cDNA clone analysis of six co-regulated mRNAs encoding skeletal muscle contractile proteins
(coordinate gene regulation/myogenesis)

KENNETH E. M. HASTINGS AND CHARLES P. EMERSON, JR.

Department of Biology, Culver Hall, University of Virginia, Charlottesville, Virginia 22901

Communicated by O. L. Miller, Jr., November 12, 1981

ABSTRACT A cDNA cloning approach was used to investigate muscle gene regulation during differentiation of cultured embryonic quail myoblasts. A cDNA clone library of cultured myofiber poly(A)⁺ RNA was constructed and screened by colony hybridization with cDNA probes of myoblast and myofiber RNA. Twenty-eight myofiber-specific cDNA clones were identified and, by cross-hybridization analysis, these clones were found to represent, at most, 18 different myofiber-specific RNAs. Six of these RNAs were identified by sequence analysis of the cDNA clones. These six RNAs encode the contractile proteins α-actin, α-tropomyosin, myosin heavy chain, myosin light chain 2, troponin C, and troponin I. The embryonic muscle contractile protein sequences are identical with, or closely match, those of adult skeletal muscle proteins and include both fast fiber (myosin light chain 2 and troponin I) and slow fiber (troponin C) isotypes. RNA gel transfer hybridization analysis showed that the cellular abundances of these contractile protein mRNAs increase 20- to 30-fold or more during myoblast differentiation. These findings indicate that coordinate activation of contractile protein synthesis during myogenesis is controlled by mechanisms that direct the accumulation of contractile protein mRNAs rather than their translational utilization. Furthermore, with the possible exception of myosin heavy chain, the contractile protein genes expressed by cultured embryonic muscle encode adult muscle proteins of both fast and slow fiber types, consistent with a co-activation-sequence selective regression model of gene regulation during fiber type differentiation in developing skeletal muscle.

The molecular mechanisms that coordinate the expression of tissue-specific gene sets pose an unsolved but fundamental problem of eukaryotic gene regulation (1). Much recent work on this problem has been devoted to the study of gene families—i.e., sets of sequence-related genes that apparently share a common ancestor, such as the globins (2), the histones (3), and insect egg chorion proteins (4). An alternative approach, which is likely to generate complementary information, is to study the coordinate regulation of apparently unrelated genes that encode functional sets of proteins. An excellent gene system for this approach is the set of genes encoding the muscle-specific contractile proteins.

The major contractile proteins, actin, myosin, tropomyosin, and troponin, are incorporated in a defined stoichiometry into highly organized myofibrils (5). The role of each of these proteins in the process of contraction is understood, their subunit compositions are known, and a great deal of amino acid sequence information is available. Muscle differentiation and contractile protein synthesis also have been studied extensively. Dividing skeletal muscle myoblasts in culture synthesize extremely small quantities, if any, of the muscle-specific contractile proteins (6-10). Yet when myoblasts withdraw from the cell cycle and fuse to form multinucleate myofibers, the synthesis of these proteins is coordinately activated to high levels (6-12). This phenomenon identifies a set of coordinately regulated genes encoding a variety of structurally diverse, but functionally related, polypeptides.

It became clear to us and other investigators (13-15) that further study of the mechanisms coordinating muscle gene expression requires the production of gene-specific hybridization probes by recombinant DNA cloning. Here we report the isolation, identification, and analysis of cloned cDNA copies of six contractile protein mRNAs. The cDNA clones in this set encode α-actin, α-tropomyosin, myosin heavy chain, myosin light chain 2, troponin C, and troponin I. In addition to preparing gene-specific probes for studies of the structure and coordinate expression of these co-regulated genes, our results provide structural information on the contractile protein genes expressed in embryonic muscle and on the mechanisms regulating contractile protein synthesis during myogenesis.

