The chicken fast skeletal troponin I gene: exon organization and sequence

William Nikovits, Jr., Gerald Kuncio¹ and Charles P. Ordahl*

Department of Anatomy, University of California, San Francisco, CA 94143 and ¹Department of Anatomy, Temple University, School of Medicine, Philadelphia, PA 19140, USA

Received 20 November 1985; Accepted 18 March 1986

ABSTRACT

The gene encoding the fast skeletal isoform of the chick troponin I (sTnI) protein has been sequenced and its organization into exons and introns established. The gene is approximately 4.5 kb in length and composed of 8 exons, the first of which contains solely 5' untranslated sequence. In addition to its major mRNA product, there is evidence that the sTnI gene encodes a second mRNA, present at low abundance levels in embryonic skeletal muscle. S1 nuclease protection and primer extension experiments indicate that the low abundance mRNA is initiated approximately 47 nucleotides upstream of the major transcriptional initiation site. Both mRNAs appear to encode identical sTnI polypeptides. A comparison of nucleotide sequence in the 5' flanking region of several muscle-specific genes, including the sTnI gene, reveals a heptanucleotide consensus sequence, 5'-CATTCCT-3', which is conserved in the 5' flanking regions of many vertebrate contractile protein genes.

INTRODUCTION

One of the major problems in developmental biology is the mechanism by which expression of complex gene sets is orchestrated during embryonic cell differentiation. During striated muscle differentiation, for example, a large number of contractile proteins are incorporated into the sarcomeric myofilaments in precise stoichiometric ratios (1). Moreover, the myofilament sub-structures themselves undergo dynamic, highly ordered, changes throughout development; in which specific myofilament proteins are replaced with similar, but distinct, isoproteins (2,3). The underlying mechanisms responsible for the coordination of contractile protein expression during myogenesis is not well understood. Some contractile protein mRNAs accumulate in a coordinate fashion (4,5) suggesting that common mechanisms may simultaneously govern expression of many genes. Other contractile protein genes, however, appear not to be regulated in a coordinate fashion (6,7,8). Thus, the highly integrated expression of contractile protein gene expression during myogenesis may involve diverse mechanisms operating via both gene-specific, as well as global, mechanisms.

To investigate the mechanisms by which expression of muscle-specific genes is coordinated we have isolated and characterized clones of several chick muscle contractile protein genes which are regulated according to both common and diverse

patterns during myogenesis (9). Sequence analysis of members of the vertebrate actin gene family revealed conserved nucleotide sequence motifs which may be important in the regulated expression of these genes during development (10,11,12). The complete sequence structure of additional muscle-specific genes, particularly in putative regulatory regions, is necessary to allow broader comparison at the nucleotide sequence level as well as to provide characterized reagents for experimental analysis of gene regulation by gene transfer.

Here, we report the sequence/structure of the gene encoding the fast isoform of chick skeletal troponin I (sTnI), a thin filament regulatory protein. The sTnI gene is one member of the troponin I gene family which also includes cardiac and slow skeletal muscle isoforms (13). sTnI is of particular interest, from a gene-regulatory point of view, because in addition to being a component of the thin filament of mature skeletal muscle, it may also serve as an embryo-specific TnI isoprotein in early embryonic cardiac muscle (14). Thus, the sTnI gene may be activated in two striated muscle cell types but its expression modulated very differently at late developmental stages of these two distinct cell types. In embryonic cardiac muscle, the observed down-regulation of the sTnI gene may be triggered by extrinsic signals resulting from neuro-muscular interaction (15). The sTnI gene is likely, therefore, to possess diverse regulatory elements in cis which allow it to respond to different regulatory signals in these two cell types. In addition, our analysis of the nucleotide sequence of the sTnI promoter region reveals short blocks of homology with comparable regions of other cardiac and skeletal contractile protein genes.

METHODS

Recombinant DNA Techniques

Protocols for restriction endonuclease mapping, plaque hybridization, DNA labeling by nick-translation, DNA/RNA blot hybridization, etc. were essentially as described in standardized cloning manuals such as Maniatis et al. (16). A chicken genomic library (17) was screened for the sTnI gene. RNA was isolated as described by Auffray & Rougeon (18). The S1 nuclease digestion (19) and primer extension (20) characterization of the 5' boundary of exon 1 were performed as described elsewhere. The sequence of each exon was determined by the dideoxynucleotide termination method (21) using subclones of genomic DNA in M13 vectors (22). All DNA sequence shown in Figure 4 was confirmed by determination of the sequence of both strands of DNA, by matching computer-generated, conceptual translations of DNA to amino acid sequence with published amino acid sequence, or both. As previously (23) we estimate the error-rate to be 1%, or less.

