MOLECULAR AND CELLULAR BIOLOGY, Dec. 1987, p. 4377–4389
0270-7306/87/124377-13$02.00/0

Vol. 7, No. 12

Multiple Positive and Negative 5′ Regulatory Elements Control the Cell-Type-Specific Expression of the Embryonic Skeletal Myosin Heavy-Chain Gene

PATRICE F. BOUVAGNET, EMANUEL E. STREHLER,† GLENN E. WHITE, MARIE-ANTOINETTE STREHLER-PAGE,† BERNARDO NADAL-GINARD, AND VIJAK MAHDAVI*

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

Received 19 May 1987/Accepted 8 September 1987

To identify the DNA sequences that regulate the expression of the sarcomeric myosin heavy-chain (MHC) genes in muscle cells, a series of deletion constructs of the rat embryonic MHC gene was assayed for transient expression after introduction into myogenic and nonmyogenic cells. The sequences in 1.4 kilobases of 5′-flanking DNA were found to be sufficient to direct expression of the MHC gene constructs in a tissue-specific manner (i.e., in differentiated muscle cells but not in undifferentiated muscle and nonmuscle cells). Three main distinct regulatory domains have been identified: (i) the upstream sequences from positions −1413 to −174, which determine the level of expression of the MHC gene and are constituted of three positive regulatory elements and two negative ones; (ii) a muscle-specific regulatory element from positions −173 to −142, which restricts the expression of the MHC gene to muscle cells; and (iii) the promoter region, downstream from position −102, which directs transcription initiation. Introduction of the simian virus 40 enhancer into constructs where subportions of or all of the upstream sequences are deleted (up to position −173) strongly increases the level of expression of such truncated constructs but without changing their muscle specificity. These upstream sequences, which can be substituted for by the simian virus 40 enhancer, function in an orientation-, position-, and promoter-dependent fashion. The muscle-specific element is also promoter specific but does not support efficient expression of the MHC gene. The MHC promoter in itself is not muscle specific. These results underline the importance of the concerted action of multiple regulatory elements that are likely to represent targets for DNA-binding-regulatory proteins.

Sarcomeric myosin heavy chain, like other contractile proteins, exists as a family of distinct isoforms that are predominantly expressed during a given stage of muscle development or in a specific muscle tissue or both (11, 12, 36, 39, 69, 76, 77). In vertebrates, the different myosin heavy-chain (MHC) isoforms are encoded by a multigene family of distinct but closely related members (38, 39, 53, 57, 78). Structural studies on a variety of MHC genes from a number of different organisms have shown that these genes, at least in higher vertebrates, are comprised of approximately 25 kilobases (kb) of DNA and display a highly complex exon-intron organization (38, 51, 68). The rat embryonic skeletal muscle MHC gene (43, 44, 57, 66, 68, 78) has recently been sequenced in its entirety and was found to be split into 41 exons distributed over 24 kb of DNA (68). In addition, most sarcomeric MHC genes are clustered in the genome (13, 35, 38, 75), and several of them are tightly linked with only a few kb of intergenic sequence separating one gene from the next (38, 39). However, the relevance of these structural features for the tightly controlled mode of expression of the MHC multigene family remains unclear. Differential MHC gene expression seems to be largely regulated at the level of transcription (3, 36, 43, 51, 72, 74), but the mechanisms that control the tissue-specific and developmental stage-specific pattern of MHC gene expression have not yet been elucidated.

In the present work, we have begun to define the *cis*-regulatory elements involved in the cell-type-specific expression of the rat embryonic MHC gene. To this end, different MHC minigenes and CAT reporter genes were constructed, and their transcriptional activity was tested in transient expression assays upon transfection into nonmyogenic and myogenic cells. The results show that 1.4 kb of 5′-flanking sequence is sufficient to direct the tissue-specific expression of this gene, whereas intragenic sequences do not appear to be of major importance. In addition to the promoter, which in itself is not muscle specific, a minimum of two functionally distinct 5′ sequence elements controls the expression of the rat embryonic MHC gene. The sequences from −173 to −142 base pairs (bp) confer tissue specificity to the gene by inhibiting its expression in nonmuscle cells. This muscle-specific negative regulatory sequence needs upstream elements located between −1413 and −174 bp for full activity. The upstream elements can be functionally replaced by the simian virus 40 (SV40) enhancer but do not act like typical enhancers (31). These results demonstrate that cell-type-specific MHC gene expression is controlled by the concerted action of multiple regulatory elements.

MATERIALS AND METHODS

Restriction endonucleases and DNA modifying enzymes were purchased from New England Biolabs, Inc., Bethesda Research Laboratories, Inc., and New England Nuclear Corp. Plasmid vectors pUC18 and pUC19 and M13 sequenc-

* Corresponding author.
† Present address: Department of Biochemistry III, Swiss Federal Institute of Technology, ETH Zentrum, CH 8092 Zurich, Switzerland.

ing vectors M13 mp18 and mp19 as well as the M13 sequencing kits were obtained from New England Biolabs, Inc. [α-^{32}P]dATP ($\geq$400 Ci/mmol), [α-^{32}P]dCTP ($\geq$400 Ci/mmol), [γ-^{32}P]ATP (3,000 Ci/mmol), [α-^{35}S]dATP (650 Ci/mmol), and [^{14}C]chloramphenicol (60 mCi/mmol) were from New England Nuclear. Dulbecco modified Eagle medium (DMEM) was purchased from GIBCO Laboratories.

Plasmid constructions. The isolation and characterization of rat embryonic MHC genomic DNA clones have been described earlier (44, 57, 66, 68). All MHC minigenes and reporter chloramphenicol acetyltransferase (CAT) genes were constructed by established procedures (40) by ligating subfragments of the rat MHC gene into appropriately restricted pUC plasmid vectors and transforming competent *Escherichia coli* HB101 cells. Screening of colonies and plasmid DNA isolation were done by standard techniques (40). Construct DNA to be used for transfections was purified through two cycles of centrifugation to equilibrium in cesium chloride-ethidium bromide density gradients (40). Constructs containing the CAT gene were made in pA10CAT (33) and pSV2CAT (20) vectors or in a pUC9 derivative (pUC9CAT) containing the CAT gene on a *Bam*HI-*Hin*dIII fragment (positions 3369 to 5003 in pSV2CAT). To fuse the MHC 5'-end sequences precisely at the end of exon 1 to the *Hin*dIII site at the 5' end of the CAT gene, a 0.3-kb *Pst*I-*Bgl*I MHC fragment (containing 0.3 kb of 5'-flanking DNA and ending 16 bp upstream of the cap site) was isolated. This fragment was then ligated into *Pst*I-*Hin*dIII-restricted pUC19, together with a double-stranded synthetic oligodeoxynucleotide representing positions -16 to $+24$ of the rat embryonic MHC gene (68) and ending with a polylinker sequence including *Bam*HI, *Sma*I, and *Hin*dIII sites. The sequences of these oligonucleotides read 5' GGGCGGCTCTCCTCTGTGCTTCACTCTGCCACAGTCAGAGGGATCCCGGGAAGCTTGG 3' (58-mer) and 3' TCCCCCGCCGAGAGGAGACACGAAGTGAGACGGTGTCAGTCTCCCTAGGGCCCTTCGAACC 5' (61-mer). The resulting plasmid with 303 bp of MHC upstream sequence was named pE303. Then 2.1 kb of 5'-flanking sequence (positions -2600 to -303) was cloned into the *Eco*RI and *Pst*I sites of pE303. From this construct, a 1.4-kb *Hin*dIII fragment (positions -1413 to $+24$) was then cloned into the *Hin*dIII site of pUC9CAT, thereby creating a vector carrying the CAT gene linked at its 5' end to exon 1 and approximately 1.4 kb of 5'-flanking sequence of the rat MHC gene (pE1413CAT). All other MHC CAT constructs were obtained by deleting the MHC 5'-end sequences upstream of various restriction sites. The SV40 enhancer, including the two 72-bp repeats and all three 21-bp repeats, was isolated as a *Bgl*II fragment, corresponding to positions 35 to 272 of the SV40 genome from pA10CAT3M (24), and was ligated into the *Bam*HI site of the MHC minigene constructs at position $+261$, or in the *Bam*HI site at the 3' end in pUC9CAT and in the MHC CAT gene constructs.

Transient expression assays. DNA was introduced into the cells by the calcium phosphate coprecipitation method (20, 21). Mouse myogenic C2C12 cells (5, 79) (hereafter referred to as C2 cells) and human HeLa cells were seeded at an initial density of approximately 0.5×10^6 cells in 10-cm Falcon tissue culture dishes and were grown in 10 ml of medium at 37°C in 5% CO_2. DMEM–20% fetal calf serum–0.5 to 1% chicken embryo extract was used for C2 cells, and DMEM–10% fetal calf serum was used for mouse fibroblasts and HeLa cells. All media were supplemented with penicillin and streptomycin. Four hours before transfection (approximately 24 h after the initial plating) the cells were fed with 10 ml of fresh medium. For each plate, 15 μg of DNA was suspended in 438 μl of H_2O followed by the addition of 62 μl of 2 M $CaCl_2$–161 mM Tris hydrochloride (pH 7.6). The DNA was added drop by drop to 500 μl of 2× HEBS (10× HEBS is 8.18% [wt/vol] NaCl–5.95% [wt/vol] *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid–0.11% [wt/vol] Na_2HPO_4) and kept for 30 min at room temperature. The suspension was then added to the cells. After 15 to 18 h, the cells were shocked with 2 ml of 15% glycerol in 1× HEBS for 2.5 min at 37°C, washed with DMEM, and supplemented with fresh medium. Growing (undifferentiated) C2 myoblasts were refed 24 h later; for induction of differentiation, cells were switched to DMEM–10% horse serum within 4 h after glycerol shock. Efficiency of transfection was internally controlled by cotransfecting pSV2CAT DNA together with the MHC minigenes or the β-galactosidase gene (3 μg per plate), under the control of the MSV promoter, together with the MHC CAT gene constructs. All experiments were carried out at concentrations of DNA (15 μg per plate) in the linear range of the transfection assay for the constructs analyzed.

RNA preparation and analysis. Cells were washed three times and scraped in 5 ml of phosphate-buffered saline solution, pelleted for 5 min at $1,000 \times g$, and suspended in 3 volumes of ice-cold lysis buffer (50 mM Tris hydrochloride [pH 8.3], 50 mM NaCl, 5 mM $MgCl_2$, 2% [vol/vol] Nonidet P-40) containing 10 mM ribonucleoside-vanadyl complex. DNA and cell debris were pelleted by centrifugation for 3 min in an Eppendorf microfuge, and the supernatant was transferred to a new tube and extracted twice with an equal volume of hot (65°C) phenol-chloroform-isoamyl alcohol (50:48:2). After two ether extractions, the RNA was precipitated by the addition of 2 volumes of 6 M guanidine hydrochloride (pH 7.4) and 1.5 volumes of ethanol at -20°C overnight. After centrifugation, the RNA was precipitated in the presence of 0.3 M sodium acetate (pH 5.5) and 2.5 volumes of ethanol and kept at -20°C until use.