MATERIALS AND METHODS

Muscle Cultures and RNA Extraction. Myoblasts were obtained from dispersed breast muscle of day 10 Japanese quail embryos (Coturnix coturnix) and cultured as described (6, 16). Secondary cultures were fed fresh medium 12-24 hr after plating. Myoblast RNA was extracted from cultures 48 hr after plating, during which time myoblasts undergo rapid cell division. Myoblast fusion commences ~60 hr after plating, and by 120 hr, 80-90% of the nuclei are found in multinucleated myofibers (6, 16). Myofiber RNA was extracted on the fifth day (120 hr) after plating. Whole cell RNA was extracted in a procedure involving lysis with NaDodSO₄, digestion with Pronase, phenol/chloroform extraction, and precipitation with 2 M LiCl (17). Poly(A)⁺ RNA was isolated by chromatography of whole cell RNA on oligo(dT)-cellulose (18).

cDNA Cloning, Colony Hybridization, and Plasmid Preparation. Double-stranded cDNA was synthesized from myofiber poly(A)⁺ RNA and cloned by G-C tailing into the Pst I site of plasmid pBR322 and by transforming into Escherichia coli strain χ1776 (19, 20). Tetracycline-resistant transformants were analyzed by differential colony hybridization (21) using [32P]cDNAs synthesized from myoblast and myofiber poly(A)⁺ RNA templates. These [32P]cDNA probes were made under the same conditions as the first cDNA strand for cloning, except that the reaction mixture contained 1.2 μCi of [α-32P]dCTP/ml (200-300 Ci/mmol; 1 Ci = 3.7 × 10¹⁰ becquerels; Amersham). Of 890 colonies screened, 28 contained myofiber-specific sequences; i.e., they hybridized much more with myofiber [32P]cDNA than with myoblast [32P]cDNA. Recombinant plasmids were isolated from liquid cultures of transformant clones (22) by isopycnic banding (23) of cleared lysates (24). Plasmid preparations were phenol extracted and stored at

Abbreviations: bp, base pair(s); kb, kilobase(s).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U. S. C. §1734 solely to indicate this fact.
Authentic contractile proteins were analyzed by precipitated muscle (by three precipitations from various other vectors DNAs) then fluorographed. In every case, Pst I restriction enzymes sites were digested and recombinant plasmids 32P-labeled with polynucleotide kinase (28). The labeled DNA was then digested with Pst I or another appropriate enzyme and the resulting two insert fragments were resolved by gel electrophoresis and analyzed by the method of Maxam and Gilbert (28).

**DNA Sequence Analysis Strategy.** To determine restriction sites suitable for DNA sequence analysis, purified recombinant plasmid DNAs were digested with Pst I in combination with various other restriction enzymes (Bethesda Research Laboratories). In every case, Pst I released the cDNA insert from the plasmid vector DNA so that restriction sites in any cDNA sequence could be established by gel electrophoresis. The S' termini at internal insert restriction sites were 32P-labeled by the method of Maxam and Gilbert (28).

**Gel Transfer Hybridization.** RNA samples were subjected to electrophoresis in a 1.2% agarose/5 mM methylmercuric hydroxide gel (29) and transferred to diazotized paper (30). Diazophenylthioether paper (B. Seed; see ref. 31) was used instead of diazobenzoylomethyl paper. Paper-bound RNA was hybridized (32) with recombinant plasmids 32P-labeled by nick-translation (33).

**RESULTS**

**Identification of Contractile Protein cDNA Clones.** Rather than attempting to purify and clone individual contractile protein mRNAs, we elected to shotgun clone the entire population of poly(A)+ mRNAs present in differentiated myofiber cultures. We initially identified developmentally regulated sequences by screening the myofiber cDNA clone library by colony hybridization using myoblast 32P-labeled DNA and myofiber 32P-labeled DNA probes. Of 890 randomly chosen transformant colonies analyzed by this method, 28 clearly hybridized to a greater extent with the myofiber probe than with the myoblast probe. These colonies therefore harbored recombinant plasmids containing cDNA copies of RNA species whose abundances increased substantially during myogenesis. For convenience, we refer to these 28 cloned sequences as myofiber-specific cDNAs.

The myofiber-specific set of 28 cDNA clones did not represent 28 different RNA species. Analysis by DNA dot hybridization (34) and restriction mapping indicated that a maximum of 17 or 18 different RNA species were represented, several by two or three independently cloned cDNAs. By hybridization translation and DNA sequence analyses of myofiber-specific cDNAs, we have identified cloned partial copies of six different contractile protein mRNAs, as summarized below.