Direct Sequencing of the Troponin I mRNA

Single-stranded, synthetic oligonucleotides complementary to portions of the troponin I mRNA were prepared by the UCSF Biomolecular Resource Center by means of solid phase-phosphoramidite chemistry on a Beckman system 1-5 DNA Synthesizer. These oligonucleotides (designated BN 1, BN 2, and BN 3; Figure 4) were used as primers for primer extension and mRNA sequencing experiments. Direct sequencing of the troponin I mRNA was performed essentially as previously described (23) except for the following modifications. First, the concentrations of the dideoxynucleotide triphosphates were increased 3-5 fold. Second, the sequencing reactions were not performed using radiolabeled nucleotide triphosphates, instead the primer was labeled with gamma ^{32}P -ATP using polynucleotide kinase (24). Finally, the synthesis reactions were phenol extracted to reduce background during autoradiographic exposure of the sequencing gels.

Figure 1. Map of Troponin I cDNA 98E7

The 98E7 cloned cDNA insert is shown as it is oriented in the pBR322 plasmid, and the 209 bp of 3' untranslated sequence in exon 8 is shown in stippling. The pBR322 sequences are shown as the thin line. The 5' end starts with amino acid 86 (nucleotide 339, Figure 4) in the 6th exon. The insert is 498 nucleotides including 13 nucleotides of poly A tail and the G-C tails used in cloning. The restriction sites shown are: R = Eco RI, V = Pvu II, P = Pst I, B = Bam HI, S = Sal I, H = Hind III, (H) = Hind III destroyed in cloning. The figure is not drawn to scale.

Figure 2. Blot of Embryonic mRNAs probed with Troponin I cDNA 98E7

Each lane contains 5 ug of poly (A) RNA. Lanes: A) day 18 leg muscle, B) day 18 cardiac muscle, and C) day 10 leg muscle. RNAs were electrophoresed and blotted according to standard methods (15). The blot was hybridized with a nick-translated probe, the Eco RI/Hind III fragment of the cDNA 98E7 (see Figure 1). The size markers are a composite of labeled Hae III digests of M13 and ØX174.

RESULTS

Identification of the Fast Skeletal Muscle Troponin I cDNA clone 98E7

The cloning and identification of several cDNAs complementary to muscle-specific mRNAs has been previously reported (9). One such cDNA clone, 98E7 (Figure 1), hybridized to a 950 nucleotide mRNA present in late embryonic skeletal muscle (Figure 2, lane A). This cDNA does not hybridize to RNA from late embryonic heart muscle (figure 2, lane B) nor to RNA from non-muscle tissues (not shown). Determination of the nucleotide sequence of the cDNA 98E7 (Figure 4, nucleotide residues 339 to 836), and conceptual translation into amino acid sequence, revealed a match with the carboxy terminal portion (amino acid residues 85 to 182) of the fast form of chicken skeletal muscle troponin I (25).

Isolation of Troponin I Genomic Clones

Using the cDNA insert of 98E7 as a hybridization probe, four independent phage clones were isolated from a chicken genomic library. The restriction maps of three of the phage clones indicated that they were derived from overlapping regions of the same genomic locus, while that of a fourth was similar but not identical to the others. The dissimilar phage clone was not included in the present study but may represent an allelic

Figure 3. Organization of the Fast Skeletal Troponin I Gene

An approximately 12 kb insert of chicken genomic DNA was cloned into a lambda phage vector designated lambda A. The subcloned Eco RI/Xho I fragment, pBXA1 insert, is expanded in the center figure. Restriction sites and length in kilobases are as shown, with the beginning of the first exon positioned at 0. The position of the exons and introns are shown on the bottom line colinear with the pBXA1 restriction map. Open boxes represent untranslated sequence, solid boxes are coding sequences. The restriction sites are as in Figure 1 with X = Xho I.

form of the troponin I gene, or possibly a related gene. One of the three overlapping clones, designated lambda A and containing 12kb of genomic DNA, was chosen for further analysis. An 8kb Eco RI/Xho I fragment from this phage was subcloned into a modified pBR322 plasmid, containing an Xho I linker replacing the Cla I site. This plasmid, pBXA1 (figure 3), containing the entire troponin I gene, was used for all subsequent analysis of the genomic structure of this gene.