For Northern blotting (RNA blotting) analysis (71), the RNA samples (10 μg per slot) were run on 3% formaldehyde–1% agarose gels in MOPS buffer (20 mM morpholinepropanesulfonic acid–1 mM EDTA, pH 7.4) after denaturation of the RNA at 65°C for 5 min in 70% formamide–8% formaldehyde–6% glycerol in MOPS buffer. The RNA was then transferred to nitrocellulose filters in 10× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate, pH 7.0) and processed for hybridization. ^{32}P-labeled DNA probes were prepared by nick translation and were used at 0.5×10^6 cpm per lane. Prehybridization, hybridization, and washing of the nitrocellulose blots were done as described previously (40). S1 nuclease protection assays were performed by a modification of the method of Berk and Sharp (4) as described previously (38). End-labeled probes were prepared by treating appropriate DNA fragments with [γ-^{32}P]ATP and T4 polynucleotide kinase followed by strand separation and polyacrylamide gel electrophoresis (40).

Primer extension analysis. Synthetic oligodeoxynucleotide primers were 5' end labeled and purified as described previously (68). Annealing of the primer to cytoplasmic RNA from transfected cells (approximately 5 ng yields 10^6 cpm per reaction containing 20 μg of RNA) and cDNA synthesis by reverse transcriptase were done as reported previously (73). The extension products were purified and run on polyacrylamide–8 M urea gels (68).

CAT assays. Cell extracts were prepared 48 h after glycerol shock through three cycles of freezing and thawing (-80°C and 37°C), followed by centrifugation (20). One-third

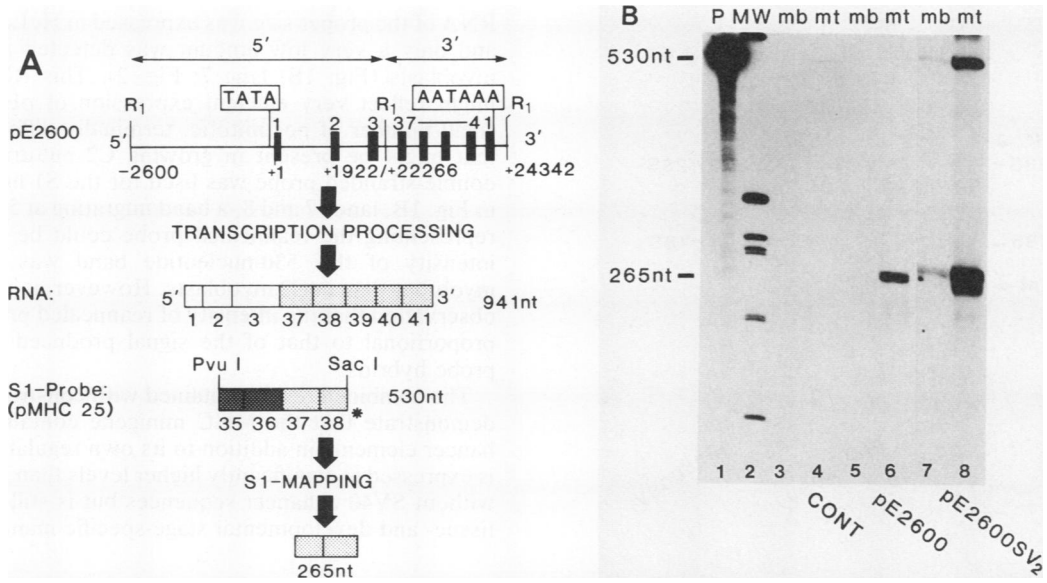


FIG. 1. Structure and expression of a rat embryonic MHC minigene. (A) Schematic representation of the pE2600 gene and of its transcript. A map of the 6.6-kb long embryonic MHC minigene in pE2600 is shown on top. Exons are represented by black boxes and are numbered according to their position in the intact MHC gene. Compared with the complete MHC gene (68) this minigene contains a 20,344-bp deletion at the middle *Eco*RI (R1) site. TATA, TATA box; AATAAA, polyadenylation signal. Upon transcription and proper RNA processing, minigene pE2600 produces a 941-nucleotide (nt)-long mRNA. Splicing of exon 3 to exon 37 was detected by S1 nuclease mapping using as a probe the 5' end-labeled, 530-nucleotide-long anti-mRNA strand of the *Pvu*I-*Sac*I fragment of cDNA pMHC25 (57) covering exons 35 to 38. Only 265 nucleotides of this probe can be protected by pE2600 RNA. (B) Detection of the minigene transcript by S1 nuclease mapping analysis. Parallel cultures of mouse myogenic C2 cells were either mock transfected (control, lanes 3 and 4) or transfected with rat MHC minigene constructs pE2600 (lanes 5 and 6) or pE2600SV2 (lanes 7 and 8). Cells were then either kept as myoblasts (mb) in growth medium (lanes 3, 5, and 7) or allowed to differentiate into myotubes (mt) in fusion medium (lanes 4, 6, and 8). Total cellular RNA was prepared 60 h after transfection and assayed (20 μg per lane) by S1 nuclease mapping with the probe (lane 1) depicted in A. Note that the minigenes are expressed at high levels only in differentiated C2 cells as evidenced by the presence of a 265-nucleotide-long partially protected fragment (A). A single-stranded probe was used for lanes 3 to 6, and a double-stranded probe was used for lanes 7 and 8. For maps of constructs pE2600 and pE2600SV2 see panel A and Fig. 2. MW, Marker lane with *Hae*III-digested φX174RF DNA.

volume of total cell extracts was incubated at 37°C for 30 min in a final volume of 180 μl containing 109 μl of 0.25 M Tris hydrochloride (pH 8.0), 20 μl of 4 mM acetyl coenzyme A, and 0.1 μCi of [^{14}C]chloramphenicol. Extraction with ethyl acetate, thin-layer chromatography, and autoradiography were done as described previously (20).

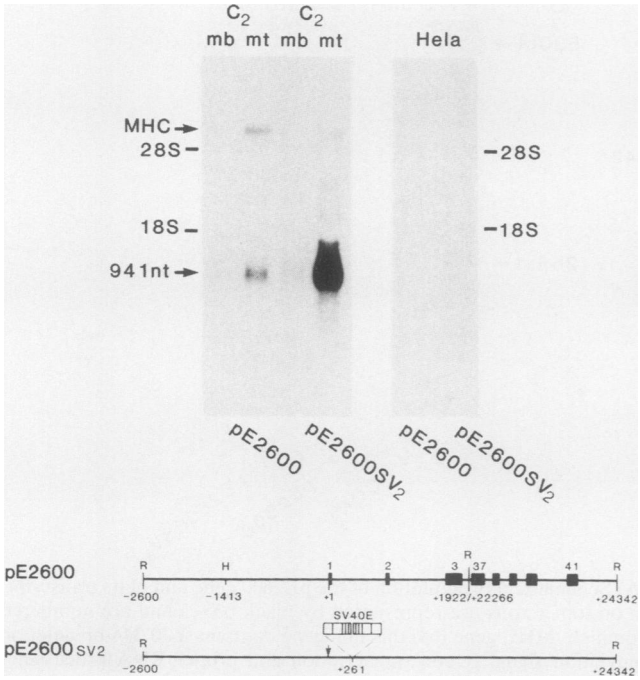
DNA sequencing. DNA sequences were determined by the dideoxynucleotide chain-termination method (60) with the M13 vector system (46). Software of the SEQ-System (9) and of IntelliGenetics, Palo Alto, Calif., was used for sequence analysis.

RESULTS

Specific expression of an embryonic MHC minigene with an internal deletion of 20 kb in differentiated muscle cells. Mammalian sarcomeric MHC genes are approximately 25 kb in size and display a highly complex exon-intron structure (38, 51, 68, 78). The complete sequence and structure of the gene encoding the rat skeletal embryonic MHC have recently been elucidated (68). This gene is normally expressed in fetal and newborn rat skeletal muscles as well as in differentiated myogenic cells in culture (43, 44, 51, 57). Because the large size of the MHC gene precludes its easy handling, plasmids containing only parts of the original gene were constructed, so that the function of putative *cis*-acting regulatory elements could be analyzed in transient transfection assays.

The MHC minigene pE2600 (Fig. 1A) consists of two joined genomic *Eco*RI fragments containing 2.6 kb of 5'-flanking DNA with the first three exons and the last five exons of the original gene with 560 bp of 3'-flanking DNA (57, 68). This construct comprises 6.6 kb of MHC gene DNA, including the original promoter region and polyadenylation signal, but carries an internal deletion of 20,344 bp encompassing exons 4 to 36. Upon introduction into permissive cells, this construct should give rise to a 941-nucleotide-long mRNA, provided that proper RNA splicing occurs (Fig. 1A). In contrast, the normal embryonic MHC gene produces a 6,035-nucleotide-long mRNA (68). The difference in size and in splicing pattern between the MHC transcripts from the endogenous gene and the exogenously added minigene can be exploited to discriminate the two MHC RNA products by Northern blotting and S1 nuclease protection assays.

When pE2600 DNA was introduced into mouse myogenic C2 cells and assayed for transient expression, no MHC transcripts could be detected by S1 nuclease mapping when the cells were kept in an undifferentiated state (Fig. 1B, lane 5). However, when the cells were allowed to differentiate in fusion medium, large amounts of RNA were produced from the rat embryonic MHC minigene (Fig. 1B, lane 6). This RNA appears to be properly spliced at the exon 3-exon 37 junction (Fig. 1A). The size of the RNA produced from the rat MHC minigene in differentiated muscle cells corresponded to the predicted length of 941 nucleotides (Fig. 2). The amount of minigene-derived RNA was comparable to or exceeded the level of endogenous MHC mRNA (Fig. 2). In contrast to its efficient expression in differentiated C2 cells, MHC minigene pE2600 was not expressed in nonmyogenic cells such as HeLa cells (Fig. 2) or rat or mouse fibroblasts



myogenic and nonmyogenic cells. Northern blots of total RNA (10 μg per lane) prepared from C2 and HeLa cell cultures 60 h after transfection with MHC minigene constructs pE2600 and pE2600SV2. A nick-translated genomic fragment containing the third exon (68) was used to detect MHC mRNA sequences. Mouse myogenic C2 cells were either kept as myoblasts (mb) or allowed to form myotubes (mt) after transfection. Arrows indicate the positions of endogenous MHC mRNA (MHC) and RNA produced from minigenes containing only exons 1 to 3 and 37 to 41 (941 nucleotides [nt]) of the rat embryonic MHC gene. The positions of 28S and 18S RNAs are also shown. Maps of the constructs used for transfections are indicated below the Northern blots. Exons are represented by black boxes for pE2600 and are numbered according to their order in the intact MHC gene. pE2600SV2 is identical to pE2600 except for an insertion at position +261 of two SV40 enhancer elements (SV40E). The site of transcription initiation in pE2600SV2 is indicated by an arrow. Restriction sites: R, *Eco*RI; H, *Hin*dIII.

(data not shown), although efficient transfection into these different types of cells was documented in cotransfection experiments with pSV2CAT DNA as a control (data not shown). Taken together, the above results show that a rat embryonic MHC minigene containing its own promoter with 2.6 kb of 5′-flanking DNA but carrying a large intragenic deletion is correctly expressed in differentiated muscle cells but not in undifferentiated muscle or nonmyogenic cells.

