a clue to the possible evolution of these genes. It is likely that alternative splicing played an important role in the evolution of this gene family by utilizing exon duplication and allowing deletion of whole exons. The data suggest that new exons generated through intragenic duplication were first alternatively spliced. This behavior would have had the selective advantage of allowing the generation of a new protein while the old ones continued to be produced, even in the organisms that have become homozygous for the gene rearrangement. The exons generated in this manner might have remained alternative if they provided a selective advantage. On the other hand, if one of the combinations proved particularly advantageous, they might have evolved into constitutive exons through mutations in the *cis*-elements that resulted in the generation of stronger splice sites.

From these arguments it is possible to postulate that alternative splicing either preceded or was concomitant with the appearance of constitutive splicing and played a role in facilitating gene evolution. This role might continue at the present time. Genes that remain in flux and need to function in different and variable environments would have retained alternative splicing, whereas those whose product has been optimized for a given cell environment or function would have made all exons constitutive or eliminated their introns altogether. For genes that are evolving and/or have to produce cell- or stage-specific isoforms or both, which have to assemble in very precise multiprotein organelles, such as the sarcomere or the cytoskeleton, alternative splicing could outweigh the advantages of gene duplication and subsequent specialization. This would allow the precise conservation of certain domains involved in intermolecular interaction, while permitting the evolution of others that might modulate function. This hypothesis is consistent with the low level of splicing in unicellular organisms and its increasing complexity in higher metazoans.

ACKNOWLEDGMENTS

The assistance of Elena Porro in obtaining the sequence of the missing exon is gratefully appreciated.

This work was partially supported by grants from the National Institutes of Health and the Muscular Dystrophy Association of America.

LITERATURE CITED

1. **Andreadis, A., M. Gallego, and B. Nadal-Ginard.** 1987. Generation of protein isoform diversity by alternative splicing: mechanistic and biological implications. Annu. Rev. Cell Biol. **4**:207–242.
2. **Basi, G. S., M. Boardman, and R. Storti.** 1984. Alternative splicing of a *Drosophila* tropomyosin gene generates muscle tropomyosin isoforms with different carboxy-terminal ends. Mol. Cell. Biol. **4**:2828–2836.
3. **Berk, A. J., and P. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell **12**:721–732.
4. **Bernstein, A., C. Hansen, K. Becker, D. Wassenberg, E. Roche, J. Donady, and C. Emerson.** 1986. Alternative RNA splicing generates transcripts encoding a thorax-specific isoform of *Drosophila melanogaster* myosin heavy chain. Mol. Cell. Biol. **6**:2511–2519.
5. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
6. **Breitbart, R., A. Andreadis, and B. Nadal-Ginard.** 1987. Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from a single gene. Annu. Rev. Biochem. **56**:467–495.