**α-Tropomyosin:** Two cross-hybridizing cDNA clones, cC101 and cC102, yielded hybridization-translation products that comigrated on two-dimensional gels with quail skeletal muscle α-tropomyosin. The sequence of a region of the 520-base pair (bp) cDNA insert of cC101 decoded into what is, with one exception, the amino acid sequence of rabbit skeletal muscle α-tropomyosin (35) at residues 134–192 (Fig. 1). The sequence of a region of the 580-bp cDNA insert of cC102 was identical with that determined for cC101, including the codon for residue 157, which is aspartate in the rabbit protein but glutamate in the decoded quail DNA sequences. Thus, this one conservative amino acid substitution is not due to an error in cloning or DNA sequence analysis but is presumably a primary structure difference between homologous quail and rabbit proteins. Clearly, the cDNA inserts of cC101 and cC102 encode α-tropomyosin and not the most closely related known protein, β-tropomyosin, since the latter differs at residues 158, 172, 174, 175, and 179, among others (41).

**cC101, cC102**...GCT AAG CAC ATT GCT GAA GAG GCT GAC GCC AAG TAT GAA GAG GTC GCT GCT AAG CTT GTC ATC ATT GAG AGT GAC... 

**α-tropomyosin**...ala lys his ile Ala glu GLU ala asp arg lys tyr glu glu val ala arg lys leu val ile ile glu ser asp... 

160 170

**cC118**...GCC TGC CCT GAA ACT TTA TTC CAC ACC ATT GCT GAA TCT GGC GCT GTC GAA ACT GAC ACC TAC AAC... 

**α-actin**...arg cys pro glu thr leu phe gln pro ser phe ile ile glu met glu ser ala gly ile his glu thr thr tyr asn... 

260 270

**cC127**...CTA GAG GAG GCT CTC GTG ACC ACC GAC TGC GAC GCC ACC ACC ACC CCC GGC GAG GAC GAG AGC AGT TGC GCT TCC... 

myosin light chain 2...leu glu leu leu leu thr pro thr glu cys asp arg phe thr pro glu ile asp ala gly ile his glu thr thr tyr... 

120 130 140

**cC111**...AAA ACT GAA GAG GAA CTC CTC GAT CCT TTC AGC ACC ACC ACC ACC CCC GGC GAG GAG ATG ATG TGG GGC GCT TCC... 

**troponin C**...lys THR glu glu glu leu ser asp leu phe arg met phe asp glu lys ala asp glu tyr ile asp leu glu glu... 

100 110

**cC106, cC120**...GAG AAA AAG AGG GCA GCC ACC GCC CCC CCG CAG CAC CTC AGT ATG ATG CTC CAG GCT TCC... 

**troponin I**...glu lys lys arg ala ala thr ala arg arg glu his leu lys ser ala met leu glu leu val thr glu... 

10 20

**cC107, cC128**...GCT GAG GAG CTC CAC GTC AAG GCA GCC ACC GCC CCC CCG CAG CAC GAG AGT ATG ATG GTC CAG GCT GTC ACT GAA... 

**troponin heavy**...ala glu glu LEU ser asp val asp leu ser lys arg phe arg lys ILE glu glu leu glu glu glu alu glu arg... 

50 40 30

**FIG. 1.** DNA and decoded protein sequences. Selected regions of DNA sequence are presented for various cDNA clones. These sequences illustrate specific aspects of the analysis of individual clones. Where two clones are indicated, the sequence shown was determined in both and was identical. Protein sequences were deduced from DNA sequences. Except for capitalized amino acids, these match perfectly with sequences established for the proteins indicated, and residues are numbered accordingly. These protein sequences are rabbit α-tropomyosin (35), rabbit α-actin (36), chicken myosin light chain 2 (37), rabbit slow muscle troponin C (38), chicken fast muscle troponin I (39), and rabbit myosin heavy chain (40). Residues in the myosin heavy chain are numbered from the COOH terminus. More extensive DNA sequence information will be presented elsewhere.
α-Actin: cC114, cC115, cC116, and cC118 yielded hybridization-translation products that comigrated on two-dimensional gels with quail skeletal muscle α-actin. The sequence of a region of the 590-bp cDNA insert of cC118 decoded exactly into the amino acid sequence of rabbit muscle α-actin (36) at residues 250–291 (Fig. 1). This insert is clearly a partial copy of α-actin mRNA since the amino acid sequences of β- and γ-actins differ at positions 259, 266, 271, 278, and 286, in addition to others (42). cC118 hybridizes strongly with cC115 but not with cC114 and cC116 (not shown). It is possible that cC114 and cC116 include α-actin mRNA sequences that do not overlap with the cDNA segment cloned in cC118.

