Multiple gene products are produced from a novel protein kinase transcription region

D. Martin Watterson[a,*], Mark Collinge[a], Thomas J. Lukas[a], Linda J. Van Eldik[b], Konstantin G. Birukov[c], Olga V. Stepanova[c], Vladimir P. Shirinsky[a,c]

[a]*Department of Molecular Pharmacology and Biological Chemistry, Northwestern University Medical School, 303 E. Chicago Ave., Chicago, IL 60611-3008, USA*
[b]*Department of Cell and Molecular Biology, Northwestern University Medical School, 303 E. Chicago Ave., Chicago, IL 60611-3008, USA*
[c]*Cardiology Research Center, Russian Academy of Medical Sciences, 3-rd Cherepkovskaya Str. 15a., Moscow 121552, Russian Federation*

Received 21 August 1995

Abstract The nonmuscle/smooth muscle myosin light chain kinase (MLCK) and the kinase related protein (KRP) that lacks protein kinase activity are myosin II binding proteins encoded in the vertebrate genome by a true gene within a gene relationship. The genomic organization and expression result in the same amino acid sequence in different molecular contexts from two different sizes of mRNA. We report here the identification and characterization of a third size class of gene products. The protein appears to be a higher molecular weight form of MLCK with additional amino terminal tail sequence which might provide differential subcellular targeting characteristics.

Key words: Gene expression; Amino acid sequence; Protein kinase; Myosin; Calmodulin

1. Introduction

The calmodulin:myosin light chain kinase (MLCK) complex from vertebrate nonmuscle and smooth muscle tissues is a prototype calcium signal transduction complex (for a recent review see [1]). The MLCK and the kinase related protein (KRP) are encoded by a novel gene within a gene relationship in which the TATA box and start of transcription for the KRP mRNA are within an MLCK intron, but the start of KRP mRNA translation is at an ATG sequence that is an internal methionine codon for MLCK [2,3]. This results in the production of the same amino acid sequence in two different molecular contexts from two different mRNA size classes. While MLCK has protein kinase activity and KRP does not [2,3], both proteins have myosin II binding activity [4]. The KRP domain of MLCK appears to function as a myosin targeting signal for the protein kinase, while the KRP protein appears to be a stabilizer of myosin minifilaments under conditions indicative of the relaxed state of the cell [4]. Although many details remain to be established in terms of cell physiology, it is clear that the same section of genomic DNA contains at least two genes encoding at least two size classes of mRNA and proteins with distinct regulation and function [2–4].

The physical clustering of the MLCK and KRP genes in the genome [2] and the apparent involvement of both proteins in the regulation of cell structure and function through the modulation of the myosin II molecular motor system [4] are reminiscent of an operon-like arrangement of cell functions found in prokaryotes, or the multiple use of DNA sequences seen with viruses. This raises the question: does this same region of genomic DNA encode additional gene products that are potential regulators of myosin II structure and function? As a first step in addressing this question and further exploring the details of this novel genetic locus, we pursued the observation that a higher molecular weight mRNA (9 kb) hybridized with probes complementary to the mRNAs encoding MLCK (5.5 kb) and KRP (2.7 kb). We isolated and sequenced additional cDNA clones and engineered antisera against synthetic peptide antigens encoded by the region of unique amino acid sequence found in the open reading frame (ORF). We report here the identification and characterization of this additional size class of gene products encoded by the MLCK/KRP transcription region. The amino acid sequence, protein immunochemical properties and mRNA distribution suggest that the newly described protein is a higher molecular weight form of MLCK with additional amino terminal tail sequence. Based on these properties and the computed protein mass, we have called this protein MLCK-210.

