MOLECULAR AND CELLULAR BIOLOGY, Nov. 1985, p. 3168–3182
0270-7306/85/113168-15$02.00/0

Vol. 5, No. 11

Myosin Light-Chain 1 and 3 Gene Has Two Structurally Distinct and Differentially Regulated Promoters Evolving at Different Rates

EMANUEL E. STREHLER, MUTHU PERIASAMY, MARIE-ANTOINETTE STREHLER-PAGE, AND BERNARDO NADAL-GINARD*

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

Received 14 February 1985/Accepted 19 August 1985

DNA fragments located 10 kilobases apart in the genome and containing, respectively, the first myosin light chain 1 ($MLC1_f$) and the first myosin light chain 3 ($MLC3_f$) specific exon of the rat myosin light chain 1 and 3 gene, together with several hundred base pairs of upstream flanking sequences, have been shown in runoff in vitro transcription assays to direct initiation of transcription at the cap sites of $MLC1_f$ and $MLC3_f$ mRNAs used in vivo. These results establish the presence of two separate, functional promoters within that gene. A comparison of the nucleotide sequence of the rat $MLC1_f/3_f$ gene with the corresponding sequences from mouse and chicken shows that: (i) the $MLC1_f$ promoter regions have been highly conserved up to position −150 from the cap site while the $MLC3_f$ promoter regions display a very poor degree of homology and even the absence or poor conservation of typical eucaryotic promoter elements such as TATA and CAT boxes; (ii) the exon/intron structure of this gene has been completely conserved in the three species; and (iii) corresponding exons, except for the regions encoding most of the 5′ and 3′ untranslated sequences, show >75% homology while corresponding introns are similar in size but considerably divergent in sequence. The above findings indicate that the overall structure of the $MLC1_f/3_f$ genes has been maintained between avian and mammalian species and that these genes contain two functional and widely spaced promoters. The fact that the structures of the alkali light chain gene from *Drosophila melanogaster* and of other related genes of the troponin C supergene family resemble a $MLC3_f$ gene without an upstream promoter and first exon strongly suggests that the present-day $MLC1_f/3_f$ genes of higher vertebrates arose from a primordial alkali light chain gene through the addition of a far-upstream $MLC1_f$-specific promoter and first exon. The two promoters have evolved at different rates, with the $MLC1_f$ promoter being more conserved than the $MLC3_f$ promoter. This discrepant evolutionary rate might reflect different mechanisms of promoter activation for the transcription of $MLC1_f$ and $MLC3_f$ RNA.

Myosin, the major protein component of striated muscle tissues, consists of two heavy and four light chains. The light chains are divided into two classes which, in vertebrates, are termed alkali- and dithiobis-nitrobenzoic acid-removable light chains, or light chains 1 and 2, respectively (67). Both the heavy and the light myosin chains exist as multiple isoproteins that are tissue specific or developmental stage specific or both (22, 24, 48, 71, 72). Vertebrate fast skeletal muscles contain two alkali myosin light chain isoforms, termed $MLC1_f$ and $MLC3_f$ (M_r = 21,000 for $MLC1_f$; M_r = 16,000 for $MLC3_f$). Amino acid sequences available for $MLC1_f$ and $MLC3_f$ from different higher vertebrate species (21, 22, 37, 52) show that these two proteins have identical sequences of 141 carboxy-terminal amino acids and differ from each other only at the amino-terminal ends, where $MLC1_f$ and $MLC3_f$ have additional isoform-specific sequences (49 and 8 amino acids, respectively). The mRNAs for $MLC1_f$ and $MLC3_f$ have been shown to be identical in their sequences specifying the 141 carboxyterminal amino acids and the 3′ untranslated regions in both the chicken (42) and the rat (52). Based on these studies it was speculated that the two alkali light chain isoforms might be encoded by the same gene (37, 56). Recently, a single gene locus has indeed been identified to carry all the coding and noncoding sequences found in $MLC1_f$ and $MLC3_f$ mRNAs (43, 52, 55). The unique organization of $MLC1_f$- and $MLC3_f$-specific exons necessitates differential splicing of the primary transcript(s) to produce mature $MLC1_f$ and $MLC3_f$ mRNAs (43, 52). It has also been suggested but not yet formally proven by functional assays that the gene contains two different transcription initiation sites specific for either $MLC1_f$ or $MLC3_f$ mRNA transcription (43, 52). This hypothesis is supported by the pattern of expression of this gene showing that the biosynthesis of its two mRNA species is differentially regulated in a developmental- and tissue-specific manner (52).

Two other examples of eucaryotic genes yielding two different mRNAs by alternative transcription from two separate promoters have recently been characterized. They include the genes for α-amylase-1^a of the mouse (59, 74) and for alcohol dehydrogenase of *Drosophila melanogaster* (5). The two promoters of these genes, although directing the synthesis of mRNAs coding for identical proteins, are differentially used during development (5) and regulated in a tissue-specific manner (59, 74).

The work described here demonstrates unequivocally that the single $MLC1_f/3_f$ gene does, in fact, contain two functional promoters, which are separated from each other by approximately 10 kilobases (kb) of DNA. In addition, we report the nucleotide sequence organization of the 5′ flanking regions corresponding to these two promoters, of all the exons, and of parts of the adjoining introns of the rat

* Corresponding author.

MLC1f/3f gene and compare them with the corresponding sequences of the chicken and mouse MLC1f/3f gene. Interestingly, although the exon-intron organization is totally conserved in these genes, the flanking sequences (promoter region) upstream of the first MLC1f-specific exon show a striking degree of similarity in the rat and mouse as compared with the chicken genes, while the sequences upstream of the first MLC3f-specific exon are significantly more divergent, suggesting differences in the selective pressure acting upon the two promoters during evolution and possibly different mechanisms of regulation of promoter activation.

MATERIALS AND METHODS

Restriction endonucleases and DNA-modifying enzymes were purchased from New England Biolabs, Inc., Beverly, Mass., New England Nuclear Corp., Boston, Mass., and Bethesda Research Laboratories, Inc., Gaithersburg, Md. Double-stranded M13 sequencing vectors M13mp8 and M13mp9 were obtained from New England Nuclear Corp., and the M13 sequencing kit was from New England Biolabs, Inc. [α-^{32}P]dATP (>600 Ci/mmol), [α-^{32}P]UTP (>600 Ci/mmol), and [γ-^{32}P]ATP (approximately 3,000 Ci/mmol) were from New England Nuclear Corp.

Source of MLC1f/3f genomic DNA. The isolation and partial characterization of the rat MLC1f/3f gene sequences have been described previously (52). λ Phage DNA was prepared by the method of Maniatis et al. (33).

Preparation of templates for in vitro transcription. A plasmid containing the adenovirus major late promoter (pFLBH [19]), kindly provided by U. Hansen, Dana Farber Cancer Inst., Boston, served as a control template. Since this plasmid has a single *Pst*I restriction site 974 base pairs (bp) downstream of the major transcription initiation site, a *Pst*I digest of pFLBH was used for runoff in vitro transcription assays.

Transcription driven by the MLC1f promoter was tested by using an M13 clone carrying a 2.2-kb *Eco*RI-*Bam*HI insert containing the first exon and approximately 1.6 kb of 5′ flanking DNA (see Fig. 1). MLC3f promoter activity was tested by using two different constructs containing a 1.7-kb *Pst*I-*Sac*I insert and a 1.2-kb *Pst*I-*Eco*RI insert, respectively, from the MLC1f/3f gene in pUC18 or M13 mp8 vectors. Both the 1.7- and 1.2-kb fragments contain the first MLC3f-specific exon and approximately 0.8 kb of 5′ flanking DNA; in addition, the 1.7-kb fragment includes MLC3f-specific exon 3 (see Fig. 1).

Double-stranded template DNA was prepared by the modified alkali lysis method of Birnboim and Doly (7) as described by Maniatis et al. (33) and purified through two cycles of centrifugation to equilibrium in cesium chloride-ethidium bromide density gradients (33). For runoff in vitro transcription assays, the circular DNA constructs were linearized at unique restriction sites downstream of the putative promoters (see Fig. 1) and precipitated with ethanol. The pellets were dried and then suspended in 10 mM Tris hydrochloride (pH 8)–1 mM EDTA at a concentration of 1 mg/ml.

In vitro transcription assays. The whole-cell HeLa extract was prepared by the method of Manley et al. (34). Transcription reaction volumes of 25 μl were incubated at 30°C for 60 min. For runoff transcription, the HeLa cell extract constituted 50% of the total volume, and the final reaction mixture contained 10 mM HEPES *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (pH 7.5), 50 mM KCl, 6 mM $MgCl_2$, 0.05 mM EDTA, 1 mM dithiothreitol, 9% glycerol, 4 mM creatine phosphate, 200 μM ATP, 200 μM CTP, 200 μM GTP, 40 μM UTP containing 15 to 20 μCi of [α-^{32}P]UTP, and 20 to 120 μg of template DNA per ml. Control reaction mixtures contained, in addition, 1 μg of α-amanitin per ml.

The RNA synthesized in vitro was purified essentially as described by Manley et al. (34, 35): an equal volume (25 μl) of 100 mM Tris hydrochloride (pH 7.5)–20 mM EDTA–2% sodium dodecyl sulfate–400 μg of proteinase K per ml was added to the in vitro transcription reaction mixture and incubated at 37°C for 15 min. The mixture was diluted to 250 μl with 10 mM Tris hydrochloride (pH 8)–10 mM EDTA–0.5% sodium dodecyl sulfate and extracted with phenol-sevac (24 parts chloroform, 1 part isoamyl alcohol) followed by sevac alone. RNA was precipitated by the addition of 1 volume of 2 M ammonium acetate (pH 8) and 2 volumes of ethanol. The pellet was redissolved in 10 mM Tris hydrochloride (pH 8)–1 mM EDTA–0.2% sodium dodecyl sulfate and precipitated again with 2.5 volumes of ethanol in the presence of 0.3 M sodium acetate (pH 5.5). The ethanol precipitation was repeated, and the pellet was washed with 70% ethanol, dried, and finally suspended in 20 μl of 85% formamide–0.05% dyes (bromophenol blue and xylene cyanol). The samples were heated at 90°C for 2 to 3 min, quick-cooled on ice, and loaded (10 μl per slot) onto pre-electrophoresed 6% polyacrylamide–8.3 M urea gels.

S1 nuclease mapping. Total cytoplasmic RNA from L_6E_9 myoblasts and myotubes (44) was isolated as described previously (2). Total cellular RNA from rat skeletal muscle was prepared by a hot phenol procedure described previously (63). S1 nuclease mapping experiments were performed by a modification of the method of Berk and Sharp (6) as described by Mahdavi et al. (32).

For S1 nuclease mapping of RNA synthesized in vitro, a single-stranded, 5′-end-labeled probe (39) was used and hybridized in aqueous conditions (31) to unlabeled RNA. RNA was synthesized and purified as described above except that 200 μM of cold UTP replaced [α-^{32}P]UTP in the in vitro transcription reactions. S1 nuclease digestion was performed with 250 U of the enzyme at 30°C for 1 h as described previously (31).

DNA sequencing. λ phage DNA containing MLC1f/3f sequences (52) was digested with appropriate restriction enzymes under the conditions suggested by the supplier, and individual DNA fragments were isolated from low-melting-point agarose gels as described previously (33) and subcloned into M13mp sequencing vectors (41). Single-stranded M13 template DNA was prepared (41), and the sequence was determined by the dideoxy chain termination method (58) with the commercially available universal M13 pentadecamer primer to prime second-strand synthesis. The sequence of the middle part of exon 1 was confirmed by using a synthetic oligodeoxynucleotide primer made on a Sam One automatic DNA synthesizer (Biosearch Inc., San Raphael, Calif.) and purified by gel electrophoresis as suggested by the supplier. For sequencing by the method of Maxam and Gilbert (38, 39) appropriate fragments were dephosphorylated with bacterial alkaline phosphatase and 5′ end labeled with [γ-^{32}P]ATP by using T4 polynucleotide kinase. These fragments were internally cleaved to produce single radiolabeled ends and then subjected to base-specific chemical cleavage reactions and polyacrylamide-urea gel electrophoresis.

Sequence analysis. Nucleotide sequences were analyzed on a Radio Shack TRS-80 model II micro computer with programs of the SEQ sequence analysis system by J.