Effect of insertion of the SV40 enhancer elements into the embryonic MHC minigene pE2600 on tissue-specific pattern of expression. To challenge the tissue-specific properties exhibited by the MHC minigene, the SV40 enhancer sequences were added to pE2600. The resulting minigene pE2600SV2 (Fig. 2), containing two copies of the SV40 enhancer placed in opposite orientation into the first intron (position +261) of the embryonic MHC gene pE2600, was assayed for its transient expression in C2 and HeLa cells. Large amounts of the 941-nucleotide MHC mRNA were detected in differentiated C2 cells (Fig. 1B, lane 8; Fig. 2), independently of the position and orientation of the enhancer (data not shown). This RNA accumulated to levels estimated to be at least 5- to 10-fold higher than those of the pE2600 transcript. No MHC RNA of the proper size was expressed in HeLa cells (Fig. 2), and only a very low amount was detected in growing C2 myoblasts (Fig. 1B, lane 7; Fig. 2). This RNA, however, might reflect very efficient expression of pE2600SV2 in a small number of postmitotic, terminally differentiated cells that might be present in growing C2 cultures. Because a double-stranded probe was used for the S1 nuclease assays in Fig. 1B, lanes 7 and 8, a band migrating at 530 nucleotides representing the reannealed probe could be detected. The intensity of this 530-nucleotide band was higher in C2 myotubes than in C2 myoblasts. However, it is our empirical observation that the intensity of reannealed probes is always proportional to that of the signal produced by the RNA-probe hybrids.

The combined results obtained with construct pE2600SV2 demonstrate that the MHC minigene containing SV40 enhancer elements in addition to its own regulatory sequences is expressed at significantly higher levels than its counterpart without SV40 enhancer sequences but is still regulated in a tissue- and developmental stage-specific manner.

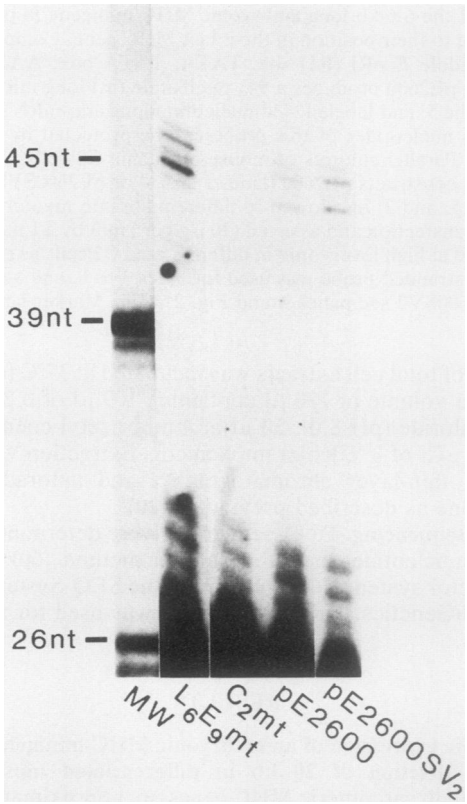


FIG. 3. Primer extension on RNA transcribed from rat embryonic MHC minigenes in differentiated mouse C2 cells. RNA was isolated from untransfected differentiated rat L6E9 (L6E9mt) and mouse C2 (C2mt) myotubes as well as from differentiated C2 cells previously transfected with rat MHC minigene constructs pE2600 or pE2600SV2 and was subsequently annealed to a 5′-end-labeled 21-mer primer representing nucleotides +26 to +46 of the rat embryonic MHC mRNA. Upon extension with reverse transcriptase, products of the expected size of 45 to 48 nucleotides (nt) were detected in rat control (L6E9mt) cells, which expressed the endogenous rat MHC gene, and in mouse C2 cells transfected with constructs pE2600 and pE2600SV2. MW, Marker lane with ^{32}P-labeled synthetic 26-mer and 39-mer oligodeoxynucleotides. For maps of pE2600 and pE2600SV2, see Fig. 2.

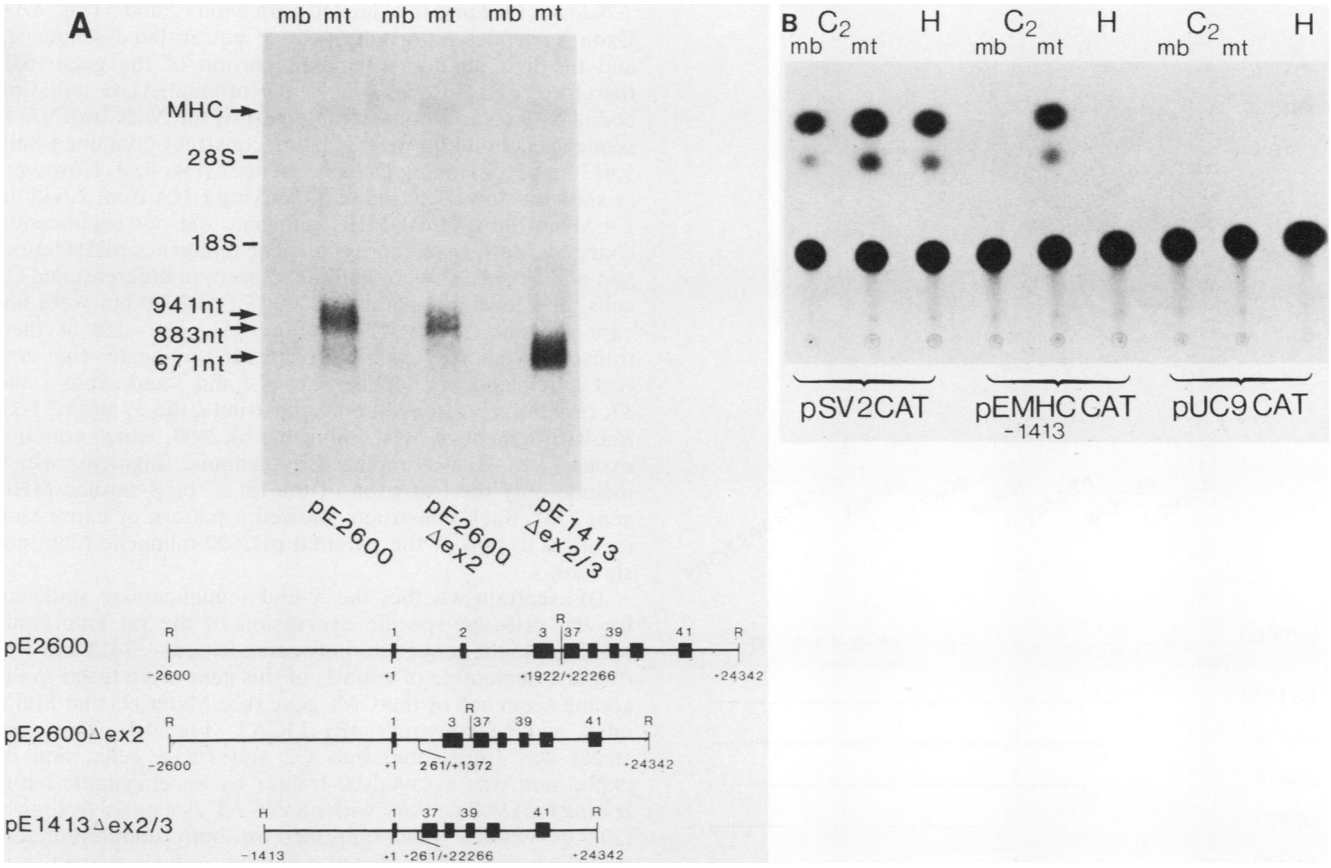
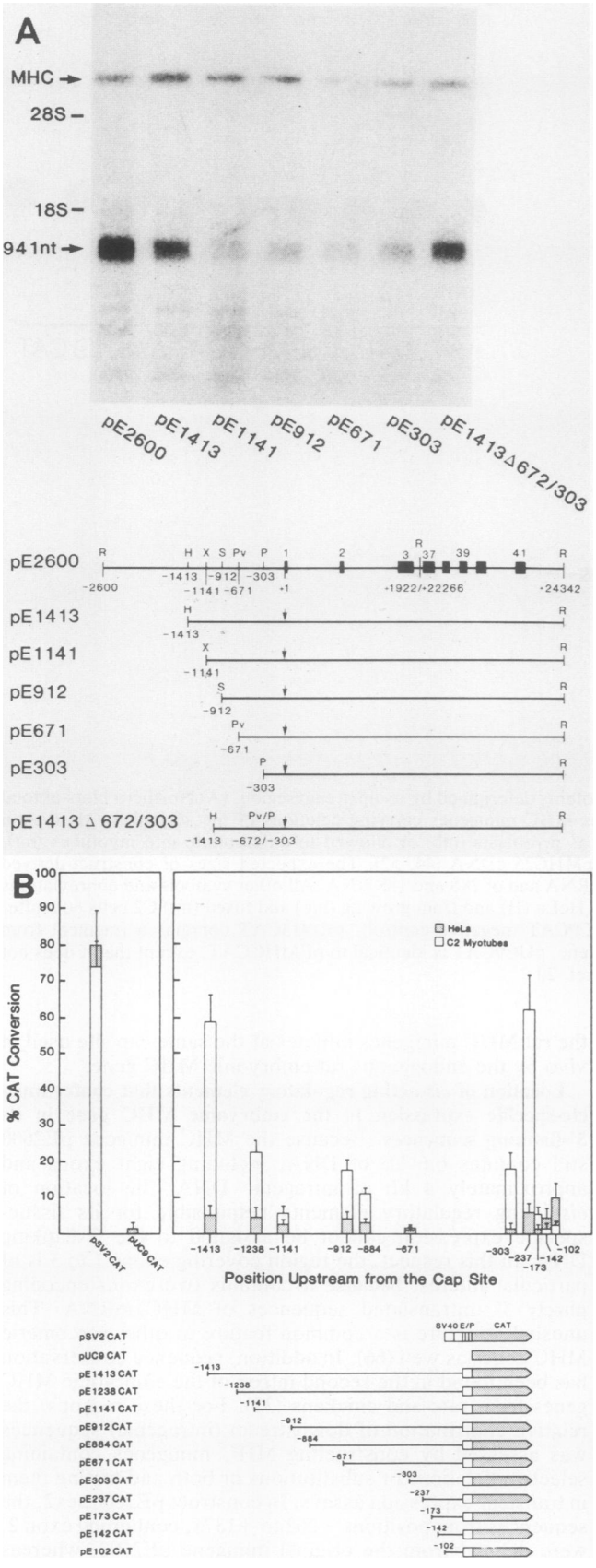


FIG. 4. Tissue-specific expression of the embryonic MHC gene promoter is determined by its upstream region. (A) Northern blots of total RNA (10 μg per lane) prepared from C2 cells transfected with various MHC minigenes carrying deletions of the second- or third-exon sequence constructs (schemes on bottom). The cells were either kept as myoblasts (mb) or allowed to differentiate into myotubes (mt). Detection of MHC-specific RNA sequences was with nick-translated pMHC25 cDNA (44, 57). The expected sizes of construct-derived mRNAs are indicated together with the positions of endogenous MHC mRNA and of 28S and 18S RNA. All other symbols and abbreviations are as in Fig. 2. (B) CAT assays (20) were performed with extracts from HeLa (H) and from growing (mb) and fused (mt) C2 cells 60 h after transfection with pSV2CAT (positive control), pE1413 CAT, and pUC9CAT (negative control). pE1413CAT contains a fragment from positions −1413 to +24 of the embryonic MHC gene fused to the CAT gene. pUC9CAT is identical to pEMHCCAT, except that it does not contain any MHC gene sequences. For a description of pSV2CAT see ref. 20.