7. **Breitbart, R., and B. Nadal-Ginard.** 1987. Developmentally induced, muscle specific trans factors control the differential splicing of alternative and constitutive troponin-T exons. Cell **49**:793–803.
8. **Breitbart, R. E., H. Hguyen, R. Medford, A. 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. **Broschat, K. D., and D. R. Burgess.** 1986. Low M_r tropomyosin isoforms from chicken brain and intestinal epithelium have distinct actin-binding properties. J. Biol. Chem. **261**:13350–13359.
10. **Buckingham, M., S. Alonso, G. Bugaisky, P. Barton, A. Cohen, P. Daubas, A. Minty, B. Robert, and A. Weydert.** 1985. The actin and myosin multigene families, p. 333–344. *In* R. Strohman and S. Wolf (ed.), Gene expression in muscle. Plenum Publishing Corp., New York.
11. **Burgess, D. R., K. O. Broschat, and J. M. Hayden.** 1987. Tropomyosin distinguishes between the two actin-binding sites of villin and affects actin-binding properties of other brush border proteins. J. Cell Biol. **104**:29–40.
12. **Cech, T.** 1986. The generality of self-splicing RNA: relationship to nuclear mRNA splicing. Cell **44**:207–210.
13. **Cech, T., and B. Bass.** 1986. Biological catalysis by RNA. Annu. Rev. Biochem. **55**:599–629.
14. **Chu, F., G. Maley, F. Maley, and M. Belfort.** 1984. Intervening sequence in the thymidate synthase gene of bacteriophage T4. Proc. Natl. Acad. Sci. USA **81**:3049–3053.
15. **Chu, F., G. Maley, D. West, M. Belfort, and F. Maley.** 1986. Characterization of the intron in the phage T4 thymidylate synthase gene and evidence for its self-excision from the primary transcript. Cell **45**:157–166.
16. **Cote, G. P.** 1983. Structural and functional properties of the non-muscle tropomyosins. Mol. Cell. Biochem. **57**:127–146.
17. **Cote, G., and L. Smillie.** 1981. Preparation and some properties of equine platelet tropomyosin. J. Biol. Chem. **265**:11004–11010.
18. **Cummins, P., and S. Perry.** 1973. The subunits and biological activity of polymorphic forms of tropomyosin. Biochem. J. **133**:765–777.
19. **Cummins, P., and S. Perry.** 1974. Chemical and immunochemical characteristics of tropomyosins from striated and smooth muscle. Biochem. J. **141**:43–49.
20. **Dhoot, G., and S. Perry.** 1979. Distribution of polymorphic forms of troponin components and tropomyosin in skeletal muscle. Nature (London) **278**:714–720.
21. **Endo, T., and B. Nidal-Ginard.** 1987. Three types of muscle-specific gene expression in fusion-blocked rat skeletal muscle cells: translational control in EGTA-treated cells. Cell **49**:515–526.
22. **Eperon, L., J. Estibeiro, and L. Eperon.** 1986. The role of nucleotide sequences in splice site selection in eukaryotic pre-messenger RNA. Nature (London) **324**:280–282.
23. **Favaloro, J., R. Treisman, and R. Kamen.** 1980. Transcriptional maps of polyoma virus-specific RNA: analysis by two-dimensional nuclease S1 gel mapping. Methods Enzymol. **65**:718–749.
24. **Fiszman, M., E. Kardami, and M. Leomonnier.** 1986. A single gene codes for gizzard and skeletal muscle α tropomyosin, p. 457–471. *In* C. Emerson, D. A Fishman, B. Nadal-Ginard, and M. A. Q. Siddiqui (ed.), Molecular biology of muscle development. Alan R. Liss, Inc., New York.