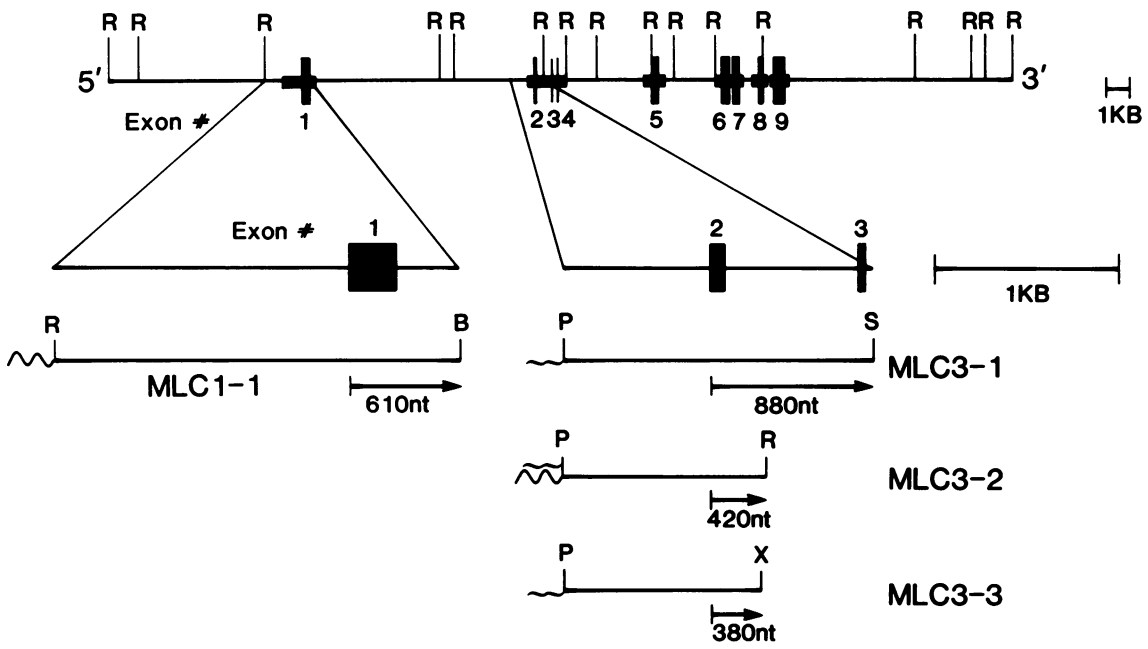

FIG. 1. Structure of the rat MLC1$_f$/3$_f$ gene and source of templates used for in vitro transcription assays. A structural map (including the *Eco*RI restriction sites [R]) of the gene is shown at the top of the figure. Exons (black boxes) are numbered in the 5'-to-3' direction; their relative size is not at scale due to the extremely small size of exons 3 and 4. Thick black horizontal bars delineate sequenced regions. The genomic region subcloned into M13 or pUC vectors used for in vitro transcription assays is shown at an enlarged scale at the bottom, together with schematic representations of the linearized templates and their expected runoff transcripts. Symbols and abbreviations: B, *Bam*HI; P, *Pst*I; S, *Sac*I; X, *Xba*I; ∿∿, pUC18 vector, sequence; ᴍ, M13mp8 vector sequence; nt, nucleotides.

Clayton, P. Friedland, L. Kedes, and D. Brutlag from Stanford University.

RESULTS

Characterization of two functional promoters in the MLC1$_f$/3$_f$ gene. Based on preliminary sequence information it has been suggested that transcription of MLC1$_f$ and MLC3$_f$ mRNAs is directed from two separate promoters located 5' to exons 1 and 2, respectively (43, 52) (Fig. 1). To localize the putative start of transcription sites for MLC1$_f$ and MLC3$_f$ mRNA, we performed an S1 nuclease mapping experiment with total RNA isolated from the rat myogenic cell line L_6E_9 (44) and from rat skeletal muscles at different stages of development (Fig. 2). MLC1$_f$ mRNA was assayed by using a 5' end-labeled 550-bp genomic *Msp*I fragment as probe (Fig. 2A, bottom). This probe, cut within the first exon of the gene and extending towards the 5' direction, yielded a single protected fragment of approximately 180 bp when RNA from late fetal, neonatal, and adult skeletal muscles was used (Fig. 2A, lanes 4 through 6). From this result and from sequencing data on the MLC1$_f$-specific cDNA clones (52) and on the genomic DNA (see below), the MLC1$_f$-specific cap site has been assigned to a position 126 nucleotides upstream of the AUG initiation codon of MLC1$_f$ mRNA (see Fig. 5). As expected from earlier observations (52), both L_6E_9 myoblasts and L_6E_9 myotubes did not accumulate detectable levels of MLC1$_f$ mRNA (Fig. 2A, lanes 2 and 3).

A similar S1 mapping experiment was performed to define the 5' start site of MLC3$_f$ mRNA; a 3.8-kb *Eco*RI-*Eco*RI genomic fragment that includes the first MLC3$_f$-specific exon, exon 2 of the MLC1$_f$/3$_f$ gene (Fig. 2B, bottom), was used as a probe. The probe (derived from an M13 clone) was uniformly labeled on its anti-mRNA strand and hybridized to the same RNA species used for mapping the MLC1$_f$ mRNA 5' end. RNA from neonatal and adult skeletal muscles and, to a lesser extent, from late fetal skeletal muscle protected approximately 70 nucleotides of the DNA probe (Fig. 2B, lanes 4 through 6). There is a considerable size heterogeneity of protected fragments, due to S1 nuclease nibbling on the relatively short DNA-RNA hybrids or to heterogeneity in the MLC3$_f$ RNA transcription start site or both. Again, no protected fragments were observed with RNA from L_6E_9 myoblasts or myotubes (Fig. 2B, lanes 2 and 3), in agreement with the earlier observation that these cells express very low levels of MLC3$_f$ mRNA (52).

The fact that no canonical acceptor splice site precedes exon 2 (52) (see Fig. 5) and that MLC3$_f$ cDNA sequences do not extend beyond the sequence found in this exon (52) supports the assumption that the 5' end of exon 2 indeed corresponds to the cap site of MLC3$_f$ mRNA. However, these data do not formally prove the existence of an independent functional promoter specific for MLC3$_f$ mRNA transcription. This is more so because the presence of an internal promoter downstream from amino acid coding sequences has not yet been documented for any other gene. Therefore, to test this assumption, we performed in vitro transcription assays with different genomic DNA fragments that include either the first MLC1$_f$-specific exon or the first MLC3$_f$-specific exon together with several hundred bp of 5' flanking sequences as templates (Fig. 1). The genomic DNA fragments, cloned in M13 or pUC vectors, were linearized at unique restriction sites downstream of the putative MLC promoters and tested for their ability to direct initiation of RNA transcription in a HeLa whole-cell extract. A MLC1$_f$-specific template containing the first exon together with approximately 1.6 kb of 5' flanking sequence and 364 bp of 3' intervening sequence yielded a 610-nucleotide transcript

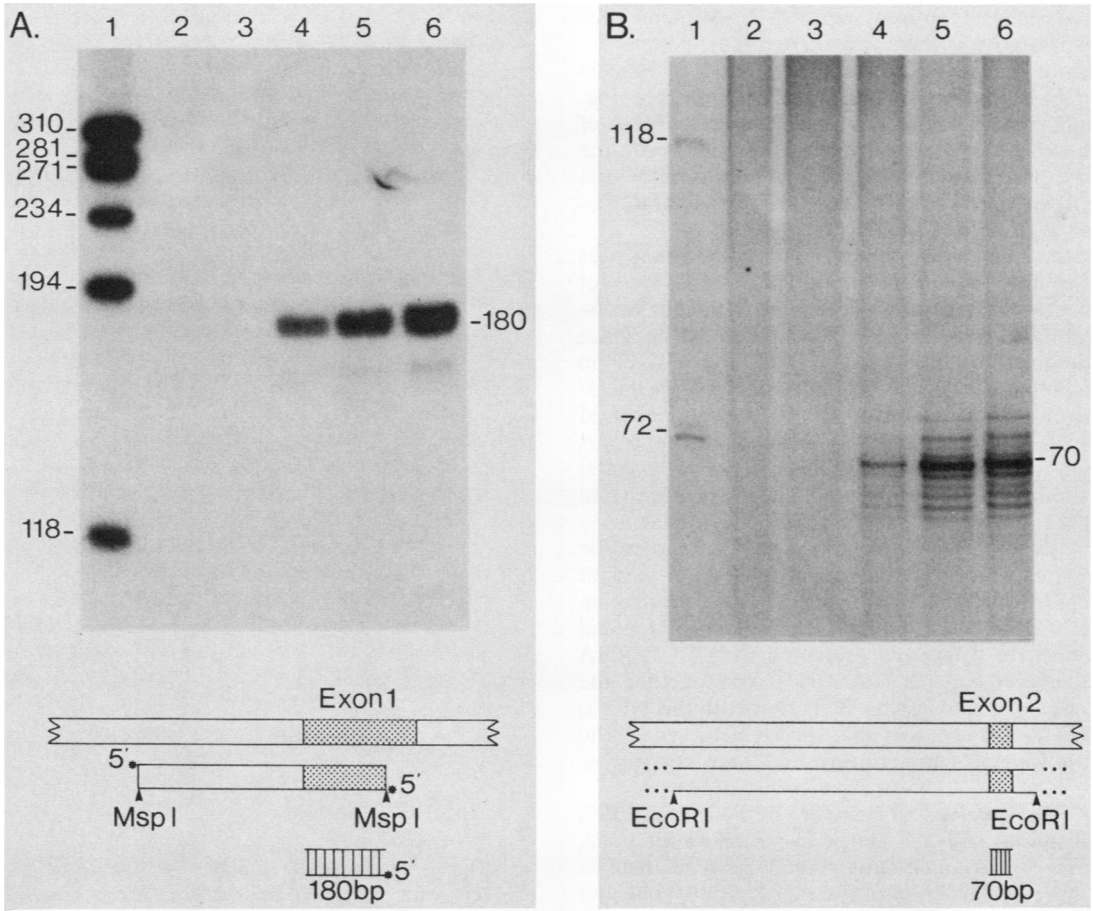

FIG. 2. Determination of the 5′ ends of $MLC1_f$ and $MLC3_f$ mRNAs. S1 nuclease mapping experiments were performed with total RNA from L_6E_9 myoblasts (lanes 2), L_6E_9 myotubes (lanes 3), fetal (20 days of gestation) rat skeletal muscle (lanes 4), and from 2 day old neonatal (lanes 5) and adult skeletal muscles (lanes 6). Lanes 1, size markers with sizes indicated in base pairs. (A) $MLC1_f$ mRNA was probed with a 550-bp 5′-end-labeled *Msp*I-*Msp*I genomic fragment ending within exon 1 of the $MLC1_f/3_f$ gene (scheme on bottom) and yielded a 180-bp protected fragment upon S1 nuclease digestion. (B) $MLC3_f$ mRNA was probed with a 3.8-kb *Eco*RI-*Eco*RI genomic fragment subcloned into M13mp8 (:::) and uniformly labeled on its anti-mRNA strand. The probe (scheme at bottom) includes exon 2 of the $MLC1_f/3_f$ gene and yielded a major protected fragment of approximately 70 bp upon hybridization to $MLC3_f$ mRNA and S1 nuclease digestion.

corresponding to the expected size of a runoff transcript starting at the putative cap site of $MLC1_f$ mRNA (Fig. 3, lanes 2 and 3). At a higher template concentration, an additional RNA transcript of approximately 700 nucleotides was also detected (Fig. 3, lane 3); however, this RNA species was also produced in the presence of 1 μg of α-amanitin per ml whereas synthesis of the 610-nucleotide RNA was completely inhibited under these conditions (Fig. 3, lane 4), suggesting that only the shorter transcript is a bonafide polymerase II product.

To test for $MLC3_f$-specific promoter activity, a template containing exon 2 (the first $MLC3_f$-specific exon) together with approximately 0.8 kb of 5′ flanking sequence and 331 bp of 3′ intervening sequence was used for runoff in vitro transcription. This template produced an approximately 420-nucleotide major RNA species that corresponds to the size expected for a runoff RNA initiated at the 5′ end of exon 2, i.e., at the cap site of $MLC3_f$ mRNA. Synthesis of this RNA species was completely inhibited in the presence of α-amanitin (Fig. 3, lane 7), showing that it represents a bona fide polymerase II transcript. Additional α-amanitin-sensitive RNA transcripts of similar but heterogeneous size were also made from the same template, particularly at a higher than optimal template concentration (Fig. 3, lane 6). These RNAs are either due to degradation products of the 420-nucleotide RNA or to RNA products transcribed from initiation sites several bases downstream of the major site of transcription initiation.

The above results demonstrate unambiguously that both the first $MLC1_f$-specific exon and the first $MLC3_f$-specific exon of the $MLC1_f/3_f$ gene are preceded by functional promoters which can independently direct initiation of transcription in vitro. To compare the relative strengths of these two promoters with a promoter known to be highly active in vitro, runoff transcription was also performed with *Pst*I-cut pFLBH DNA (19) containing the adenovirus major late promoter. This template yielded the expected RNA transcript of 974 nucleotides (19) (Fig. 3, lane 8) and its promoter was at least 10- to 20-fold stronger than either of the MLC promoters under the conditions used in these experiments. Of the two $MLC1_f/3_f$ gene promoters, we have consistently observed that the internal $MLC3_f$ promoter is more active in vitro than its $MLC1_f$-specific counterpart (compare, e.g., Fig. 3, lanes 2 and 5).

