Volume 16 Number 21 1988 Nucleic Acids Research

Promoter analysis of myosin alkali light chain genes expressed in mouse striated muscle

Arlette Cohen, Paul J.R.Barton*, Benoît Robert, Ian Garner, Serge Alonso and Margaret E.Buckingham

Département de Biologie Moléculaire, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris cedex 15, France

Received August 15, 1988; Revised and Accepted October 6, 1988 Accession nos X12971-X12973 (incl.)

ABSTRACT.

There are three principal myosin alkali light chain (MLC) genes expressed in mouse striated muscle. The skeletal muscle gene $MLC1_F$ /$MLC3_F$, the ventricular muscle / slow skeletal muscle gene $MLC1_V$($MLC1_S$), and the atrial muscle / foetal striated muscle gene $MLC1_A$(MLC1emb). $MLC1_V$ and $MLC1_A$ are expressed in both cardiac and skeletal muscle, and we show here that these genes use a single site of initiation of transcription, and therefore the same proximal promoter region, in both muscle types, and in myogenic cell lines in culture. We have previously shown that for the $MLC1_F$/$MLC3_F$ gene, 1200bp of upstream sequence from the $MLC1_F$ promoter is sufficient to allow tissue specific and developmentally regulated expression. We have therefore isolated, characterised, and sequenced over 1200bp upstream of each of the three MLC genes in order to look for elements which may be involved in their regulation. Detailed comparison of their promoter sequences, as well as those of the cardiac and skeletal muscle α-actin genes, reveals a number of common elements. Amoung these is an "MLC-sequence" (CCTTTTATAG) common to all MLC genes, including those of chick and rat, and a "cardiac sequence" common to the mouse $MLC1_A$, $MLC1_V$ and α-cardiac actin genes expressed in the heart.

INTRODUCTION.

Myosin alkali light chains (MLC) are encoded by a multigene family with different isoforms present in both muscle and non-mucle tissues (see 1). In mouse there are three principal MLC genes expressed in striated muscle. These are (i) the $MLC1_F$/$MLC3_F$ gene (2) which encodes the two isoforms present in adult fast skeletal muscle, $MLC1_F$ and $MLC3_F$. These isoforms differ only in their NH_2 terminal sequences of 42 and 8 amino acids respectively and the gene uses two independent promoters (3,4) to generate two types of primary transcript which, following splicing, form mRNAs carrying $MLC1_F$ or $MLC3_F$ specific 5' sequences. (ii) The $MLC1_V$ ($MLC1_S$) gene (5) which is expressed in cardiac ventricular muscle where it encodes the ventricular isoform $MLC1_V$, and in slow skeletal muscle where the isoform is known as the slow skeletal muscle isoform $MLC1_S$. (iii) The $MLC1_A$/MLC1emb gene (6,7) which is expressed in cardiac atrial muscle where it encodes the major atrial isoform $MLC1_A$ and in foetal striated muscle where this isoform is known as MLC1emb (8).
Genes of the striated muscle MLC1-type are characterised by the presence of a 5'

exon encoding the NH_2 terminal sequence which distinguishes these isoforms from those of $MLC3_F$ and the smooth muscle and non-muscle isoforms (see 1). Evolutionary considerations suggest that the MLC1-like genes were generated from an ancestral gene similar in construction to the present day $MLC1_F/MLC3_F$ gene, and that they retained the $MLC1_F$-like 5' exon in common. The cis-acting regulatory sequences present in the proximal promoter regions of these genes probably arose in the same way and may therefore have retained common regulatory elements, as well as evolving new ones required for their independent regulation. In addition, these genes are co-expressed in certain tissues. For example, $MLC1_F$ and $MLC1_A$ are co-expressed in foetal skeletal muscle, $MLC1_V$ and $MLC1_A$ are co-expressed in foetal ventricular muscle (see 1). This co-expression may be effected through the use of common regulatory mechanisms using common sequence elements.

We have previously demonstrated that in the case of the mouse $MLC1_F$ promoter, 1200bp of sequence immediately upstream of the cap site is sufficient to direct tissue (muscle cell) specific and developmentally regulated transcription (with respect to the myoblast to myotube transition) of an $MLC1_F$-CAT fusion gene (3). We have therefore sequenced and compared over 1200bp of upstream sequence of the $MLC1_A$, $MLC1_V$ and $MLC1_F$ genes in order to look for elements which may be involved in their regulation. We have previously isolated, characterised and sequenced the mouse $MLC1_F$ (2) and $MLC1_A$ (6) genes. Here we have extended these data and present the sequence of up to -1623bp and -1311bp form their respective cap sites. We also present the isolation, characterisation and sequence of exon 1 and 1297bp of upstream sequence of the mouse $MLC1_V$ gene. We show that the $MLC1_A$ and $MLC1_V$ genes both use a single site of initiation of transcription in cardiac and skeletal muscle, and that in the case of the $MLC1_A$ gene, this site is also used in myogenic cells in culture. Comparison of the upstream regions of these genes reveals sequence elements which may be involved in their regulation; these include a common "MLC-element" and an element common to genes expressed in cardiac muscle.

MATERIALS AND METHODS

Isolation and characterisation of the MLC1V gene promoter.

Genomic recombinant phage λVLC2 was isolated by screening a mouse genomic DNA library (generated by the insertion of Sau3A partially digested mouse (C3H) DNA into the BamH1 site of phage EMBL3, and kindly provided by Laurence Amar) using, as probe, a 900bp Bgl II fragment from recombinant λVLC1 (located at the end of the insert corresponding to the most 5' part of the gene) previously characterised as containing part of the mouse $MLC1_V$ gene (5) . The restriction map of λVLC2 was determined by partial restriction digests following methods described previously (2).

The position of exon 1 was determined by hybridization with the 5' coding region of the human $MLC1_V$ (MLC1Sb) sequence contained in a 380bp Sac I - Pst I fragment of cDNA clone pHMLC1-Sb (9, kindly provided Robert Wade and Peter Gunning) .

Sequencing of promoter regions.

We have previously characterised and sequenced the mouse $MLC1_A$ (6) and $MLC1_F$/$MLC3_F$ (2) genes. Additional sequence from the upstream regions of these genes has .been obtained by sequencing of M13 recombinants generated by sonication of cloned genomic DNA as described (10) and by the use of synthetic oligonucleotides for priming sequencing reactions on suitably long M13 recombinants. For the $MLC1_V$ gene, we have sequenced the 2.1kb Bam HI fragment identified as containing exon 1. Sequencing was carried out using a library of sonicated λVLC 2 DNA fragments cloned in M13 using the 2.1kb Bam HI fragment as probe to identify relevant clones.The exact position of the exon within this fragment was deduced from sequence homology with other MLC genes and confirmed by S1 nuclease analysis as described below.

S1 nuclease and primer extension analysis.