24a. **Flach, J., G. Lindquester, S. Berish, K. Hickman, and R. Devlin.** 1986. Analysis of tropomyosin cDNAs isolated from skeletal and non-muscle mRNA. Nucleic Acids Res. **14**:9193–9211.

25. **Hallauer, P., K. Hastings, A. Baldwin, S. Pearson-White, P. Merrifield, and C. Emerson.** 1987. Closely related α-tropomyosin mRNAs in quail fibroblasts and skeletal muscle cells. J. Biol. Chem. **262**:3590–3596.
26. **Hayward, L. J., and R. J. Schwartz.** 1986. Sequential expression of chicken actin genes during myogenesis. J. Cell Biol. **102**:

1485–1493.
27. **Helfman, D. M., S. Cheley, E. Juismanen, L. 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.
28. **Helfman, D. M., J. Feramisco, W. Ricci, and S. Hughes.** 1984. Isolation and sequence of a cDNA clone that contains the entire coding region for chicken smooth-muscle α-tropomyosin. J. Biol. Chem. **259**:14136–14143.
29. **Kaine, B., R. Gupta, and C. Woese.** 1983. Putative introns in tRNA genes of prokaryotes. Proc. Natl. Acad. Sci. USA **80**: 3309–3312.
30. **Karlik, C., and E. Fyrberg.** 1986. Two *Drosophila melanogaster* tropomyosin genes: structural and functional aspects. Mol. Cell. Biol. **6**:1965–1973.
31. **Lau, S., C. Sanders, and L. Smillie.** 1985. Amino acid sequence of chicken gizzard γ-tropomyosin. J. Biol. Chem. **260**:7257–7263.
32. **Leavitt, J., G. Latter, L. Lutomski, D. Goldstein, and S. Burbeck.** 1986. Tropomyosin isoform switching in tumorigenic human fibroblasts. Mol. Cell. Biol. **7**:2721–2726.
33. **Leff, S., R. Evans, and M. Rosenfeld.** 1987. Splice commitment dictates neuron-specific alternative RNA processing in calcitonin/CGRP gene expression. Cell **48**:517–524.
34. **Leff, S., M. Rosenfeld, and R. Evans.** 1986. Complex transcriptional units: diversity in gene expression by alternative RNA processing. Annu. Rev. Biochem. **55**:1091–1117.
35. **Lewis, W., G. Cote, A. Mak, and L. Smillie.** 1983. Amino acid sequence of equine platelet tropomyosin. Correlation with interaction properties. FEBS Lett. **156**:269–273.
36. **MacLeod, A., C. Houlker, F. Reinach, L. Smillie, K. Talbot, G. Modi, and F. Walsh.** 1985. A muscle-type tropomyosin in human fibroblasts: evidence for expression by an alternative RNA splicing mechanism. Proc. Natl. Acad. Sci. USA **82**:7835–7839.
37. **MacLeod, A., C. Houlker, F. Reinach, and K. Talbot.** 1986. The mRNA and RNA-copy pseudogenes encoding $TM30_{nm}$, a human cytoskeletal tropomyosin. Nucleic Acids Res. **14**:8413–8426.
38. **Mak, A., L. Smillie, and G. Stewart.** 1980. A comparison of the amino acid sequences of rabbit skeletal muscle α and β tropomyosins. J. Biol. Chem. **255**:3647–3655.
39. **Marston, S. B., and C. W. J. Smith.** 1985. The thin filaments of smooth muscles. J. Muscle Res. Cell Motil. **6**:669–708.
40. **Matsamura, F., S. Yamashiro-Matsumara, and J. J.-C. Lin.** 1983. Isolation and characterization of tropomyosin-containing microfilaments from cultured cells. J. Biol. Chem. **258**:6636–6644.
41. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
42. **McLachlan, A. D., and M. Stewart.** 1976. The 14-fold periodicity in α-tropomyosin and the interaction with actin. J. Mol. Biol. **103**:271–298.
43. **Medford, R., H. Nguyen, A. 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.
44. **Mitchell, P., G. Urlaub, and L. Chasin.** 1986. Spontaneous splicing mutations at the dihydrofolate reductase locus in Chinese hamster ovary cells. Mol. Cell. Biol. **6**:1926–1935.
45. **Nadal-Ginard, B.** 1979. Commitment, fusion, and biochemical differentiation of a myogenic cell line in the absence of DNA synthesis. Cell **15**:855–864.
46. **Nadal-Ginard, B., R. Breitbart, M. Gallego, A. Andreadis, T. Endo, and C. Smith.** 1987. Alternative splicing: splice junctions are not functionally homologous, their use is regulated by cis- and trans-acting elements. J. Cell. Biochem. **11c**:42.
47. **Nadal-Ginard, B., R. Breitbart, E. Strehler, N. Ruiz-Opazo, M. Periasamy, and V. Mahdavi.** 1986. Alternative splicing: a common mechanism for the generation of contractile protein diversity from single genes, p. 387–410. *In* C. Emerson, D. Fischman, B. Nadal-Ginard, and M. Siddiqui (ed.), Molecular biology of muscle development. Alan R. Liss, Inc., New York.
48. **Parry, D.** 1975. Analysis of the primary sequence of α-tropomyosin from rabbit skeletal muscle. J. Mol. Biol. **98**:519–535.