2. Materials and methods

2.1. Molecular biology

By previously described protocols [2,3], cDNA libraries derived from adult chicken brain and chicken embryo fibroblast (CEF) polyadenylated RNA were screened, overlapping cDNA clones isolated, and sequences determined for both strands of DNA. The minimal clones and sequence required for the open reading frame (ORF) for MLCK-210 were found in two overlapping clones termed (from 5′ to 3′ end) 7n-32-R and λ11, the latter clone being described previously by Shoemaker et al. [3]. Briefly, the approach was a 5′ walk strategy in which libraries were sequentially screened with the more 5′ end of previously isolated clones. Initially, this allowed the isolation of two indistinguishable cDNA clones from the CEF and adult brain libraries by using probes corresponding to nucleotides 722–922 and 1229–1584 of the MLCK cDNA sequence previously described [3]. These clones, designated 45t-26-BBB (CEF) and 9e-1 (adult brain), contained nucleotides 1–1760 of the cDNA sequence of Shoemaker et al. [3] plus an additional 746 bp of more 5′ sequence. This 746 bp unique sequence was used as a probe to re-screen the CEF library and isolate cDNA clone 7n-32-R, which overlaps clone λ11. The composite DNA sequence encodes an ORF beginning with an ATG codon towards the 5′ end of 7n-32-R (the A is designated as nucleotide #1 in the EMBL database entry with accession number X52876) that is continuous with the open reading frame described by Shoemaker et al. [3] and encodes a protein of 1906 amino

*Corresponding author. Fax: (1) (312) 503-0796.
E-mail: m-watterson@nwu.edu

Abbreviations: CEF, chicken embryo fibroblasts; KRP, kinase related protein; MLCK, myosin light chain kinase.

SSDI 0014-5793(95)01048-3

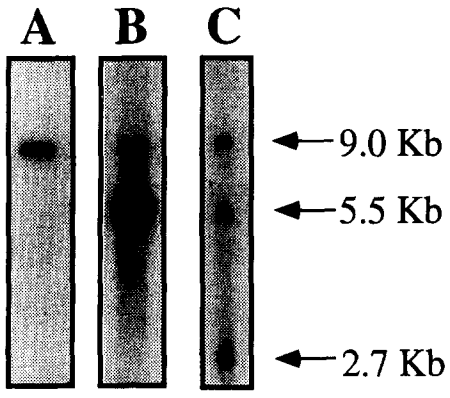

Fig. 1. Northern blot analysis of chicken mRNA with probes based on the MLCK/KRP transcription region. Poly(A^+) RNA (5 μg per lane) from CEF (lanes A and B) or adult chicken brain (lane C) were subjected to Northern blot analysis using three probes based on different regions of the MLCK/KRP transcription region. Lane A: Selective detection of 9.0 kb mRNA size class using a 91-bp probe corresponding to the unique region of the MLCK-210 cDNA sequence (nucleotides 544–634 of EMBL X52876; encompassing amino acids 181–211 of Fig. 2). Lane B: Selective detection of 9.0 kb and 5.5 kb mRNA size classes using a 199-bp probe corresponding to a common region of cDNA sequence for MLCK-210 and MLCK-108 not found in KRP (nucleotides 2668–2866; encompassing amino acids 889–955). Lane C: Detection of all three mRNA size classes (9.0, 5.5 and 2.7 kb) using a 681-bp probe located within the common 3′-noncoding region (nucleotides 5362–6042 of EMBL accession number M88284).

acids and a computed mass of 210,465 (designated MLCK-210). PCR amplification and labeling of probes, preparation of RNA and Northern blot analyses were performed as previously described [2].

2.2. Immunochemical methods

To prepare antibodies selective for MLCK-210, a peptide immunogen (AKKETFYTSREAKDGK) corresponding to amino acids 306–321 of the MLCK-210 ORF was synthesized and conjugated to keyhole limpet hemocyanin (KLH) as previously described [5]. KRP was purified from chicken gizzard tissue as previously described [2,4]. Rabbit antibodies against the peptide-KLH immunogen or against purified KRP were generated by East Acres Biologicals, Inc (Southbridge, MA) using previously described protocols [5].