Total RNA was prepared from mouse (C3H) cardiac atrial and ventricular muscle, soleus muscle (11), and from 18 day foetal skeletal (hind leg) muscle and from the myogenic cell line T984-Cl10 (12) as previously described (5). S1 nuclease and primer extension analysis was carried out using probes derived from M13 recombinants from the libraries of sonicated DNA used for sequence determination (see below).

For the $MLC1_V$ gene, probes were derived from M13 clone (D8) containing over 200 bp of upstream sequence, exon 1, and 51bp of DNA of the first intron. The insertion in this recombinant is such that synthesis from ssDNA template generates molecules complementary to the coding strand of $MLC1_V$ mRNA. To define the 3' end of exon 1, probe was prepared using the M13 universal primer and synthesis carried out in the presence of ^{32}P-dCTP and ^{32}P-dTTP with cold dATP and dGTP. Following synthesis, DNA was digested with Pvu II to generate a 190bp probe composed of M13-derived sequence, 51bp of intron sequence and 72bp complementary to the 3' end of exon 1. For defining the 5' end of this exon, a probe was prepared using, as primer, an oligonucleotide (5'-GGGCCATTGGAGCTG-3'; complementary to nucleotides +60 to +74 of the $MLC1_V$ gene) ^{32}P-labelled at the 5' end with polynucleotide kinase, and synthesis was carried out using cold dNTPs. Following digestion with Ava II this generates a 190bp probe complementary to the first 74bp of exon 1 and 116bp of upstream sequence.

For the $MLC1_A$ gene, a probe was synthesised from M13 clone mBG1 using an oligonucleotide (5'-CTCCTTCTTAGGCTC- 3' ; complementary to nucleotides +94 to

+108 of the $MLC1_A$ gene) as primer, in the presence of ^{32}P-dCTP and ^{32}P-dTTP with cold dATP and dGTP. Clone mBG1 contains a 980bp Bgl II fragment of the $MLC1_A$ gene including upstream sequence up to -739, exon 1 of this gene and intron sequence up to nucleotide +255 relative to the cap site. Following digestion with Pvu II this generates a 228bp probe complementary to the first 108bp of exon 1 and 120bp of upstream sequence.

Conditions for S1 nuclease analysis were as described in (6). Briefly, probes generated as described above, were separated on 6% acrylamide gels, electroeluted, precipitated with 10-40µg of RNA, resuspended in 25µl of 80% formamide, 0.4M NaCl, 40mM HEPES pH6.4, 1mM EDTA, heat denatured at 65°C and hybridized at 42°C for 16h. Hybrids were then diluted to 200µl with 30mM sodium acetate pH4.5, 250mM NaCl, 1mM $ZnCl_2$ and digested with 45-450 units of S1 nuclease (Amersham UK) at 25°C for 1h. Digested products were analysed on acrylamide gels and the size of protected products determined by reference to the products of sequencing reactions run on the same gels.

Primer extension analysis was carried out as described in (6). As primer, we used the same $MLC1_V$ oligonucleotide (0.2picomoles) as used for S1 protection analysis. This was ^{32}P-labelled using polynucleotide kinase, and hybridized with 20µg of RNA in 10µl of 100mM NaCl, 20mM Tris-HCL pH8.3, 0.1mM EDTA, at 60°C for 3h. Hybridized probe was extended by addition of 10µl of 80mM Tris-HCl pH8.3, 10mM $MgCl_2$, 4mM dithiothreitol, 4mM each of dATP, dCTP, dGTP, and dTTP and 17 units of AMV reverse transcriptase (Genofit). Extended products were analysed on an 8% acrylamide sequencing gel together with the products of sequencing reactions using the same oligonucleotide as primer on M13 clone D8 containing exon 1 of the $MLC1_V$ gene.

RESULTS.

Isolation and characterisation of the 5' end of the $MLC1_V$ gene: definition of exon 1.

We had previously isolated a recombinant genomic phage, λVLC1, containing most of the mouse $MLC1_V$ gene (5) but which lacks the 5' end. We have used a 900bp fragment of genomic DNA from this recombinant (see materials and methods) to isolate a second overlapping recombinant : λVLC2. The restriction map of this phage (data not shown) indicates that it contains 13kb of insert DNA and overlaps with λVLC1 over 3kb. We identified the presence of exon 1 of the $MLC1_V$ gene within a 2.1kb Bam HI fragment of λVLC2 by cross-hybridization with the 5' end of the human $MLC1_V$ cDNA clone pHMLC1-Sa (9, kindly provided by Robert Wade and Peter Gunning). This was confirmed by sequencing of the entire 2.1kb Bam HI fragment

