Avian tropomyosin gene expression

Gary J.Lindquester⁺⁺, James E.Flach[§], Donald E.Fleenor^φ, Kathryn H.Hickman and Robert B.Devlin^ο

Emory University, Biology Department, Atlanta, GA 30322, USA

Received September 16, 1988; Revised and Accepted January 31, 1989

ABSTRACT

Sequence analysis of overlapping fragments from a quail genomic library has revealed a tropomyosin gene consisting of 13 exons spaced over about 18 kilobase pairs of DNA. Skeletal muscle and smooth muscle transcripts share the same 5' untranslated sequence and may initiate from the same promoter. However, the regions encoding amino acids 39-80 and 258-284 are specific to each muscle type. The two sets of exons encoding these regions undergo mutually exclusive alternative splicing in a tissue-specific manner as determined by Northern blots and S1-nuclease protection. Similarly, the 3' ends of the transcripts are different in skeletal muscle and smooth muscle, and each contains two polyadenylation signals which appear to be utilized *in vivo*. The avian alpha-tropomyosin gene is not expressed in cardiac muscle. The sequence of the gene shows great homology with other muscle-specific tropomyosins and includes a region homologous to the amino terminus of nonmuscle tropomyosins.

INTRODUCTION

Contractile proteins are the primary components of muscle cells and are major components of the cytoskeleton of nonmuscle cells. During myogenesis, different developmental and muscle-specific isoforms of these proteins are expressed. Thus, formation of specific muscle types such as smooth, skeletal, and cardiac muscle is associated with the differential accumulation of specific contractile protein isoforms.

One of these contractile proteins, tropomyosin, is a member of a closely related family of proteins which are bound to actin and troponin and which serve to mediate the effect of calcium on skeletal muscle contraction (1-3). Skeletal muscle contains two tropomyosin subunits, alpha and beta, which are present in different molar ratios in individual muscle types and at different times during development (4,5). Although mammalian cardiac muscle apparently contains only an alpha type subunit identical to the skeletal alpha subunit (5), avians exhibit a unique cardiac isoform (6, manuscript in preparation). Avian smooth muscle contains

alpha, beta, and gamma type subunits (7), distinct from the skeletal subunits, but very little is known about the function of these proteins since contraction in smooth muscle is thought to be regulated independently of tropomyosin (8,9).

In *Drosophila*, there are three tropomyosin genes, two which are expressed in muscle and one which is expressed only in nonmuscle tissue (10,11,12). One of the muscle genes, mTM I, utilizes differential splicing to express two sets of tissue-specific mRNAs - one in embryos and cell cultures and one in thoracic flight muscle of the adult.

Evidence that multiple vertebrate tropomyosin isoforms are expressed from the same gene has been presented from our laboratory and others (13-17). The rat skeletal beta-tropomyosin gene has been isolated and shown to express a 284 amino acid rat embryonic fibroblast tropomyosin by alternative RNA splicing and polyadenylation (18). Similarly, the rat alpha-tropomyosin gene has been isolated and shown to alternatively express alpha-smooth muscle tropomyosin (19).

Here we report the isolation of a quail tropomyosin gene which may express at least three distinct protein isoforms - skeletal muscle alpha-tropomyosin, smooth muscle gamma-tropomyosin, and a nonmuscle tropomyosin. Like the rat alpha-tropomyosin gene, mRNAs initiated from a common promoter undergo alternative splicing of mutually exclusive, isotype-specific exons coding for amino acids 39-80 and 258-284 including the 3' untranslated region. These mRNAs encode the 284 amino acid skeletal and smooth muscle isoforms. Each mutually exclusive 3' exon contains two polyadenylation signals, all four of which appear to be utilized *in vivo*. We also present evidence which suggests that the quail tropomyosin gene may be transcribed from an alternative promoter in nonmuscle tissue to encode a 248 amino acid protein with a unique amino terminus.

MATERIALS AND METHODS

Isolation and Characterization of Genomic Clones

Quail genomic libraries created by ligation of fragments from a partial EcoRI digest of quail liver DNA were screened essentially as described by Benton and Davis (20). Isolation and preparation of nick-translated tropomyosin cDNA probes was described previously (13). Hybridizations were carried out at 42°C overnight in a hybridization buffer containing 50% formamide, 4x SET (10x SET is 1.5M NaCl, 0.3M Tris pH8.0, 10mM EDTA), 0.2% SDS, 0.1% sodium pyrophosphate, 50ug/ml yeast RNA (Type III, Sigma), and 50ug/ml heparin. Final washes were performed at 42°C in 2x SET.

A restriction map of each genomic clone was generated by analysis of single and double restriction enzyme digests. Southern blot analysis was used to identify portions of the clone which hybridize to the cDNA probes.

DNA Sequencing

Genomic fragments were subcloned for sequencing into M13mp18 or M13mp19 by standard procedures (13). A series of nested deletions was generated from fragments larger than 500 bases as described by Dale, McClure, and Houchins (21). Dideoxy sequencing was performed as described earlier (13,22). Sequence analysis was facilitated by the use of IBI Sequence Analysis Software (International Biotechnologies, Inc.).

RNA Isolation and Analysis

Tissue used for RNA extraction was dissected, rinsed in saline, and stored at -80°C until used. The frozen tissue was then fragmented in a Waring blender with 20 volumes of 20mM sodium acetate, pH5.1, 250mM NaCl, 10mM EDTA, 1% SDS, 50% phenol, and 500ug/ml heparin. The homogenate was centrifuged at 12,000g for 20 minutes, and the supernatant was extracted twice with equal volumes of phenol and chloroform. After precipitation in ethanol, the RNA was dissolved in 10mM Tris, pH8.0, 1mM EDTA and precipitated in 2M LiCl. In some cases Poly(A) RNA was isolated as described earlier (23).

Northern blot analysis was done as described previously (23,24).

S1-nuclease mapping studies were done essentially as described by Favaloro, Treisman, and Kamen (25). Fragments were 5' or 3' end-labeled (26,27) or uniformly labeled (28). Labeled fragments were hybridized with 30ug of total RNA in 80% formamide, 0.4M NaCl, 40mM Pipes, pH6.4, 1mM EDTA. Hybridization was at 48-56°C (experimentally optimized) for 16 hours, followed by digestion with 150 units of S1-nuclease. Products were analyzed on 8% polyacrylamide/8M urea gels.

RESULTS

Isolation of Alpha-Tropomyosin Genomic Clones

We previously reported the isolation of several clones containing alpha-tropomyosin sequences from cDNA libraries prepared from quail skeletal or smooth muscle RNA (13). These clones contained identical sequences coding for amino acids 81-257 of skeletal muscle alpha-tropomyosin where they overlap, strongly indicating they were derived from the same gene. However, differences were seen in the coding regions for the final 27 amino acids and in 3' untranslated sequences. S1-nuclease and Northern blot analysis indicated that these sequences, as well as sequences at the 5' end of the tropomyosin mRNA, were possibly expressed by alternative splicing of a primary transcript from a single gene.