Initiation of the MHC minigene transcripts at the proper cap site. To determine whether the RNA transcripts from the MHC minigene constructs pE2600 and pE2600SV2 initiate at the cap site used by the endogenous embryonic MHC gene in vivo, primer extension experiments were performed. A ^{32}P-labeled synthetic 21-mer oligodeoxynucleotide with a sequence complementary to nucleotides +26 to +46 of the rat embryonic MHC mRNA (68) was used as the primer. This part of the mRNA is in the second exon of the rat embryonic MHC gene and contains only 5′ untranslated sequences that are known to be highly gene specific. When this primer was used with RNA from differentiated rat L6E9 cells, which express the rat embryonic MHC gene (43, 44, 51, 57), the major extension products upon reverse transcription were, as expected (68), 45 to 48 nucleotides long (Fig. 3, lane 2). In contrast, when RNA from differentiated mouse C2 cells was used in the same type of primer extension assay, no specific extension product was made (Fig. 3, lane 3). However, when the RNAs from differentiated C2 cells previously transfected with the MHC minigene pE2600 or pE2600SV2 were assayed in the same fashion, the pattern of the major extension products was identical to that seen with L6E9 RNA (Fig. 3, lanes 4 and 5). Thus, transcription from the rat MHC minigenes initiates at the same cap site used in vivo by the endogenous rat embryonic MHC gene.

Location of *cis*-acting regulatory elements that confer muscle-specific expression to the embryonic MHC gene in its 5′-flanking sequences. Because the MHC minigene pE2600 still contains 6.6 kb of DNA, including eight exons and approximately 4 kb of intragenic DNA, the location of *cis*-acting regulatory elements responsible for its tissue-specific expression cannot be assigned to the 5′-flanking DNA. In this respect, the region covering exons 1 to 3 is of particular interest because it contains two exons encoding purely 5′ untranslated sequences of MHC mRNA. This unusual structure is a common feature in other sarcomeric MHC genes as well (66). In addition, sequence conservation has been found in the second intron of the embryonic MHC genes from rats and chickens (32). For these reasons, the relative contribution of downstream (intragenic) sequences was assessed by constructing MHC minigenes containing selective deletions or substitutions or both and testing them in transient expression assays. In construct pE2600Δex2, the sequences from positions +262 to +1371, containing exon 2, were deleted from the original minigene pE2600, whereas construct pE1413Δex2/3 carried a deletion from positions



+262 to +1922 and thus lacked both exons 2 and 3 (Fig. 4A). Exon 3 encodes a portion of the 5′ untranslated sequences and the first amino-acid-coding portion of the gene (68); therefore, pE1413Δex2/3 lacks the original AUG initiation codon as well as a considerable part of the MHC mRNA 5′ sequences. In addition, this latter construct contained only 1,413 bp of 5′-flanking DNA sequences (Fig. 4A). However, as shown below, deletion of 5′-flanking DNA from 2.6 kb to 1.4 kb in the original MHC minigene did not significantly change its pattern of expression. The minigenes pE2600Δex2 and pE1413Δex2/3 were both expressed, in differentiated C2 cells, at a level comparable to that of pE2600 but were not expressed in C2 myoblasts (Fig. 4A). The size of their transcripts, shorter than that of pE2600, indicates that correct splicing occurs between exons 1 and 3 and exons 1 and 37, respectively. In additional constructs, the 3′-most 2.1-kb *Eco*RI fragment of MHC minigene pE2600, which contains exons 37 to 41, was replaced by genomic fragments which include only the last exon of the rat α- or β-cardiac MHC gene (38). Such constructs showed a pattern of expression identical to that of the parental pE2600 minigene (data not shown).

To ascertain whether the 5′-end sequences are sufficient for the cell-type-specific expression of the rat embryonic MHC gene, the sequences between positions −1413 and +24 (the last nucleotide of exon 1) of this gene were fused to the coding sequence of the CAT gene (see Materials and Methods), making construct pE1413CAT (Fig. 4B). This construct was transfected into C2 and HeLa cells, and its expression was assayed 60 h later by an enzymatic assay testing for CAT activity with pSV2CAT as a positive control (20). pSV2CAT was expressed in both undifferentiated myoblasts and differentiated myotube C2 cells as well as in HeLa cells (Fig. 4B). In contrast, pE1413CAT was exclusively expressed in differentiated C2 cells but remained silent in growing C2 myoblasts or in fibroblasts. Thus, it can be concluded from these experiments that the muscle-tissue-specific expression of the embMHC gene is governed by its promoter and 5′ upstream flanking sequences.

Involvement of multiple interspersed positive and negative *cis*-acting regulatory elements in the regulated expression of the embryonic MHC gene. To delineate the *cis*-acting element(s) conferring tissue specificity to the MHC gene more precisely, a series of constructs resembling pE2600 but containing progressive deletions of the 5′ flanking DNA was made (Fig. 5A) and tested for transient expression in C2 cells. As mentioned above, the level of expression of the MHC minigene containing 1,413 bp of 5′-flanking DNA (construct pE1413) was similar to, although slightly lower than, that of pE2600 (Fig. 5A, lanes 1 and 2). However, when the sequences from positions −1413 to −1141 were

FIG. 5. Transient expression of MHC gene constructs with progressive deletions of 5′-flanking DNA. (A) Northern blots of total RNA isolated from differentiated C2 cells transfected with various MHC minigene constructs. All constructs contain the same coding sequences as the original MHC minigene pE2600 but differ in their amount of 5′-flanking DNA (schemes at bottom). All labels, symbols, and abbreviations are as in Fig. 2. (B) CAT activity in C2 myotubes transfected with various MHC CAT plasmids with decreasing 5′-end upstream sequences (schemes at bottom). Transfection efficiency was normalized to the β-galactosidase activity in the samples. CAT activity was calculated as percent conversion from nonacetylated to acetylated form. Standard deviations of the mean (n = 4 to 8) are given. Restriction sites: R, *Eco*RI; H, *Hin*dIII; X, *Xho*II; S, *Sph*I; Pv, *Pvu*II; P, *Pst*I.

deleted, the expression of the corresponding MHC gene construct (pE1141) decreased significantly. A very low level of expression was also exhibited by MHC minigene constructs with successive deletions up to position −303 (Fig. 5A, lanes 3 to 6). From these results, we conclude that the 5′-flanking DNA of the rat embryonic MHC gene contains regulatory sequences in its far upstream region, between positions −1413 and −1141, that are necessary for its efficient transcription in muscle cells. These sequences exert their effect in a relatively distance-independent manner, since deletion of a portion of the 5′-end region, from positions −672 to −303, did not affect the level of expression of the corresponding construct (pE1413Δ672/303 compared with pE1413).

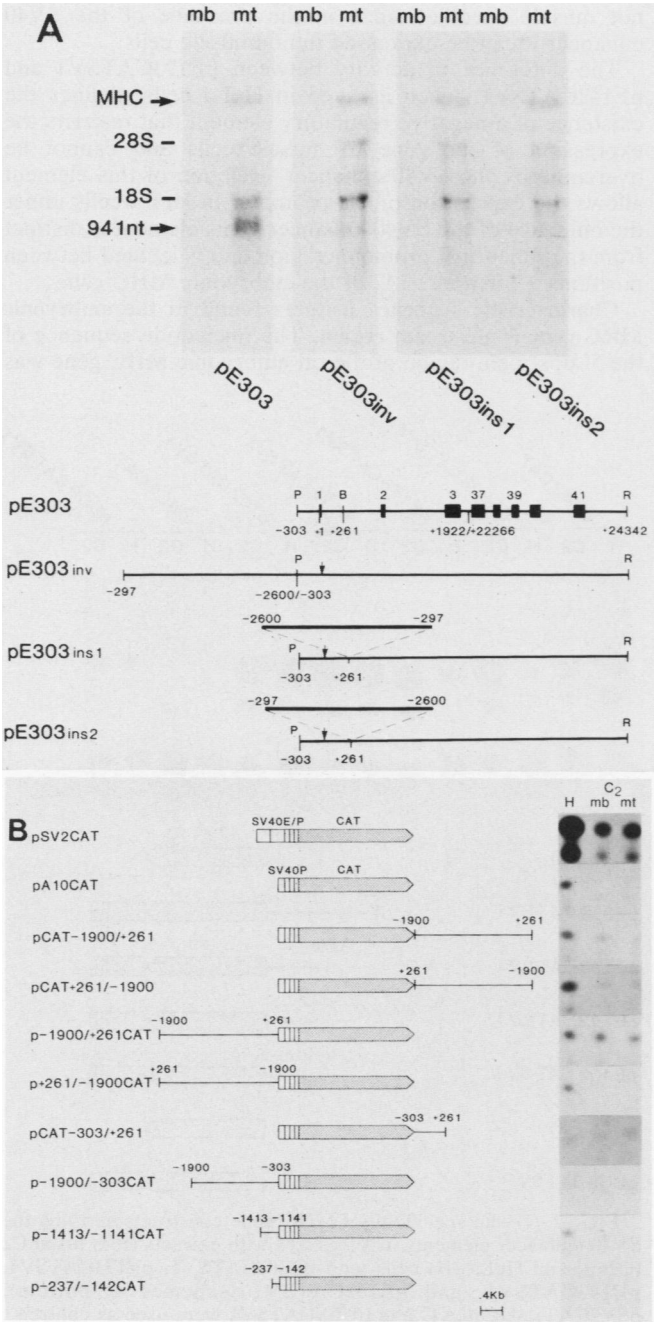


FIG. 6. Test for enhancerlike activity in the 5′ upstream sequences of the embryonic MHC gene. (A) Northern blots of total RNA prepared from C2 myoblasts (mb) and myotubes (mt) transfected with the constructs indicated at the bottom of the panel. In pE303inv, the region between positions −2600 and −297 was inverted and then fused to pE303, whereas pE303ins1 and pE303ins2 contain the same piece of DNA in opposite orientations at position +261. Symbols and abbreviations are as in Fig. 2. (B) CAT assays were performed with extracts from HeLa cells (H), C2 myoblasts (mb) and C2 myotubes (mt) transfected with CAT constructs containing 5′-end sequences of the rat embryonic MHC gene. Schematic maps of the relevant region of each construct are shown on the left. pSV2CAT (20) contains the CAT gene under the control of the SV40 enhancer and promoter (SV40E/P) and serves as positive control (top line). pA10CAT lacks most of the 72-bp repeat SV40 enhancer sequences and acts as control for CAT activity of enhancerless constructs. All other constructs are derived from pA10CAT and carry additional sequences from the MHC gene 5′-flanking region at their 5′ or 3′ ends as indicated in the schemes at the left.