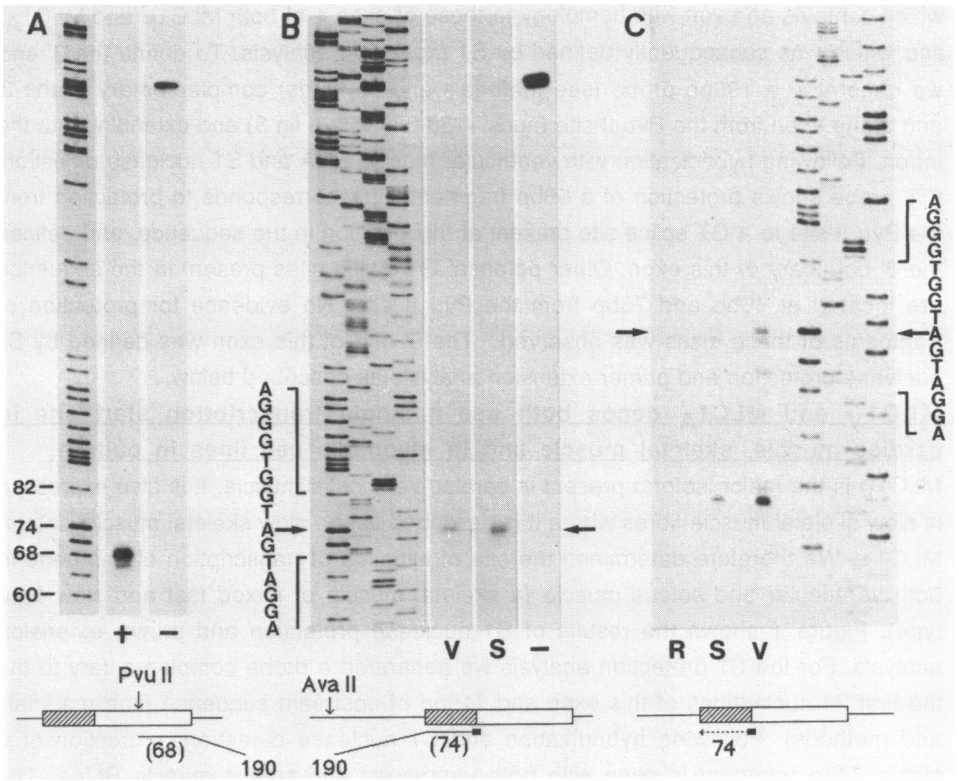

Figure 1: Definition of the 5' and 3' boundaries of exon 1 of the MLC1$_V$ gene. Top. (A); S1 protection analysis using a 190bp probe complementary to the 3' end of exon 1. Probe was hydridized with 10μg of ventricular RNA, treated with nuclease S1 (+) and compared to undigested probe (-). Size markers were derived from sequencing reactions as in B. **(B)**; S1 protection analysis using a 190bp probe complementary to the 5' end of exon 1 hybridized with 10μg of ventricular (V) or soleus (S) muscle RNA digested with nuclease S1 (+) and compared to undigested probe (-). For size markers sequence reactions (T,C,G,A left to right) corresponding to the probe (and complementary to the mRNA strand of the gene shown 5' to 3' top to bottom beside A and C) were used. **(C);** Primer extension analysis using an oligonucleotide primer with ventricular (V), soleus (S), and control ribosomal (R) RNAs. Extended products were run on a sequencing gel with markers as in B. **Bottom** Schematic representation of the probes used in A, B and C. Open boxes represent exon 1, hatched boxes represent non-coding sequence and the small filled box represents the oligonucleotide used in probe preparation and sequencing reactions. Restriction sites used, and the length of the probes, protected fragments (in brackets) and longest extended products are indicated.

which contains an exon with homology to those of exon 1 of both $MLC1_F$ and $MLC1_A$, and which was subsequently defined by S1 protection analysis. To define the 3' end we generated a 190bp probe (see materials and methods) complementary to the 3' end of the exon from the Pvu II site (nucs +153 to +158 in fig 3) and extending into the intron. Following hybridization with ventricular muscle RNA and S1 nuclease digestion, this probe shows protection of a 68bp fragment. This corresponds to protection from the Pvu II site to a GT splice site present at this position in the sequence, and defines the 3' boundary of this exon. Other potential GT splice sites present in the sequence are located at 46bp and 76bp from the Pvu II site. No evidence for protection of fragments of these sizes was observed. The 5' end of this exon was defined by S1 nuclease protection and primer extension analysis as described below.

$MLC1_V$ and $MLC1_A$ genes both use a single transcription start site in cardiac muscle, skeletal muscle and in myogenic cell lines in culture.

$MLC1_V$ is the major isoform present in cardiac ventricular muscle. It is also expressed in slow skeletal muscle fibres where it is described as the slow skeletal muscle isoform $MLC1_S$. We therefore determined the site of initiation of transcription of this gene in both ventricular and soleus muscle (a skeletal muscle of mixed fast and slow fibre type). Figure 1 shows the results of S1 nuclease protection and primer extension analysis. For the S1 protection analysis we generated a probe complementary to the the first 74 nucleotides of this exon and 119bp of upstream sequence (see materials and methods). Following hybridization and S1 nuclease digestion, protection of a single 74bp fragment is seen with both ventricular and soleus muscle RNAs. The sequencing reactions shown on this gel are derived from the same DNA template and oligonucleotide primer as used for probe preparation. The sequence therefore corresponds to that of the protected fragment, and the nucleotide at the site of initiation of transcription can be read directly from the gel. This result is confirmed by primer extension analysis (fig 1) using 5'-labelled oligonucleotide to prime reverse transcriptase synthesis on both ventricular and soleus muscle RNAs. The same primer was used in sequencing reactions with M13 recombinant D8 (see materials and methods), such that the sequence on the gel corresponds to that of the extended products. Both ventricular and soleus muscle RNAs give rise to the same pattern of extension. The longest product is 74 nucleotides ending on the same A residue identified as the transcription start site by S1 protection. The other, shorter, extension products cannot represent *bona fide* mRNAs of different length as these would have also been revealed by the S1 nuclease protection assay. They must therefore represent partially extended products. Together, these results identify the site of initiation of transcription of the $MLC1_V$ gene, and demonstrate that the same site is used in both ventricular and soleus muscle.

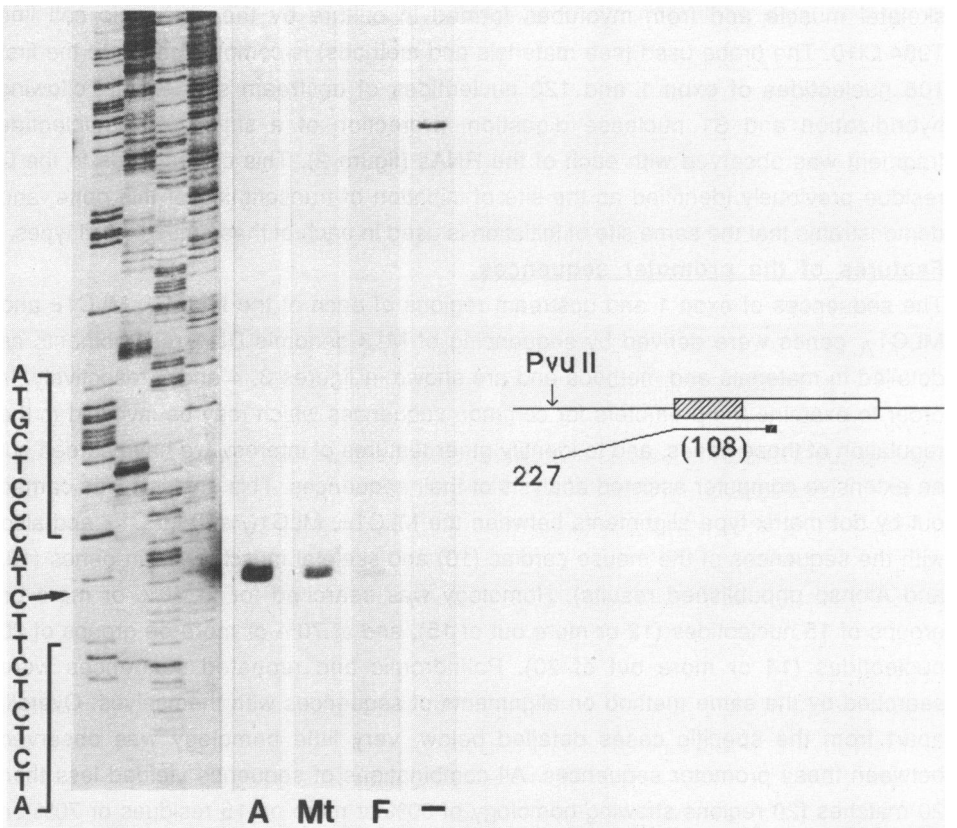

Figure 2 : Definition of the transcription start site of the MLC1$_A$ gene in atrial muscle fetal skeletal muscle and in myotubes formed in culture. Left: S1 protection analysis using a 227bp probe complementary to 108bp of the 5' end of exon 1 and 120bp of upstream sequence/ Probe was hybridized to 10µg of atrial (A) or T984-Cl10 myotube (Mt) RNA or 40µg of fetal leg muscle (F) RNA, treated with nuclease S1 (+) and compared to undigested probe (-). Sequencing reactions (T,C,G,A left to right) corresponding to the probe (and complementary to the mRNA strand of the gene shown 5' to 3', top to bottom in the figure) were used to determine the size of extended products. **Right:** Schematic representation of the probe used. Open box represents exon 1 of the MLC1A gene, the hatched box represents non-coding sequence and the small filled box represents the oligonucleotide primer. Restriction sites used and the length of probe and of the protected fragment (in brackets) are indicated.