The identity of the pBXA1 genomic clone as encoding fast skeletal TnI was established by its nucleotide sequence homology with the cDNA clone 98E7. Sequences matching 98E7 were found to be contained in three exons (# 6-8), the last of which includes the entire 3' untranslated region of the gene (Figures 3 and 4). Consensus splice sequences (26) were located at these, and all other, exon boundaries (not shown) and a putative poly (A) addition sequence, AAATAA, is found 18 nucleotides upstream from the site of poly (A) addition.

There are three nucleotide differences found between the genomic and cDNA clones (see legend to Figure 4), none of which affects the amino acid sequence of the encoded protein. Two of these changes are in the 3' untranslated sequence, while the third is a silent base change at position 590. These differences between cDNA clone 98E7 and genomic clone pBAX1 may be due to allelic variation since the cDNA and genomic libraries were constructed using material from different flocks.

Determination of Amino Acid Encoding Exons

Determination of the remainder of the amino acid encoding exons, 2 - 6 (filled boxes in Figure 3), was made by comparing conceptual translations of pBXA1 nucleotide sequence to the published amino acid sequence for the chicken troponin I protein (25). These translations yielded blocks of perfect homology which allowed unambiguous

```

-150                                -100
TGCAGGGGCTGCCTGCCCGCCAGGAGGTGCATGGGGCTCTTCACACCATTTGGTCTGAGAAGGTCTCCAT
*****

-50*                                ☆
GCCCTTAAAGAGAAGGTGGATTTTCTGCTGAGCCACCGAGAAGCCCTTTTATACCTCTCTCCCGGCAGCTGCTTGGC
*****

1|2
GCTTGCCAGTCAGTCTCTGGCGGCCAGGGGGCTCTCTGTTCTCTCTCGGACTCGGGTTCACAACCATCTAAAGCAAG
BN 3

2|3    3|4                                4|5
-1 MET SER ASP GLU GLU LYS LYS ARG ARG ALA ALA THR ALA ARG ARG GLN HIS LEU LYS SER
84 ATG TCT GAT GAA GAG AAA AAG AGG AGG GCA GCC ACC GCC CGG CGG CAG CAC CTG AAG AGT
BN 2                                BN 1

20 ALA MET LEU GLN LEU ALA VAL THR GLU ILE GLU LYS GLU ALA ALA ALA LYS GLU VAL GLU
144 GCT ATG CTC CAG CTT GCT GTC ACT GAA ATA GAA AAA GAA GCA GCT GCT AAA GAA GTG GAA

40 LYS GLN ASN TYR LEU ALA GLU HIS CYS PRO PRO LEU SER LEU PRO GLY SER MET GLN GLU
204 AAG CAA AAC TAC CTG GCA GAG CAT TGC CCT OCT CTG TCC CTC CCA GGA TOC ATG CAG GAA

5|6
60 LEU GLN GLU LEU CYS LYS LYS LEU HIS ALA LYS ILE ASP SER VAL ASP GLU GLU ARG TYR
264 CTT CAG GAA CTG TGC AAA AAG CTT CAT GCC AAG ATA GAC TCA GTG GAT GAG GAA AGG TAT

D                                6|7
80 ASP THR GLU VAL LYS LEU GLN LYS THR ASN LYS GLU LEU GLU ASP LEU SER GLN LYS LEU
324 GAC ACA GAG GTG AAG CTA CAG AAG ACT AAC AAG GAG CTG GAG GAC CTG AGC CAG AAG CTG

100 PHE ASP LEU ARG GLY LYS PHE LYS ARG PRO PRO LEU ARG ARG VAL ARG MET SER ALA ASP
384 TTT GAC CTG AGG GGC AAG TTC AAG AGG CCA CCT CTG CGC CGG GTG CGC ATG TCT GCT GAT

120 ALA MET LEU ARG ALA LEU LEU GLY SER LYS HIS LYS VAL ASN MET ASP LEU ARG ALA ASN
444 GCC ATG CTG OCT GCC CTG CTG GGC TCC AAG CAC AAG GTC AAC ATG GAC CTC CGG GCC AAC

7|8
140 LEU LYS GLN VAL LYS LYS GLU ASP THR GLU LYS GLU LYS ASP LEU ARG ASP VAL GLY ASP
504 CTG AAG CAA GTC AAG AAG GAG GAC ACG GAG AAG GAG AAG GAC CTC CGC GAT GTG GGT GAC

160 TRP ARG LYS ASN ILE GLU GLU LYS SER GLY MET GLU GLY ARG LYS LYS MET PHE GLU ALA
564 TGG AGG AAG AAC ATT GAG GAG AAA TCC GGC ATG GAG GGC AGG AAG AAG ATG TTT GAG GCC

180 GLY GLU SER ### GCACCTGGTCTCTCCACTCTTGCCACTACCAACCCTCTTCCATCCCTCTGAGCATGGCCACAG
624 GGC GAG TCC TAA

699 CTGTGATCATGGCCACCTCCCTGCCCTGAACCTCAACATGCTGTCACCATGATTGAACCACTGACTGGGTGCTGTTTC

779 TCTGATTTTGTGAAGAGCTGCAGTCTGGAGCGCAGTGTAAATAAAGCTTTTCATGGGAGAGGGGGATGTGGCTCGCGT

859 GGTGGGGCTGAGGGTGCCTTAGGGCTGTGGGAACACACTGAGGACA