To further demonstrate that the internal ($MLC3_f$-specific) promoter, located approximately 10 kb downstream of the 5′ end of the $MLC1_f/3_f$ gene, is indeed capable of directing correct initiation of transcription, we tested several addi-

tional $MLC3_f$-specific templates (Fig. 4A). All templates contained the first $MLC3_f$-specific exon (exon 2) and approximately 0.8 kb of 5′ flanking sequence, but they differed in the length of DNA downstream of the $MLC3_f$ promoter (Fig. 1). Each template produced an α-amanitin-sensitive RNA of the exact size expected for a runoff transcript initiated at the $MLC3_f$ cap site (Fig. 4A), thus confirming the presence of a functional polymerase II-dependent, $MLC3_f$-specific promoter in front of exon 2 of the $MLC1_f/3_f$ gene.

Because of the lack of convenient unique restriction sites downstream of the $MLC1_f$ promoter, the same type of experiment could not be easily performed to further investigate this promoter. Instead, we performed an S1 nuclease mapping experiment with unlabeled RNA synthesized in vitro from template MLC1-1 (Fig. 1 and 3) with the 5′ end-labeled anti-mRNA strand of a 550-bp genomic *Msp*I fragment (Fig. 2A, bottom) as probe. The same probe was also hybridized to total RNA isolated from adult rat skeletal muscle, and the hybrids were digested with S1 nuclease and run on an 8% denaturing polyacrylamide gel next to a Maxam and Gilbert (39) sequencing ladder of the probe alone. The major protected fragments with RNA from rat skeletal muscle (in vivo RNA) and with RNA synthesized in vitro from $MLC1_f$-specific template MLC1-1 were identical and corresponded to fragments protected by $MLC1_f$ RNA initiated at the in vivo cap site (Fig. 4B). Taken together, the above results demonstrate conclusively that both the $MLC1_f$ and the $MLC3_f$ promoters are able to direct, in vitro, the accurate initiation of transcription of their respective mRNAs.

Determination of the detailed structure of the rat $MLC1_f/3_f$ gene and comparison with the mouse and chicken $MLC1_f/3_f$ and the *D. melanogaster* alkali light chain gene structures. To further determine the structure of the rat $MLC1_f/3_f$ gene and its two promoters and to compare it with the recently reported data on the $MLC1_f/3_f$ gene from chicken (43) and mouse (55) and on the single alkali light chain gene from *D. melanogaster* (18), the nucleotide sequences of the promoter regions and of all the exons and parts of the adjoining introns were determined as described in Materials and Methods. The results are displayed in Fig. 1 and 5.

Typical features of eucaryotic promoter regions like the TATA and the CAT boxes (10) are found 23 and 71 bp upstream of the $MLC1_f$-specific transcription initiation site of the rat $MLC1_f/3_f$ gene (Fig. 5; see also reference 52). These features were not clearly recognized in the sequence upstream of the $MLC3_f$ transcription initiation site (Fig. 5). However, there were TATA- and CAT-like sequences at −21 and −46, respectively. With one exception, the sequences found in the exons corresponded exactly to the previously published cDNA sequences for $MLC1_f$ and $MLC3_f$ (52). The exception concerns the middle region of exon 1, where two amino acid codons, specifying an alanine and a proline residue (the sequence GCT CCT encoding amino acids 13 and 14 (or 15 and 16) of the published $MLC1_f$ cDNA [52]), were deleted in the genomic DNA when compared with the sequence represented in the $MLC1_f$ cDNA. Because all the other sequences, including the 5′ and 3′ untranslated regions of the $MLC1_f$ mRNA, were identical in the genomic and cDNA clones, the observed differences are most likely due to allelic polymorphism of the $MLC1_f/3_f$ gene. Indeed, S1 nuclease protection experiments (D. Wieczorek, unpublished results) show the presence of two forms of $MLC1_f$ mRNA diverging exactly at the position described above in the fast skeletal muscles of some individual rats. Altogether, the rat $MLC1_f/3_f$ gene contains nine

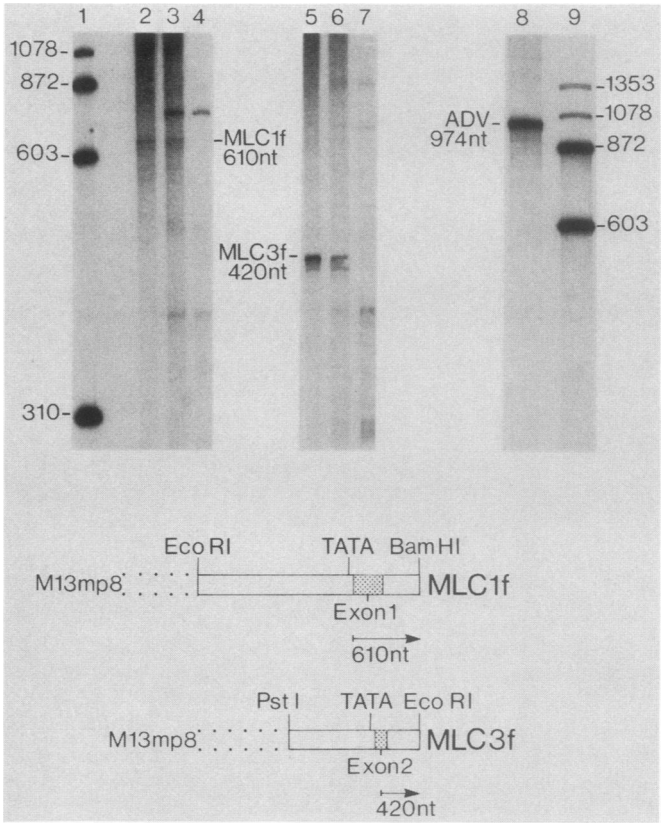

FIG. 3. Runoff in vitro transcription to test for $MLC1_f$- and $MLC3_f$-specific promoter activity. Lanes 1 and 9, Molecular size markers with sizes indicated in base pairs. Lanes 2 through 4, $MLC1_f$-specific template MLC1-1 (see scheme on bottom and Fig. 1), linearized at the unique *Bam*HI site approximately 610 nucleotides (nt) downstream of the $MLC1_f$ mRNA cap site, was used at a concentration of 50 μg/ml (lane 2) or 100 μg/ml (lanes 3 and 4) for in vitro transcription assays in the absence (lanes 2 and 3) or presence (lane 4) of 1 μg of α-amanitin per ml. A $MLC1_f$-specific promoter (TATA) therefore directs the synthesis of a 610-nucleotide runoff RNA in vitro. Lanes 5 through 7, $MLC3_f$-specific template MLC3-2 (see scheme on bottom and Fig. 1), linearized at the unique *Eco*RI-site approximately 420 nucleotides downstream of the $MLC3_f$ mRNA cap site, was used at a concentration of 50 μg/ml (lane 5) or 100 μg/ml (lanes 6 and 7) for in vitro transcription assays in the absence (lanes 5 and 6) or presence (lane 7) of 1 μg of α-amanitin per ml. A $MLC3_f$-specific promoter (TATA) therefore directs the synthesis of a 420-nucleotide runoff RNA in vitro. Lane 8, In vitro transcription with 20 μg of *Pst*I-restricted plasmid pFLBH (19) per ml as template. This template contains the adenovirus major late promoter (ADV) and yields a 974-nucleotide runoff RNA in vitro.

exons spread over more than 20 kb of DNA (Fig. 1, 5, and 6) with the $MLC1_f$- and $MLC3_f$-specific exons arranged in a sequence that necessitates differential splicing of the primary transcript(s) to produce each of the two mature mRNAs (52).

The structural organization and partial DNA sequence of the chicken (43) and the mouse (55) $MLC1_f/3_f$ genes and of the alkali light chain gene from *D. melanogaster* (18) have recently been published. While the chicken gene is somewhat smaller than its rat and mouse counterparts (18 versus 21 kb), the arrangement and the number of the $MLC1_f$-specific, $MLC3_f$-specific, and common exons are identical in all three vertebrate genes (Fig. 6). Chicken $MLC1_f$ is a protein of 191 amino acids (43) as compared with the 188 and

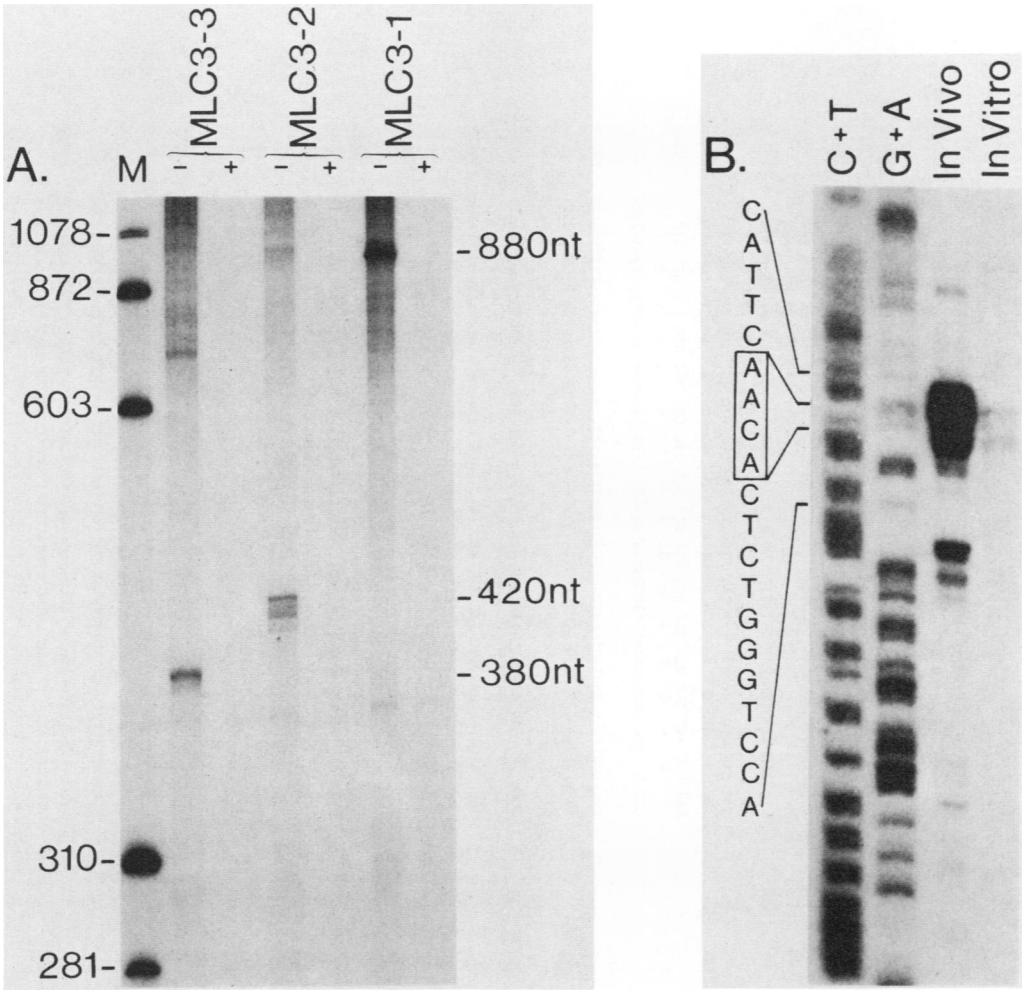

FIG. 4. $MLC1_f$- and $MLC3_f$-specific in vitro transcription initiates at the in vivo cap sites of the corresponding mRNAs. (A) Runoff in vitro transcription assays were performed with 75 μg of $MLC3_f$-specific templates per ml linearized 380 (template MLC3-3), 420 (MLC3-2), and 880 (MLC3-1) nucleotides (nt) downstream of the $MLC3_f$ mRNA start site in the absence (−) or presence (+) of 1 μg of α-amanitin per ml. For details of template constructions, see Fig. 1. Lane M contains size markers indicated in base pairs. (B) Total RNA from adult rat skeletal muscle (in vivo RNA) and RNA transcribed in vitro from $MLC1_f$-specific template MLC1-1 (see Fig. 1) was hybridized to the 5′-end-labeled anti-mRNA strand of a 550-bp genomic *Msp*I fragment (see Fig. 2A, bottom), and the hybrids were digested with S1 nuclease and run on a 8% denaturing polyacrylamide gel adjacent to a sequencing ladder of the probe after pyrimidine-specific (C+T) or purine-specific (A+G) chemical cleavage (39). The sequence around the $MLC1_f$ mRNA cap-site (boxed) is shown on the left; for the complete sequence of the fragment used as probe, see Fig. 5, nucleotides 437 through 989.