For the MLC1$_A$ gene, we have previously characterised and sequenced exon 1. To confirm that a single site of transcription initiation is used in different cell types we have carried out S1 nuclease protection analysis with RNA from cardiac atrial muscle, foetal

skeletal muscle and from myotubes formed in culture by the myogenic cell line T984-Cl10. The probe used (see materials and methods) is complementary to the first 108 nucleotides of exon 1 and 120 nucleotides of upstream sequence. Following hybridization and S1 nuclease digestion, protection of a single, 108 nucleotide fragment was observed with each of the RNAs (figure 2). This corresponds to the C residue previously identified as the site of initiation of transcription of this gene, and demonstrates that the same site of initiation is used in each of these muscle cell types.

Features of the promoter sequences.

The sequences of exon 1 and upstream regions of each of the $MLC1_V$, $MLC1_F$ and $MLC1_A$ genes were derived by sequencing of M13 genomic DNA recombinants as detailed in materials and methods and are shown in figures 3, 4 and 5 respctively. In order to examine the promoters for common sequences which may be involved in the regulation of these genes, and to identify other features of interest, we have carried out an extensive computer assisted analysis of their sequences. This analysis was carried out by dot matrix type alignments between the $MLC1_F$, $MLC1_V$ and $MLC1_A$ and also with the sequences of the mouse cardiac (10) and skeletal muscle α-actin genes (13, and Alonso unpublished results). Homology was searched for at 80% or more on groups of 15 nucleotides (12 or more out of 15), and at 70% or more on groups of 20 nucleotides (14 or more out of 20). Palindromic and repeated sequences were searched by the same method on alignments of sequences with themselves. Overall, apart from the specific cases detailed below, very little homology was observed between these promoter sequences. All combinations of sequence yielded less than 20 matches (20 regions showing homology of 80% or more on 15 residues or 70% on groups of 20) that are located at a similar position ± 400bp. Since, in general, elements which are involved in gene regulation are located in similar positions in different genes, the sequences presented here are those located at ± 200bp.

General features. In each case numbering refers to the most 5' nucleotide of the sequence and is relative to the cap site. Each of the MLC promoters shows the presence of a TATA-like sequence located at -30bp (± 2bp) although in the case of the $MLC1_A$ gene the sequence is atypical (TAAAA). The $MLC1_F$ and $MLC1_V$ genes have CCAAT box elements located close to the classic (-80) position at -75bp and -92bp respectively. The $MLC1_A$ promoter has no CCAAT box in this region but has two CCAAT sequences at -266 and -248. A feature of interest in the $MLC1_V$ promoter, is the presence of a 49bp repeat of nucleotides -1000 to -952 at position -1085 to -1037. This repetition occurs in a region of (GT) rich sequence which flanks both repeats and separates them by the 36 nucleotide sequence $-(GT)_{10}\,G\,(GT)_7\,G-$. The fact that the

```
                                                    5'-AACTGGTGAGAGCCTTAGATCAGAGCCTAGGGGCAAA
-1260 CCTATCCAGCACTTTGGGGTTCCTCAGAGATGCAGGATCCCTGGAGCAGAAGGAAACATGGATAACAGGA
      GTAGTAGAGGAAGACTTCAGAGAGGGCAAATGGAGGGGTGAGCAGGAAGGGACAGAGGGTGCAATGAGTT
-1120 CTAGGGGTGAGCAAGAGGCTGGACCCTGGAGTGTGCAGGGGAAGATAGAGACTGGACATCGCATAGAGGG
      GTGGAGAGAAGCCAGTGTGTGTGTGTGTGTGTGTGGTGTGTGTGTGTGTGCAGAGGAAGATAGAGACTGG
-980  GCATCACGTGGAGGGGTGGAGAGAATCCAGTGTGTATGTATGTGTGTGTGCACACATGTGTGTGTGTGTG
      CGCGCGTGTGCGAGTGTGGGTGTGTAAAACCTGAGACACTAGTCTTAGGAGTTTTGTAAAAGGAGCCTGT
-840  TGAACTAGTCTGGAGGACAGTCCTCGAGGGAGTGCCACGTGTGCTTGGGGTGTTCCTGGAGACCCTGGGA
      CCCTGTGAGTGGACTGCAGCTGCCTGCTGTATCCAGACACTCATACACTAGTGCATACAAACACCCAGAC
-700  ATAAATATACAGACACTCACTCAGACCCACACAGACATACATGAATGTGAACACCCGGAAACGGGACCAC
      AGAGACAAGACACACGACAGCCTCTTCCCCTAGACCCAGCTGGGGCACTGACAGGCCCTGAACAGCGACA
-560  CCAGACAATGCATACAAATGTCCAGAGACCTCCCACCATGCACACAGCCAGGGACCCTCCCTACTTGAAA
      AGGTGCCTCAACACAGATCCTACCACCCCTCGCTCTTAATCAGGGGACCTGGGCCACCCAGGTGCGAAGT
-420  TATTCAAGAACGACCCTTTGGTCATTCCAGCCCCTACTATGGTGTATAGCATGTGGTACACACTTTTCTA
      ACCTTTACCTGTCCAAATTTATTGTGCAGAAGAGTTGCTCCACGTGTGTATGTGTCTGTCTGTCTGTCCA
-280  TCTGTCTGTCTGTCAGTCTGTGTCTGTGTGCTCAGAGCACAGGCTGAGCAGAGGCCGAGGCCTCTGTCTC
      CCTCTCCCTCCCTCTCCCCAACAGGTGACTGGGGGAATGGAGCACACTCCCCTCCCCCACCCATTCCTGG
-140  GCTTGAACAATGCCCTCTCCAGGACCCAAAATAGCCCCTTGAGCAGAGCCAATGCCCTTTTATGGCCCTG
      TCCCTATTGTGCATGGTAGGGGTGGGGCTGGGGTCATGAGGTATATGAGCAGGATAAAGCCCAGGGTGGT
                                                                        Met
+1    AGTAGGACTCACCTATCTTACTCTCTTCTCCTTTCCTTCTTCCCTCCTCTTTGTGATTACAGCTCCAATG

      AlaProLysLysProGluProLysLysAspAspAlaLysAlaAlaAlaProLysAlaAlaProAlaProA
      GCCCCCAAAAAGCCAGAGCCCAAGAAGGATGATGCCAAAGCGGCTGCCCCCAAAGCAGCTCCAGCTCCTG

      laAlaAlaProAlaAlaAlaProAlaAlaAlaProGluProGluArgProLysGluAlaGluPheAspAl
+141  CGGCTGCACCTGCAGCTGCACCTGCGGCTGCACCAGAGCCTGAGCGCCCCAAGGAAGCCGAGTTTGATGC

      aSerLysIleLys
      CTCCAAGATTAAGGTGAGCGGGTAGCTGGGGTGTAGAAGGCACAGAGCCAATGAACGAGTCTCAGAGTGG

+281  AACCTACCATGGCAGGTAAGGCCCTGGACCCGAGCTAAAGCTGGCAGTTTCTGTGCAGAGGTTAGACAGG
      ACCTGTTTGTTTCAGACATCTGCAGGACCACAGGAGGAGGGAGGGTGGTCAGATAGCTGGTGAGACTCCC
+421  CTGAGGCTGACGATACACTCTGAGTGTAGAGTCTGGACCACGAAATGCTCTTTGGTGATTTTGAAGATCG
      ACCCCCTTTGCTCTCTCTGGATTTCATATCTTTAGGGACTCCTGGAGCCAGCCGTCCCCGAGTCTTTAGG
+561  GCACTTCCTTCTCTGGCTGTGGGGGTTCCCAAGTCCATAGCAAACATCCTTAG-T-GATGTAAGGGAAAT
      CAGGGCTTGCAGGCTGAGTGACATCCCTGCCAAGATCACAAGCCTAGAGCTAGCAGAGCCCT-CTCGAGT
+701  GTCATGTTCAGCTGGGAAAATGGAGGAACGGGCAGGAGAGGGAGGCTGTCTCTGAAAGTCATGCAGGAAG
      CTGGTTGTCGAGCACGACCAAGCAGCACCACGAGGCCTTTCCTCGCCCGGCTGTCGTTCCCCACAGGTCG
+841  TAA-3'
```