```

Figure 4. Sequence of the Fast Skeletal Troponin I

The Troponin I DNA (mRNA) nucleotide sequence is shown with additional 5' and 3' genomic nucleotide sequence. The predominate transcriptional start site is marked with a star and the putative minor start site with an asterisk. The conceptual translation of the cDNA nucleotide sequence into amino acid sequence is shown above. The putative promoter elements, TATA, CCATT, and CCAT are underscored, as are the possible TATA-like sequence, TTAAA, and the poly (A) addition site, AATAAA. Sequences complementary to the oligonucleotide primers BN 1, BN 2, and BN 3 are indicated. The exon boundaries are marked by vertical lines with the exon number on either side. The three nucleotide differences found between the cDNA 98E7 and the genomic sequence are indicated by closed circles. The 5' limit of the sequence contained in cDNA clone 98E7 is indicated by the open arrowhead at nucleotide 339, and the position of poly (A) addition in the cDNA is indicated with a closed arrowhead between nucleotides 836 and 837. The CATTCCCT consensus sequence, and homologies to the cardiac troponin T gene are overscored.

assignment of exon/intron boundaries (Figure 4). However, the extremely small size of exon 3 (7 nucleotides), and the small amount of amino acid coding region in exon 2 (8 nucleotides), necessitated confirmation of their identity and nucleotide sequence by direct sequencing of the mRNA. An oligonucleotide sequencing primer (BN 1, figure 4), complementary to the mRNA in the region of exon 4 (nucleotides 123 to 140), allowed the determination of the mRNA sequence to within approximately 40 nucleotides of the mRNA 5' end (data not shown). The homology between mRNA sequence and the genomic DNA sequence completely accounted for all exonic material and established the 5' and 3' boundaries of exons 2 and 3, as well as delimiting the 3' boundary of exon 1.

Determination of Untranslated Regions of Exons 1 and 2

Exons 1 and 2 contain predominately 5' untranslated sequence of the sTnI mRNA (Figure 4) and, therefore, amino acid sequence homology could not be used to identify these exons. To establish the nucleotide sequence of these two exons, mRNA sequence was again directly determined by primer extension sequencing. First, an oligonucleotide (primer BN 2, Figure 4), with sequence complementary to the amino acid coding portions of exons 2 and 3 (nucleotides 79 to 98), was used to establish the sequence of the untranslated region of exon 2 and the 3' region of exon 1 (data not shown). A second oligonucleotide (primer BN 3, Figure 4), complementary to the 3' portion of exon 1 (nucleotides 46 to 63), was then used to determine the nucleotide sequence of the 5' portion of exon 1 and the position of the start of transcription (Figure 5A).

Independent confirmation of the position of the 5' boundary of exon 1 was obtained from S1 nuclease protection experiments. An isolated 100 bp Ava I genomic fragment (Figure 5C), spanning the putative cap site, was labeled at its 5' end with polynucleotide kinase and annealed to poly(A) RNA from embryonic day 18 leg muscle. A 60 nucleotide fragment is protected from S1 nuclease digestion (Figure 5B, lane 2) which positions the sTnI transcriptional initiation site to the same nucleotides as determined by direct mRNA sequencing (Figure 5A). No protection was observed when poly(A) RNA from embryonic day 10 brain tissue was used (Figure 5B, lane 3).

In addition to the major transcriptional start site indicated above (star in Figure 4), a minority of troponin I message in embryonic day 18 leg muscle may initiate transcription at a second upstream site. A minor band is observed on longer exposure of the gel of the S1 nuclease protection experiment (Figure 5B, lanes 2 and 4) which maps to nucleotides -47 thru -49 (asterisk in Figure 4) upstream from the main cap site. Primer extension experiments (as diagrammed in Figure 5C) also suggest the existence of a longer minor transcript (Figure 5A and additional data not shown) mapping to the same position determined by S1 nuclease experiments. It will be necessary to perform functional tests of the ability of this putative distal promoter to support transcription via gene fusion experiments.