To confirm this possibility, genomic fragments containing tropomyosin sequences were isolated from a genomic library prepared by partial Eco R1 digestion of quail DNA. A series of 7 adjacent and overlapping clones comprising over 31 kb of contiguous genomic sequences were isolated and analyzed by standard procedures using the skeletal

alpha-tropomyosin cDNA clones described previously (13). As described below, these clones contain sequences coding for a tropomyosin gene differentially expressed in skeletal and smooth muscle, and possibly in some nonmuscle cells. Several phage containing sequences coding for other tropomyosin genes were also isolated and will be described elsewhere.

Localization of Skeletal Muscle Alpha-Tropomyosin Sequences

Each of the Eco R1 fragments shown in Figure 1a was subcloned into pBR322, subjected to single and double digestion with several restriction enzymes, transferred to nitrocellulose and probed with the alpha-tropomyosin cDNA clones derived from skeletal muscle. Ten distinct regions of homology with the cDNA clones were located within these fragments. These regions were subcloned into mp18 and mp19 and sequenced. The results of the sequence analysis are presented in Figure 1. Ten exons were identified by homology with corresponding sequences on the skeletal muscle alpha-tropomyosin cDNA clones and consensus splice junctions (29). They encode 5' untranslated sequences and amino acids 1-38 (Exon 1), 39-80 (Exon 3), 81-125 (Exon 5), 126-164 (Exon 6), 165-188 (Exon 7), 189-213 (Exon 8), 214-234 (Exon 9), 235-257 (Exon 10), 258-284 (Exon 11), and 3' untranslated sequences (Exon 12). The last base for the final codon as well as the termination codon are located on this latter exon. Two polyadenylation signals (AATAAA) are contained within the 3' untranslated sequences, and both appear to be utilized in vivo, creating mRNAs with different 3' untranslated regions (13). The sequence encoded by these 10 exons is identical to the sequence of the skeletal alpha-tropomyosin cDNA clone, SK74, presented elsewhere (13), except the cDNA clone did not contain the entire 5' untranslated region.

Localization of Smooth Muscle-Specific Tropomyosin Sequences

In an earlier report we described a cDNA clone containing sequences coding for amino acids 81-257 of skeletal muscle alpha-tropomyosin contiguous with sequences coding for an additional 27 amino acids (258-284) and 3' untranslated sequences expressed only in smooth muscle, suggestive of alternative splicing of mRNA from a single gene (13). Also, northern blot and S1 protection experiments suggested differential splicing at the 5' end of the mRNA.

To determine the location of exons coding for smooth muscle-specific sequences, poly(A) RNA from quail smooth muscle (gizzard) and skeletal muscle (breast) was used to identify two regions from the genomic clones which hybridize only to the smooth muscle RNA probes. One of these regions is contained in the 7.5 kb Eco R1 fragment (which also contains the 3' end of the skeletal mRNA sequences), and the other is found on the 7.2 kb Eco R1 fragment (which contains two exons coding for the first 80 amino acids of the skeletal muscle tropomyosin).

Appropriate fragments containing these sequences were subcloned and sequenced. The 7.5 kb Eco R1 fragment contains sequences identical with a portion of SM18, a clone isolated from a smooth muscle cDNA