The low level of expression of the minigene deletions downstream from position −1413 made it difficult to quantitate the relative role of the sequences between positions −1413 and −303 as well as to further dissect sequences downstream from position −303. For these reasons, a more extensive set of 5′ deletions with a common 3′ end at position +24 was linked to the CAT gene and analyzed in transient transfections into C2 myotubes and HeLa cells (Fig. 5B). As expected, no significant expression of the MHC constructs was detected in HeLa cells. The results obtained with the CAT constructs in differentiated C2 myotubes were consistent with the data obtained with the MHC minigenes (Fig. 5A) but revealed a more complex functional organization of the 5′-flanking region of the gene between positions −1413 and −102. There are a minimum of three positive regulatory elements separated by two negative ones (Fig. 5B). The two major positive regulatory elements are located between positions −1413 and −1238 and between positions −237 and −173, with a minor one between positions −912 and −884. Each positive element is flanked by a negative one that reduced dramatically its level of expression. No significant expression of any of the deletions downstream of position −173 was detected in these muscle or nonmuscle cells. It is striking that sequences included in the pE237CAT construct were able to produce the maximum level of regulated expression of the gene, which was comparable to that of all the sequences included in the pE1413CAT construct. Yet, it is clear from the behavior of the intermediate deletions between these two points that important regulatory elements are located upstream from position −237. The inhibitory effect of the sequences between positions −671 and −237 is particularly impressive.

Behavior of the upstream regulatory region of the embryonic MHC gene compared with that of a typical enhancer element. An important feature of typical enhancer elements is their ability to stimulate transcription from their target promoter in a relatively orientation- and distance-independent manner (2, 31). To test for the presence in the 5′-end portion of the embryonic MHC gene of regulatory element(s) with the properties of (tissue-specific) enhancers, these sequences were tested in various combinations for their ability to enhance the expression of the homologous and the SV40 promoters in the pA10CAT configuration (33). None of the sequences tested behaved like an enhancer, since they were unable to increase the expression of the

homologous promoter when placed downstream from it or in an inverted orientation (Fig. 6A). Likewise, they did not enhance the expression of the heterologous promoter in any orientation (Fig. 6B). The same result was obtained in muscle and nonmuscle cells.

The DNA fragment corresponding to positions −2600 to −297 in the reversed orientation in the pE303 minigene was silent in the muscle cells (Fig. 6A). In fact, RNA expression of pE303 inv. appeared to be even lower than that of the truncated MHC minigene pE303 itself. The same result was obtained when the fragment between positions −2600 and −297 was inserted, in either orientation, into the first intron (at position +261) of MHC minigene pE303. As expected, the above constructs were not expressed in HeLa cells (data not shown). The same results were obtained when similar fragments were tested in the enhancerless pA10CAT vector (Fig. 6B), with either the entire MHC sequence from positions −1900 to +261 or two subfragments spanning this region (positions −1900 to −303 and −303 to +261).

All of the constructs described above contain a combination of positive and negative *cis*-acting transcriptional elements (Fig. 5B), and the failure to act as enhancers could be due to competition between these elements. To test this possibility, the two fragments that behaved as positive regulators in the deletion experiments were tested individually in the pA10CAT enhancer assay (Fig. 6B). Neither the fragment between positions −1413 and −1141 nor the one between positions −237 and −102 was able to stimulate the expression of the test promoter above control levels in any cell tested. Similar results were obtained when the same sequences were tested with the α- and β-MHC promoters (data not shown). Thus, the 5′ region of the rat embryonic MHC gene does not appear to include sequences able to function like typical enhancer elements. It appears, therefore, that regulatory elements in the 5′ upstream region may exert their function not only in an orientation-dependent fashion but also in a promoter-linked fashion.

Use of the SV40 enhancer to uncover a sequence (positions −173 to −143) that determines muscle-specific expression. From the experiments described above, it is clear that all of the MHC constructs tested exhibit a tissue-regulated and developmentally regulated pattern of expression. Yet from these results it is not possible to determine whether this regulated phenotype is conferred by the upstream sequences, the promoter element, or a combination of both. It is also not possible to test the tissue specificity of the upstream elements, since they do not function in combination with heterologous promoters. However, as shown above, the SV40 enhancer increases the expression of MHC gene construct pE2600, containing 2.6 kb of upstream sequence without affecting its tissue specificity (Fig. 1B and 2). For these reasons it was of interest to determine whether the positive upstream elements could be replaced by the SV40 enhancer and thus ascertain the tissue specificity of the promoter. To this end, serial deletions downstream from position −303 were challenged with the SV40 enhancer located at the 3′ end of the CAT gene and tested in muscle and nonmuscle cells. The SV40 enhancer increased significantly the expression of the pE303CAT and pE173CAT without loss of tissue specificity (Fig. 7). The level of muscle-specific expression of these constructs was comparable to that of pE1413CAT and much higher that of pE303CAT and pE173CAT (Fig. 5B). These results demonstrate that the SV40 enhancer is able to not only substitute for the far and near upstream positive elements but also overcome the role of the negative regulatory element located between positions −303 and −237 when expressed in differentiated muscle cells but not in nonmuscle cells.

The data outlined above indicate that either the MHC gene promoter is tissue and development specific or there is a *cis*-acting element downstream from position −173 that restricts the expression of the MHC gene to differentiated muscle cells. The ability of the SV40 enhancer to rescue the expression of constructs with deleted upstream elements was exploited to distinguish between these two possibilities. Additional deletion constructs with only 142 or 102 bp of flanking DNA and their counterparts with the SV40 enhancer sequences were then tested for expression. No expression of pE142CAT and pE102CAT was detectable (Fig. 5B). In contrast, pE142CATSV1 and pE102CATSV1 were expressed in both C2 myotubes and HeLa cells (Fig. 7). This result clearly demonstrates that the MHC promoter is not muscle specific, since in the presence of the SV40 enhancer it can be expressed in nonmuscle cells.

The difference in activity between pE173CATSV1 and pE142CATSV1, when assayed in HeLa cells, defines the existence of a negative regulatory element that restricts the expression of the gene to muscle cells and cannot be overcome by the SV40 enhancer. Deletion of this element allows the expression of the promoter in HeLa cells under the influence of the SV40 enhancer. This element is distinct from the functional promoter region and is located between positions −173 and −142 of the embryonic MHC gene.

Characteristic sequence features found in the embryonic MHC gene 5′ upstream region. The nucleotide sequence of the 5′ upstream region of the rat embryonic MHC gene was

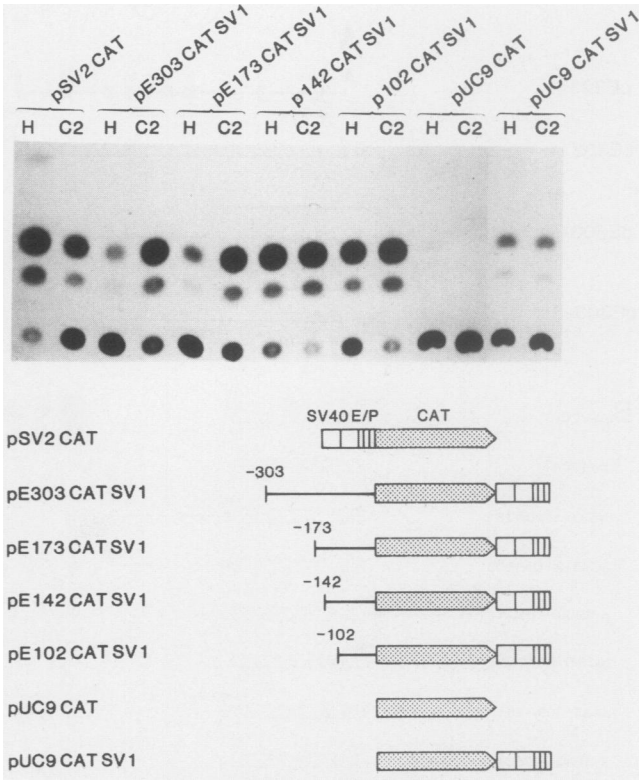


FIG. 7. Tissue specificity of MHC gene constructs carrying the SV40 enhancer elements. CAT assays with extracts from fused C2 (C2mt) and HeLa (H) cells with pE303CATSV1, pE173CATSV1, pE142CATSV1, and pE102CATSV1 (schemes at bottom). pSV2CAT, pUC9CAT, and pUC9CATSV1 were used as controls.

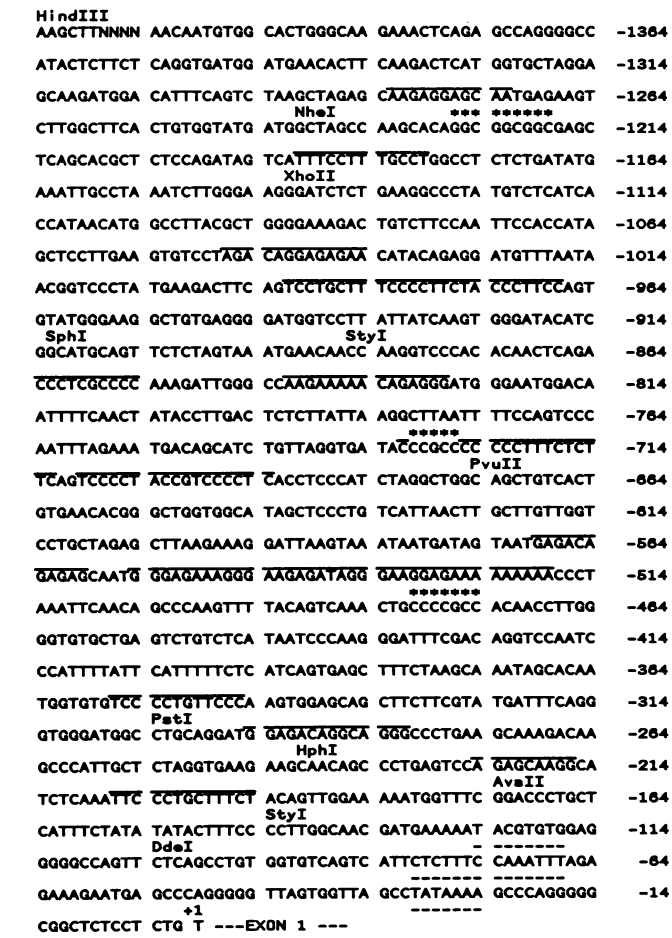


FIG. 8. Nucleotide sequence of the 5′-flanking region of the rat embryonic MHC gene. The sequence of the rat embryonic MHC gene 5′-flanking region is shown between a *Hin*dIII site 1413 nucleotides upstream from the site of transcription initiation and the cap site at position +1. The sequence downstream of the *Pst*I site at position −303 has been reported previously (68). TATA and CCAAT sequences are boxed. Stretches of pyrimidine-rich and of purine-rich sequences are overlined. Three GC box-like sequences are highlighted with stars. The locations of restriction sites used for the deletion constructs are also indicated.

determined (Fig. 8). Noticeable features in the far upstream region include the presence of a CCCAATTTAGAAATGA motif at positions −766 to −750 with an almost identical sequence to the one including the CCAAT box starting at position −74. Three GC-box repeats similar to the core sequence of the binding site for the Sp1 factor (15) are found at positions −1226 to −1217, −730 to 725, and −479 to −474. The sequence CAAGA is repeated four times within the region that included the far upstream regulatory elements (positions −1413 to −1143). Interestingly, these pentameric sequences are preceeded by 7- to 8-nucleotide-long motifs that can form hairpin loop structures. Fourteen stretches of purine (or pyrimidine)-rich sequences are interspersed throughout the 5′ region of the embMHC gene (overlined in Fig. 8). These sequences are 10 to 37 nucleotides long and are interrupted by a single pyrimidine (or purine). This feature is one of the few traits which is common to all three MHC (emb, α, and β) genes that have been characterized so far and is reminiscent of the structure of S1 nuclease-hypersensitive sites that have been mapped in several eucaryotic genes (63).