Myosin light chain 2: cC127 yielded a hybridization-translation product that comigrated on two-dimensional gels with quail skeletal muscle myosin light chain 2. Both fast and slow fiber isotypes of this protein were observed, although the fast fiber isotype greatly predominated. The sequence of a region of the 590-bp cDNA insert of cC127 decoded into an amino acid sequence identical to residues 73–151 of the presumably fast isotype of chicken myosin light chain 2, as analyzed by Matsuda et al. (37). That the cC127 sequence encodes the fast, and not the slow, isotype of myosin light chain 2 is further supported by the identification of a cysteine residue at residue 125 of the decoded quail DNA sequence (Fig. 1). Cysteine is present at this position in chicken and rabbit fast myosin light chain 2 but is absent in rabbit slow isotype (43).

Troponin C: cC111 yielded a hybridization-translation product that migrated on two-dimensional gels to a position close to that of quail skeletal muscle troponin C. The sequence of a region of the 420-bp cDNA insert of cC111 (Fig. 1) decoded, with two exceptions, into the amino acid sequence of the slow fiber isotype of rabbit troponin C (38) at residues 47–152. In the rabbit protein, residues 93 and 115 are serine and aspartate, respectively, whereas the DNA sequence decodes into threonine and glutamate at these positions. These few very-conservative substitutions could be species differences or cloning errors, but it is clear that the cDNA insert of cC111 encodes the slow fiber isotype of troponin C. The decoded quail sequence differs at >20 residues from the fast fiber isotype of rabbit troponin C (38) and the closely homologous chicken protein analyzed by Wilkinson (44) but matches the sequence of the rabbit slow fiber isotype of troponin C at all but one of these slow-specific residues.

Troponin I: The cDNA inserts of cC106, cC112, and cC120 cross-hybridized strongly (not shown). None of them gave a hybridization-translation product detectable by one- or two-dimensional gel electrophoresis. However, their DNA sequences decoded exactly into the amino acid sequence of the fast fiber isotype of chicken troponin I (39) at residues 1–127 and 154–182 (Fig. 1). In these decoded regions, the rabbit fast and slow fiber isotypes of troponin I differ at 69 residues (39). Thus, it is clear that cC106, cC112, and cC120 encode the fast isotype of this protein.

Myosin heavy chain: The sequence of a region of the 590-bp insert of cC128 decoded into a stretch of 61 amino acids that was a close match with the COOH-terminal sequence of rabbit skeletal muscle myosin heavy chain (40) (Fig. 1). The decoded DNA sequence differed from the rabbit protein sequence at nine positions and contained an extra glutamate residue 29 amino acids from the COOH terminus. The sequence of a region of the 430-bp insert of cC107 corresponding to the first 44 of these 61 amino acids was identical with that determined for cC128 and confirmed the extra glutamate and all four of the quail/rabbit amino acid differences in this region (Fig. 1). Since only one rabbit myosin heavy chain protein sequence has been determined, we cannot yet say whether these substitutions reflect species differences or isotype differences. However, the overall good sequence match with rabbit myosin heavy chain and the fact that cC128 hybridizes to a 7-kilobase (kb) myofiber RNA (not shown) identifies the cDNA inserts of cC128 and cC107 as partial copies of the 3′ terminus of myosin heavy chain mRNA.