ccctacaattctgtgattctgatgaattctgaagggtccgtcttaagactctggtattgttttctaactactcttctgtgagtgaaagacgctt	4783
cagcacaatacatagttcttaaaataggttagttcccggtccctgccatggagtgctgtgacacctctggagcgacgagcgagctgtgagggtgac	4883
tcccactccgtctggaagtctgactgaggaagctggccgtgggtgaggtctagcatctcaaaagggtctgcatctgctgcagctgcagcag	4983
gtttatgtgattgatggaactctgcagctgtgtttgtgtgacacagttttgtttggagaggatcagccc (...150bp...) gggcacaatacagag	5218
gtgcttatttgcactgcagccttgacatggtatttttactgtcacagctgattttctgcatgactggaaggctgtgaaacatggcgagc	5318
aagctgaggtgtgaacctgacccctcccttagtgcagggctgactgtgacacagctgacccctctgctatgctcttgggtgtttttttaaatttt	5418
atttcttctgtgcaggaacagatagcaaacagtgcaaacctgcgcgaacagcatcacagcttcaactgctgtgtgtgttaaacagatg	5518
gggaagggtctgggtgtctgcagcgccagcgtgtgctctactaaaagctgtggctgtgtctcactggcgtgtgactgcatcttctaataagttgtcc	5618
tctctgcagctgattgattagctgaaca (...100bp...) ttacgccataaacgcgtgtgtgttacgcactgacgtctacactgccacgcccacgccc	5803
gtccctcttgccttttcccttcccttctcgcgaacttgcccttcccccactcaataacggggctccctttaggttccatttattgctttactcacc	5903
tcgaccccaaaaactgatttgggtgaggttttccattggccatgcccctgagatagcggttttttcgccccttgaggttcacagctctttaaattg	6003
tggactctgtgtccaacgtgaacaacactcaacctatctgggtgattcttttgattataaaggattttgcgatttcggaacacacattcaaaagg	6103
atttttgcctgtggggcaaacacagctgagacgctgtgtgcacactctcaggggcagcggtgaagggtacacagctgttgcctgtctactgtgaa	6203
agaaaacacacccctgtgccecaatcgcacactctctccgcggttggccagttcatlaagcagctggcagacagagtttccgcagctgaaagcgggc	6303
agtgaacgcacgcgaataattgtgagttagctc (...175bp...) ctgcagatattgtctattttactacttctgaggtacagctggctactacca	6563
gttgtgcttaacacgcgataagaacacacttaggcactataaaacaacctcagtgagagcagtatcactgctactggagctgctagcagtaaacacgactc	6663
gggaagcacaataaagtattgaaacttaattgggtactagagcagcagacactgacgaggagtggaaggagacaactctttacacttccatttcaagct	6763
gcatataaactcttccctgattgtgtgttaatacaagcagtgcatctccaacactgcaggtttggaatgcacactgttatcagcagtgagaagaattaa	6863
gtcagcgtgaaatggaagtgaagaaggtgtgtctccctcactctcgctcagtgctgcctctagaccactaaagtgttcaaatatcatattttgtc	6963
tcccagctgtgtgttactgtcagcagctccagtgacgtgatactgctcactgaggtgtgtttatagctcaagtgttcttctgtgtgttagcgca	7063
catctggtagtagcagcagctgactcagaaagttaaaatgcagaacatctgcagg (...50bp...) cctgcagctccttagacattttagaagaag	7199
aaagcactcaagctgactcttttgggggaagatttccctgtgtgtgtttttaaaccctccccctcagctcagcttctgtgtgtgtgtgtgtgtgtgt	7299
atagcaagatcatttgcagcttgatttcaaatagctgttaacttgggaagcatttaactatttagatataagttatttggcttattagcttcttcagaa	7399
gaacttaacttaataacaggaattgtacaggaagagctggatttctgcacacttcagaaagtgtgtgttaagtaagttataagtaaaagcttataaag	7499
aggcagagattgtgtgagttgttataaaaaaagaaatagcaagattgtgtgcagagagagatggatttgggtggcagcattgggtggggatgaactgt	7599
tttaagcaggtgtgtgtataggcatttgggtgttcaaacgaggaaggttcaatgaagggttcaaggtaaatgtagggtgtgtgtgtgtgtgtgtgt	7699
tgatttcttaggaagggaagagattgtattgt	7799
atatacattgtgcacatttcttcaatcagcttaataagaaaaagacatcagaagaacatgcacactctgtctatggacttctagtgaaagaacac	7899
agcaattaaagtgtccattcagtagtactagaaggtacatgggttaactctcataagaggtcagctgctgctctgtgtgtgtgtgtgtgtgtgtgt	7999
gtgtctag GCT GAG AGT GAA GTA GCT TCC CTG AAC AGA CGC ATC CAA CTG GTT GAG GAA GAG TTG GAT CGG GCT GAG	8076
Ex 5 (81) A E S E V E S E A C L A N R I Q L V E E L D R A Q	
GAG CGC TTG GCT ACT GCC CTG CAG AAG CTG GAA GAG GCT GAG AAG GCT GCA GAT GAG AGT GAA AG gtgggtga	8150
E R L A T A L Q K L E E A E K A A D E S E R (125)	
ggttactcttgaagcagattaggcggtgctgtatgacgaaggtggttaattgctagggtgtgggttagggaaataaataaacgcgtcttatgtctct	8250
atgcattatgtataccactctgcttcttggactcagtcacacagcttctgcagctctgcagatggatggatgcagatgcacacataggtttaaaat	8350
gcttgtgtcagctcagtttgaagagataaagtagagcgtgggtgagcttttctcattatcatcagcttctcggggaaacgttggagccagctgag	8450
accttatttgaagcactatcagagggtgattgt	8550
tttttttttctgtcttacc A GA ATG AAG GTC ATT GAA AAT AGA GCC CAG AAG GAG AAG ATT GAA AIT C Q	8735
Ex 6 (126) G M K V I E N R A Q K D E E K M E A I C Q	8810
GAG ACT CAG CTT AAA GAA GCT AAG CAT ACT GCT GAA GAG GCT GAC CGC AAG TAT GAA GAG gtcaagtttggggcag	8886
E I Q L K E A K H I A E E A D R K Y E E (164)	
agggaacggcctgaagcactttaccacactttgggttctaggggacggtggt	8986
tag GTG GCT CBT AAG CTT GTG ATT GAG GGT GAG CTG GAG CAG GCT GAG GAA CBT GCT GAA CTA TCA GAA	9058
Ex7 (165) V A R K L V I I E G D L E R A E E R A C L S E	
AG gtaagtgctcaattgccccctgacatgctgtgaggtgctacaagacagaactgttaacagcatggaacattggcagtcactgtgtgttca	9152
S (188)	
gtcacaagctatgtttgtctttaaagttgaagtaaatggtcaacaccaagtaaatgtgtttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	9252
cttcagcactgattttgtgaagcttttggctatttcatgtgttctcaaaagcagaactgcagacagctggaagacagcttaagaataatggatacaact	9352
tggaacacttnttggctgcagaggaataagctactgatactaaataaacagtttttagtttaactgcacccagcttcttgcacttccatagccatagc	9452
tcactcagtagcaaaataacttctcactgtcttctactcccttgcatcttctgttgggtgttttttcttggctcttttttgaatactttttttacc	9552
tttaaaactgactctcctgtgcagccaaaagaagctgaactgtctgtatttgaagctctgtgaaatattctgtgtaaaccaagactgtaaaactcta	9652
tgttggtgtggcattgttgaagtgtcacttaaaatggaatttggcatcacaacaaatgtcttcagctgtgtagtgatcagctgttcaaaagccgt	9752
tttttgtctgtagttgttttcaatttttctctactcttttcccttttttttttcttatttttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	9852
tgcttctcttcgaataacaag C AAA TGT GCT GAG CTT GAA GAG TTG AAA ACT GTG AAC AAC CTG AAG TCG	9929
Ex8 (189) K C A E L E E L K T V T N M N L K S	
CTG GAG GCT CAG GCT GAG AAG gtaggcataagcttaccatagagcagtgcttctatgctggccataggttggctgtgtgtgtgtttta	10018
L E A Q A E K (213)	
tat (...220BP...) gaattctgcttttgaatgaatgt	10323
actcccttttatacagttgattactcaggagacaggtgttttatgtttttagataactggtctaaagcagtggttttcttcttaattacag	10417
TAC TCG CAG AAA GAA GAA TAT GAA GAG GAG ATT AAA GTT CTA GAG GAG gta	10483
Ex9 (214) Y S Q K E D K Y E D E I K V L T D K L K E (234)	