190 amino acids of the rat protein as deduced from the genomic and cDNA sequences, respectively (Fig. 5 and reference 52), and to the 187 amino acids reported for the mouse protein (55). This size heterogeneity is due to differences among the three $MLC1_f/3_f$ genes in the number of codons in exon 1 (41, 38 or 40, and 37 in chicken, rat, and mouse, respectively). All the other amino acid-encoding exons in the three $MLC1_f/3_f$ genes are identical in size. In all three species, exon 2 encodes the 5′ untranslated region of $MLC3_f$ mRNA and precisely ends after the ATG initiation codon. Interestingly, this exon corresponds to the first exon of the single *D. melanogaster* alkali light chain gene (Fig. 6). Furthermore, only one miniexon, encoding amino acids 1 to 9, was found in the *D. melanogaster* gene (Fig. 6). This miniexon (exon 2 of the *D. melanogaster* gene) is more closely related to $MLC1_f$-specific miniexon 4 than to $MLC3_f$-specific miniexon 3 of vertebrate $MLC1_f/3_f$ genes (nucleotide sequence homologies of 29 and 46% between exon 2 of the *D. melanogaster* gene and exons 3 and 4, respectively, of the rat gene; Fig. 5 and reference 18). Exon 3 of the *D. melanogaster* gene corresponds to the combined exons 5 and 6 of the vertebrate $MLC1_f/3_f$ genes; i.e., the *D. melanogaster* gene lacks the intron separating these two exons in the rat, mouse, and chicken genes (Fig. 6). All the other intron positions were precisely conserved in the insect and vertebrate genes (Fig. 6). All the exons which end with a split codon (in all cases a guanine) in the rat gene (exons 3, 4, 5, 6, and 7; Fig. 5) also end with the same split codon in the mouse and chicken genes. Even the corresponding exons of the *D. melanogaster* gene (exons 2, 3, and 4) end with split codons; however, they end with different nucleotides (adenine in exon 2, guanine in exon 3, and cytosine in exon 4 [18]). In the vertebrate $MLC1_f/3_f$ genes, exon 8 contains the nine carboxy-terminal amino acid codons and part of the 3′ untranslated sequence, 17 bp in rat (Fig. 5), 17 bp in mouse (55), and 27 bp in chicken (43); the corresponding exon of the *D. melanogaster* alkali light chain gene encodes 14 carboxy-terminal amino acids and 6 bp of 3′ untranslated sequences

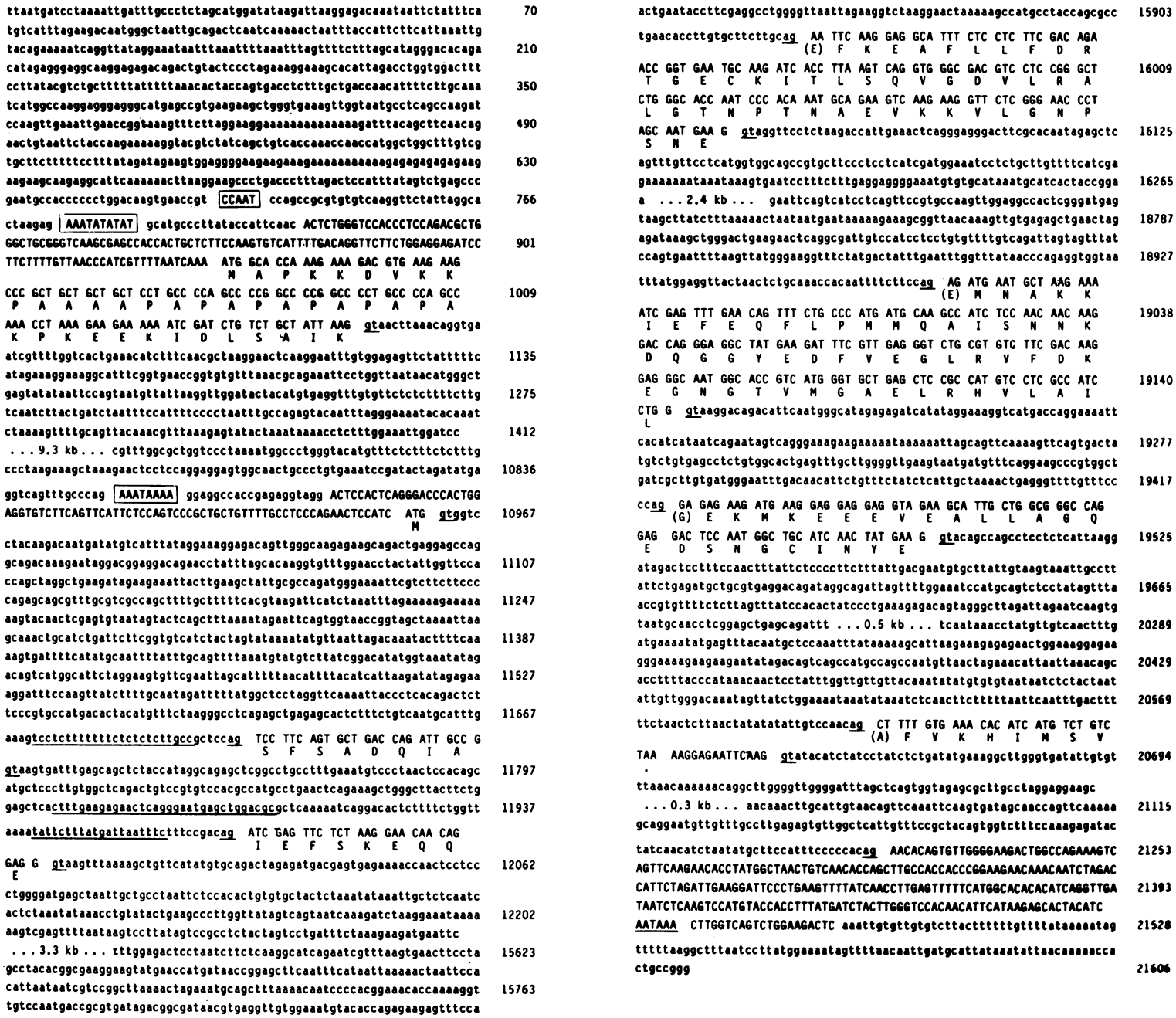
ttaatgatcctaaaattgatttgccctctagcatggatataagattaaggagacaaataattctatttca 70
tgtcatttagaagacaatgggctaattgcagactcaatcaaaaactaatttaccattcttcattaaattg
tacagaaaaatcaggttataggaaataatttaaatttttaaatttagttttctttagcatagggacacaga 210
catagagggaggcaaggagagacagactgtactccctagaaaggaaagcacattagacctggtggacttt
ccttatacgtctgcttttttatttttaaacactaccagtgacctctttgctgaccaacattttcttgcaaa 350
tcatggccaaggagggagggcatgagccgtgaagaagctgggtgaaagttggtaatgcctcagccaagat
ccaagttgaaattgaaccggtaaagtttcttaggaaggaaaaaaaaaaaaaaagatttacagcttcaacag 490
aactgtaattctaccaagaaaaaggtacgtctatcagctgtcaccaaaccaaccatggctggctttgtcg
tgcttctttttccttttatagatagaagtggaggggaagaagaaagaaaaaaaaaaagagagagagagaag 630
aagaagcaagaggcattcaaaaaacttaaggaagccctgacccttttagactccatttatagtctgagccc
gaatgccacccccctggacaagtgaaccgt CCAAT ccagccgcgtgtgtcaaggttctattaggca 766
ctaagag AAATATATAT gcatgcccttataccattcaac ACTCTGGGTCCACCCTCCAGACGCTG
GGCTGCGGGTCAAGCGAGCCACCACTGCTCTTCCAAGTGTCATTTTGACAGGTTCTTCTGGAGGAGATCC 901
TTCTTTTGTTAACCCATCGTTTTAATCAAA ATG GCA CCA AAG AAA GAC GTG AAG AAG
M A P K K D V K K
CCC GCT GCT GCT GCT CCT GCC CCA GCC CCG GCC CCG GCC CCT GCC CCA GCC 1009
P A A A A P A P A P A P A P A P A
AAA CCT AAA GAA GAA AAA ATC GAT CTG TCT GCT ATT AAG gtaacttaaacaggtga
K P K E E K I D L S A I K
atcgttttggtcactgaaacatctttcaacgctaaggaactcaaggaatttgtggagagttctatttttc 1135
atagaaaggaaaggcatttcggtgaaccggtgtgtttaaacgcagaaattcctggttaataacatgggct
gagtatataattccagtaatgttattaaggttggatactacatgtgaggtttgtgttctctcttttcttg 1275
tcaatcttactgatctaatttccatttttcccctaatttgccagagtacaatttagggaaaatacacaaat
ctaaaagttttgcagttacaaacgtttaaagagtatactaaatataaacctctttggaaattggatcc 1412
...9.3 kb... cgtttggcgctggtcccctaaaatggccctgggtacatgtttctctttctctttg
ccctaagaaagctaaagaactcctccaggaggagtggcaactgccctgtgaaatccgatactagatatga 10836
ggtcagtttgcccag AAATAAAA ggaggccaccgagaggtagg ACTCCACTCAGGGACCCACTGG
AGGTGTCTTCAGTTCATTCTCCAGTCCCGCTGCTGTTTTGCCTCCCAGAACTCCATC ATG gtggtc 10967
M
ctacaagacaatgatatgtcatttataggaaaggagacagttgggcaagagaagcagactgaggagccag
gcagacaaagaataggacggaggacagaacctatttagcacaaggtgtttggaacctactattggttcca 11107
ccagctaggctgaagatagaagaaattacttgaagctattgcgccagatgggaaaattcgtcttcttccc
cagagcagcgtttgcgtcgccagcttttgcttttcacgtaagattcatctaaatttagaaaaagaaaaa 11247
aagtacaactcgagtgtaatagtactcagctttaaaatagaattcagtggtaaccggtagctaaaattaa
gcaaactgcatctgattcttcggtgtcatctactagtataaaatatgttaattagacaaatacttttcaa 11387
aagtgattttcatatgcaatttttatttgcagttttaaatgtatgtcttatcggacatatggtaaatatag
acagtcatggcattctaggaagtgttcgaattagcatttttaacattttacatcattaagatatagagaa 11527
aggatttccaagttatctttttgcaatagatttttatggctcctaggttcaaaattaccctcacagactct
tcccgtgccatgacactacatgtttctaagggcctcagagctgagagcactctttctgtcaatgcatttg 11667
aaagtcctcttttttttctctctcttgccgctccag TCC TTC AGT GCT GAC CAG ATT GCC G
S F S A D Q I A
gtaagtgatttgagcagctctaccataggcagagctcggcctgcctttgaaatgtccctaactccacagc 11797
atgctccctttgtggctcagactgtccgtgtccacgccatgcctgaactcagaaagctgggcttacttctg
gagctcactttgaagagaactcagggaatgagctggacgcgctcaaaaatcaggacactctttttctggtt 11937
aaaatattctttatgattaatttctttccgacag ATC GAG TTC TCT AAG GAA CAA CAG
I E F S K E Q Q
GAG G gtaagtttaaaagctgttcatatgtgcagactagagatgacgagtgagaaaaccaactcctcc 12062
E
ctggggatgagctaattgctgcctaattctccacactgtgtgctactctaaatataaattgctctcaatc
actctaaatataaacctgtatactgaagcccttggttatagtcagtaatcaaagatctaaggaaataaaa 12202
aagtcgagttttaataagtccttatagtccgcctctactagtcctgatttctaaagaagatgaattc
...3.3 kb... tttggagactcctaatcttctcaaggcatcagaatcgtttaagtgaacttccta 15623
gcctacacggcgaaggaagtatgaaccatgataaccggagcttcaatttcataattaaaaactaattcca
cattaataatcgtccggcttaaaactagaaatgcagctttaaaacaatccccacggaaacaccaaaaggt 15763
tgtccaatgaccgcgtgatagacggcgataacgtgaggttgtggaaatgtacaccagagaagagtttcca
actgaataccttcgaggcctggggttaattagaaggtctaaggaactaaaaagccatgcctaccagcgcc 15903
tgaacaccttgtgcttcttgcag AA TTC AAG GAG GCA TTT CTC CTC TTC GAC AGA
(E) F K E A F L L F D R
ACC GGT GAA TGC AAG ATC ACC TTA AGT CAG GTG GGC GAC GTC CTC CGG GCT 16009
T G E C K I T L S Q V G D V L R A
CTG GGC ACC AAT CCC ACA AAT GCA GAA GTC AAG AAG GTT CTC GGG AAC CCT
L G T N P T N A E V K K K V L G N P
AGC AAT GAA G gtaggttcctctaagaccattgaaactcagggagggacttcgcacaatagagctc 16125
S N E
agtttgttcctcatggtggcagccgtgcttccctcctcatcgatggaaatcctctgcttgttttcatcga
gaaaaaaatataaataaagtgaatcctttctttgaggaggggaaatgtgtgcataaatgcatcactaccgga 16265
a ...2.4 kb... gaattcagtcatcctcagttccgtgccaagttggaggccactcgggatgag
taagcttatctttaaaaactaataatgaataaaaagaaagcggttaacaaagttgtgagagctgaactag 18787
agataaagctgggactgaagaactcaggcgattgtccatcctcctgtgttttgtcagattagtagtttat
ccagtgaatttaagttatgggaaggtttctatgactatttgaatttggtttataacccagaggtggtaa 18927
tttatggaggttactaactctgcaaaccacaattttcttccag AG ATG AAT GCT AAG AAA
(E) M N A K K
ATC GAG TTT GAA CAG TTT CTG CCC ATG ATG CAA GCC ATC TCC AAC AAC AAG 19038
I E F E Q F L P M M Q A I S N N K
GAC CAG GGA GGC TAT GAA GAT TTC GTT GAG GGT CTG CGT GTG TTC GAC AAG
D Q G G Y E D F V E G L R V F D K
GAG GGC AAT GGC ACC GTC ATG GGT GCT GAG CTC CGC CAT GTC CTC GCC ATC 19140
E G N G T V M G A E L R H V L A I
CTG G gtaaggacagacattcaatgggcatagagagatcatataggaaaggtcatgaccaggaaaatt
L
cacatcataatcagaatagtcagggaaagaagaaaaataaaaaattagcagttcaaaagttcagtgacta 19277
tgtctgtgagcctctgtggcactgagtttgcttggggttgaagtaatgatgtttcaggaagcccgtggct
gatcgcttgtgatgggaatttgacaacattctgtttctatctcattgctaaaactgagggttttgtttcc 19417
ccag GA GAG AAG ATG AAG GAG GAG GAG GTA GAA GCA TTG CTG GCG GGC CAG
(G) E K M K E E E V E A L L A G Q
GAG GAC TCC AAT GGC TGC ATC AAC TAT GAA G gtacagccagcctcctctcattaagg 19525
E D S N G C I N Y E
atagactcctttccaacttattctccccttctttattgacgaatgtgcttattgtaagtaaattgcctt
attctgagatgctgcgtgaggacagataggcagattagttttggaaatccatgcagtctcctatagttta 19665
accgtgttttctcttagtttatccacactatccctgaaagagacagtagggcttagattagaatcaagtg
taatgcaacctcggagctgagcagattt ...0.5 kb ... tcaataaaccctatgttgtcaacttttg 20289
atgaaaatatgagtttacaatgctccaaatttataaaaagcattaagaaagagagaactggaaaggagaa
gggaaaagaagaagaatatagacagtcagccatgccagccaatgttaactagaaacattaattaaacagc 20429
acctttttacccataaacaactcctatttggttgttgttacaaatatatgtgtgtaataatctctactaat
attgttgggacaaatagttatctggaaaataaatataaatctcaacttctttttaattcaatttgacttt 20569
ttctaactcttaactatatatattgtccaacag CT TTT GTG AAA CAC ATC ATG TCT GTC
(A) F V K H I M S V
TAA AAGGAGAATTCAAG gtatacatctatcctatctctgatatgaaaggcttgggtgatattgtgt 20694
ttaaacaaaaaacaggcttgggggttgggatttagctcagtggtagagcgcttgcctaggaggaagc
...0.3 kb... aacaaacttgcattgtaacagttcaaattcaagtgatagcaaccagttcaaaaa 21115
gcaggaatgttgtttgccttgagagtgttggctcattgtttccgctacagtggtctttccaaagagatac
tatcaacatctaatatgcttccatttcccccacag AACACAGTGTTGGGGAAGACTGGCCAGAAAGTC 21253
AGTTCAAGAACACCTATGGCTAACTGTCAACACCAGCTTGCCACCACCCGGAAGAACAAACAATCTAGAC
CATTCTAGATTGAAGGATTCCCTGAAGTTTTATCAACCTTGAGTTTTTCATGGCACACACATCAGGTTGA 21393
TAATCTCAAGTCCATGTACCACCTTTATGATCTACTTGGGTCCACAACATTCATAAGAGCACTACATC
AATAAA CTTGGTCAGTCTGGAAGACTC aaattgtgttgtgtcttactttttttgttttataaaaatag 21528
tttttaaggctttaatccttatggaaaatagttttaacaattgatgcattatataaatattaacaaaaacca
ctgccggg 21606