Figure 5. Start of Transcription of the Troponin I Gene
Panel A. Single-stranded genomic DNA cloned in M-13 Vectors and poly (A) RNA were sequenced in parallel using the oligonucleotide BN 3 as a primer. On this gel, strong stops in the mRNA sequence prevent reading some nucleotides. These nucleotides were discernable on other films. The mRNA sequence positions the primary start of transcription (starred) at the T as indicated in Figure 4. The asterisk marks the approximate position in the genomic sequence identified by S1 nuclease digestion to be the putative transcriptional start position for the longer, low abundance message.
Panel B. The start of transcription is positioned by S1 nuclease mapping using the exon 1

specific 100 bp *Ava* I fragment as shown in panel C. The primary transcript maps to the same group of nucleotides as in panel A (TTG). The longer, low abundance transcription product (*) is clearly evident and maps to the nucleotides -47 to -49, relative to the major cap site. Lane 1 - kinase-labelled *Ava* I fragment used as hybridization probe. Lane 2 - S1 treated sample protected with RNA from embryonic day 18 leg muscle. Lane 3 - S1 treated sample protected with RNA from embryonic day 10 brain. Lane 4 - longer exposure of lane 2 to enhance the signal from the low abundance transcription product.

Panel C. Primer extension sequencing and S1 nuclease mapping methods are graphically represented. Primer BN 3, labelled at the 5' end, was hybridized to either mRNA or single-stranded DNA from the 3' region of exon 1, and extended. For S1 protection experiments, the 100 bp *Ava* I fragment shown on the bottom line, was kinase-labelled at the 5' end and used as described in the text.

5' Flanking Sequence of the Troponin I Gene

Sequences related to the canonical transcriptional promoter sequences, CAAT and TATA (27), are located 5' to the start of transcription (Figure 4). The sequence, TATA, is located at position -26, and there are at least two candidates for CAAT sequences: CCAT at -80 and CCATT at -100. While neither CAAT-like sequence perfectly fits the canonical CAAT homology, non-canonical CAAT boxes may be a feature of muscle-specific genes, since cardiac troponin T (19) and the alpha-actin genes (23,28) also have non-canonical CAAT sequences.

A second TATA-like sequence (TTAAA) is located at position -72 (Figure 4). Two observations are consistent with the notion that this may serve as the TATA motif for the putative distal transcription initiation site. First, it lies 23-25 nucleotides upstream from the distal initiation site; and, second it matches the TATA motif of the cardiac troponin T gene (19).

DISCUSSION

Exon Organization of the Troponin I Gene

The sequence organization of the chick fast skeletal troponin I (sTnI) gene (Figures 3 and 4) was determined by comparing the nucleotide sequence of the sTnI genomic clone (pBXA1) to; i. nucleotide sequence of the cDNA clone, 98E7; ii. published amino acid sequence of chick sTnI protein (25); and, iii. direct sequencing of the sTnI mRNA by primer extension. The amino acid sequence of sTnI, as determined from the nucleotide sequence of the gene, precisely matches its published protein sequence, except for the presumptive initiator methionine residue which is absent from the mature sTnI protein. Since sTnI, like most contractile proteins, has a posttranslationally modified N-terminus (25), it is likely that this methionine residue is removed during the modification process.

The sTnI gene is organized into 8 exons spanning approximately 4.5 kb of genomic DNA. The first exon is composed entirely of untranslated sequence similar to other vertebrate contractile protein genes (19,23,28,29, and see 30 for additional references).

The second exon contains untranslated sequence plus the methionine initiator codon and 5 of the 6 nucleotides encoding the first two amino acids of the mature sTnI protein. The third exon is only 7 nucleotides in length making it one of the shortest exons discovered to date (see Cooper & Ordahl (19) for discussion of this point).