Figure 3 : Sequence of exon 1 and proximal promoter of the MLC1v gene. Underlined in the sequence are the 49bp repeat, CCAAT and TATA-like elements, the MLC-element and the whole of exon 1. The amino acid sequence of exon 1 is also shown.

```
                                                                    5'-GGATCCACATTAA
      TGTGGTAGAAGAGCCTCAACTCCACAAGGTTCTTTGTCCTCTGACATCTACATGCAAGCCGCACCGACAA
-1540 ATGTGCATTGCATAAACACACATGCATACAGACACACATACAATCATGATGATGGTGAGTAATTATATAT
      GATAGTTAATGGTCATAATAAAATGTAAATGCTCATTAATTAAATGTCAAGCTGAGCAACAAACGTGAGA
-1400 GTGACGCTGGACAATCTTATAATTGCAGTACCCTCTCTGCTGGGATTTGTCACTTAGAAAGTTATACCCA
      TTGGTTTTTTGTTTGGTTGTTTTTTATTTCTTTATTTTTATTTATTTATTTATTTATTTATTTATTTATT
-1260 TTAATTAATTTATTTATTTATTTATTTATTATATGTAAGTACACTGTAGCTGTCTTCAAACACCGTTACA
      GATGGTTGTGAGCCACCATGTGGTTGTTGGGATTCGAACTCAGGACCTTTGGAAGAGTAGTCGGCGCTCT
-1120 TAACCGCTGAGCCATCTCACCAGCCCCCTACCCATTGGTCTTTATGAATTGCTTTTTTGCTTGTTTTCAA
      GAGAGGGTCTCACTAGATAGCTCAGGCTGTCTTGAAACTCACTGTGTAGACCAGACTGGCCTTGAACTCA
-980  CAGAGAGCTATCTGTCTCTGTCCCAGGTACTCAAATTAAAGCTATACATTAATACACCCAGTTTTGGTTT
      TCATATATTACTTAATCTGAATCTAACCTTGCAAACTAGAAACCAAGGACAAAAATAAGAAAAAGCAAAT
-840  ATCAAATTGCCAGCAAAACATACACAGTATTGATAATTAAGTCTTTTTTCAAATTTATTTCCCTAAAAAA
      TGTATTTGTTCCTCTCGCATGGATATAAGATTAAGGAGACAAATAATTCTATTTCATGTCATTTAGAAGG
-700  CAATGGGCTAATTGCAGACTTCAACTGAAAACTAATTCACCATCCTTCATTAAATTGTACAGAAAATCAA
      GTTATAGGAAATAATTTTAAATTAAGTTTTCTTTAGTGTACGGACACAGACATAGAGAGAGGGAGGCAAG
-560  GAGAGACACAAACTGTACTCCCTAGAAAGGAAACCACTTTAAGACTTGGTGGACTTTCCTTATACGTCTG
      CTTTTTATTTTTAACCACTACCAGTGACCTCTTTGCTGACCAACATTTTTCTTGCAAACCATGGCCAAGG
-420  AGGGAGGGGCATGAGCCGTGAAGAAGCTGGGTGAAAGTTGGGTGATGGCCTCAGCCAAAATTCCAAGTTG
      AAATTGAATCCGGTAAAGCTTCTTAGGAAGTGGAAAAAAGAAAAAGAAGACTTACAGCTTCAACAGAACT
-280  GTAATTCTACCAAGAAAAAGGTACGTCTGTCAGCTGTCACCAAACCAACCATGACTGGCTTTGTCATGCT
      TCTTTCTTCTTTACAGATAGGGGTGGAGGGAAAAAAAAAAGCAAGAGGCGGTCAAAAAACTTGAGGAAGC
-140  CCTGACCCTTTAGATTCCATTTATAGTCTGAGCCGGAATGCCACCCCCCTGGACAGGCAAACCGTCCAAT
      CCAGCCGCGTGTGTCAAGGTTCTATTAGGCACTAAGGGAAATATATATGCATGCCCTTATACCATTTAAC
+1   ACTCTGGGTCCACCCTCCAGACGCTGGGCTGCGGGTCAAGCGAGTCACACTGCTCTTCCAAGTGTCACTC
                                                                    MetAlaProLysLys
      TGACAGGTTCTTCTGGAGGAGATCCTTCTTTTGTTAACCCATCCTTTTAATCAAAATGGCACCAAAGAAA
                                                                    
      AspValLysLysProAlaAlaAlaProAlaProAlaProAlaProAlaProAlaProAlaLysProLysG
+141  GACGTGAAGAAGCCCCGCTGCTGCGCCTGCCCCAGCCCCGGCCCCGGCCCCTGCCCCAGCCAAACCTAAGG
      
      luGluLysIleAspLeuSerAlaIleLys
      AAGAAAAAATCGATCTGTCTGCTATTAAGGTAACTTAAACAGGTGAATCGTTTTGGTCACTGAAACGTCT
+281  TTCAACGCTAAAGAACTCAAGGAATTTGTGGAGCATTCTATTTTTCACTGAAAGGAAAGGCATTTCTCAG
      TGAACCGGTGTGTTTTAAAGCCGGGAATCCTGCTTAATAATATGGGCTGGAGCATACA-3'
```