Comparisons of the 5′ upstream sequences of the rat embryonic MHC gene with the corresponding regions of different rat skeletal and cardiac muscle genes encoding MHCs (38), myosin light chains 1, 3 (67), and 2 (54), α-actin (81), and troponin T (8) revealed the presence of several stretches of up to 65% sequence similarity between these genes. These stretches range in size from 20 bp to over 100 bp and are widely dispersed over the 5′-flanking regions of the different muscle genes but in general do not map in the regulatory regions identified above. Segments of the rat embryonic MHC gene that show significant sequence similarity with upstream sequences from several other muscle genes are found around position −1300, between positions −650 and −550, and between positions −300 and −200. Short stretches of sequence identity are present between the rat embryonic MHC gene at positions −112 to −101 (GGGCCAGTTCTC) and the skeletal muscle α-actin gene at positions −181 to −170 (81) and between the MHC gene at positions −259 to −250 (GTGAAGAAGC) and the myosin light chain 1 gene (67) at positions −427 to −418. Although the existence of muscle gene-specific sequence elements has been postulated on the basis of short sequence similarities found in the 5′-flanking region of several sarcomeric genes (1a, 22, 30, 45, 48), no characteristic sequence motif with a defined functional role has so far been found to be shared by all (or most) sarcomeric protein genes. However, muscle-gene-specific sequence elements may be rather short and only partially homologous to each other and may be differently located with respect to the transcriptional start site of the different genes.

DISCUSSION

Tissue-specific and developmentally regulated expression of MHC gene constructs. In vivo studies have previously shown that the embryonic MHC gene is expressed exclusively in striated skeletal muscle cells (57, 76). The data presently available strongly suggest that this member of the MHC multigene family is first induced during in vivo and in vitro myogenic differentiation of skeletal muscle precursor cells in mammals (43, 57) and is not dependent on innervation. Expression of this gene is rapidly deinduced around the time of birth in most muscle cells, where it is replaced by neonatal and adult isoforms (39, 76). In some muscle fibers such as extraocular muscle, however, this gene continues to be expressed throughout adult life (77). In the slow but not in the fast leg muscles, it is expressed in adult hypothyroid animals (29).

This selective pattern of expression indicates that although neural and hormonal stimuli might play a role in the expression of the embryonic MHC gene, its heterogeneous phenotype in different muscles and fiber types points toward cell-specific factors as the most important regulators of this gene. As a first step toward the identification of such molecules, we have initiated the characterization of the *cis*-acting regulatory elements of embryonic MHC gene that presumably constitute binding sites for such specific factors.

The transient expression studies reported here with MHC gene 5′-flanking sequences in minigene and in CAT reporter gene constructs unambiguously demonstrate that the information necessary for the correct expression of this gene in cultured tissue cells is contained within its 5′-upstream region. The sequences between positions −1413 and +24 are responsible for the correct initiation of transcription and

expression of this gene in differentiated muscle cells but not in undifferentiated and nonmuscle cells. These results are consistent with the information on the endogenous embryonic MHC gene summarized above.

Control of the tissue-specific expression of the embryonic MHC gene by multiple positive and negative *cis*-acting regulatory elements. The deletion, insertion, and recombination mapping analyses indicate that the regulatory elements of the embryonic MHC gene span a minimum of 1,400 bp of DNA and have an unusual structural organization as well as distinct functional properties. Three main regulatory domains have been defined: (i) the promoter element that directs transcription initiation; (ii) the muscle-specific restrictive region that, when present, allows expression only in differentiated muscle cells; and (iii) the upstream regulatory region that determines the level of expression of the promoter and is composed of positive and negative elements.

Lack of muscle specificity of the embryonic MHC promoter. One of the intriguing features of the embryonic MHC gene is that all 5'-flanking regulatory elements detected so far appear to be functional only when linked to the homologous promoter. This promoter, however, is not in itself muscle specific. Sequences downstream from position −102 are sufficient for proper initiation of transcription in in vitro assays (80) and in transient transfections. However, this proximal regulatory region is not an efficient transcriptional unit unless it is provided with additional *cis*-acting regulatory elements. Such elements are included in the natural gene itself (between positions −1413 and −173) or can be substituted for by the enhancer sequences of SV40. All of the constructs that contain the sequences between positions −173 and −142 are expressed only in differentiated muscle cells, independently of whether their level of expression is determined by the homologous upstream sequences or by the SV40 enhancer. In contrast, when sequences downstream from position −142 are used, the SV40 enhancer directs the expression of the MHC promoter in differentiated, undifferentiated, and nonmuscle cells, demonstrating that the promoter is not, by itself, muscle specific. Serial deletions between position −143 and the CAAT box (at position −74) indicate that as long as promoter function can be detected (up to position −80; Y.-T. Yu and B. Nadal-Ginard, unpublished observations) in the presence of the SV40 enhancer, there is no evidence of tissue specificity.

Muscle-restrictive element contained in sequences between positions −173 and −142. Sequences that restrict the specificity of gene expression to certain cell types have been widely documented (16, 18, 56, 70, 73). Most of these sequences belong to the growing group of positive cell-specific enhancers which are active only in the target cells. In general, these cell-specific enhancers are able to confer cell specificity to heterologous promoters as well. On the other hand, a different type of regulation has been described in several genes, including the ones encoding insulin 1 (72), immunoglobulin heavy chain (28), α-fetoprotein (49), and human beta interferon (19). Expression of these genes in all but the expressive cells is inhibited by a negative regulator. In fact, the tissue specific expression of these genes is controlled by the concerted action of both positive and negative elements (16–19, 23, 28, 37, 41, 49, 52, 55, 61, 65, 73).

The behavior of MHC sequences between positions −173 and −142 is reminiscent of that of the negative regulator in the immunoglobulin heavy-chain gene (28), since these MHC sequences do not by themselves appear to play a role in the level of expression of the gene in muscle cells. Indeed, the level of expression of the constructs at positions −142 and −102 under the influence of the SV40 enhancer is similar in muscle and nonmuscle cells. As with the immunoglobulin heavy-chain gene, the presence of the negative regulator sequences is sufficient to block expression in nonpermissive cells, even in the presence of the SV40 enhancer. This latter effect is striking, given the lack of cell specificity of the viral enhancer. Therefore, although not definitively proven by the data presently available, the behavior of the MHC muscle-restrictive sequences suggests a negative mode of regulation in nonmuscle cells. This type of regulation suggests the existence of general *trans* repressors common to all nonmuscle cells that are able to interact with this element. Their mode of action could conceivably block transcriptional activity directly or interfere with the interaction between the MHC regulatory upstream sequences (or the SV40 enhancer) and the promoter. Experiments are in progress to determine whether these sequences might, in addition, be targets for muscle-specific positive regulatory factors and whether their effect is limited to the MHC promoter. Surprisingly, no such mode of regulation has been as yet documented for other muscle-specific genes.

Behavior of positive and negative upstream regulatory elements and determination of the level of embryonic MHC gene expression. The promoter and muscle-restrictive elements discussed above appear to be sufficient to determine the muscle-specific expression of the embryonic MHC gene, but neither element contains the sequences required for efficient expression. The deletion mapping results clearly indicate that the sequences that modulate the level of expression are spread over a long sequence upstream from the muscle-restrictive elements and extending up to −1413 and perhaps further upstream. This latter statement is suggested by the data shown in Fig. 5A documenting a higher level of expression of the constructs extending up to position −2600 than the one at position −1413. These putative elements, however, have not yet been further characterized.

The behavior of the different deletion constructs indicates that the regulation of the level of transcription involves a complex interplay of positive and negative elements that are all functional in the same muscle cell type. From the deletions analyzed, three positive and two negative elements can be discerned. This number is likely to be an underestimate, since it is possible that fragments with an overall positive or negative effect might be further subdivided into combinations of binding sites for positive and negative factors. The data shown in Fig. 5B emphasize the pitfalls of using large restriction fragments for the mapping of regulatory sequences. Analysis of only the pE1413CAT and pE237 constructs would have led to the conclusion that all the sequences involved in the regulation of this gene were located downstream from position −237, whereas in fact even the partial picture presented here is significantly more complex. Little is known at this time about the negative elements in this gene, except that their effect can be neutralized by the positive elements in the homologous gene or in the SV40 enhancer. It is the balance between these types of sequences that determines the level of expression of a given construct. It is noticeable that the SV40 enhancer is able to overcome the effect of the negative regulators in muscle cells but not in nonmuscle cells. The fact that constructs containing the SV40 enhancers do not lose their tissue-specific pattern of expression is explained by the existence of the muscle-restrictive element between positions −172 and −142.

The ability of the SV40 enhancer to replace the upstream

regulatory elements in the embryonic MHC gene raised the possibility that these sequences were bona fide enhancers (16, 18, 56, 70, 73). Based on the results obtained with the enhancer assays with the homologous promoter, the pA10CAT (Fig. 6), and other MHC gene promoters (data not shown), no such enhancer activity has been detected. Furthermore, all results presently available suggest an exclusivity of interaction between flanking sequences and the homologous promoter. This interaction is position and orientation dependent, although the ability of these sequences to productively interact with other promoters in other contexts has not been formally ruled out.

From the data presently available it is not possible to determine whether these positive or negative regulatory elements are muscle specific or can function in other cell types. The availability of MHC promoter deletion constructs able to be expressed in undifferentiated and nonmuscle cells should make it possible to test this alternative as long as no strict position dependence is involved.

It is likely, therefore, that the precise spatial organization of various regulatory *cis*-acting elements in the embryonic MHC gene is a determinant in the binding and interaction of specific transcription-stimulating factors. Indeed, the activation of multiple *cis*-acting sequences appears to be a prerequisite in a number of eucaryotic genes (10, 14, 16, 18, 23, 26, 41, 58, 62, 64, 70). Enhancers and other upstream (or downstream) regulatory sequence elements are targets for the binding of constitutive or tissue-specific protein factors (15, 27, 34, 55, 59, 65). Regulatory elements with a modular organization of different sequence motifs can be independently recognized by separate factors (19, 26, 27) and might be envisaged as condensed complex regulatory regions. The wide range of tissue specificity exhibited by different genes could be then the result of the large number of combinatorial possibilities offered by a relatively small number of distinct sequence elements.