The troponin complex is composed of three subunits (I, C and T) and regulates the interaction between myosin and actin during muscle contraction. These regulatory subunits are products of three apparently unrelated gene families. Troponin I binds to actin, thereby blocking myosin-actin interaction in relaxed muscle (31). The actin binding domain of sTnI has been localized to amino acid residues 105-114 (32), which are encoded within an internal region of the large exon #7 (amino acid residues 92-152). During contraction, sTnI-actin interaction is disrupted allosterically by an interaction between troponin C and the actin binding site of sTnI (33). A second region of sTnI (residues 1-40) (34) which interacts with troponin C is encoded in 4 short exons (#2-5, Figure 4). If the troponin C binding region can be defined more precisely it may be possible to correlate it to a specific exon(s). The extremely small amount of amino acid coding in exons 2 and 3 suggest, however, that the troponin C binding domain is likely to reside in exon 4 and/or 5. sTnI also has binding sites for troponin T (35), but that region has not yet been localized within the sTnI protein. Unlike troponin T, the TnI protein shows no sequence variability, and we have found no evidence of alternative splicing either by S1 nuclease mapping or primer extension analysis.

It will be interesting to see whether other TnI isogenes share exonic organization similar to that of the chick sTnI gene. While the exon organization of many gene families is highly conserved, this has not been the rule for contractile protein genes. For example, the cytoplasmic and muscle actin genes differ substantially in exon organization (23) as do the skeletal and cardiac troponin T genes (19).

5' Flanking and Putative Promoter Region

We have determined the position of the transcriptional initiation for the sTnI gene both by primer extension run-off as well as by S1 protection (Figure 5). By these criteria, there are two transcription initiation sites; a proximal site whose usage in skeletal muscle vastly predominates over a second, distal site 47-49 nucleotides upstream from the proximal site (Figure 4). Both sites have TATA-like sequences 26 and 23-25 nucleotides upstream from the proximal and distal transcription initiation sites, respectively, suggesting the existence of two separate promoter/initiation sites for this gene.

There are two CAAT-like sequences, 100 and 80 nucleotides upstream from the major (proximal) transcriptional initiation site (Figure 4). These positions correspond reasonably well with those typically found for CAAT homologies in other genes but neither fits the canonical CAAT sequence perfectly. The putative CAAT motifs of other

TABLE 1.
Homology in the 5' Flanking Sequences of Muscle-Specific Genes

<u>GENE</u>	<u>SEQUENCE</u>	<u>POSITION</u> (from cap)	<u>REFERENCE</u>
Chick skeletal a-actin	C A T T C C T	-63	(23)
Rat skeletal a-actin	C A T T C C T	-65	(28)
Chick cardiac troponin T	C A T T C C T	-96 & -73	(19)
Chick cardiac a-actin	C A T T C a T C A T T C C g	-117 -261	(30)
Chick skeletal troponin I	C A T T g g T	-101	(this paper)
Quail skeletal troponin I	C A T T g g T	-100	(47)
Rat cardiac a-MHC	C A a a C C T	-172	(41)
Chick (type 5) cytoplasmic actin	C A T T C C g $\nearrow$	-160	(42)
Chick cytoplasmic b-actin	C A T T C C c $\nearrow$	-240	(43)
Chick skeletal MLC1	C A a T C C a	-72	(44)
Chick skeletal MLC3	C A a T C C a $\nearrow$	-60	
Mouse skeletal MLC1	C A a T C C a	-73	(45)
Mouse skeletal MLC3	a A c T C C T	-79	
Rat skeletal MLC1	C A a T C C a	-74	(46)
Rat skeletal MLC3	C A a T g C c $\nearrow$	-79	
Consensus Sequence:	C A T T C C T		

muscle-specific genes also show divergence from the canonical CAAT homology (19,23,28), and multiple CAAT-like sequences have also been observed in the cardiac troponin T gene (19), and cardiac a-actin gene (29). It remains to be determined whether these two CAAT-like motifs in the sTnI gene serve either or both putative transcriptional initiation sites.

Comparison of Troponin I Sequence to Other Muscle-Specific Genes

A comparison of the promoter regions of the skeletal alpha actin genes of chick and rat has revealed conserved sequences which might be involved in the tissue- and stage-specific expression of these genes (10,11). It is not yet clear, however, which, if any, of those conserved sequence motifs might be general elements involved in the

expression of other muscle-specific genes, because only partial homology with these sequence elements has been found in other muscle-specific genes (29, and C. Emerson, personal communication).

Several homologies were noted between putative regulatory regions of the sTnI gene and similar regions of the cardiac troponin T (cTnT) gene (19,36). First, the TATA motif of the putative distal sTnI promoter (TTAAA; underscored in Figure 4) matches the TATA motif of the cTnT gene. Second, 11 of the 15 nucleotides immediately downstream of the proximal sTnI TATA motif (overscored in Figure 4) match the same region of the cTnT promoter. Finally, the sTnI poly(A) addition signal is imbedded within a longer sequence, 5'-TGTAATAAG-3', which matches perfectly the poly(A) addition sequence of the cTnT gene.