Myogenic increase in contractile protein mRNA levels. cDNA clones were used as probes to assay contractile protein mRNAs in myoblast and myofiber RNA by gel transfer hybridization (30). The α-tropomyosin clone cC102 hybridized intensely with a 1.2-kb RNA in 10 μg of myofiber total RNA; as expected, the hybridization signal with 1 μg of myofiber RNA was much reduced, but the signal with 1 μg of myofiber RNA was still stronger than that obtained when this α-tropomyosin cDNA probe was hybridized to 10 μg of myoblast total RNA (Fig. 2A). Therefore, the abundance of α-tropomyosin mRNA

![Fig. 2. Gel transfer hybridization of cDNA clones (cC102 (A), cC118 (B), cC127 (C), and cC129 (D)) with myoblast and myofiber RNAs. RNA samples were subjected to electrophoresis in adjacent wells of a methylmercury-agarose gel, transferred to diazotized paper, and hybridized with 32P-labeled cDNA clones. Lanes: 1, 10 μg of myoblast total RNA; 2 and 3, 1 and 10 μg, respectively, of myofiber total RNA. Size markers were E. coli 23S RNA (2904 nucleotides (45)), E. coli 16S RNA (1541 nucleotides (46)), and rabbit globin mRNA (650 nucleotides (47)).](image-url)
increases at least 10-fold after myoblast fusion. Similar increases in RNA abundance were observed for α-actin mRNA (1.6 kb; Fig. 2B) and myosin light chain 2 mRNA (0.75 kb; Fig. 2C) and for troponin C mRNA (0.85 kb) and troponin I mRNA (1.0 kb) (not shown). The RNA preparations analyzed in Fig. 2A–C were also hybridized with a control cDNA clone, cC129. cC129 is one of three cloned myoblast-specific sequences identified in our initial colony hybridization screen because they hybridized to a greater extent with myoblast [32P]cDNA than with myofiber [32P]cDNA. As shown in Fig. 2D, cC129 hybridized with a 2.4-kb RNA that was much more abundant in myoblast RNA than in myofiber RNA. This result demonstrated the integrity of the myoblast RNA preparation used to assay contractile protein mRNA abundance and, in conjunction with Fig. 2A–C, further confirmed the validity of the initial colony hybridization screen used to identify developmentally regulated RNA sequences.

**DISCUSSION**

Using a differential colony hybridization screening procedure, we have identified a set of cloned myofiber-specific sequences in a recombinant cDNA library representing cultured embryonic myofiber RNA. DNA sequence analysis has identified members of this set encoding six different contractile proteins: α-actin, α-tropomyosin, myosin heavy chain, myosin light chain 2, troponin C, and troponin I. The decoded cDNA sequences corresponding to troponin I, α-actin, and myosin light chain 2 match perfectly with known adult muscle contractile protein amino acid sequences. The few conservative amino acid substitutions noted for α-tropomyosin (aspartate → glutamate) and troponin C (serine → threonine and aspartate → glutamate) are consistent with the slight gull/rabbit species divergences expected among these highly conserved proteins. The possibility that these substitutions reflect isotype differences rather than species variation is virtually excluded by the high degree of divergence among the various isotypes of these proteins within a given species and the excellent match between the decoded cDNA sequences and established amino acid sequences of known adult isotypes. Our results, therefore, indicate that embryonic myofiber muscle cultures express adult genes for these contractile proteins rather than a special set of genes encoding embryonic isotypes, as is the case for the embryonic β-globins (48).

The greatest departure from perfect agreement between decoded embryonic cDNA and established adult protein sequences was in the case of myosin heavy chain. Although the amino acid sequence match at 51 out of 61 residues unambiguously identifies myosin heavy chain cDNA clones, the disagreement at 10 residues still remains to be explained. Only one myosin heavy chain protein sequence has been determined. Sequence analysis of additional forms of myosin heavy chain will probably be necessary to resolve whether the substitutions we observe are species differences in exactly homologous proteins or whether the cDNA sequences correspond to a different isotype of the protein, perhaps an embryonic form (49).