FIG. 5. Nucleotide sequence of the rat MLC1f/3f gene. DNA sequences corresponding to the 5′ flanking region, theexons (capital letters), and parts of the intervening sequences (lower case letters) were determined as described in Materials and Methods. Amino acids encoded within the exons are labeled in the one-letter code beneath the second nucleotide of each codon. Letters in parentheses designate amino acids encoded by split codons; the TAA translational termination codon is represented by a period. The AATAAA polyadenylation signal and the concensus splice sequences at intron boundaries are underlined. Three intron sequences showing more than 80% homology in the rat and chicken (43) gene are also underlined. TATA and CAT sequences are boxed. Unsequenced gaps are included with their approximate lengths, and the numbering system incorporates them.

(18). In all four genes, the last exon contains purely sequences of the 3′ untranslated region.

The above comparison strongly suggests that the present-day vertebrate MLC1f/3f genes originated from an ancestral gene resembling the actual alkali light chain gene from *D. melanogaster* in that they contained only one promoter (the vertebrate internal MLC3f promoter) and only one miniexon (the vertebrate MLC1f-specific exon 4). Duplication of the miniexon and the addition of a far upstream promoter and first exon were later events, perhaps occurring together with the development of promoter choice-dependent differential splicing pathways for the two different primary transcripts.

It is also remarkable that not only intron positions have been precisely conserved in the genes studied above but, at least in the vertebrates, even the sizes of corresponding introns have been maintained to a large extent (Fig. 6). Of particular interest is the conservation of the large size (over 9 kb) of the first intron of the MLC1f/3f genes because this intron contains the MLC3f-specific internal promoter found within these genes.

DNA sequence comparisons. A comparison of the 5′ flanking sequences corresponding to the MLC1f and MLC3f promoter regions immediately upstream of exons 1 and 2, respectively, is of particular interest because it may reveal additional data on the fine structure of these promoters as well as on the evolutionary conservation of their regulatory features.

Rat (Fig. 5) and mouse (55) sequences showed homologies of over 90% in both the MLC1f and the MLC3f promoter regions, extending over more than 100 bp upstream of the respective cap sites. While this high degree of homology may not be surprising, given the close evolutionary relationship of rat and mouse species, it is, however, significantly higher than the degree of similarities found in the available intron sequences (80 to 85%) and only slightly lower than the homologies between the exons specifying the 5′ untranslated region of MLC1f mRNA and the amino acid codons of both MLC1f and MLC3f mRNAs (95%). Interestingly, exons 2 (specifying the 5′ untranslated region of MLC3f mRNA), 8, and 9 (containing the sequences of the 3′ untranslated region) showed homologies of only 80% (exon 2), 90% (exon 8) and 87% (exon 9) between the rat and the mouse genes, i.e., a degree of divergence comparable to that of the intron sequences. Apparently, a higher selective pressure acts upon the promoter regions than on the intron sequences and on the exons encoding the 5′ untranslated region of MLC3f mRNA and the 3′ untranslated regions of both mRNAs.

A similar comparison between the rat and the chicken sequences yields even more striking results. Because of the much larger evolutionary distance between birds and mammals (2.5×10^8 to 3.0×10^8 years [53]), the differences between evolutionarily conserved and divergent sequences become more evident (Fig. 7 and 8). Starting at position −170 (position +1 being the 5′ terminus of MLC1f mRNA [Fig. 7A]), there is a large stretch of purines (Fig. 7A) that is located at similar positions in the rat and chicken genes and that may correspond to a region of altered DNA conformation (11, 17) important for promoter function. Further downstream, from position −145 down to the cap site, a striking degree of sequence homology was found between the MLC1f-specific promoter regions of the rat and the chicken gene (Fig. 7A and 8A). In fact, the degree of homology found in this region (which includes the TATA and the CAT box) is higher (90%) than between the adjoining exon sequences encoding the 5′ untranslated region of MLC1f mRNA, where homology is about 70% in the 5′ terminal part but only 40%

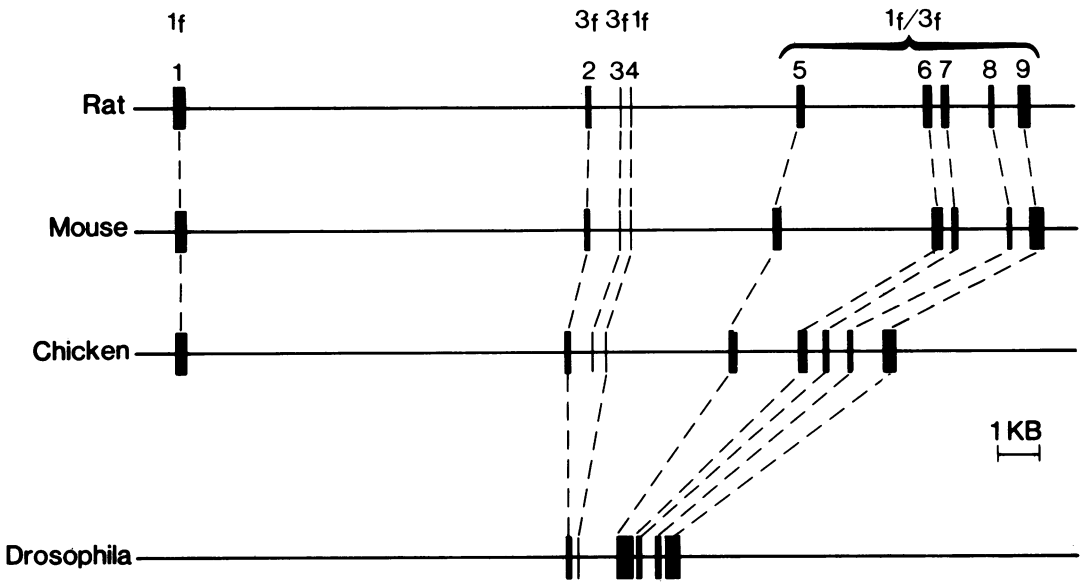

FIG. 6. Comparison of the structural organization of rat, mouse, and chicken $MLC1_f/3_f$ genes. Linear maps of the rat, mouse, and chicken $MLC1_f/3_f$ genes and of the single *D. melanogaster* alkali light chain gene are shown. Exons (black boxes) are numbered from 1 to 9 for the rat gene, corresponding exons in the genes are connected by stippled lines. In mature transcripts from the $MLC1_f/3_f$ genes, exons 1 and 4 are only represented in $MLC1_f$ (1_f) mRNA, exons 2 and 3 only in $MLC3_f$ (3_f) mRNA, while exons 5 to 9 are common to both $MLC1_f$ and $MLC3_f$ mRNAs ($1_f/3_f$). The structural maps of the mouse, chicken, and *D. melanogaster* genes are from Robert et al. (55), Nabeshima et al. (43), and Falkenthal et al. (18), respectively.

in the 3′ part immediately upstream of the ATG initiation codon even when gaps are introduced to maximize homology (Fig. 7A and 8A). There is a deletion of six nucleotides between the CAT and the TATA box of the chicken as compared to the rat sequence (positions −60 to −65 in the rat sequence [Fig. 7A]). A different chicken sequence, however, shows the presence of six nucleotides at that position, four of which are identical to the ones found in the rat sequence (R. Billeter, personal communication). Thus, the 5′ region upstream of the $MLC1_f$ cap site is more conserved than any of the intragenic sequences, with the exception of some of the amino acid-encoding exons.

In sharp contrast to the results obtained with the $MLC1_f$ promoter regions, a comparison of the 5′ flanking sequences upstream of exon 2, i.e., of the $MLC3_f$-specific promoter regions, revealed only poor sequence conservation between the rat and chicken genes (Fig. 7B and 8B). Not even the TATA sequence itself has been completely retained in the two genes, and while there is evidence for a CAT box in the chicken gene, the corresponding sequence in the rat gene (positions −46 to −49) has been changed into an ATA sequence (Fig. 7B). Altogether, sequence conservation in the $MLC3_f$ promoter region of the rat and chicken genes amounts to only 50% and is comparable to the degree of homology found in the adjoining exon specifying the 5′ untranslated region of $MLC3_f$ mRNA (Fig. 7B and 8B). Interestingly, however, there is a stretch of highly conserved DNA sequence in the rat and chicken $MLC3_f$ promoter in the far upstream region at positions −72 to −121 (Fig. 7B and 8B) where homology reaches more than 80%.