We have also found a consensus heptanucleotide sequence (5'-CATTCCT-3') in the 5' flanking regions of a number of different contractile protein genes from several species (Table I). The position of this consensus sequence was found to be somewhat variable (60-261 nucleotides upstream from the transcription initiation site), but in the majority of cases resided less than 120 nucleotides upstream from the transcription initiation sites of the genes. The CATTCCT consensus may, in fact, serve as the CAAT motif in several of the genes noted.

Although this consensus sequence is short, and imperfectly conserved in the various genes; short, imperfectly conserved sequence motifs have been implicated as being the active sequence components of viral and cellular promoter elements (37,38,39). It is possible, therefore, that a short consensus sequence element, such as we describe here, is important in the regulated expression of contractile protein genes indicated in Table I. This sequence cannot, however, be considered to be absolutely limited to muscle-specific genes because the 5' flanking regions of two cytoskeletal actin genes were also found to possess a variation on the consensus sequence far upstream of their transcription initiation sites. Since the cytoskeletal and contractile actin genes are both related and strongly conserved at the nucleotide sequence level, it is possible that conservation of this sequence motif reflects that conservation.

Cell culture systems which permit transfer of cloned genes into skeletal and cardiac myocytes in primary culture should allow functional analysis of such potential regulatory sequences (40). In addition, the ability of cultured cardiac myocytes to down-regulate activity of the sTnI gene in response to extrinsic signals (15) may offer an experimental system for analyzing the cis elements involved in modulation of sTnI gene expression.

ACKNOWLEDGEMENTS

We would like to thank Dr. Charles Emerson and Albert Baldwin for communicating results on the sequence of the Quail skeletal troponin I gene prior to

publication. That sequence was published (ref. 47) while this manuscript was in review. We also thank Dr. Charles Craik and Ms. Jennifer Barnett for synthesizing the oligonucleotide primers, and Drs. Thomas Cooper and Janet Mar for helpful suggestions in preparation of this manuscript. This work was supported by NIH grants GM-32018 and HL 35561 to CPO.

*To whom correspondence should be addressed

REFERENCES

1. Potter, J.D. (1974) Arch. Biochem. Biophys. 162, 436-441.
2. Fischman, D.A. (1982) in Muscle Development: Molecular and Cellular Control, Pearson, M.L. and Epstein, H.F. Eds., pp. 397-404, Cold Spring Harbor Laboratory, New York.
3. Caplan, A.L., Fisman, M.Y. and Eppenberger, H.M. (1983) Science 221, 921-927.
4. Devlin, R.B. and Emerson, C.P. Jr (1978) Cell 13, 599-611.
5. Hastings, K.E.M. and Emerson, C.P. Jr (1982) Proc. Natl. Acad. Sci. USA 79, 1553-1557.
6. Buckingham, M.E., Alonso, S., Barton, P., Bugaisky, G., Cohen, A., Daubas, P., Garner, I., Minty, A., Robert, B. and Weydert, A. (1984) in Molecular Biology of Development, Davidson, E.H. and Firtel, R.A. Eds., pp. 275-292, Alan R. Liss, Inc., New York.
7. Weydert, A., Daubas, P., Caravatti, M., Minty, A., Bugaisky, G., Cohen, A., Robert, B. and Buckingham, M. (1983) J. Biol. Chem. 258, 13867-13874.
8. Wicczorek, D.F., Periasamy, M., Butler-Browne, G.S., Whalen, R.G. and Nadal-Ginard, B. (1985) J. Cell Biol. 101, 618-629.
9. Ordahl, C.P., Cooper, T., Ovitt, C.E., Fornwald, J.A., Calman, A.F. and Caplan, A.L. (1982) in Muscle Development: Molecular and Cellular Control, Pearson, M.L. and Epstein, H.F., Eds., pp. 225-236, Cold Spring Harbor Laboratory, New York.
10. Ordahl, C.P. and Cooper, T.A. (1983) Nature 303, 348-349.
11. Nudel, U., Greenberg, D., Ordahl, C.P., Saxel, O., Neuman, S. and Yaffe, D. (1985) Proc. Natl. Acad. Sci. USA 82, 3106-3109.
12. Yaffe, D., Nudel, U., Mayer, Y. and Neuman, S. (1985) Nucleic Acids Res. 13, 3723-3737.
13. Wilkinson, J.M. and Grand, R.J.A. (1978) Nature 271, 3-35.
14. Toyota, N. and Shimada, Y. (1981) J. Cell Biol. 91, 497-504.
15. Toyota, N. and Shimada, Y. (1983) Cell 33, 297-304.
16. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, New York.
17. Dodgson, J.B., Strommer, J. and Engel, J.D. (1979) Cell 17, 879-887.
18. Auffray, C. and Rougeon, F. (1980) Eur. J. Biochem. 107, 303-314.
19. Cooper, T.A. and Ordahl, C.P. (1985) J. Biol. Chem. 260, 11140-11148.
20. McKnight, S.L., Gavis, E.R. and Kingsbury, R. (1981) Cell 25, 385-398.
21. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
22. Messing, J. (1983) Methods in Enzymol. 101, 20-78.
23. Fornwald, J.A., Kuncio, G., Peng, I. and Ordahl, C.P. (1982) Nucleic Acids Res. 10, 3861-3876.
24. Chaconas, G. and van de Sande, J.H. (1980) Methods in Enzymol. 65, 75-85.
25. Wilkinson, J.M. and Grand, R.J.A. (1978) Eur. J. Biochem. 82, 493-501.
26. Mount, S.M. (1982) Nucleic Acids Res. 10, 459-472.
27. Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383.
28. Zakut, R., Shani, M., Givol, D., Neuman, S., Yaffe, D. and Nudel, U. (1982) Nature 298, 857-859.