Figure 4 : Sequence of exon 1 and proximal promoter of the $MLC1_F/MLC3_F$ gene Underlined in the sequence are CCAAT and TATA-like elements, the MLC-element, and the whole of exon 1. The amino acid sequence of exon 1 is also shown.

repeat is not identical (6 substitutions) makes it highly unlikely that this represents a cloning artefact, but rather suggests that it represents a duplication probably provoked by the presence of the flanking (GT) sequences: This type of sequence is known to promote recombination (eg. 14) and, since it flanks the repeated element, could in this

```
                .         .         .         .         .         .         .
                                    5'-GTGAAAAAAAAGATGTTTCTGCCGTTCTCCCTCTCACACCTTAGGTCCAGG
-1260 GGTGTGGTTGTGAACAGTTGATTCCCACCTCCCCACCTCCACCTCCCCAGCCTTCGTTCCTTTCCTCGCT
      TTCTCTCTGTTTTTTTTTTTTTTTTTTGTTGTTGTTGTTTTTTTTTCCGGAATTAAGATCTTGCTACAT
-1120 AACCCAGACGGGTTTCAAACTTACGGTCCCTCTGTTTCTTCTCCTGAGAGTCGGGATGACCCTGGGCCTC
      ATCCTGCTCTATTTCTATCCTGCCCTTCCCGATGACTGGTCCACCCCACTTTTCTTTCTCATATAAATGA
-980  GACTGTAGTATAAATTATTCACCTACACACATTAGGACTAGTCCTACCGGTCTTCCAATCCCAACTCATG
      ATTAATCATTTAAGTGACCTTGGCCAAGTTATTCTAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
-840  GTTATCTGAGTAATGATGGTATCCGTAGAATTATGGAAGGTTTAATGATATTTTATAGAAATAGTGTCTA
      GTAGTATCTGGCACATAGTAAGCAACAGACAGATCTTGAGTTTCCTAACCTCCAGCCTCCCAACTTGTTG
-700  CCATATCTCTGTTTCTTACCCAGATTTCCACCCTCGGCACAAATGTGCAATACTTGTTACAGAACAACTG
      CAGGTAGAGGGTGGATCTAGGAAGCTGAAGTCTGTGGAGTTCTTGGGATGCTTGCCCTGGACAACACAGG
-560  GCTTTGCCTGCGGTTGGGAGAGTGGGAATAGCAGAGCCAGCTCCACCAGCTGTTGCTTCAGGGCTCCATG
      CCTGAACTGTTGAGCATCAGGGCAGGGCCTAGGGCTAGCTTAAAGGCTCTTGGTATAAAAGGAATGTGTA
-420  GTGGGCTAGCCCCTGCTCAGGAGGGACTAAAAAAAGCCTCAACCTCCCCTGGGCTTTGTGTGAGGGTTAT
      CAACGGCTCAAGTCGGCTCATCCCTCTCTGGCTGCTCTGGCCCCTTGGAGAGGGGCTCTTTTCCCACCCC
-280  TCTCTGGGTTTCCACCAATTGGCAGGAAAGGACCAATCTGTCCCAGAGGTGCAAGCCTTGAGATGAGTGA
      GGTGTGTGCATGGGGGGTGGGGGTGGGCTGGGCTGGGGTGGGGGAGGATGGATTGGTGGCTCAAACCTGG
-140  TGATAATACTGAGATGTCAGCTGCACCCGGCTGGTGTCTCTTCCTTTTATAGTCAGCAGCAGTTGCTGCT
      GCCCTCCCCAGCCCCTCTGTGGGGGCTCCTACCCAGAATAAAAGCAGGGGCAGGCCTTCCAGTCTCCCAT
+1   CTTCTCTCTCACGAGCCACCTCTCCTCCTTTGGTTCCTTTCTAGGTCCCACCTCCACTGGAGAGCCTAAA
         MetProProLysLysProGluProLysLysGluAlaAlaLysProAlaAlaAlaProAlaProAl
     GCAACATGCCTCCCAAGAAACCCGAGCCTAAGAAGGAGGCTGCCAAGCCGGCTGCAGCCCCTGCTCCAGC
     aAlaSerAlaAlaProGluProLeuLysAspSerAlaPheAspProLysSerValLys
+141 TGCATCCGCAGCTCCGGAGCCCCTCAAGGACTCTGCCTTTGACCCAAAGAGTGTGAAGGTAAGTCAAACC
      CACCACACCGACAAGGACCATCAGGCTGAGAGATCTTATTTCTCTCAGGCCAGACAGACATGAGGACCAG
+281 GGCCTGTATGCCCTAGATCATAGCCCGAGGCAGCAGAACAGTACGAGATTGGGGTCAGAGGC-3'
```

Figure 5 : Sequence of exon 1 and proximal promoter of the $MLC1_A$ gene. Underlined in the sequence are CCAAT and TATA-like elements, the MLC-element as well as the whole of Exon 1. The amino acid sequence of exon 1 is also shown.

case have given rise to an unequal cross-over. The $MLC1_A$ gene has a $(GT)_{16}$ sequence located at -871 to -840. The $MLC1_F$ gene has no such sequence within the 1623bp of upstream sequence analysed here.

Identification of an "MLC-sequence" and other motifs. Sequence comparisons between $MLC1_F$, $MLC1_V$ and $MLC1_A$ have revealed a number of shared sequence elements (see table 1). Of these however, only one was found to be common to all three genes. This element contains a core sequence of 10 nucleotides (CCTTTTATAG) and is located at a similar position in each promoter ($MLC1_F$; -124, $MLC1_A$; -99 and $MLC1_V$; -85). Furthermore this sequence is also found, at the same position, in the $MLC1_F$ promoters of rat (4, 15) and chicken (16, 17) which are the only

A : An element common to myosin alkali light chain genes

MLC1v	-85	CCTTTTATGG	MLC1F -Mouse	-124	CCATTTATAG
MLC1A	-99	CCTTTTATAG	MLC1F -Rat	-124	CCATTTATAG
MLC1F	-124	CCATTTATAG	MLC1F -Chick	-122	CCATTTATAG
Consensus		CCTTTTATAG			

B : An element common to MLC1v MLC1A and α-cardiac actin.

```
MLC1A            (-286)   CTCTTTTCCCACCCCTCTCTGGGTT
                           * * *  ******* *  ***** *
MLC1v            (-161)   CCCCTCCCCCACCCATTCCTGGGCT
                           *** ********** ** **** **
α-cardiac actin  (-278)   CCCGTCCCCCACCCCTTGCTGGCCT
```