The homology between the 5′ flanking and untranslated mRNA sequences has to be compared with the respective degrees of homology found in the remaining exon and intron sequences. The average degree of homology between the amino acid-encoding sequences amounts to 75 to 80%, while the overall degree of sequence conservation between the 3′ untranslated regions is less than 30% (Fig. 5 and reference 43). The available intron sequences showed no similarities (25%) in the regions that could be examined (Fig. 5 and reference 43). However, while sequence identity does not extend far beyond the GT^G_A at the 5′ end and the CAG at the 3′ end of the introns, the donor sequence immediately following $MLC3_f$-specific miniexon 3, the donor sequence following $MLC1_f$-specific miniexon 4, and the acceptor sequence immediately upstream of the first common exon 5 are conserved over a stretch of 8, 7, and 7 bases, respectively, in the rat and chicken gene (Fig. 5 and reference 43). This conservation may be significant because miniexons 3 and 4, which are functionally separated from each other by a pseudointron (52), are both spliced to exon 5 in the differential processing steps leading to formation of the mature $MLC1_f$ and $MLC3_f$ mRNAs (43, 52). In addition, the pseudointron between exons 3 and 4 and the intron preceding exon 3 contain additional sequence stretches with a high degree of similarity (80%) in the rat and chicken gene (Fig. 5, underlined areas). Thus, while most of the intron sequences seem to evolve under no obvious constraints, some stretches have been conserved, suggesting their importance in gene structure or function or both. The extremely high degree of conservation of the $MLC1_f$ promoter regions in the rat, mouse, and chicken $MLC1_f/3_f$ genes is outstanding. On the other hand, the $MLC3_f$ promoter region, which is structurally part of intron 1 in the $MLC1_f$ primary transcript, has evolved at a higher rate. Yet, the degree of sequence conservation in this region is still significantly higher than that in the remaining intron and 3′ untranslated sequences.

DISCUSSION

The $MLC1_f/3_f$ gene has two functional and differentially regulated promoters. Although previous work on the structure of the $MLC1_f/3_f$ gene from the chicken (43), the rat (52) and the mouse (55) has suggested the presence of two

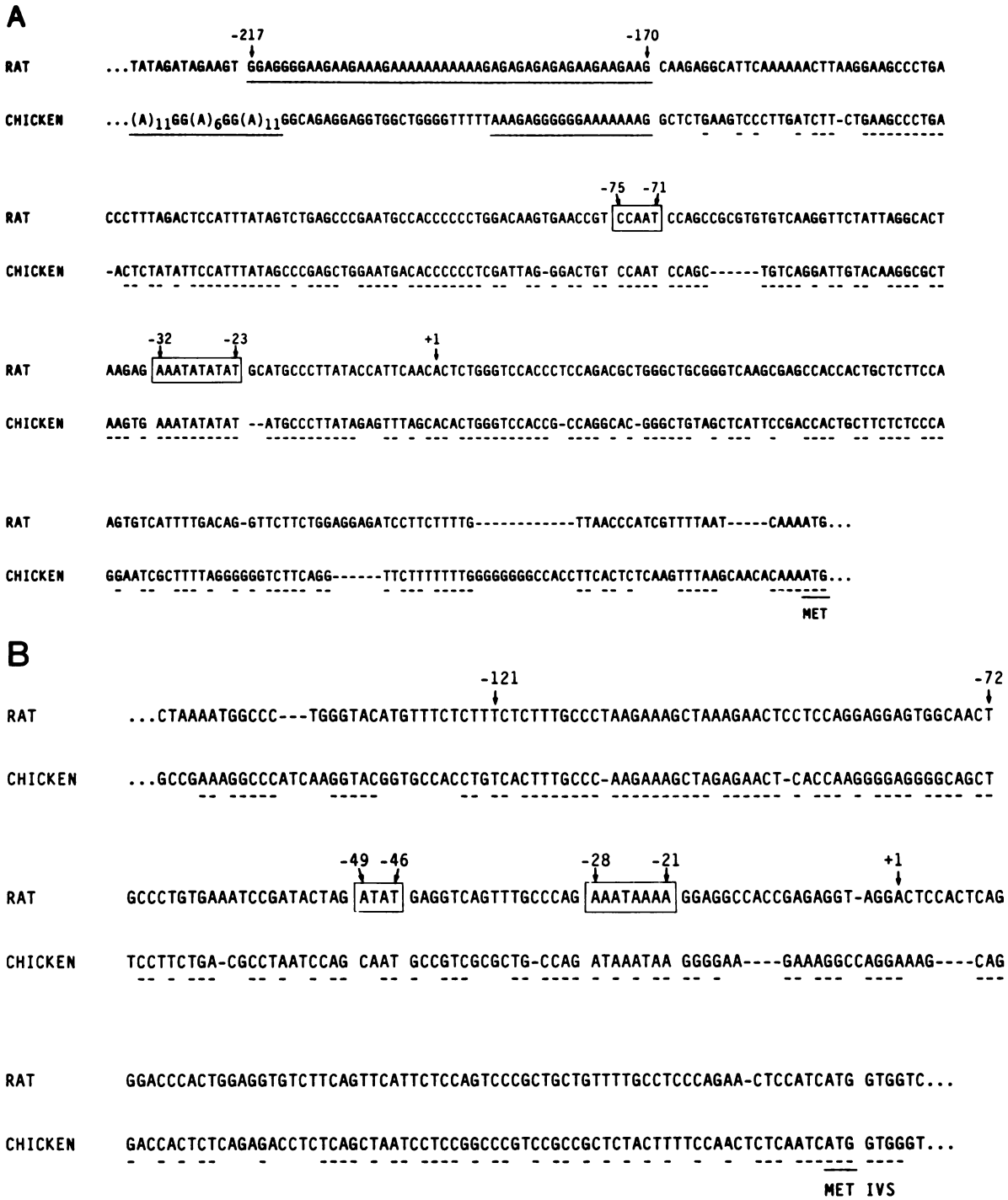

FIG. 7. Comparison of rat and chicken DNA sequences corresponding to the $MLC1_f$ and $MLC3_f$ promoter regions. (A) The 5′ flanking region upstream of $MLC1_f$-specific exon 1 is displayed together with the beginning of exon 1 down to the methionine (MET) translational initiation codon ATG (underlined). Rat and chicken sequences have been aligned; dashes are introduced to optimize homologies. Conserved nucleotides are underlined. Stretches of homopurine (G-A) sequences in the far upstream region are also underlined. The position of the 5′ start site of rat $MLC1_f$ mRNA is labeled +1; TATA and CAT sequences are boxed for the rat gene. (B) Rat and chicken sequences corresponding to the 5′ flanking region upstream of $MLC3_f$-specific exon 2, exon 2 itself (specifying the 5′ untranslated region and the methionine (MET) translational initiation codon (ATG), and the beginning of the adjoining intervening sequence (IVS) are displayed and have been aligned to optimize homology. The major start site of rat $MLC3_f$ mRNA is labeled +1. The boxed rat sequence from position −21 to −28 corresponds to the TATA region. A second rat ATA sequence (−47 to −49) is also boxed. All other symbols are as described for panel A. Rat sequences are derived from Fig. 5; chicken sequences are from Nabeshima et al. (43).

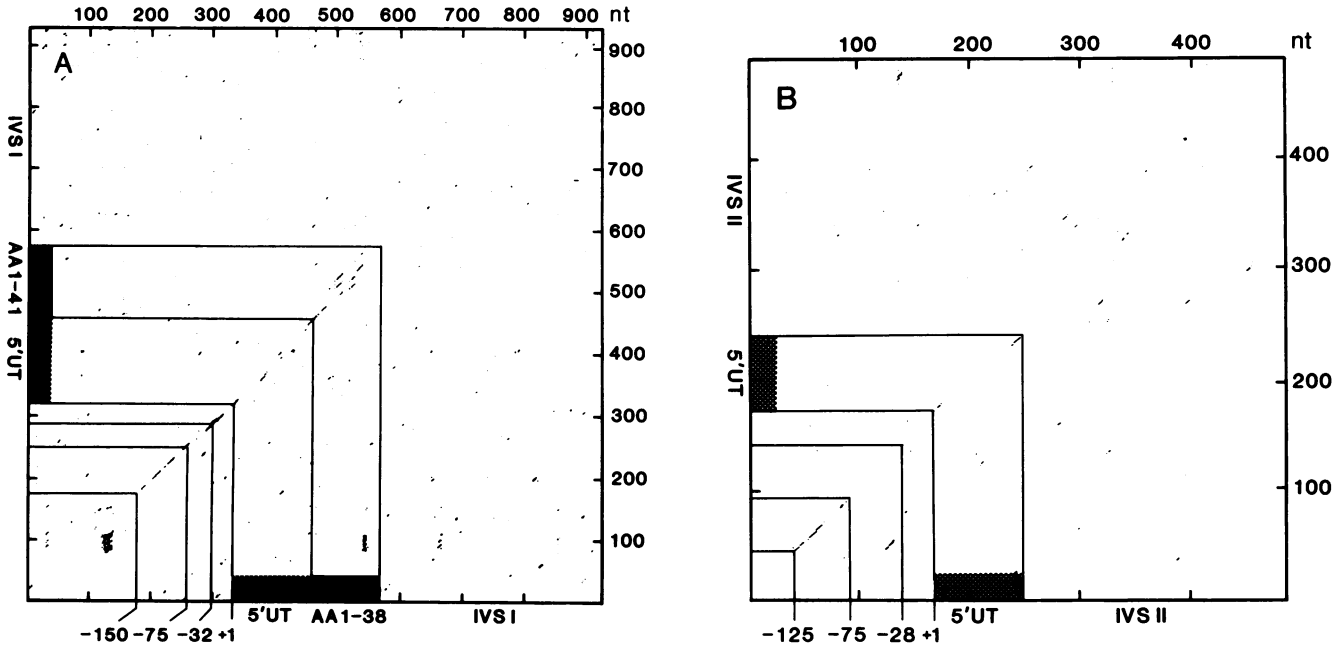

FIG. 8. Dot plots illustrating regions of extended sequence homology in the $MLC1_f$- and $MLC3_f$-specific promoter and adjoining exon sequences of the rat and chicken $MLC1_f/3_f$ gene. (A) Comparison of rat (abscissa) and chicken (ordinate) sequences comprising the 5′ flanking region upstream of exon 1 ($MLC1_f$ promoter), exon 1 (specifying the 5′ untranslated region [5′ UT, shaded box]), amino acids (AA) 1 to 38 (rat) and 1 to 41 (chicken) of $MLC1_f$ mRNA (black box), and part of the first intron (IVS I). The positions of the cap site (+1), the beginning of the TATA box (−32), the CAT box (−75), and the conserved far upstream sequences (−150) are indicated for the rat sequence. (B) Comparison of rat (abscissa) and chicken (ordinate) sequences comprising the 5′ flanking region upstream of exon 2 ($MLC3_f$ promoter), exon 2 (specifying the 5′ untranslated region of $MLC3_f$ mRNA, shaded box), and part of the adjoining intron 2 (IVS II). The positions of the cap site (+1), the beginning of the TATA box (−28), and the beginning (−125) and end (−75) of the conserved far upstream sequences are indicated for the rat sequence. A dot is made wherever there is homology amounting to eight or more matches of 10 consecutive bases between the two sequences compared with each other. Rat sequences are derived from Fig. 5; chicken sequences are from Nabeshima et al. (43).

promoters within the single $MLC1_f/3_f$ gene, the data presented here demonstrate that this gene indeed contains two promoters capable of independently directing transcription initiation of $MLC1_f$ and $MLC3_f$ mRNA, respectively.

In the in vitro transcription system and under the conditions used in this study the $MLC1_f$-specific promoters and the $MLC3_f$-specific promoters are both active in promoting accurate initiation of transcription. It has been shown that a wide variety of different eucaryotic promoter DNAs can be used as template in this system and will be efficiently transcribed from the in vivo cap site provided that some minimal sequence requirements are fulfilled (30, 34, 68). The $MLC1_f$- and $MLC3_f$-specific promoter templates used in this study contained 1.6 and 0.8 kb, respectively, of 5′ flanking sequences, thus including several hundred nucleotides upstream of the TATA box that has been shown to be important for correct positioning of the transcription initiation complex (for a recent review, see reference 66). The differences in the efficiency of in vitro transcription between the two MLC promoters and between them and the highly active adenovirus major late promoter may reflect true differences in promoter strength, although the possibility cannot be ruled out that additional regulatory upstream or downstream sequence elements may be important for efficient transcription from the $MLC1_f$ or the $MLC3_f$ promoter or both (9, 27, 30, 73).

Among the eucaryotic genes giving rise to more than one mature mRNA species, the gene for the mouse liver and parotid gland α-amylase is an example in which regulated differential transcription from two promoters (separated by approximately 3 kb) leads to the production of two different mRNAs coding for the same protein (59, 74). Of these two promoters, the one responsible for transcription of the parotid gland-specific mRNA is stronger (about 30-fold) than the liver-specific promoter. While the parotid gland-specific promoter is exclusively active in the parotid, the other, weaker promoter is active in both the parotid and the liver at similar rates (59). A similar organization and differential regulation has been described for the two closely spaced (approximately 600 bp) promoters of the *D. melanogaster* alcohol dehydrogenase gene (5). In contrast, the two promoters of the $MLC1_f/3_f$ gene are 10 kb apart, give rise to two mRNAs coding for different proteins, and seem to be simultaneously and equally active in adult fast skeletal muscle (52). However, the two promoters of the $MLC1_f/3_f$ gene are also differentially regulated. Initiation of transcription from the $MLC1_f$-specific promoter occurs before the activation of the $MLC3_f$-specific promoter during fast skeletal muscle development as reflected at the mRNA (52) and protein levels (24, 25, 70). The possibility that only one species of alkali light chain RNA is synthesized in a given fast skeletal muscle nucleus cannot be formally ruled out. The fact that both $MLC1_f$ and $MLC3_f$ mRNAs are present in the same amount (52) and that corresponding proteins are synthesized at similar rates in adult muscle and have been shown to be present in the same fiber, the same myofibril, and even within the same thick filament (23, 61) is suggestive of their production in the same nucleus. Whatever the case, a merely

stochastic usage of the two promoters of the $MLC1_f/3_f$ gene can hardly explain the differentially regulated expression of $MLC1_f$ and $MLC3_f$ mRNAs both during development and in specific adult fast skeletal muscle tissues.