29. Chang, K.S., Rothblum K.N. and Schwartz R.J. (1985) *Nucleic Acids Res.* 13, 1223-1237.
30. Buckingham, M.E. and Minty, A.J. (1983) in *Eucaryotic Genes: Their Structure, Activity, and Regulation*, McLean, N., Gregory, S.P. and Flavell, R.A., Eds., pp. 365-395, Butterworths, London.
31. Potter, J.D. and Gergely, J. (1974) *Biochemistry* 13, 2697-2703.
32. Talbot, J.A. and Hodges, R.S. (1981) *J. Biol. Chem.* 256, 2798-2802.
33. Cachia, P.J., Sykes, B.D. and Hodges, R.S. (1983) *Biochemistry* 22, 4145-4152.
34. Syska, H., Wilkinson, J.M., Grand, R.J.A. and Perry, S.V. (1976) *Biochem. J.* 153, 375-387.
35. Pearlstone, J.R. and Smillie, L.B. (1980) *Can. J. Biochem.* 58, 649-654.
36. Cooper, T.A. and Ordahl, C.P. (1984) *Science* 226, 979-982.
37. Davidson, E.H., Jacobs, H.T. and Britten, R.J. (1983) *Nature* 301, 468-470.
38. Dynan, W.S. and Tjian, R. (1985) *Nature* 316, 774-778.
39. Payvar, F., DeFranco, D., Firestone, G.L., Edgar, B., Wrange, O., Okret, S., Gustafsson, J.-A. and Yamamoto, K.R. (1983) *Cell* 35, 381-392.
40. Ordahl, C.P., Mar, J., Clyne, J. and Simpson, P. (1985) in *Molecular Biology of Muscle Development*, Vol. 29., Emerson, C., Fischman, D.A., Nadal-Ginard, B. and Siddiqui, M.A.Q., Eds., Alan R. Liss, Inc., New York. In press.
41. Mahdavi, V., Chambers, A.P. and Nadal-Ginard, B. (1984) *Proc. Natl. Acad. Sci. USA* 81, 2626-2630.
42. Bergsma, D.J., Chang, K.S. and Schwartz, R.J. (1985) *Molec. Cell. Biol.* 5, 1151-1162.
43. Kost, T.A., Theodorakis, N. and Hughes, S.H. (1983) *Nucleic Acids Res.* 11, 8287-8301.
44. Nabeshima, Y., Fujii-Kuriyama, Y., Muramatsu, M. and Ogata, K. (1984) *Nature* 308, 333-338.
45. Robert, B., Daubas, P., Akimenko, M., Cohen, A., Garner, I., Guenet, J. and Buckingham, M. (1984) *Cell* 39, 129-140.
46. Periasamy, M., Strehler, E.E., Garfinkel, L.I., Gubits, R.M., Ruiz-Opazo, N. and Nadal-Ginard, B. (1984) *J. Biol. Chem.* 259, 13595-13604.
47. Baldwin, A.S. Jr, Kittler, E.L.W. and Emerson, C.P. Jr (1985) *Proc. Natl. Acad. Sci. USA* 82, 8080-8084.