Are regulatory signals likely to be common for all muscle-specific genes? The complex mode of regulation described above for the embryonic MHC gene does not appear to be generalized to other muscle-specific genes. Besides, no sequence motif with a defined functional role appears to be shared by most sarcomeric protein genes. This fact is unexpected, since the major regulatory constraint on these genes is the need for the continuous production of a complex functional unit, the sarcomere, which is constituted by several proteins assembled and organized with a precise stoichiometry. However, most proteins constituting the sarcomere have multiple isoforms which are produced by selective transcription of a given member of a multigene family (16a), as well as in some cases by selective alternative splicing of an individual gene transcript (1, 7). The choice of isoform is tissue specific, reversible, and made in response to developmental, hormonal, and physiological stimuli. In addition, all regulatory signals act on a cell which has irreversibly withdrawn from the cell cycle (50). One of the basic principles that derives from the data presently available is that the various muscle genes are not equally regulated given the same external stimulus. For instance, in higher vertebrates there are two sarcomeric α-actin genes (13, 25, 42, 47), whereas the myosin heavy-chain genes constitute a family of a minimum of seven members (39, 53). Both α-actin isoforms are expressed in skeletal and cardiac tissues (25, 42, 47), but some MHC isoforms are specific for cardiac or fast or slow skeletal myofibers (39). During skeletal muscle development, the α-actins undergo a single isoform transition (25, 42, 47), whereas the MHCs are subjected to a complex set of switches that involve a minimum of five distinct genes (39, 76, 77). In addition, thyroid hormone modulates, in a tissue-specific manner, the expression of all MHC genes (29) but does not affect the expression of a number of other sarcomeric genes (S. Izumo, B. Nadal-Ginard, and V. Mahdavi, Proc. Natl. Acad. Sci. USA, in press).

Most intriguing, however, is that the specific response to developmental, hormonal, and physiological stimuli can vary not only among the different sarcomeric genes or among the members of a gene family but also in a given gene when it is expressed in different muscle types. As mentioned above, expression of the embryonic skeletal MHC gene described here is restricted to the fetal and very early postnatal stages of most skeletal muscle cells (57) but is constitutive in the extraocular muscle fibers (77). This gene is also induced by hypothyroidism in adult slow skeletal myofibers but not in fast muscle fibers (29).

In addition to these examples, taken out of many which illustrate specific regulatory modes for individual genes, it is also possible to observe a response shared by several if not all muscle genes. Such is the case of the reinduction of the fetal isogenes during cardiac hypertrophy (Izumo et al., in press) and the *trans* induction of the myogenic program in various nonmuscle cells when fused to differentiated myotubes (5, 6).

In this context, it is obvious that transcriptional control of muscle gene expression is only one of the steps involved in the production of the muscle phenotype. Production of this complex phenotype is likely to involve multileveled mechanisms, where interactions between a relatively small number of *cis*-acting sequences and *trans*-acting factors are used in a combinatorial mode. This would allow different modes of regulation of the same set of genes in different muscles. Several distinct regulatory regions, some of which might be common to the members of a given multigene family or to several distinct sarcomeric genes and whose existence is shown here for the embryonic MHC gene, would be a requisite for this type of regulation.

ACKNOWLEDGMENTS

This work was supported in part by grants from the National Institute of Health, the Muscular Distrophy Association of America and the American Heart Association. P.F.B. is supported by the Institut National de la Santé et de la Recherche Médicale (France) and by research fellowships from the North Atlantic Treaty Organization and the Claude Pompidou Foundation. V.M. is an established investigator of the American Heart Association. B.N.G. is an investigator of the Howard Hughes Medical Institute.

ADDENDUM IN PROOF

The sequence data reported in this work have been submitted to the EMBL/GenBank libraries under the accession number X05004.

LITERATURE CITED

1. **Andreadis, A., M. E. Gallego, and B. Nadal-Ginard.** 1987. Generation of protein isoform diversity by alternative splicing: mechanistic and biological implications. Annu. Rev. Cell Biol. 3:207–242.

1a. **Baldwin, A. S., Jr., E. L. W. Kittle, and C. P. Emerson, Jr.** 1985. Structure, evolution, and regulation of a fast skeletal muscle troponin I gene. Proc. Natl. Acad. Sci. USA 82:8080–8084.

2. **Banerji, J., S. Rusconi, and W. Schaffner.** 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell 27:299–308.

3. **Benoff, S., and B. Nadal-Ginard.** 1980. Transient induction of poly(A)-short myosin heavy chain messenger RNA during terminal differentiation of L6E9 myoblasts. J. Mol. Biol. **140:** 283–298.
4. **Berk, A., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids. Cell **12:**721–732.
5. **Blau, H. M., C.-P. Chiu, and C. Webster.** 1983. Cytoplasmic activation of human nuclear genes in stable heterocaryons. Cell **32:**1171–1180.
6. **Blau, H. M., G. K. Pavlath, E. C. Hardeman, C. P. Chiu, L. Silberstein, S. G. Webster, S. C. Miller, and C. Webster.** 1985. Plasticity of differentiated state. Science **230:**758–766.
7. **Breitbart, R. E., A. Andreadis, and B. Nadal-Ginard.** 1987. Alternative splicing: an ubiquitous mechanism for the generation of multiple protein isoforms from single genes. Annu. Rev. Biochem. **56:**467–495.
8. **Breitbart, R. E., and B. Nadal-Ginard.** 1986. Complete nucleotide sequence of the fast skeletal troponin T gene. Alternatively spliced exons exhibit unusual interspecies divergence. J. Mol. Biol. **188:**313–324.
9. **Brutlag, D. L., J. Clayton, P. Friedland, and L. H. Kedes.** 1982. SEQ: a nucleotide sequence analysis and recombination system. Nucleic Acids Res. **10:**279–294.
10. **Buetti, E., and B. Kuehnel.** 1986. Distinct sequence elements involved in the glucocorticoid regulation of the mouse mammary tumor virus promoter identified by linker scanning mutagenesis. Mol. Biol. **190:**379–389.
11. **Butler-Browne, G. S., and R. G. Whalen.** 1984. Myosin isozyme transitions occurring during the postnatal development of the rat soleus. Dev. Biol. **102:**324–334.
12. **Caplan, A. I., M. Y. Fiszman, and H. M. Eppenberger.** 1983. Molecular and cell isoforms during development. Science **221:**921–927.
13. **Czosnek, H., U. Nudel, Y. Mayer, M. Shani, P. F. Barker, D. D. Pravtcheva, F. H. Ruddle, and D. Yaffe.** 1982. The genes coding for the muscle contractile proteins, myosin heavy chain, myosin light chain 2, and skeletal muscle actin are located on three different mouse chromosomes. EMBO J. **1:**1299–1305.
14. **Diese, D., and J. W. Casey.** 1986. Two elements in the bovine leukemia virus long terminal repeat that regulate gene expression. Science **231:**1437–1440.
15. **Dynan, W. S., and R. Tjian.** 1985. Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. Nature (London) **316:**774–778.
16. **Edlund, T., M. D. Walker, P. J. Barr, and W. J. Rutter.** 1985. Cell-specific expression of the rat insulin gene: evidence for role of two distinct 5′ flanking elements. Science **230:**912–916.

16a. **Emerson, C., D. A. Fischman, B. Nadal-Ginard, and M. A. Q. Siddiqui (ed.)** 1986. Molecular biology of muscle development. UCLA Symp. Mol. Cell. Biol. New Ser. **29:**1–928.

17. **Foster, J., J. Stafford, and C. Queen.** 1985. An immunoglobulin promoter displays cell-type specificity independently of the enhancer. Nature (London) **315:**423–425.
18. **Godbout, R., R. Ingram, and S. M. Tilghman.** 1986. Multiple regulatory elements in the intergenic region between the α-fetoprotein and albumin genes. Mol. Cell. Biol. **6:**477–487.
19. **Goodbourn, S., H. Burnstein, and T. Maniatis.** 1986. The human β-interferon gene enhancer is under negative control. Cell **45:** 601–610.
20. **Gorman, C. M., L. F. Moffat, and B. H. Howard.** 1982. Recombinant genomes which express chloramphenicol acetyl transferase in mammalian cells. Mol. Cell. Biol. **2:**1044–1051.
21. **Graham, F. L., and A. J. van der Eb.** 1973. A new technique for assay of infectivity of human adenovirus 5 DNA. Virology **52:**456–461.
22. **Grichnik, J. M., D. J. Bergsma, and R. J. Schwartz.** 1986. Tissue restricted and stage specific transcription is maintained within 411 nucleotides flanking the 5′ end of the chicken α skeletal actin gene. Nucleic Acids Res. **14:**1683–1701.
23. **Grosschedl, R., and D. Baltimore.** 1985. Cell-type specificity of immunoglobulin gene expression is regulated by at least three DNA sequence elements. Cell **41:**885–897.
24. **Gruss, P., R. Dhar, and G. Khoury.** 1981. Simian Virus 40 tandem repeated sequences as an element of the early promoter. Proc. Natl. Acad. Sci. USA **78:**943–947.
25. **Gunning, P., P. Ponte, H. Blau, and L. Kedes.** 1983. α-Skeletal and α-cardiac actin genes are coexpressed in adult human skeletal muscle and heart. Mol. Cell. Biol. **3:**1985–1995.
26. **Haslinger, A., and M. Karin.** 1985. Upstream promoter element of the human metallothionein IIA gene can act like an enhancer element. Proc. Natl. Acad. Sci. USA **82:**8572–8576.
27. **Herr, W., and J. Clarke.** 1986. The SV40 enhancer is compared of multiple functional elements that can compensate for one another. Cell **45:**461–470.
28. **Imler, J.-L., C. Lemaire, C. Wasylyk, and B. Wasylyk.** 1987. Negative regulation contributes to tissue specificity of the immunoglobulin heavy-chain enhancer. Mol. Cell. Biol. **7:** 2558–2567.
29. **Izumo, S., B. Nadal-Ginard, and V. Mahdavi.** 1986. All members of the myosin heavy chain multi-gene family respond to thyroid hormone in a highly tissue specific manner. Science **231:**597–600.
30. **Jaynes, J. B., J. S. Chamberlain, J. N. Buskin, J. E. Johnson, and S. D. Hauschka.** 1986. Transcriptional regulation of the muscle creatine kinase gene and regulated expression in transfected mouse myoblasts. Mol. Cell. Biol. **6:**2855–2864.
31. **Khoury, G., and P. Gruss.** 1983. Enhancer elements. Cell **33:** 313–314.
32. **Kropp, K., J. Gulick, and J. Robbins.** 1986. A canonical sequence organization at the 5′-end of the myosin heavy chain genes. J. Biol. Chem. **261:**6613–6618.
33. **Laimins, L. A., G. Khoury, C. Gorman, B. Howard, and P. Gruss.** 1982. Host specific activation of transcription by tandem repeats from simian virus 40 and moloney murine sarcoma virus. Proc. Natl. Acad. Sci. USA **79:**6453–6457.
34. **Learned, R. M., T. K. Learned, M. M. Haltiner, and R. T. Tjian.** 1986. Human rRNA transcription is modulated by the coordinate binding of two factors to an upstream control element. Cell **45:**847–857.
35. **Leinwand, L. A., R. E. K. Fournier, B. Nadal-Ginard, and T. B. Shows.** 1983. Multigene family for sarcomeric myosin heavy chain in mouse and human DNA: localization on a single chromosome. Science **221:**766–769.
36. **Lompre, A.-M., B. Nadal-Ginard, and V. Mahdavi.** 1984. Expression of the cardiac ventricular α- and β-myosin heavy chain genes is developmentally and hormonally regulated. J. Biol. Chem. **259:**6437–6446.
37. **Maeda, H., D. Kitamura, A. Kudo, K. Araki, and T. Watanabe.** 1986. Trans acting nuclear protein responsible for induction of rearranged human immunoglobulin heavy chain gene. Cell **45:** 25–33.
38. **Mahdavi, V., A. P. Chambers, and B. Nadal-Ginard.** 1984. Cardiac α- and β-myosin heavy chain genes are organized in tandem. Proc. Natl. Acad. Sci. USA **82:**2626–2630.
39. **Mahdavi, V., E. E. Strehler, M. Periasamy, D. Wiezorek, S. Izumo, S. Grund, M. A. Strehler, and B. Nadal-Ginard.** 1986. Sarcomeric myosin heavy chain gene family: organization and pattern of expression. UCLA Symp. Mol. Cell. Biol. **29:** 345–361.
40. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
41. **Mason, J. O., G. T. Williams, and M. S. Neuberger.** 1985. Transcription cell type specificity is conferred by an immunoglobulin V_H gene promoter that includes a functional consensus sequence. Cell **41:**479–487.
42. **Mayer, Y., H. Czosnek, P. E. Zeelon, D. Yaffe, and U. Nudel.** 1984. Expression of the genes coding for the skeletal muscle and cardiac actins in the heart. Nucleic Acids Res. **12:**1087–1100.
43. **Medford, R. M., H. T. Nguyen, and B. Nadal-Ginard.** 1983. Transcriptional and cell cycle-mediated regulation of myosin heavy chain gene expression during muscle differentiation. J. Biol. Chem. **258:**11063–11073.
44. **Medford, R. M., R. M. Wydro, H. T. Nguyen, and B. Nadal-Ginard.** 1980. Cytoplasmic processing of myosin heavy chain