The existence of two developmentally regulated but widely spaced promoters that drive overlapping transcriptional units in the same cell raises interesting questions about their sequential activation and mutual interaction. In the case of the $MLC1_f/3_f$ gene, it is clear that an active chromatin conformation exists in the $MLC3_f$ promoter region from the moment the $MLC1_f$ promoter is activated, since transcription initiated at the upstream cap site must run throughout the downstream promoter region. However, this conformation is not sufficient for $MLC3_f$-specific transcription initiation. To prevent the immediate activation of the internal ($MLC3_f$) promoter a specific but reversible mechanism must operate which allows readthrough of the transcriptional complex initiated at the upstream ($MLC1_f$) promoter while preventing initiation of transcription from the $MLC3_f$ promoter. Suppression of one promoter by an actively transcribing nearby promoter might be explained by transcriptional interference, as has been recently described for separate but physically linked eucaryotic genes in retrovirus systems (14), or by other epigenetic *cis*-acting mechanisms (16). However, in adult muscle tissues producing both $MLC1_f$ and $MLC3_f$ mRNA, both promoters are simultaneously active; i.e., any putative transcriptional interference must have been overcome. Two features set the $MLC1_f/3_f$ gene apart from the genes for which transcriptional interference has been demonstrated: (i) Both promoters belong to the same gene (although one might formally envisage the $MLC1_f/3_f$ transcriptional unit as being made up by two overlapping genes), and (ii) the two promoters are separated by approximately 10 kb of DNA. This lattter finding, which holds true for all the $MLC1_f/3_f$ genes, may be significant in terms of a recent report suggesting that a minimal intergenic distance may be required between two linked genes for them both to be functional (45). The presence and the maintenance of a large intervening sequence between the two promoters of the $MLC1_f/3_f$ genes may therefore be of significance for the regulation and the proper functioning of transcription within these genes. Despite the fact that the mechanisms of transcriptional interference and of extragenic territorial influence are not yet understood, it seems clear that *trans*-acting factors are involved in the differential activation of the $MLC3_f$ promoter within the $MLC1_f/3_f$ gene. In fact, recent studies with stable heterokaryons of human nonmuscle and mouse muscle cells have demonstrated the involvement of diffusible factors in muscle gene activation (8, 12).

The structural organization of $MLC1_f/3_f$ genes is highly conserved, but their two promoters are evolving at different rates. All three of the $MLC1_f/3_f$ genes analyzed in detail so far, chicken (43), rat (reference 52 and this report), and mouse (55), are identical in their exon-intron structure. In all these species, the functional $MLC1_f/3_f$ gene is a single-copy gene (in addition, a processed $MLC1_f/3_f$ pseudogene has been detected in the mouse [55]), from which two mRNAs, specifying $MLC1_f$ and $MLC3_f$, are produced by a mechanism involving both differential transcription and splicing.

The complete conservation of intron positions between corresponding genes of evolutionarily distant species or between different members of a multigene family within the same species has already been shown for a number of genes, most notably for the α- and β-globin genes (15, 51). Among the muscle protein genes, chicken and rat α-actin genes also show conserved intron positions (20, 75), however, intron positions vary between α- and β-actin genes in these species or between different actin genes from vertebrates and invertebrates (20, 47, 75). A similar conservation of intron positions is apparent among myosin heavy chain genes (64). While many of the corresponding introns in the abovementioned genes vary greatly in size, they are similar in length in the $MLC1_f/3_f$ genes. Most remarkably, the unusually large first intervening sequence (approximately 10 kb), which contains the $MLC3_f$-specific promoter, has been maintained in the rat, mouse, and chicken genes, whereas the known sequences of this intron have diverged considerably. Although the role of introns in gene structure and expression is not clear, it may be speculated that a rather strict maintenance of intron location and size is essential in the $MLC1_f/3_f$ genes, which contain two differentially regulated promoters and give rise to two differentially spliced RNA products (43, 52). This is relevant because the primary transcripts originating from each of the two promoters have only one predetermined splicing pathway open to them, despite the fact that exons 2, 3, and 4 are alternatively spliced to generate $MLC1_f$ and $MLC3_f$ mRNAs. Therefore, the alternative splicing of the primary transcripts is governed by promoter choice selection and is most probably dependent on the primary structure of the respective transcript (*cis* regulated). Maintenance of a different splicing pattern for each of the two transcripts could impose specific constraints on the intron sequences involved. For example, the stretches of extended sequence homology between the rat (or mouse) gene and the chicken gene near the 3′ end of introns 2 and 3 (i.e., sequences preceding miniexons 3 and 4 [Fig. 5]) may be important for correct splicing of the specific 5′-most exon of $MLC1_f$ (exon 1) and $MLC3_f$ (exon 2) to the respective downstream exon (exon 4 for $MLC1_f$ and exon 3 for $MLC3_f$). Sequences near the 3′ end of introns have recently been shown to be involved in the formation of lariat intermediates in the initial steps of splicing (26, 50, 57, 69); it may thus be significant that the lariat consensus sequences in the $MLC1_f/3_f$ gene introns 2 and 3 lie within extended stretches of conserved sequences.

Despite the highly conserved nature of the overall structure of $MLC1_f/3_f$ genes, a close inspection of the available nucleotide sequences reveals surprising differences not only in the degree of divergence in noncoding versus coding sequences but also in the rate of accumulation of nucleotide changes in the two promoter regions. The degree and extent of sequence homology between the $MLC1_f$ promoter regions is striking and is significantly higher than the conservation of intron sequences, 5′ and 3′ untranslated sequences, and some of the coding exon sequences. Besides the TATA and CAT boxes (which have been shown to be important for correct positioning of transcription initiation and for determining the level of transcription [3, 4, 10, 13, 28, 29, 36, 40, 54, 65, 66]) large stretches of the sequences between the mRNA start site and the TATA box, between the TATA and the CAT box, and upstream of the CAT box have been highly conserved in these genes. Conservation in the corresponding regions of the α-actin (47, 49) and of α2 type I collagen (60) genes has also been described. In none of them, however, is the homology between mammalian and avian species as high as in the $MLC1_f$ promoter. Several stretches of extended sequence identity found in the far upstream region of the rat and chicken $MLC1_f/3_f$ genes (the corresponding mouse sequences are not available) may represent further important transcriptional control signals of these genes (positions −145 to −136, GAAGCCCTGA; −125 to −115, TCCATTTATAG; and −99 to −91, CACCCCCCT; in

addition to the extended stretch of homopurine-homopyrimidine sequence found in the region upstream of position −170).

In contrast to the highly conserved $MLC1_f$ promoter sequences, the $MLC3_f$ promoter regions of the avian (chicken) and mammalian (rat, mouse) $MLC1_f/3_f$ gene are significantly more divergent. The TATA box sequence differs at three of nine positions, and no CAT box homology is found between the chicken and the rat (or mouse) genes. The additional ATA sequence present at positions −46 to −49 in the rat (and mouse) gene may actually play the role of a second TATA box (see also reference 55), thereby leading to changes in the site of transcription initiation. In fact, heterogeneity at the 5′ end of $MLC3_f$ mRNAs has been noticed in the chicken (R. Billeter, personal communication) and may also occur in the rat (Fig. 2B; Fig. 3, lanes 5 and 6). The only sequence stretch of high homology between the $MLC3_f$ promoters of the avian and the mammalian $MLC1_f/3_f$ genes resides in the far upstream region (position −72 to −121 in Fig. 7B and 8B). It is tempting to speculate that these far upstream sequences are somehow involved in regulating differential transcription from the $MLC3_f$ promoter and must therefore be maintained almost unchanged during evolution.

$MLC1_f/3_f$ genes have evolved from an ancestral gene lacking a far upstream promoter and first exon. So far, the $MLC1_f/3_f$ gene is the only member of the vertebrate alkali light chain gene family known to give rise to two different mRNAs originating from different promoters. In slow skeletal, cardiac, and smooth muscle tissues, only a single alkali light chain isoform, MLC1, is expressed. Because of the lack of information on the structure of the genes encoding these other alkali light chains it is not yet clear how the organization of the $MLC1_f/3_f$ gene compares with that of the other members of this multigene family. However, the recently published nucleotide sequence and structure of the single alkali light chain gene from *D. melanogaster* (18) shows that the $MLC1_f/3_f$ gene of higher vertebrates most probably evolved from an ancestral alkali light chain gene containing only one promoter (the $MLC3_f$ promoter of the vertebrate gene), and encoding the 5′ untranslated sequences down to the methionine start codon in its first exon. This exon was followed by a single miniexon (as is found in the *D. melanogaster* alkali light chain gene) that encoded the first few amino acids (1 to 9) of the ancestral myosin light chain protein and that corresponds to miniexon 4 of the present-day vertebrate $MLC1_f/3_f$ genes. It is likely that the original vertebrate $MLC1_f/3_f$ gene arose through a duplication of the miniexon region and the addition (or activation) of a far upstream promoter and a newly acquired first exon. Since, as shown in this report, the original promoter, now lying approximately 10 kb downstream of the new 5′ end of the gene, remained functional, two widely different primary transcripts originate from this single gene. The appearance of promoter choice-dependent differential splicing patterns for these two transcripts and the duplication of the mini-exon may have been concomitant events since the inclusion of both miniexons (exons 3 and 4) in a single mature mRNA generates a shift in the reading frame, leading to premature stop codons. It is interesting to note that, although there is no promoter choice for the transcription of the *D. melanogaster* alkali light chain gene, regulated differential splicing has been shown to occur at the 3′ end of this gene (18).

The hypothesis that the ancestral alkali light chain gene did not have a far upstream promoter and first exon is further supported by the structures of the rat skeletal muscle MLC2 (46) and the chicken calmodulin gene (62). Both the MLC2 (dithiobis-nitrobenzoic acid-removable light chain) and the calmodulin genes belong to the same troponin C supergene family that also includes the alkali light chain genes among others (1). In both the rat MLC2 and the chicken calmodulin gene, as in the hypothetical ancestral alkali light chain gene, the first exon contains the 5′ untranslated sequences down to the ATG initiation codon while the remaining mRNA sequences are distributed over six exons (46, 62).

The evolutionary advantage of a newly created gene with a $MLC1_f$-specific far upstream promoter that is transcriptionally activated before the $MLC3_f$ promoter during skeletal muscle development might have put the upstream promoter under stronger selective pressure than the downstream ($MLC3_f$) one. The activation of a promoter already open to the transcriptional complex may be subject to different constraints than the activation of a transcriptionally silent one. Whether the higher accumulation of nucleotide changes in the 5′ flanking sequences of the $MLC3_f$ compared with the $MLC1_f$ promoter reflects these putative differences in promoter activation remains to be experimentally demonstrated.

ACKNOWLEDGMENTS

We thank P. Bayard for excellent technical assistance, U. Hansen for advice on the in vitro transcription experiments and for the gifts of plasmid pFLBH and HeLa whole-cell extract, and M. Hager for help with the preparation of this manuscript.

This work was supported by Public Health Service grants from the National Institutes of Health, Muscular Dystrophy Association of America, and the American Heart Association. E.E.S. was a fellow of the Swiss National Science Foundation (grant no. 83.026.082).

LITERATURE CITED

1. **Baba, M. L., M. Goodman, J. Berger-Cohn, J. G. Demaille, and G. Matsuda.** 1984. The early adaptive evolution of calmodulin. Mol. Biol. Evol. **1**:442–455.
2. **Benoff, S., and B. Nadal-Ginard.** 1979. Most myosin heavy chain mRNA in rat L_6E_9 myotubes has a short poly(A) tail. Proc. Natl. Acad. Sci. USA **76**:1853–1857.
3. **Benoist, C., and P. Chambon.** 1981. In vivo sequence requirements of the SV40 early promoter region. Nature (London) **290**:304–310.
4. **Benoist, C., K. O'Hare, R. Breathnach, and P. Chambon.** 1980. The ovalbumin gene-sequence of putative control regions. Nucleic Acids Res. **8**:127–142.
5. **Benyajati, C., N. Spoerel, H. Haymerle, and M. Ashburner.** 1983. The messenger RNA for alcohol dehydrogenase in *Drosophila melanogaster* differs in its 5′ end in different developmental stages. Cell **33**:125–133.
6. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids. Cell **12**:721–732.
7. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
8. **Blau, H., C.-P. Chiu, and C. Webster.** 1983. Cytoplasmic activation of human nuclear genes in stable heterocaryons. Cell **32**:1171–1180.
9. **Brady, J., M. Radonovich, M. Thoren, G. Das, and N. P. Salzman.** 1984. Simian virus 40 major late promoter: an upstream DNA sequence required for efficient in vitro transcription. Mol. Cell. Biol. **4**:133–141.
10. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
11. **Cantor, C. R., and A. Efstratiadis.** 1984. Possible structures of homopurine-homopyrimidine S1-hypersensitive sites. Nucleic Acids Res. **12**:8059–8072.