It is particularly interesting that the cDNA clone cC1111 decodes into the slow, and not the fast, fiber isotype of troponin C. This indicates that birds have a slow fiber isotype of troponin C, as do mammals (38). Moreover the expression of mRNA encoding a slow fiber contractile protein in cultures of embryonic presumptive fast muscle (breast) has some bearing on the mechanism of skeletal muscle fiber type differentiation in the adult. Keller and Emerson (9) have shown that both presumptive fast and presumptive slow embryonic avian muscles, as well as clonal muscle cultures, synthesize both fast and slow fiber isoforms of myosin light chains. It was concluded that all skeletal muscle myosin light chain genes are coactivated immediately after myoblast fusion in embryonic muscle and that the restricted fiber type isotype patterns evident in adult muscle must result from selective turnover of expression of inappropriate isoforms later in development. Our identification of a slow fiber troponin C clone in an embryonic presumptive fast muscle cDNA library suggests that a similar mechanism of coexpression followed by selective repression also regulates the troponin C gene family. Such a regulatory system implies the existence of two distinct and sequentially acting control mechanisms. The first mechanism is responsible for the coordinate activation during myogenesis of a large set of muscle genes, including genes encoding fast and slow fiber isotypes of contractile proteins. The second mechanism is responsible for the repression later in development of a specific subset of the genes originally activated during myogenesis. It is significant that both control mechanisms operate on functional sets of genes, for this indicates that the developing organism has two different ways by which it can recognize, as functional sets, the various members of the large group of skeletal muscle genes.

Studies of avian and mammalian myogenesis have shown a dramatic and coordinate activation of contractile protein synthesis during myoblast differentiation (8–10). Earlier results (e.g., ref. 50) indicated that contractile protein synthesis might be regulated by translational activation of a pool of inactive mRNAs accumulated in myoblasts. However, recent work indicates that dividing myoblasts contain little if any muscle-specific contractile protein mRNA and that the myogenic activation of contractile protein synthesis results from accumulation of contractile protein mRNAs during myoblast differentiation (17, 51, 52). This has been most clearly shown in mammalian systems by the use of cDNA clones corresponding to α-actin, myosin light chain 2, and myosin heavy chain (13, 14). Our results show that these observations also apply to avian myogenesis and extend them to include α-tropomyosin, troponin C, and troponin I mRNAs. First, we identified cDNA clones encoding six different contractile proteins among a set of cloned sequences that hybridized to a greater extent with myofiber than with myoblast [32P]cDNAs. This result indicates that these contractile protein mRNAs are considerably more abundant in myofiber than in myoblast poly(A)+ RNA. Second, by using the identified cDNA clones as RNA sequence probes in gel transfer hybridization analysis, we found that these contractile protein RNAs increase 10-fold or more in relative abundance in total RNA during myogenesis. Moreover, considering the 2- to 3-fold net accumulation of total RNA per myofiber nucleus after myoblast fusion (53), the cellular abundance of these mRNAs actually increases 20- to 30-fold or more. Since the gel transfer-hybridization assay is independent of translatability, presence or absence of poly(A), or subcellular localization, it seems clear that myoblasts do not contain sufficient amounts of any contractile protein-coding RNA sequences, translatable or otherwise, to account for the dramatic activation of synthesis after myoblast fusion (8). These results apply to mRNAs encoding one or more subunits of each of the major contractile proteins actin, myosin, tropomyosin, and troponin. In addition, as many as 11 or 12 as yet unidentified RNAs are also represented in our set of myofiber-specific cDNA clones. Presumably these unidentified myofiber-specific RNAs encode other proteins, including additional contractile proteins, whose synthesis is activated during myogenesis (8, 9). Thus, we can now generalize that the activation of contractile protein synthesis during myoblast differentiation is regulated by the coordinate accumulation of contractile protein mRNAs.

We do not yet know whether contractile protein mRNA accumulation is regulated by transcriptional activation of muscle genes or by posttranscriptional stabilization of their RNA tran-
scripts. However, the contractile protein cDNA clones we describe in this report provide the necessary probes to answer this question and to isolate genomic clones of these coregulated genes. A comparative study of contractile protein gene structure may reveal what these diverse genes have in common that allows them to be regulated as a set during myogenesis.

We thank Anne Glenn, Margaret Ober, Linda Schuman, Druen Robinson, and Ruth Ashforth for excellent technical assistance and Sanford Bernstein for valuable discussions. We also thank Steve McKnight, Robert Kingsbury, and Donald Brown for their generous advice and the use of their facilities at the Carnegie Institution of Washington. This work was supported by grants from the National Institutes of Health (HD07796) and the Muscular Dystrophy Association (to C.P.E.) and a Muscular Dystrophy Association Postdoctoral Fellowship (to K.E.M.H.).