Similar sequences located at

```
MLC1v            (-465)        CTACCACCCCTCGCT
α-cardiac actin  (-83)       CCCCCCACCCCTGCC
                 (-215)      GCCCCCTCCCCTTCC
                 (-319)      TCCCCCACCCA
MLC1A            (-74)     CCCTCCCCAGCCCCTCTCTG
```

C : Other shared sequence elements.

```
MLC1v   (-66)    TGTGCATGGTAGGGGTGGGGCTGGGGT
                 *********  ********* **** *
MLC1A   (-206)   TGTGCATGG--GGGGTGGGGGTGGGCT

MLC1A   (-1006)  CCCACTTTTCTTTCTCATATA
                 **** ****  ** *******
MLC1F   (-924)   CCCAGTTTTGGTTTTCATATA

MLC1v   (-76)    GCCCTGTCCCTATTG
                 ****** **** * *
MLC1F   (-142)   GCCCTGACCCTTTAG

MLC1v   (-154)   CCACCCATTCCTGGGCTTA
                 *** ****  ****  **
MLC1F   (-237)   CCAACCATGACTGGCTTTG
```

D. Elements showing homology with the E1A-enhancer sequence.

```
                          C       A
E1A consensus             AGCAGGAAGTGACTAA  (underlined is the core sequence)

                           *******  *** **
MLC1A   (-280)            GGCAGGAAAGGACCAA

                          ********* ***  *
MLC1v   (-1150)           AGCAGGAAGGGACAGA

                             ******** * **
MLC1F   (-329)            CTTAGGAAGTGGAAAA
```

Table 1 : Nucleotide sequence homologies identified in the promoter regions of the MLC1F, MLC1v and MLC1A genes. Numbering is relative to the cap site and refers to the first nucleotide of the sequence shown. α-cardiac actin sequence is from (10), rat and chick MLC1F sequences are from (4) and (21) respectively. E1A enhancer consensus is from (19).

other myosin alkali light chain genes for which promoter sequence data is currently available. The conservation of sequence and position of this element strongly implies that it is involved in the transcription of these genes.

Since both the $MLC1_A$and $MLC1_V$ genes are expressed in cardiac muscle we have looked for elements shared between these genes and with the α-cardiac actin gene. Only one such element was found (see table 1) which is centered around the motif (CCCACCCCT) and which is in fact repeated in each of these genes. The main homology is that located at -286 in $MLC1_A$, -161 in $MLC1_V$ and at -278 in the α-cardiac actin gene where homology extends beyond this core sequence to give 21 matches out of 25 nucleotides for $MLC1_V$ versus α-cardiac actin, 14 out of 18 for $MLC1_A$ versus $MLC1_V$ and 13 out of 16 for $MLC1_A$ versus α-cardiac actin. This element is found in different forms in the α-cardiac actin gene at -83 at -215 and at -319, and in $MLC1_A$ at -74, and in $MLC1_V$ at -456.

Other sequences shared between the MLC promoters include an element showing 23 out of 27 matched nucleotides between $MLC1_V$ and $MLC1_A$ in a region composed of the motifs ($GGG^G/_CT$) and ($GGGG^G/_CT$) (see table 1). It is interesting to note that this sequence forms an imperfect inversion of the (CCCACCCCT) motif noted above, and could potentialy be involved in forming loop structures in the $MLC1_A$ and $MLC1_V$ promoters. This sequence is not present in $MLC1_F$, α-cardiac or α-skeletal actin genes. Other homologies include a region containing a CCArGG box-like element ($CC(A+T\text{-rich})_6GG$: see 18). The $MLC1_V$ promoter contains one such elements at -85 which in fact corresponds to the MLC-element noted above. The $MLC1_F$ promoter contains two CCArGG box-like elements at -923 (CCAGTTTTGG), and at -540 (CCTAGAAAGG). No such elements exist in the $MLC1_A$ promoter region presented here.

We have previously noted the presence of a sequence in both the $MLC1_F$ and α-cardiac actin gene promoters (10, 17) which is similar to that of the viral E1A core enhancer sequence $^A/_CGGAAGTGA^A/_C$ (19). We have examined the $MLC1_A$, $MLC1_V$ and extended $MLC1_F$ promoter regions and have found similar sequences in each, although at rather different positions as shown in table 1. In the case of the $MLC1_A$ gene this E1A enhancer-like element is flanked by two CCAAT box elements.

DISCUSSION

In order to define the proximal promoter regions of the three myosin alkali light chain genes expressed in mouse striated muscle we have determined the site of initiation of transcription of these genes in different muscle types. The $MLC1_A$ and $MLC1_V$ genes are expressed in both skeletal and cardiac muscle. We have shown here that the same site of initiation of transcription, and therefore the same proximal promoter region, is

used by these genes in both cell types. We have sequenced and compared over 1200bp of the proximal promoter regions of the $MLC1_F$, $MLC1_V$ and $MLC1_A$ genes in order to identify potential cis-acting elements. In the case of the mouse $MLC1_F$ gene this region is sufficient to direct tissue (muscle cell) specific regulation and correct activation of transcription during the myoblast - myotube transition of an $MLC1_F$-CAT fusion gene (3) and the same is true of the chick $MLC1_F$ gene (20, 21).

Of the common sequence motifs identified between these genes, two are of particular interest. Firstly there is a motif (consenus; CCTTTTATAG) present in $MLC1_A$, $MLC1_V$ and $MLC1_F$ which is conserved in sequence and position, and which is also present at the same position in the $MLC1_F$ gene promoters of rat and chick (both CCATTTATAG). It is worth noting that sequence elements which are involved in tissue specific gene regulation, such as the immunoglobulin octamer (eg. 22), the heat shock consensus sequence (eg. 23) and regulatory elements of the albumin gene (eg. 24) are all short (less than 10 nucleotides) and are not always completely conserved. In the case of the MLC-element it is possible that the minor differences between genes plays a role since, in the case of the $MLC1_F$ gene, the sequence is conserved between species. It will be of interest to see whether other MLC genes follow this rule as data become available for other species.

In the case of the $MLC1_V$ gene, the sequence of the MLC-element (CCTTTTATGG) corresponds to that of a "CCArGG" box ($CC(A+T\text{-rich})_6GG$: see 18). These elements have been shown to be involved in regulation of α-cardiac, α-skeletal actin and c-fos gene transcription and to bind nuclear protein factors (25, 26, 27, 28, 29). CCArGG box elements identical with the mouse $MLC1_V$ sequence are also found in both the human interleukin 2 receptor gene (30) and in the β-actin (non-muscle actin) genes of mouse rat and man in a region of the first intron that has been shown to have enhancer activity (31). The exact relationship between this element and those of the MLC-genes is unclear but as the β-actin and interleukin 2 genes are expressed in non-muscle cells this element cannot, in itself, confer muscle specificity.