aaattagagctgttatgtacacagatcttccagcagctgtgtgattctttaaataaacagctagcttgaattcttaagaataatgtcttggggac	10583
agttgattattggatctcctaaagaagagctacaccttattgtctattttaaactcagctgttctcactgttcttcttctccttctcattatgt	10683
ctcctctatttgaagttagtattgtgtattctcacaatgt	10768
Ex10 (235) A E T R A E F A C	
GAG AGG TCA GTA ACC AAG CTG GAG AAG AGC ATT GAT GAT CTA GAA G gtaggtacttgttttcttcgcaacttaaaagagat	10850
E R S V T K L E E K S I D D L E (257)	
acaggcttttatttctcatgttttattcactaccacacactcactcaaatgttcagatcacataaattgggtctacttaaaaagtaaaaaggta	10950
aaatgatttcccatcacaatacatattgataattatgtgt (...150bp...) ctccaagcatttgcacttcttcttcatagcttgcagctgggaacact	11185
aaataaaataaataaggtaaatatgagataaaatagatcatgatactaaaagctgtggaagagctctgtgtaagtttttcaatgtaaaat	11285
gtcgaagcgtgaataattatgttctgataaggaggagacagacacagtgactggttttagtagagcattaggtgttaacaacattgaggaggac	11385
accttatttactcaggaatttgaagtttgaatgttttgaatgttctcagacgtcagctgtgaatgcacttcttccctggcgaataaactgagcagtttgg	11485
aatcaagtttgaatggagcttctcatttgcagtggtctatcaagcttagattttatgcagtttgaatttgcagttggcagactacatgttaagag	11685
tatgcagttatagcttaacacagtgaaatgt	11785
tgatttgcacttcccatctcaggtgttttgcgttcacatgtcttcttataagcactgaacaaatttccaggacaagagagcagattttt	11885
gttccctctcactatgctttgtgtatccagacacagctgtccacccgggcagtgagttatgaacagaaactatgtctgtgttttattcttcc	11985
ttaattttgtattttagctgttttttctattgtgtcagctgtccctcaacacttcttctgtcactcacccttctcgtctgtgtgtgtgtgtgtgt	12085
ccacaccactgcttcttctcattcccatgacgcactccatttctgtgtcactctgcttcccttggggacgggttag AT GAG CT	12175
Ex11 (258) D E L	
TAT GCT CAG AAA CTG AAG TAC AAA GCC ATG AGC GAG GAG CTC GAC CAT GCT CTC AAC GAT ATG ACT TCC AT	12246
Y A Q K L K Y K I G S T S E E L D H A L N D M T S I (284)	

```

gtaaatcttttaccagctgtgctgtgctgttccccgccctcattgattgtgtggttctgctttgtgaattatactccctgttagtggagatg 12346
cattactgtagtgaagtttagttagtacaataaccatgaaccatttattaccgtggcagagctgggtgtgtgtgtgccaagctgtgtaaaactatg 12446
gtgattttgttaagctaatccatgagctcatgtgtgtttatcagtgatcagcagctctatactcctgtgtgctcattgtgttagctctgcattta 12546
tctctcattgtgacactctgtgttcaagttattttttatcattcatttgaatttgcagaaatctactatcacaacatgccgcatcagaggtctctctg 12646
ccttcagttttatctggcacaagctttacataaagtttcagaaggactagaattagctctgttcatattgtagattgcttttcaatttccctga 12746
actgcttgaattccaccactacaattcaattttcaactgtccactacaataattacaaggtggaaggttaagacattttgtgattagagcagaaaggtg 12846
acactgaagtgtgtgtgtgtgtgtatcttctgtgaggcagcaacgaatgaatttaaatgccagaagtgatgaattgcagcttcagcttagtagggac 12946
agcacactgggggtgggggaaaaacagaaagcaatcatttagtaggagtgtagtaattcctgtgtgtgtgtatgtaccacttttccctgcattga 13046
cataaaactgcaacagaggtcagctagctaaatccctcctccatgctcatcccttccatttgcactatacatttggcgtgggaggaagattttgcttt 13146
ctcaagtgcctntgaggtgtgacgttcagttccagcatcttctttgaaggtgtgtgctttcaggcaggactgcag (...200bp...) atttttcatt 13431
ggatattgcagatggaaacaggtctcaaatgtagccctctgtgacatgaggaaatgctatgcattccaatcccagcattgtgtggcatttccactagc 13531
agtcatcccgagctcctcattttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 13631
ttgtgcag A TAA TTTCTTTTGTCTCCTCTCTCAAGACATCTTCATCGGGTGGATAAAACATCTTTCCAAGCTTTTGGATGTTTCATGGCTCC 13726
Ex12 END (3' UT)

CTGTCTGTCTTCTAGCTTACTTGGCTTATCTTCTGGCACCTGTTTCAGAACATGCTCTTTGTGCTCTATTGTACAGAAACTCTACATCTCTAAATGT 13826
AAATAAACATCCCGAGCTGATTTCTTTGGCTATCCCTGCTCTTAAATTTATTCAAATAGCTATTTAGTATTATTAATAAAATAAATGTTGTGTGG 13926
GTTtttttttgcctcagagattttgttttgcattgttgccttttttatttcccccccaacgcttttgccttaattgtaaaatgcaaaagcagcg 14026
gtagctgtctgtgggcttatgaatgcagttaaaaatagcacaataatgtgggaattttctcagttgttattgttgggcagcatctgcag (...4300bp...) 18414
acagatctcctgtgagtgatcatttgcagctgcagctgcagctgcagctgcagctgcagctgcagctgcagctgcagctgcagctgcagctgcagct 18514
attctctgaggtatcagctatttattgatcatgcatatttctaccttccagagcctttaaataagaatgatcagctgcagctgcagctgcagctgcagct 18614
tgccctggaagactgctattcattgtttgtatttctgatccccacaag AG AAA GTG GCG CAT GCC AAA GAA GAA AAC CTT 18696
Ex13 (258) E K V A H A K E E M L

AAT ATG CAT CAG ATG CTG GAT CAA ACT TTA CTG GAG TTA AAC AAC ATG TGA AAACCTTCTTAGCAGCAACCATGTTGTTATTA 18779
N H H Q M L D Q T L L E L N M N END (3' UT)

TTTTTACACCTGCTGCTGCGGAAGCCCATAAACATGTCTTTATTTAGTTTCCACCAGCTCACACAACATCAGCTAAATATCAGGCAAGGTCAGGAA 18879
GTACCAAAAAGGCGCAAGCTGTATGCGGGGTAAGCTATTTCTATTATGTTTAAATGTGCATTTCCCAACAATAATGTTACCTACACCTTTGTAGAGAGTTC 18979
AGAGATTGGGAATGCTTAGCTTTTGGAAATATCTTGCTCTGTGAGGTAGCGTCCCACTCAAGTGCCAACACAGATTTTGAACAGCATGATACATCTTC 19079
AAAAAGATCCAGCATAAAGGGTGTATTTGGATCCAAAGCTCTGTGAGGGGCTTTTATTTGATACCCGATGATGAGGACTGTTGAGCACAGTAATTT 19179
ATTTTATCAGCATTTTACTGATCTGTGAGGATTGTGACTTTGTGCTTTGATTTCAATAGAACGAAGGCTGTGTGGTGCACACAGTAACAAAGGCTTT 19279
ATGAGGTTTAAATTTGCAACTGCTGATACATAGTGTAGGTCACATTTCCGCTTGAGTATTCTTCATTGTATTTTTCATGTTTCAATGAAAG 19379
ATTCAATAAATTCAGGAGAAAACACAATTTGGAAGAGATGTTTAAATGTATTTTATTTGTGGAGCTTGGGGCGGGGGTGGAATACAAATCGATA 19479
TTCGCTCCTCATTTATTTTACACATTTCTTGAGCAGTTTAAAGTATACCTGTATGTAACACATGTATATGATATGTAATAATGATGTAATAATGATGTA 19579
TTTTTAAATCGTACctcctctggatttcatcctcataatgtctgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 19679
atggttttcagtttccccntgcttttaagctagcatataagaagaagcatgaaatgcagtggtttggactttagtggtgagctgcagctgtcagggggacc 19779
atactgtgagaatcatccccccccccatcattgtatgtgttccagatggatgcagcagcagcaccatcactcaaggaagacagacatlaagtaaa 19879
aatgtaaaagcagtaaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 19979
gcagaaaggcgtgtgaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 20079
tagcagtaaggtaaaagtgttacacgtggggtcattaaaggagctgtaggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 20179
tattttgaatgaaggaggaagtttgcagcattgcagggcctcattggctgtgtgaggtcaacctgaagtgggaaggagagatcggtggaatcacatccca 20279
tgtaaagtagcctttggaggagcaaaatgaagtgttaaaggctctgttagtccttaagcttgcatcctgcaggtcgactctagacatccccgggtac 20379
cgagctc