messenger RNA: evidence provided by using a recombinant DNA plasmid. Proc. Natl. Acad. Sci. USA **77**:5749–5753.
45. **Melloul, D., B. Aloni, J. Calvo, D. Yaffe, and U. Nudel.** 1984. Developmentally regulated expression of chimeric genes containing muscle actin DNA sequences in transfected myogenic cells. EMBO J. **3**:983–990.
46. **Messing, J., R. Crea, and P. H. Seeburg.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9**:309–321.
47. **Minty, A. J., S. Alonso, M. Caravatti, and M. E. Buckingham.** 1982. A fetal skeletal muscle actin mRNA in the mouse and its identity with cardiac actin mRNA. Cell **30**:185–192.
48. **Minty, A., and L. Kedes.** 1986. Upstream regions of the human cardiac actin gene that modulate its transcription in muscle cells: presence of an evolutionarily-conserved repeated motif. Mol. Cell. Biol. **6**:2125–2136.
49. **Muglia, L., and L. B. Rothman-Denes.** 1986. Cell type-specific negative regulatory element in the control region of the rat α-fetoprotein gene. Proc. Natl. Acad. Sci. USA **83**:7653–7657.
50. **Nadal-Ginard, B.** 1978. Commitment, fusion and biochemical differentiation of a myogenic cell line in the absence of DNA synthesis. Cell. **15**:855–864.
51. **Nadal-Ginard, B., R. M. Medford, H. T. Nguyen, M. Periasamy, R. M. Wydro, D. Hornig, R. M. Gubits, L. I. Garfinkel, D. Wieczorek, E. Bekesi, and V. Mahdavi.** 1982. Structure and regulation of a mammalian sarcomeric myosin heavy chain gene, p. 143–168. *In* M. L. Pearson and H. F. Epstein (ed.), Muscle development: molecular and cellular control. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
52. **Nir, U., M. D. Walker, and W. J. Rutter.** 1986. Regulation of rat insulin 1 gene expression: evidence for negative regulation in non-pancreatic cells. Proc. Natl. Acad. Sci. USA **83**:3180–3184.
53. **Nguyen, H. T., R. M. Gubits, R. M. Wydro, and B. Nadal-Ginard.** 1982. Sarcomeric myosin heavy chain is coded by a highly conserved multigene family. Proc. Natl. Acad. Sci. USA **79**:5230–5234.
54. **Nudel, U., J. M. Calvo, M. Shani, and Z. Levy.** 1984. The nucleotide sequence of a rat myosin light chain 2 gene. Nucleic Acids Res. **12**:7175–7186.
55. **Ohlsson, H., and T. Edlund.** 1986. Sequence-specific imteractions of nuclear factors with the insulin gene enhancer. Cell **45**:35–44.
56. **Ott, M.-O., L. Sperling, P. Herbomel, M. Yaniv, and M. C. Weiss.** 1984. Tissue-specific expression is conferred by a sequence from the 5′ end of the rat albumin gene. EMBO J. **3**:2505–2510.
57. **Periasamy, M., R. M. Wydro, M.-A. Strehler-Page, E. E. Strehler, and B. Nadal-Ginard.** 1985. Characterization of cDNA and genomic sequences corresponding to an embryonic myosin heavy chain. J. Biol. Chem. **260**:15856–15862.
58. **Rhodes, D., and A. Klug.** 1986. An underlying repeat in some trancriptional control sequences corresponding to half a double helical turn of DNA. Cell **46**:123–132.
59. **Robbins, P. D., D. C. Rio, and M. R. Botchan.** 1986. Transactivation of the simian virus 40 enhancer. Mol. Cell. Biol. **6**:1283–1295.
60. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
61. **Sarver, N., R. Muschel, J. C. Byrne, G. Khoury, and P. M. Howley.** 1985. Enhancer-dependent expression of the rat preproinsulin gene in bovine papillomavirus type 1 vectors. Mol. Cell. Biol. **5**:3507–3516.
62. **Schmidt, A., P. Rossi, and B. de Crombrugghe.** 1986. Transcriptional control of the mouse α_2 (1) collagen gene: functional deletion analysis of the promoter and evidence for cell-specific expression. Mol. Cell. Biol. **6**:347–354.
63. **Schon, E., T. Evans, J. Welsh, and A. Efstratiadis.** 1983. Conformation of promoter DNA: fine mapping of S1-hypersensitive sites. Cell **35**:837–848.
64. **Serfling, E., A. Luebbe, K. Dorsch-Haesler, and W. Schaffner.** 1985. Metal dependent SV40 viruses containing inducible enhancers from the upstream region of metallothionein genes. EMBO J. **4**:3851–3859.
65. **Singh, H., R. Sen, D. Baltimore, and P. A. Sharp.** 1986. A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. Nature (London) **319**:154–158.
66. **Strehler, E. E., V. Mahdavi, M. Periasamy, and B. Nadal-Ginard.** 1985. Intron positions are conserved in the 5′ end region of myosin heavy chain genes. J. Biol. Chem. **260**:468–471.
67. **Strehler, E. E., M. Periasamy, M.-A. Strehler-Page, and B. Nadal-Ginard.** 1985. Myosin light-chain 1 and 3 gene has two structurally distinct and differentially regulated promoters evolving at different rates. Mol. Cell. Biol. **5**:3168–3182.
68. **Strehler, E. E., M.-A. Strehler-Page, J.-C. Perriard, M. Periasamy, and B. Nadal-Ginard.** 1986. Complete nucleotide, and encoded amino acid sequence of a mammalian myosin heavy chain gene. J. Mol. Biol. **190**:291–317.
69. **Swynghedauw, B.** 1986. Developmental and functional adaptation of contractile proteins in cardiac and skeletal muscles. Physiol. Rev. **66**:710–771.
70. **Theisen, M., A. Stief, and A. E. Sippel.** 1986. The lysozyme enhancer: cell-specific activation of the chicken lysozyme gene by a far-upstream DNA element. EMBO J. **5**:719–724.
71. **Thomas, P.** 1980. Hybridisation of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77**:5201–5205.
72. **Umeda, P. K., J. E. Levin, A. M. Sinha, L. L. Cribbs, D. S. Darling, D. J. Ende, H.-J. Hsu, and S. Jakovcic.** 1986. Molecular anatomy of cardiac myosin heavy chain genes. UCLA Symp. Mol. Cell. Biol. New Ser. **29**:809–823.
73. **Walker, M. D., T. Edlund, A. M. Boulet, and W. J. Rutter.** 1983. Cell-specific expression controlled by the 5′-flanking region of insulin and chymothrypsim genes. Nature (London) **306**: 557–561.
74. **Weydert, A., P. Daubas, M. Caravatti, A. Minty, G. Bugaisky, A. Cohen, B. Robert, and M. Buckingham.** 1983. Sequential accumulation of mRNAs encoding different myosin heavy chain isoforms during skeletal muscle development in vivo detected with a recombinant plasmid identified as coding for an adult fast myosin heavy chain from mouse skeletal muscle. J. Biol. Chem. **258**:13867–13874.
75. **Weydert, A., P. Daubas, I. Lazaridis, P. Barton, I. Garner, D. P. Leader, F. Bonhomme, J. Catalan, D. Simon, J.-L. Guenet, F. Gros, and M. E. Buckingham.** 1985. Genes for skeletal muscle myosin heavy chains are clustered and are not located on the same mouse chromosome as a cardiac myosin heavy chain. Proc. Natl. Acad. Sci. USA **82**:7183–7187.
76. **Whalen, R. G., S. M. Sell, G. S. Butler-Browne, K. Schwartz, P. Bouveret, and I. Pinset-Harstrom.** 1981. Three myosin heavy chain isozymes appear sequentially in rat muscle development. Nature (London) **292**:805–809.
77. **Wieczorek, D. F., M. Periasamy, G. S. Butler-Browne, R. G. Whalen, and B. Nadal-Ginard.** 1985. Co-expression of multiple myosin heavy chain genes, in addition to a tissue-specific one, in extraocular musculature. J. Cell Biol. **101**:618–629.
78. **Wydro, R. M., H. T. Nguyen, R. M. Gubits, and B. Nadal-Ginard.** 1983. Characterization of sarcomeric myosin heavy chain genes. J. Biol. Chem. **258**:670–678.
79. **Yaffe, D., and O. Saxel.** 1977. Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscles. Nature (London) **270**:725–727.
80. **Yu, Y.-T., and B. Nadal-Ginard.** 1987. Trans-acting factor(s) interacting with rat skeletal embryonic myosin heavy chain promoters is (are) found in both HeLa and fused mouse C2 cells. J. Cell. Biochem. Suppl. C **11**:118.
81. **Zakut, R., M. Shani, D. Givol, S. Neuman, D. Yaffe, and U. Nudel.** 1982. Nucleotide sequence of the rat skeletal muscle actin gene. Nature (London) **298**:857–859.