A second sequence identified is that of a cardiac muscle element with a central consensus motif (CCCACCCCT). This sequence is also present in the α-cardiac actin gene promoter but is not found in those of $MLC1_F$ or α-skeletal actin. It is interesting that it is repeated several times in these genes and in the case of $MLC1_A$ and $MLC1_V$ it forms an imperfect inverted repeat with the $GGG^G/_CT$ motifs present in these promoters (see table 1). Further definition of this element, and of the MLC-element will require functional analysis by cell transfection and DNA binding type assays to determine their role in the regulation of these genes.

Recently, a number of enhancer-like elements have been identified which may play a specific role in muscle gene expression (20, 32). These elements are not required for

correct tissue specific, or developmentally regulated expression, but rather are involved in augmenting the level of transcription. There is no homolgy between the published sequences of these enhancer regions and the MLC genes presented here. In contrast we have found sequences in the $MLC1_F$, $MLC1_V$ and $MLC1_A$ promoters which show significant homology with the viral E1A core-enhancer sequence, and have previously noted the presence of similar elements in the mouse α-cardiac actin gene (10). Polyoma gene transcription, including that dependent on the A box which has an E1A-like enhancer element, is up regulated during myogenesis *in vitro* (33) and it is possible that these elements are involved in this type of stimulation.

ACKNOWLEDGEMENTS.

The authors wish to thank the participants of the 1987 Pasteur DNA sequencing course for their contribution to this work. We are grateful to Peter Gunning, Robert Wade and Larry Kedes for providing the human $MLC1_V$ cDNA clone pHMLC1-Sb, and to Christine Biben for help in sequence analysis. This work was supported by the Association pour la Recherche sur le Cancer, the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale and the Muscular Dystrophy Association of America. I.G. is the recipient of a fellowship from the Association Française contre les Myopathies.

*To whom correspondence should be addressed

REFERENCES

1. Barton,PJ.R. and Buckingham,M.E. (1985) Biochem. J. **231**, 249-261.
2. Robert,B. Daubas,P. Akimenko,M.A. Cohen,A. Garner,I. Guénet, J.L. Buckingham,M. (1984) Cell **39**, 129-140.
3. Daubas,P. Klarsfeld,A. Garner,I. Pinset,C. Cox,R. Buckingham,M. (1988) Nucl. Acids Res. **16**, 1251-1271.
4. Strehler,E.E. Periasamy,M. Strehler-Page,M.A. Nadal-Ginard,B. (1985) Mol. Cell Biol. **5**, 3168-3182.
5. Barton,PJ.R. Cohen,A. Robert,B. Fiszman,M.Y. Bonhomme,F. Guénet,J.L. Leader,D.P. Buckingham,M.E. (1985) J. Biol. Chem. **260**, 8578-8584.
6. Barton,PJ.R. Robert,B. Cohen,A. Garner,I. Sassoon,D. Weydert,A. Buckingham M.E. (1988) J. Biol. Chem. **263**, 12669-12676.
7. Barton,P.J.R. Robert,B. Fiszman,M.Y. Leader,D.P. Buckingham,M.E. (1985) J. Muscle Res. Cell Motil. **6**, 461-475.
8. Whalen,R.G. Butler-Browne,G.S. Gros,F. (1978) J. Mol. Biol. **126**, 415-431.
9. Wade,R. Feldman,D. Gunning,P. Kedes,L. Submitted for Publication.
10. Garner,I. Minty,A.J. Alonso,S. Barton,P.J. Buckingham,M.E. (1986) EMBO J. **5**, 2559-2567.
11. Chirgwin,J.M. Przybyla,A.E. MacDonald,R.J. Rutter.W.J. (1979) Biochemistry **18**, 5294-5299.

12. Jakob,H. Buckingham,M.E. Cohen,A. Dupont,L. Fiszman,M. Jacob,F. (1978) Expl. Cell Res. **114**, 403-408.
13. Hu, M.C-T. Sharp,S.B. Davidson,N. (1986) Mol.Cell Biol. **6**, 15-25.
14. Kmiec,E.B. Angelides,K.J. Holloman,W.K. (1985) Cell. **40**, 139-145.
15. Periasamy,M. Strehler,E.E. Garfinkel,L.I. Gubits,R.M. Ruiz-Opazo,N. Nadal-Ginard,B. (1984) J. Biol. Chem. **259**, 13595-13604.
16. Nabeshima,Y-I. Fujii-Kuriyama,Y. Muramatsu, M. Ogata,K. (1984) Nature (Lond.) **308**, 333-338.
17. Daubas,P. Robert,B. Garner,I. Buckingham,M. (1985) Nucl. Acids Res. **13**, 4623-4643.
18. Minty,A. Kedes,L. (1986) Mol. Cell Biol. **6**, 2125-2136.
19. Hearing,P. Shenk,T. (1983) Cell. **33**, 695-703.
20. Billeter,R. Quitschke,W. Paterson,B.M. (1988) Mol. Cell Biol. **8**, 1361-1365.
21. Shirakata,M. Nabeshima,Y-I. Konishi,K, Fujii-Kuryama,Y. (1988) Mol. Cell Biol. **8**, 2581-2588.
22. Falkner,F.G. Zachau,H. (1984) Nature (Lond.) **310**, 71-74.
23. Pelham,H.R.B. (1982) Cell **30**, 517-528.
24. Heard,J.M. Herbomel,P. Ott,M.O. Mottura-Rollier,A. Weiss,M. Yaniv,M. (1987) Mol. Cell. Biol. **7**, 2425-2434.
25. Miwa,T. Boxer,L.M. Kedes,L. (1987) Proc. Natl. Acad. Sci. (USA) **84**, 6702-6706.
26. Phan-Dinh-Tuy,F. Tuil,D. Schweighoffer,F. Pinset, C. Kahn,A. Minty,A. (1988) Eur. J. Biochem. **173**, 507-515.
27. Walsh,K. Schimmel,P. (1987) J. Biol. Chem. **262**, 9429-9432.
28. Muscat,G.E.O. Kedes,L. (1987) Mol. Cell. Biol. **7**, 4089-4099.
29. Treisman,R. (1985) Cell **42**, 889-902.
30. Cross,S.L. Feinberg,M.B. Wolf,J.B. Holbrook,N.J. Wong-Staal,F. Leonard,W.J. (1987) Cell **49**, 47-56.
31. Kawamoto,T. Makino,K. Niwa,H. Sugiyama,H., Kimura,S. Amemura,M. Nakata,A. Kakunaga,T. (1988) Mol. Cell Biol. **8**, 267-272.
32. Jaynes,J.B. Johnson,J.E. Buskin,J.N. Gartside,C.L. Hauschka,S.D. (1988) Mol. Cell Biol. **8**, 62-70.
33. Felsani,A. Maione,R. Ricci,L. Amati,P. (1985) Cold Spring Harbour Symp. Quant. Biol. **50**, 753-757.