```

Figure 1. Nucleotide sequence of the alpha-tropomyosin gene.

a. The diagram shows the exon arrangement of the quail tropomyosin gene. Hatched boxes represent alternatively spliced exons and solid boxes represent exons common to all the transcripts, with the exception of Exon 1, which is not found in the cytoskeletal transcript.

b. Nucleotide position +1 marks the putative transcription start site. Exon sequences are in bold capital letters, intron sequences in small letters. Amino acids are indicated using single letter code. Putative "CATAT" and "CAAT" sequences are boxed and polyadenylation sites are underlined. Portions of the introns not sequenced are represented by dots with the amount of unsequenced DNA in parentheses.

library. The sequences represent the last 27 codons and the entire 3' untranslated sequence (Exon 13). As shown in Figure 1, these sequences are contained in one exon, unlike the corresponding sequences expressed in skeletal muscle which are present in two exons - one containing information for the final 27 amino acids and the other containing 3' untranslated sequences. The 3' untranslated sequences contain two alternative polyadenylation consensus signals, separated by 168 bases. Northern blot analysis of smooth muscle RNA indicates that both of the polyadenylation signals are used (Figure 2).

The smooth muscle-specific region located on the 7.2 KB Eco R1

fragment contains an open reading frame of 126 bases bounded by consensus splice sites (Exon 2). The coding region is the size predicted of an exon encoding amino acids 39-80, which has recently been shown to be alternatively spliced in the rat (19). The derived amino acid sequence is identical to the corresponding amino acid sequence of chicken gizzard gamma-tropomyosin (29.5), but shows only 64% homology to the corresponding region reported for the rat by Ruiz-Opazo (19). In contrast, the derived amino acid sequence corresponding to amino acids 258-284 is identical between quail and rat in 25 out of 27 positions (92% homology). These findings indicate that the alpha-tropomyosin gene contains alternatively spliced exons coding for amino acids 39-80, 258-284 and 3' untranslated sequences. One set of exons is expressed in skeletal muscle and the other is expressed in smooth muscle.

Exons 2 and 13 Are Expressed Only in Smooth Muscle

To verify the tissue-specific expression of Exons 2 and 13, Northern blot and S1-nuclease analyses were performed. Oligonucleotide probes homologous to each of the exons were hybridized to RNA isolated from a variety of muscle tissue. Figure 2 shows the results when the oligonucleotide homologous with a portion of Exon 13 was used. It does not hybridize with RNA isolated from breast (skeletal muscle), heart (cardiac muscle), leg (mixed fast and slow skeletal muscle), anterior latissimus dorsi (slow skeletal muscle), or posterior latissimus dorsi (fast skeletal muscle). The probe does hybridize with two sizes of gizzard (smooth muscle) mRNA differing by about 175 bases in length.

Figure 2. Northern blot analysis of smooth muscle-specific sequences.

RNA was isolated from anterior latissimus dorsi (A), posterior latissimus dorsi (P), breast (B), heart (C), leg (L), and gizzard (G) tissues, separated on a agarose/formaldehyde denaturing gel and transferred to nitrocellulose. A 30 base oligonucleotide homologous to a portion of Exon 13 was end-labelled and hybridized with the filter. RNA standards purchased from Bethesda Research Laboratories were used to determine transcript size. RNA was isolated from each tissue at six times during development: 10 day embryo (1), 14 day embryo (2), 17 day embryo (3), 1 day post-hatch (4), 10 days post-hatch (5), and 21 days post-hatch (6).

These two sizes would be expected if both polyadenylation sites in the 3' untranslated region were used.

To determine if either RNA was preferentially expressed at different times during development, RNA was also isolated from early (10 day), mid (14 day), and late stage (17 day) embryos, and from 2 day, 10 day and 21 day post-hatching quail. As can be seen in Figure 2, both RNAs are expressed in essentially the same ratio throughout development. It should also be noted that the probe fails to hybridize with other muscle types at any stage of development. Identical results were obtained using an oligonucleotide probe homologous with a portion of Exon 2, which codes for amino acids 39-80 (data not shown).

Although Exon 13 sequences were found on cDNA clones isolated from a smooth muscle library, none of the cDNA clones was large enough to contain sequences from Exon 2. Therefore, S1-nuclease analysis was used to further verify that this exon is specifically expressed in smooth muscle. A 660 bp Pvu II fragment containing Exon 2 was uniformly labeled, hybridized with RNA from skeletal and smooth muscle, and digested with S1 nuclease. Figure 3 shows that a fragment of about 130

Figure 3. S1-nuclease protection of Exon 2.

A 660 bp Pvu II fragment containing the 126 bases coding for Exon 2 was uniformly single-strand labeled from a M13 subclone and used in an S1-nuclease protection experiment with smooth and skeletal mRNA. The size of the protected fragment in Lane G is 130 bases. Labels at the top of the figure are: 123 bp DNA ladder purchased from BRL (S); unreacted 660 pb Pvu II fragment (P); mock protection assay using tRNA instead of muscle RNA (M); protection assay using skeletal (breast) muscle (B); protection assay using smooth (gizzard) muscle (G).

Figure 4. Northern blot analysis of skeletal muscle-specific sequences.

RNA from the same tissues and times of development as described in Figure 2 was separated on an agarose/formaldehyde gel and transferred to nitrocellulose. A 30 base oligonucleotide homologous to a portion of Exon 12 was end-labelled and hybridized to the filter. A 1.2-1.3 kb RNA found only in skeletal muscle hybridized with the probe.

bp is protected in smooth muscle RNA only, a size very close to the 126 bp that would be expected if Exon 2 was expressed in smooth muscle.

Exons 3, 11, and 12 Are Expressed Only in Skeletal Muscle

Oligonucleotides were also made against Exons 3, 11, and 12 and hybridized to RNA as described above. The results, shown in Figure 4 for Exon 12, indicate that only skeletal muscle RNA (breast, leg, and latissimus dorsi) contains sequences for these three exons. These profiles are reproducible using probes to Exons 3 and 11 also (data not shown). Furthermore, these 3 exons are only expressed in the fast twitch PLD muscle, and not in the predominately slow twitch ALD muscle.

It should also be noted that the skeletal muscle-specific exons do not hybridize with cardiac mRNA. This finding is in contrast to the situation in mammals in which the cardiac tropomyosin protein is identical with its skeletal muscle counterpart (5,6). In fact, in quail there is a distinct tropomyosin gene which is expressed only in cardiac muscle (manuscript in preparation).

Exons 5-10 Are Expressed in Both Smooth and Skeletal Muscle

As reported previously (13), S1-nuclease protection and Northern blot analysis demonstrates that the exon coding for 5' untranslated sequences and amino acids 1-38 (Exon 1) is expressed in both smooth and skeletal tropomyosin mRNAs. In contrast to earlier reports (14), similar findings have now been reported for the rat alpha-tropomyosin gene (19). To determine if any of the exons coding for amino acids 81-257 (Exons 5-10) undergo alternative splicing in the tissues studied here, an S1-nuclease protection assay was performed. A 600 bp Pst I fragment from a cDNA clone containing these sequences (13) was end-labeled at the Pst I site within Exon 5 and protected from S1-nuclease digestion by hybridization with skeletal or smooth muscle mRNA. Figure 5 shows that

Figure 5. S1-Nuclease analysis of Exons 5-10.

A 600 bp Pst I fragment from cDNA clone pQTM682 (13) was end-labeled and used in an S1-nuclease protection experiment to determine if Exons 5-10 were expressed commonly in skeletal and smooth muscle. A 410 bp protected fragment should result if these exons are expressed in both transcripts since the cDNA clone does not contain Exon 11 as shown by the dashed line. As can be seen in the autoradiograph, a 410 bp band is present when either smooth or skeletal muscle RNA is used (lanes G and B). Lane designations and molecular weight markers are the same as in Figure 3.

the smallest protected fragment in both muscle types was 410 bases, demonstrating that Exons 5-10 are expressed in both smooth and skeletal muscle. This evidence is also supported by Northern blot analysis, using oligonucleotide probes specific for each exon (data not shown).