12. **Chiu, C.-P., and H. Blau.** 1984. Reprogramming cell differentiation in the absence of DNA synthesis. Cell **37**:879–887.
13. **Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kredinger, and P. Chambon.** 1980. Promoter sequences of eukaryotic protein-coding genes. Science **209**:1406–1414.
14. **Cullen, B. R., P. T. Lomedico, and G. Ju.** 1984. Transcriptional interference in avian retroviruses—implications for the promoter insertion model of leukaemogenesis. Nature (London) **307**:241–245.
15. **Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Spritz, J. K. DeRiel, B. G. Forget, S. M. Weissman, J. L. Slightom, A. E. Blechl, O. Smithies, F. E. Baralle, C. S. Shoulders, and N. J. Proudfoot.** 1980. The structure and evolution of the human β-globin gene family. Cell **21**:653–668.
16. **Emerman, M., and H. T. Temin.** 1984. Genes with promoters in retrovirus vectors can be independently suppressed by an epigenetic mechanism. Cell **39**:459–467.
17. **Evans, T., E. Schon, G. Gora-Maslak, J. Patterson, and A. Efstratiadis.** 1984. S1-hypersensitive sites in eukaryotic promoter regions. Nucleic Acids Res. **12**:8043–8058.
18. **Falkenthal, S., V. P. Parker, and N. Davidson.** 1985. Developmental variations in the splicing pattern of transcripts from the *Drosophila* gene encoding myosin alkali light chain result in different carboxyl-terminal amino acid sequences. Proc. Natl. Acad. Sci. USA **82**:449–453.
19. **Fire, A., M. Samuels, and P. A. Sharp.** 1984. Interactions between RNA polymerase II, factors, and template leading to accurate transcription. J. Biol. Chem. **259**:2509–2516.
20. **Fornwald, J. A., G. Kuncio, I. Peng, and C. P. Ordahl.** 1982. The complete nucleotide sequence of the chick α-actin gene and its evolutionary relationship to the actin gene family. Nucleic Acids Res. **10**:3861–3874.
21. **Frank, G., and A. G. Weeds.** 1974. The amino-acid sequence of the alkali light-chains of rabbit skeletal-muscle myosin. Eur. J. Biochem. **44**:317–334.
22. **Garfinkel, L. I., M. Periasamy, and B. Nadal-Ginard.** 1982. Cloning and characterization of cDNA sequences corresponding to myosin light chains 1, 2, and 3, troponin-C, troponin-T, α-tropomyosin, and α-actin. J. Biol. Chem. **257**:11078–11086.
23. **Gauthier, G. F., and S. Lowey.** 1979. Distribution of myosin isoenzymes among skeletal muscle fiber types. J. Cell Biol. **81**:10–25.
24. **Gauthier, G. F., S. Lowey, P. A. Benfield, and A. W. Hobbs.** 1982. Distribution and properties of myosin isoenzymes in developing avian and mammalian skeletal muscle fibers. J. Cell Biol. **92**:472–484.
25. **Gauthier, G. F., S. Lowey, and A. W. Hobbs.** 1981. Fast and slow myosin in developing muscle fibres. Nature (London) **274**:25–29.
26. **Grabowski, P. J., R. A. Padgett, and P. A. Sharp.** 1984. Messenger RNA splicing in vitro: an excised intervening sequence and a potential intermediate. Cell **37**:415–427.
27. **Grosschedl, R., and M. Birnstiel.** 1982. Delimitation of far upstream sequences required for maximal in vitro transcription of an H2A histone gene. Proc. Natl. Acad. Sci. USA **79**: 297–301.
28. **Grosschedl, R., and M. L. Birnstiel.** 1980. Spacer DNA sequences upstream of the TATAAATA sequence are essential for promotion of H2A histone gene transcription in vivo. Proc. Natl. Acad. Sci. USA **77**:7102–7106.
29. **Grosveld, G. C., E. deBoer, C. K. Shewmaker, and R. A. Flavell.** 1982. DNA sequences necessary for transcription of the rabbit β-globin gene in vivo. Nature (London) **295**:120–126.
30. **Hansen, U., and P. A. Sharp.** 1983. Sequences controlling in vitro transcription of SV40 promoters. EMBO J. **2**:2293–2303.
31. **Hansen, U., D. G. Tenen, D. M. Livingston, and P. A. Sharp.** 1981. T antigen repression of SV40 early transcription from two promoters. Cell **27**:603–612.
32. **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 **81**:2626–2630.
33. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
34. **Manley, J. L., A. Fire, A. Cano, P. A. Sharp, and M. L. Gefter.** 1980. DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Natl. Acad. Sci. USA **77**:3855–3859.
35. **Manley, J. L., A. Fire, M. Samuels, and P. A. Sharp.** 1983. In vitro transcription: whole-cell extract. Methods Enzymol. **101**:568–582.
36. **Mathis, D. J., and P. Chambon.** 1981. The SV40 early region TATA box is required for accurate in vitro initiation of transcription. Nature (London) **290**:310–315.
37. **Matsuda, G., T. Maita, and T. Umegane.** 1981. The primary structure of L-1 light chain of chicken fast skeletal muscle myosin and its genetic implication. FEBS Lett. **126**:111–113.
38. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74**:560–564.
39. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
40. **McKnight, S. L., and R. Kingsbury.** 1982. Transcriptional control signals of a eukaryotic protein-coding gene. Science **217**:316–324.
41. **Messing, J., R. Crea, and P. H. Seeburg.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9**:309–321.
42. **Nabeshima, Y., Y. Fujii-Kuriyama, M. Muramatsu, and K. Ogata.** 1982. Molecular cloning and nucleotide sequences of the complementary DNAs to chicken skeletal muscle myosin two alkali light chain mRNAs. Nucleic Acids Res. **10**:6099–6110.
43. **Nabeshima, Y., Y. Fujii-Kuriyama, M. Muramatsu, and K. Ogata.** 1984. Alternative transcription and two modes of splicing result in two myosin light chains from one gene. Nature (London) **308**:333–338.
44. **Nadal-Ginard, B.** 1978. Commitment, fusion and biochemical differentiation of a myogenic cell line in the absence of DNA synthesis. Cell **15**:855–864.
45. **Naora, H., and N. J. Deacon.** 1982. Clustered genes require extragenic territorial sequences. Differentiation **21**:1–6.
46. **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.
47. **Nudel, U., Y. Mayer, R. Zakut, M. Shani, H. Czosnek, B. Aloni, D. Melloul, and D. Yaffe.** 1984. Structure and expression of rat actin genes. Exp. Biol. Med. **9**:219–227.
48. **Obinata, T., K. Maruyama, H. Sugita, K. Kohama, and S. Ebashi.** 1981. Dynamic aspects of structural proteins in vertebrate skeletal muscle. Muscle & Nerve **4**:456–488.
49. **Ordahl, C. P., I. Peng, and T. A. Cooper.** 1984. Structure and expression of the chick skeletal muscle alpha-actin gene. Exp. Biol. Med. **9**:211–218.
50. **Padgett, R. A., M. M. Konarska, P. J. Grabowski, S. F. Hardy, and P. A. Sharp.** 1984. Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. Science **225**:898–903.
51. **Patient, R. K., J. A. Elkington, R. M. Kay, and J. G. Williams.** 1980. Internal organization of the major adult α- and β-globin genes of *X. laevis*. Cell **21**:565–573.
52. **Periasamy, M., E. E. Strehler, L. I. Garfinkel, R. M. Gubits, N. Ruiz-Opazo, and B. Nadal-Ginard.** 1984. Fast skeletal muscle myosin light chain 1 and 3 are produced from a single gene by a combined process of differential RNA transcription and splicing. J. Biol. Chem. **259**:13595–13604.
53. **Perler, F., A. Efstratiadis, P. Lomedico, W. Gilbert, R. Kolodner, and J. Dodgson.** 1980. The evolution of genes: the chicken preproinsulin gene. Cell **20**:555–566.
54. **Reudelhuber, T.** 1984. Upstream and downstream control of eukaryotic genes. Nature (London) **312**:700–701.
55. **Robert, B., P. Daubas, M.-A. Akimenko, A. Cohen, I. Garner, J.-L. Guenet, and M. Buckingham.** 1984. A single locus in the mouse encodes both myosin light chains 1 and 3, a second locus corresponds to a related pseudogene. Cell **39**:129–140.
56. **Robert, B., A. Weydert, M. Caravatti, A. Minty, A. Cohen, P. H. Daubas, F. Gros, and M. Buckingham.** 1982. DNA recom-

binant plasmid complementary to mRNAs for light chains 1 and 3 of mouse skeletal muscle myosin. Proc. Natl. Acad. Sci. USA **79**:2437–2441.

57. **Ruskin, B., A. R. Krainer, T. Maniatis, and M. R. Green.** 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. Cell **38**:317–331.
58. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
59. **Schibler, U., O. Hagenbuechle, P. K. Wellauer, and A. C. Pittet.** 1983. Two promoters of different strengths control the transcription of the mouse alpha-amylase gene Amy-1[a] in the parotid gland and the liver. Cell **33**:501–508.
60. **Schmidt, A., Y. Yamada, and B. de Crombrugghe.** 1984. DNA sequence comparison of the regulatory signals at the 5′ end of the mouse and chick α2 type I collagen genes. J. Biol. Chem. **259**:7411–7415.
61. **Silberstein, L., and S. Lowey.** 1981. Isolation and distribution of myosin isoenzymes in chicken pectoralis muscle. J. Mol. Biol. **148**:153–189.
62. **Simmen, R. C. M., T. Tanaka, K. Fang Ts'ui, J. A. Putkey, M. J. Scott, E. C. Lai, and A. R. Means.** 1985. The structural organization of the chicken calmodulin gene. J. Biol. Chem. **260**:907–912.
63. **Soeiro, R., H. C. Birnboim, and J. E. Darnell.** 1966. Rapidly labelled HeLa cell nuclear RNA. II. Base composition and cellular location of a heterogeneous RNA fraction. J. Mol. Biol. **19**:362–373.
64. **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.
65. **Wasylyk, B., R. Derbyshire, A. Guy, D. Molko, A. Roget, R. Teoule, and P. Chambon.** 1980. Specific in vitro transcription of conalbumin gene is drastically decreased by single-point mutation in TATA box homology sequence. Proc. Natl. Acad. Sci. USA **77**:7024–7028.
66. **Wasylyk, B., C. Wasylyk, and P. Chambon.** 1984. A novel eukaryotic promoter element: the simian virus 40 72 base pair repeat. Fed. Proc. **43**:226–234.
67. **Weeds, A. G., and S. Lowey.** 1971. Substructure of the myosin molecule. II. The light chains of myosin. J. Mol. Biol. **61**:701–725.
68. **Weil, P. A., D. S. Luse, J. Segall, and R. G. Roeder.** 1979. Selective and acurate initiation of transcription at the Ad2 major late promoter in a soluble system dependent on purified RNA polymerase II and DNA. Cell **18**:469–484.
69. **Weissmann, C.** 1984. Excision of introns in lariat form. Nature (London) **311**:103–104.
70. **Whalen, R. G., L. B. Bugaisky, G. S. Butler-Browne, S. M. Sell, K. Schwartz, and I. Pinset-Harstrom.** 1982. Characterization of myosin isozymes appearing during rat muscle development, p. 25–33. *In* M. L. Pearson and H. F. Epstein (ed.), Muscle development: molecular and Cellular Control. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
71. **Whalen, R. G., G. S. Butler-Browne, and F. Gros.** 1978. Identification of a novel form of myosin light chain present in embryonic muscle tissue and cultured muscle cells. J. Mol. Biol. **126**:415–431.
72. **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.
73. **Wildeman, A. G., P. Sassone-Corsi, T. Gundstrom, M. Zenke, and P. Chambon.** 1984. Stimulation of in vitro transcription from the SV40 early promoter by the enhancer involves specific trans-acting factors. EMBO J. **3**:3129–3133.
74. **Young, R. A., O. Hagenbuechle, and U. Schibler.** 1981. A single mouse α-amylase gene specifies two different tissue-specific mRNAs. Cell **23**:451–458.
75. **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.