49. **Periasamy, M., E. Strehler, L. Garfinkel, R. Gubits, N. Ruiz-Opazo, and B. Nadal-Ginard.** 1984. Fast skeletal muscle myosin light chain 1 and 3 are encoded by a single gene with a novel structural organization. J. Biol. Chem. **259**:13595–13604.
50. **Peterson, L., and P. Perry.** 1986. Regulated production of μ_m and μ_s mRNA requires linkage of the poly (A) addition sites and is dependent on the length of the μ_s-μ_m intron. Proc. Natl. Acad. Sci. USA **83**:8883–8887.
51. **Phillips, G.** 1986. Construction of an atomic model for tropomyosin and implications for interactions with actin. J. Mol. Biol. **192**:128–131.
52. **Phillips, G. N., Jr., J. P. Fillers, and C. Cohen.** 1986. Tropomyosin crystal structure and muscle regulation. J. Mol. Biol. **192**:111–131.
53. **Reed, R., and T. Maniatis.** 1986. A role for exon sequences and splice-site proximity in splice-site selection. Cell **46**:681–690.
54. **Reinach, F., and A. MacLeod.** 1986. Tissue-specific expression of the human tropomyosin gene involved in the generation of the *trk* oncogene. Nature (London) **322**:648–650.
55. **Reynolds, G., S. Basu, T. Osborne, D. Chin, G. Gil, M. Brown, J. Goldstein, and K. Luskey.** 1984. HMG CoA reductase: a negatively regulated gene with unusual promoter and 5′ untranslated regions. Cell **38**:275–285.
56. **Robert, B., P. Barton, S. Alonso, A. Cohen, P. Daubas, I. Garner, A. Weydert, and M. Buckingham.** 1986. The structure and organization of actin and myosin genes and their expression in mouse striated muscle, p. 487–506. *In* C. Emerson, D. Fischman, B. Nadal-Ginard, and M. Siddiqui (ed.), Molecular biology of muscle development. Alan R. Liss, Inc., New York.
57. **Ruiz-Opazo, N., and B. Nadal-Ginard.** 1987. α-Tropomyosin gene organization. J. Biol. Chem. **262**:4755–4765.
58. **Ruiz-Opazo, N., J. Weinberger, and B. Nadal-Ginard.** 1985. Comparison of α-tropomyosin sequences from smooth and striated muscle. Nature (London) **315**:67–70.
59. **Salviati, G., R. Betto, and D. Betto.** 1982. Polymorphism of myofibrillar proteins of rabbit skeletal-muscle fibers: an electrophoretic study of single fibers. Biochem. J. **207**:261–272.
60. **Sanders, C., and L. Smillie.** 1985. Amino acid sequence of chicken gizzard beta-tropomyosin. J. Biol. Chem. **260**:7264–7275.
61. **Sanger, F., S. Nicklen, and A. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
62. **Schachat, F., D. Bronson, and O. McDonald.** 1985. Heterogeneity of contractile proteins. J. Biol. Chem. **260**:1108–1113.
63. **Smillie, L.** 1979. Structure and functions of tropomyosins from muscle and non-muscle sources. Trends Biochem. Sci. **4**:151–155.
64. **Sodek, J., R. Hodges, and L. Smillie.** 1978. Amino acid sequence of rabbit skeletal muscle alpha-tropomyosin. J. Biol. Chem. **253**:1129–1136.
65. **Soeiro, R., H. Birnboim, and J. Darnell.** 1966. Rapidly labelled HeLa cell nuclear RNA. II. Base composition and cellular localization of a heterogeneous RNA fraction. J. Mol. Biol. **19**:362–372.
66. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
67. **Stone, D., and L. Smillie.** 1978. The amino acid sequence of rabbit skeletal muscle alpha-tropomyosin. J. Biol. Chem. **253**: 1137–1148.
68. **Strehler, E.** 1985. Multigene families, differential transcription and differential splicing: different origin of contractile isoproteins in muscle, p. 345–355. *In* R. Strohman and S. Wolf (ed.), Gene expression in muscle. Plenum Publishing Corp., New York.
69. **Strehler, E., M. Periasamy, M.-A. Strehler-Page, and B. Nadal-Ginard.** 1985. Myosin light chain 1 and 3 gene has two structurally distinct and differentially regulated promoters evolving at different rates. Mol. Cell. Biol. **5**:3168–3182.
70. **Taubman, M., J. Grant, and B. Nadal-Ginard.** 1987. Cloning

and characterization of mammalian myosin regulatory light chain (RLC) cDNA: the RLC gene is expressed in smooth, sarcomeric, and non-muscle tissues. J. Cell Biol. **104**:1505–1513.

71. **Thomas, P. S.** 1980. Hybridization of dentured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77**:5201–5205.
72. **Walker, M., T. Edlund, A. Boulet, and W. Rutter.** 1983. Cell-specific expression controlled by the 5′-flanking region of insulin and chymotrypsin genes. Nature (London) **306**:557–561.
73. **Wieczorek, D., M. Periasamy, G. Butler-Browne, R. 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.
74. **Yamawaki-Kataoka, Y., and D. Helfman.** 1985. Rat embryonic fibroblast tropomyosin 1. J. Biol. Chem. **260**:14440–14445.