Identification of a Putative Exon Which May Be Expressed in Nonmuscle Cells

During the course of sequencing the exons, a considerable amount of intron sequences was also obtained. Computer analysis of these sequences revealed an open reading frame in the intron following Exon 3

Figure 6. Comparison of tropomyosin sequences. Several tropomyosin sequences are shown in comparison to the quail alpha-skeletal sequence. The complete sequence in bold type is that of the quail skeletal alpha-tropomyosin (QSk). Partial sequences in bold type are those of the putative quail nonmuscle and the quail smooth muscle tropomyosin, as labeled. Chicken skeletal alpha-tropomyosin is identical to the quail skeletal sequence (13). Below the quail skeletal sequence is the sequence for rat and rabbit (RSk), which are also identical (19,38), and for the human skeletal alpha-tropomyosin (HSk) (40). Smooth muscle sequences for the identical quail and chicken (29.5) gamma-tropomyosin (QSm) and rat smooth alpha-tropomyosin (RSm) (19) are shown. Finally, nonmuscle tropomyosins from quail (QBrn), human (HFib) (40), and equine platelet (EPla) (30) are shown. Dots represent identity to the bold-typed sequence immediately above. Amino acid differences are indicated by the correct amino acid. The number of the last amino acid in each line is designated. Only the sequence for the first 44 amino acids of the suggested quail nonmuscle tropomyosin are shown. Above each line of the quail sequence, alternating diagonal and vertical stripped bars represent adjacent exons.

which showed 64% homology to the N-terminal 45 amino acids of the human fibroblast tropomyosin gene and equine platelet tropomyosin (17,30). The comparison is shown in Figure 6. This region is referred to as Exon 4 in Figure 1. The open reading frame of Exon 4 begins with a methionine codon and ends with a consensus donor site. The methionine codon is contained within a Kozak consensus sequence of eukaryotic translation initiation sites (31). Initiation of translation at the

methionine codon followed by splicing of Exon 4 to Exon 5 would yield a protein of 248 amino acids. This length is identical to the length of human fibroblast and equine platelet tropomyosins (17,30) as well as low molecular weight tropomyosin isoforms detected in chicken brain and intestinal epithelium (32). These data suggest that an additional alternative splicing event may occur in this gene, in which Exon 4 is spliced to Exon 5, producing a 248 amino acid protein.

Preliminary S1-nuclease protection data suggest that mRNA isolated from both muscle (skeletal and smooth) and nonmuscle (brain) is homologous to the genomic sequences for approximately 264bp upstream of the methionine codon of Exon 4. We have sequenced 313bp upstream of the methionine codon (Figure 1) but have been unable to identify potential CCAAT or TATA consensus sequences. However, these preliminary data suggest that Exon 4 may be differentially expressed in a cytoskeletal isoform of the alpha-tropomyosin gene containing a unique 5' untranslated leader sequence.

Analysis of the 5' End of the Alpha-Tropomyosin Gene

One cDNA clone, SK74, extends 39 bases 5' from the initiation ATG codon (13). These 39 bases are identical to genomic sequences contiguous with the initiation ATG codon of Exon 1, indicating that both coding and 5' untranslated sequences are present on this exon. To determine the extent of the 5' untranslated sequences present on Exon 1, an M13 genomic subclone beginning 23 bp into the coding region of Exon 1 and extending for 1280 bases in the 5' direction was uniformly labeled and hybridized to skeletal or smooth muscle mRNA. Following S1-nuclease digestion a fragment of approximately 180 bases was protected in both skeletal and smooth muscle mRNA (Figure 7). The band seen in the cardiac lane is not reproducible. This result indicates that approximately 160 bp of 5' untranslated sequences are contained on the first coding exon. A tropomyosin mRNA containing 160 bases of 5' untranslated RNA, 852 bases of coding RNA, and 286 bases of 3' untranslated RNA is consistent with the size determined by Northern blot data (13).

Although it is possible that a very small exon containing the transcription start site and additional 5' untranslated sequences has not yet been identified, four lines of evidence argue against this possibility. First, no consensus acceptor splice site can be identified near the 5' boundary of the exon, implying that we have identified the 5'-most exon. Second, using labeled mRNA from both skeletal and smooth muscle as probes, we have not been able to detect homologous sequences in the 6000 bases located upstream from Exon 1. Third, putative CCAAT and TATA consensus sequences are located 76 and 24 bases upstream from the S1-nuclease cleavage site. Finally, after these experiments were completed, a description of the rat skeletal muscle tropomyosin gene was published (19), containing an identical 5' transcription start site (see Figure 8 and Discussion).

Figure 7. S1-nuclease protection of the 5' untranslated region contiguous with Exon 1 coding sequences.

A 1280 bp PstI/EcoRI fragment extending in the 5' direction from 23 bp within the coding region of Exon 1 was uniformly single-strand labeled and hybridized to RNA isolated from brain (N), slow-twitch anterior latissimus dorsi (S), fast-twitch posterior latissimus dorsi (F), breast (B), leg (L), gizzard (G) and heart (H). After S1-nuclease digestion, the products were separated on acrylamide gels. The locations of marker fragments are as indicated. A 180 bp band was protected using skeletal (S,F,L,B) and smooth (G) RNA.

DISCUSSION

Tissue-Specific Expression of an Alpha-Tropomyosin Gene by Alternative Exon Splicing

The quail alpha-tropomyosin gene consists of 13 exons spaced over approximately 18 kb of genomic DNA. A diagram showing the location and tissue-specificity of each exon is shown in Figure 1. Exons 1, 3, and 5-12 are expressed as the primary alpha-tropomyosin mRNA in quail skeletal muscle as exemplified in leg and breast tissues. Exons 1, 2, 5-10, and 13 are expressed as the primary alpha-tropomyosin mRNA in smooth muscle as exemplified in gizzard tissue. In these arrangements, Exon 1 and Exons 5-10 are constitutively expressed. Exons 2 and 3 are mutually exclusive alternative exons. Exons 11/12 and 13 also act as mutually exclusive alternative exons generating different 3' untranslated sequences with multiple polyadenylation sites. Figure 9 shows a schematic representation of the major transcripts.

The exon arrangement and tissue-specific expression outlined above is identical to that described recently for the rat alpha-tropomyosin gene (19). Intron positions are also extremely conserved, not only to the amino acid, but also to the basepair. Several exon/intron junctions represent

```
-59      aaaagtattgctgtcctaaggaaatcggtcgcccccttgggaagtacatatcaggga      Rat
      ***** ***** ***** **  **** ***** ***** **
-63      CCAAAAAAGTATTGGATGTCCTGAGGAATGCAGTCAGCCCCCTGGGAAAGGCCATATCAGTGA      Quail

      +1
-17      gcagcaggcagctcc..gc...g.ctc..gcact...ccggtctgtccaccctaccgcgtc
      ***  **  **  *  ***  *****  *****
-14      .....GGCTC.TCCCCGCC...GCTCCCGCACTCCACCGGCTCCCGTCCCGCACTCGTCGG

+35      gcgtcxxxctgggtctctctccaagg.....

+41      CCCACGCGCTCCTCGGGCGCGCGCTCGGATTCACTCGGGGCTCTCCGCGCGTTTCCTCG

+61      .....gcctctg.ccaccgccacc atg gac gcc
      ***** ***** **  ***  **
+104     TCTGATTTCGGGTGTACTTTGCACGGGGACGGGCCCTAACCCACCGCGGCC ATG GAT GCC
      Met Asp Ala
```

Figure 8. Comparison of the 5' flanking and 5' untranslated sequences between the quail and rat alpha-tropomyosin genes.

Rat sequences (19) are shown on top in small letters, quail sequences on the bottom in capital letters. Putative CCAAT, TATA, and transcriptional start sequences are underlined, and the putative transcription start site is designated as +1. Asterisks indicate bases which are identical between the two organisms, and dots represent locations where there are no corresponding rat sequences.

split codons. Class O junctions (those between codons) occur following Exons 1, 2, 3, 4, 6, 8, and 9. A Class I junction (following the first nucleotide in a codon) follows Exon 10, and Class II junctions (following the second nucleotide in a codon) follow Exons 5, 7, and 11. The

A. SKELETAL MUSCLE TRANSCRIPT

B. SMOOTH MUSCLE TRANSCRIPT

Figure 9. Splicing pathways involved in the production of smooth and skeletal isoforms from a single alpha-tropomyosin gene.

The common and isotype specific exons are indicated by differently shaded boxes.

corresponding exons of the rat alpha-tropomyosin gene have identical junctions. There does not appear to be a correlation between the type of junction and the constitutive or alternative nature of an exon.

Analysis of the 5' End of the Gene

Figure 8 compares the 5' flanking and 5' untranslated regions of the quail and rat alpha-tropomyosin genes. The unusual promotor sequence CATAT is located 20 bases upstream of the putative quail transcription start site and 22 bases upstream of the putative rat start site. A putative CCAAT sequence, CCAAAA, identical to that found 77 bases upstream of the rat insulin I gene (33), is located 71 bases upstream from the quail start site. A comparison of the 5' flanking sequences of the quail and rat genes reveals a striking 78% conservation of sequences, particularly those sequences between the CCAAT and TATA motifs. Although this conservation extends about 20 bases into the 5' untranslated region of the mRNA and the 20 bases immediately prior to the first codon, the remainder of the 5' untranslated sequences have almost no similarity.

Additional Diversity in the Avian Tropomyosins

There are several important features of the quail alpha-tropomyosin gene which differ from the current information available about the rat gene. The first is the presence of multiple polyadenylation signals in each of the isotype specific 3' terminal exons.

Exons 12 and 13, containing 3' untranslated sequences for skeletal and smooth muscle mRNA respectively, each terminate approximately 12 bp after one of two polyadenylation consensus sequences (AATAAA). cDNA clones have been identified from skeletal muscle mRNA libraries which utilize both poly(A) signals present on Exon 12 (13). Although both of the cDNA clones which contain Exon 13 utilize the first poly(A) site, the small sample size precludes the assignment of this site as the primary, or only, site used in smooth muscle. The fact that Northern blots display two smooth muscle mRNAs differing in size by about 175 bases (Figure 2), the distance between the Exon 13 poly(A) signals, indicates that both sites are likely used. Selective use of poly(A) sites sometimes occurs in a growth regulated or tissue-specific manner (34,35). However, our Northern blot data do not suggest a developmental or tissue-specific regulation of polyadenylation, since both mRNAs are found in roughly equal proportions throughout quail development.

Second, a putative, alternative exon encoding amino acids 189-213 has been identified in the rat alpha-tropomyosin gene (36). A region of almost exact homology to this putative exon is present in the quail sequence shown in Figure 1b beginning at nucleotide 9307, with one significant difference. In the quail sequence, a guanine residue is missing between nucleotides 9348 and 9349 which results in a frame shift to a stop codon (TGA) beginning at nucleotide 9453. Interestingly, a polyadenylation consensus signal (AATAAA) is found 110 nucleotides further downstream. However, no evidence exists indicating that this arrangement is utilized in vivo. In fact, the results presented in figure 5

show that no such RNA is detectable in smooth or skeletal muscle.

Next, we have identified a region of extensive amino acid homology to the published human fibroblast and equine platelet nonmuscle tropomyosin sequences (17,30). We have designated this region Exon 4. There is reason to believe that the exon represents an alternative promoter generating a mRNA coding for a unique amino terminus resulting in a 248 amino acid protein. This would be the first identification of a tropomyosin gene initiated from different promoters and coding for proteins with different amino termini. Preliminary S1-nuclease mapping studies provide evidence that this exon is expressed in brain tissue and contains 5' untranslated sequences as well as coding sequences. Although our sequencing data do not identify promoter-like sequences, many constitutively expressed genes, as may be expected of ubiquitous cytoskeletal genes, do not exhibit consensus CCAAT and TATA sequences.

Sequence comparisons of the derived amino acid sequence of Exon 4 and the human fibroblast and equine platelet tropomyosins are shown in Figure 6. Exon 4 is 64% and 62% homologous at the amino acid level with the human and horse sequences, respectively. Although sequencing of the equine platelet protein does not indicate that the terminal methionine is present (30), the human and quail sequences are of the same length. Replacement of Exons 1 and 2 with Exon 4 would generate an mRNA capable of coding for a 248 amino acid protein. The length of this protein is consistent with the predicted length of the human fibroblast tropomyosin, the length of the equine platelet tropomyosin if the methionine were included, and the estimated length of low molecular weight tropomyosin isoforms detected in chicken brain and intestinal epithelium (32).

The sequences of the 3' ends of human and equine platelet nonmuscle tropomyosins are also shown in Figure 6. These amino acid sequences are 54% and 63% homologous, respectively with the quail smooth muscle specific final 27 amino acids. If similar events occur in the quail and mammals to generate the nonmuscle mRNA from the alpha-skeletal gene, this homology would lead one to predict that mRNAs expressing Exon 4 would also express Exon 13.

Of course, more detailed S1-nuclease analysis and primer extension experiments or isolation and sequencing of full length cDNAs are necessary before any conclusions can be made regarding the initiation and expression of this putative exon.

Recently, it has been shown that the rat alpha-tropomyosin gene is expressed in nonmuscle tissue such as brain but not in liver (36). Although a nonmuscle-specific amino terminal coding exon has not been identified in the rat, the reported lack of the smooth or skeletal exon encoding amino acids 39-80 is consistent with such a model. The rat nonmuscle isoform also appears to express the smooth muscle type 3'-terminal exon, corresponding the quail Exon 13, as predicted.

Finally, in mammals it has been shown that the alpha-cardiac

tropomyosin isoform has the same amino acid sequence as the alpha-skeletal isoform (37). From this observation it has been concluded that these mammalian isoforms are encoded by the same gene. However, in the chicken, the skeletal and cardiac isoforms exhibit electrophoretic and immunologic differences (6). In fact, our laboratory has identified another tropomyosin gene which is expressed predominantly in cardiac tissue. The mRNA from this quail cardiac alpha-tropomyosin gene is also differentially spliced and also contains an exon which shows 66% homology to the Exon 4 presented here. Details of the cardiac gene and its expression will be presented elsewhere (manuscript in preparation).

Interspecies Sequence Homology of Alpha-Tropomyosin

The amino acid sequences of proteins encoded by many multigene families are highly conserved throughout evolution. Tropomyosin is no exception. Several derived or verified amino acid sequences have been published including those for rabbit skeletal muscle tropomyosin (38), chicken skeletal muscle tropomyosin (39), rat smooth and skeletal muscle tropomyosin (19), human alpha-skeletal and fibroblast tropomyosin (40), equine platelet tropomyosin (30), and Drosophila melanogaster thoracic flight muscle tropomyosin (10). Comparisons of several of the amino acid sequences are shown in Figure 6.

The amino acid sequences of the invertebrate Drosophila tropomyosin and the vertebrate rabbit sequence show only a 50 percent homology. However, the proteins are of the same apparent molecular weight and contain 284 amino acids each. The closely related chicken and quail skeletal muscle alpha-tropomyosin sequences differ at only six positions in the coding region at the nucleotide level while the amino acid sequences are identical. Interestingly, even the 3' untranslated sequences are highly homologous showing only 11 nucleotide differences between the chicken and quail. Similarly, the mammals, rat and rabbit, have identical skeletal tropomyosin amino acid sequences. The quail/chicken sequence and the rat/rabbit sequence differ only at 10 amino acid positions. Rat and human sequences differ in 24 amino acids.

Within the smooth muscle-specific sequences, rat and quail exhibit 14 differences in the first alternative exon and only two substitutions in the final 27 amino acids. Similar homologies would be expected to occur between nonmuscle-specific alternative exons as discussed above. Selective pressure (possibly exerted by the functional interactions of tropomyosin with the other contractile proteins) to maintain this extent of homology over such evolutionary distances must be great.

ACKNOWLEDGEMENTS

This work was supported by grant HD18174 from the National Institutes of Health and by grant 1-851 from the March of Dimes Birth Defects Foundation. We thank Sally Berish and Pat Conklin for expert technical assistance and Brian Holloway of the Centers for Disease Control for synthesizing the oligonucleotides used in some experiments.

*To whom correspondence should be addressed

Present addresses: ⁺Rhodes College, 2000 North Parkway, Memphis, TN 38112, [§]Department of Biology, Yale University, New Haven, CT, [¶]Department of Medicine, Duke University, Durham, NC and [°]US Environmental Protection Agency, University of North Carolina, Chapel Hill, NC, USA

REFERENCES

1. Adelstein, R. S. and Eisenberg, E. (1980). *Ann. Rev. Biochem.* 49, 921-956.
2. Perry, S. V. (1979). *Biochem. Soc. Trans.* 7, 593.
3. Taylor, E. W. (1979). *CRC Crit. Rev. Biochem.* 6, 103-164.
4. Cummins, P. and Perry, S. V. (1973). *Biochem. J.* 133, 765-777.
5. Cummins, P. and Perry, S. V. (1973). *Biochem. J.* 141, 43-49.
6. Hyashi, J. I., Ishimida, T., and Hirabayashi, T. (1979). *J. Biochem.* 81, 1487-1495.
7. Segura, M. and Saborio, J. L. (1982). *J. Biol. Chem.* 257, 8867-8871.
8. Chacko, S., Conti, M. A., and Adelstein, R. S. (1977). *Proc. Natl. Acad. Sci. USA* 74, 129-133.
9. Sobieszek, A. and Small, J. V. (1976). *J. Mol. Biol.* 101, 75-92.
10. Basi, G. S., Boardman, M., and Storti, R. V. (1984). *Mol. Cell. Biol.* 4, 2828-2836.
11. Bautch, B. and Storti, R. V. (1983). *Proc. Natl. Acad. Sci. USA* 80, 7123-7127.
12. Karlick, C. C., Mahaffey, J. W., Coutu, M. D., and Fyrberg, E. A. (1984). *Cell* 37, 469-481.
13. Flach, J., Lindquister, G., Berish, S., Hickman, K., and Devlin, R. (1986). *Nucleic Acids Res.* 14, 9193-9211.
14. Ruiz-Opazo, N., Weinberger, J., and Nadal-Ginard, B. (1985) *Nature* 315, 67-70.
15. Reinach, F.C. and MacLeod, A.R. (1986). *Nature* 322, 648-650.
16. Hallauer, P.L., Hastings, K.E.M., Baldwin, A.S., Pearson-White, S., Merrifield, P.A., and Emerson, C.P. Jr., (1987). *J. Biol. Chem.* 262, 3590-3596.
17. MacLeod, A.R., Houlker, C., Reinach, F.C., Smillie, L.B., Talbot, K., Modi, G., and Walsh, F.S. (1985). *Proc. Natl. Acad. Sci. USA* 82, 7835-7839.
18. Helfman, D.M., Cheley, S., Kuusmanen, E., Finn, L.A., and Yamawaki-Katoaka, Y. (1986). *Mol. Cell. Biol.* 6, 3582-3596.
19. Ruiz-Opazo, N. and Nadal-Ginard, B. (1987). *J. Biol. Chem.* 262, 4755-4765.
20. Benton, W.D. and Davis, R.W. (1977). *Science* 196, 180-182.
21. Dale, R.M.K., McClure, B.A., and Houchins, J.P. (1985). *Plasmid* 13, 31-40.
22. Sanger, F., Nicklen, S. and Coulson, A.R. (1977). *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
23. Devlin, R.B., and Haskell, S.G. (1983). *Develop. Biol.* 95, 476-483.
24. Covington, M., Fleenor, D., and Devlin, R.B. (1983). *Nucleic Acids Res.* 12, 4559-4573.

25. Favaloro, J.R., Treisman, R., and Kamen, R. (1980). *Methods in Enzymol.* 68, 718-749.
26. Maxam, A.M. and Gilbert, W. (1980). *Methods in Enzymol.* 65, 499-560.
27. O'Farrell, P. (1981). Bethesda Research Labs, Focus 3, 1.
28. Hu, N. and Messing, J. (1982). *Gene* 17, 271-277.
29. Mount, S. (1982). *Nucleic Acids Res.* 10, 459-472.
- 29.5 Lau, S.T.M., Sanders, C., and Smillie, L.B. (1985). *J. Biol. Chem.* 260, 7257-7263.
30. Lewis, W.G., Cote, G.P., Mak, A.S., and Smillie, L.B. (1983). *FEBS Lett.* 156, 269-272.
31. Kozak, M. (1981). *Nucleic Acids Res.* 9, 5233-5252.
32. Broschat, K.O. and Bruggess, D.R. (1986). *J. Biol. Chem.* 261, 13350-13359.
33. Cordell, B., Bell, G., Tischer, E., DeNoto, F.M., Ullrich, A., Pictet, R., Rutter, W.J., and Goodman, H.M. (1979). *Cell* 18, 533-543.
34. Kaufman, R.J. and Sharp, P.A. (1983). *Mol. Cell. Biol.* 3, 1598-1608.
35. Amara, S.G., Jonas, V., Rosenfeld, M.G., Ong, E.S., and Evans, R.M. (1982). *Nature* 298, 240-244.
36. Wieczorek, D.F., Smith, C.W.J., and Nadal-Ginard, B. (1988). *Mol. Cell. Biol.* 8, 679-694.
37. Lewis, N.G. and Smillie, L.B. (1980). *J. Biol. Chem.* 255, 6854-6859.
38. Stone, D. and Smillie, L.B. (1978). *J. Biol. Chem.* 253, 1137-1148.
39. MacLeod, A.R. (1982). *Eur. J. Biochem.* 126, 293-297.
40. MacLeod, A.R., Houlker, C., Reinach, F.C., and Talbot, K. (1986). *Nucleic Acids Res.* 14, 8413-8426.