Smooth muscle cells from the medial layer of adult chicken aorta were prepared using the approach of Birukov et al. [6] and grown in DMEM containing 10% fetal calf serum. Confluent cells of passages

```
›MLCK-210
MGDVKLVTSTRVSKTSLTLSPSVPAEAPAFTLPPRNIRVQLGATARFEGKVRGYPEPQITWYRNGHPLPEGDHYVVDHSI      80

RGIFSLVIKGVQEGDSGKYTCEAANDGGVRQVTVELTVEGNSLKKYSLPSSAKTPGGRLSVPPVEHRPSIWGESPPKFAT     160

KPNRVVVREGQTGRFSCKITGRPQPQVTWTKGDIHLQQNERFNMFEKTGIQYLEIQNVQLADAGIYTCTVVNSAGKASVS     240

AELTVQGPDKTDTHAQPLCMPPKPTTLATKAIENSDFKQATSNGIAKELKSTSTELMVETKDRLSAKKETFYTSREAKDG     320

KQGQNQEANAVPLQESRGTKGPQVLQKTSSTITLQAVKAQPEPKAEPQTTFIRQAEDRKRTVQPLMTTTTQENPSLTGQV     400

SPRSRETENRAGVRKSVKEEKREPLGIPPQFESRPQSLEASEGQEIKFKSKVSGKPKPDVEWFKEGVPIKTGEGIQIYEE     480

DGTHCLWLKKACLGDSGSYSCAAFNPRGQTSTSWLLTVKRPKVEEVAPCFSSVLKGCTVSEGQDFVLQCYVGGVPVPEIT     560

WLLNEQPIQYAHSTFEAGVAKLTVQDALPEDDGIYTCLAENNAGRASCSAQVTVKEKKSSKKAEGTQAAKLNKTFAPIFL     640

KGLTDLKVMDGSQVIMTVEVSANPCPEIIWLHNGKEIQETEDFHFEKKGNEYSLYIQEVFPEDTGKYTCEAWNELGETQT     720

QATLTVQEPQDGIQPWFISKPRSVTAAAGQNVLISCAIAGDPFPTVHWFKDGQEITPGTGCEILQNEDIFTLILRNVQSR     800

HAGQYEIQLRNQVGECSCQVSLMLRESSASRAEMLRDGRESASSGERRDGGNYGALTFGRTSGFKKSSSETRAAEEEQED     880
                                                                ›MLCK-108
VRGVLKRRVETREHTEESLRQQEAEQLDFRDILGKKVSTKSFSEEDLKEIPAEQMDFRANLQRQVKPKTLSEEERKVHAP     960

QQVDFRSVLAKKGTPKTPLPEKVPPPKPAVTDFRSVLGAKKKPPAENGSASTPAPNARAGSEAQNATPNSEAPAPKPVVK    1040

KEEKNDRKCEHGCAVVDGGIIGKKAENKPAASKPTPPPSKGTAPSFTEKLQDAKVADGEKLVLQCRISSDPPASVSWTLD    1120

SKAIKSSKSIVISQEGTLCSLTIEKVMPEDGGEYKCIAENAAGKAECACKVLVEDTSSTKAAKPAEKKTKKPKTTLPPVL    1200

STESSEATVKKKPAPKTPPKAATPPQITQFPEDRKVRAGESVELFAKVVGTAPITCTWMKFRKQIQENEYIKIENAENSS    1280

KLTISSTKQEHCGCYTLVVENKLGSRQAQVNLTVVDKPDPPAGTPCASDIRSSSLTLSWYGSSYDGGSAVQSYTVEIWNS    1360

VDNKWTDLTTCRSTSFNVQDLQADREYKFRVRAANVYGISEPSQESEVVKVGEKQEEELKEEEAELSDDEGKETEVNYRT    1440

VTINTEQKVSDVYNIEERLGSGKFGQVFRLVEKKTGKVWAGKFFKAYSAKEKENIRDEISIMNCLHHPKLVQCVDAFEEK    1520

ANIVMVLEMVSGGELFERIIDEDFELTERECIKYMRQISEGVEYIHKQGIVHLDLKPENIMCVNKTGTSIKLIDFGLARR    1600

LESAGSLKVLFGTPEFVAPEVINYEPIGYETDMWSIGVICYILVSGLSPFMGDNDNETLANVTSATWDFDDEAFDEISDD    1680
                                                                ›KRP
AKDFISNLLKKDMKSRLNCTQCLQHPWLQKDTKNMEAKKLSKDRMKKYMARRKWQKTGHAVRAIGRLSSMAMISGMSGRK    1760

ASGSSPTSPINADKVENEDAFLEEVAEEKPHVKPYFTKTILDMEVVEGSAARFDCKIEGYPDPEVMWYKDDQPVKESRHF    1840

QIDYDEEGNCSLTISEVCGDDDAKYTCKAVNSLGEATCTAELLVETMGKEGEGEGEGEEDEEEEEE                  1906
```

Fig. 2. Amino acid sequence of MLCK-210. The sequence represents the extended open reading frame (ORF) of the cDNA sequence found in EMBL database accession number X52876. Amino acid sequence numbering is shown on the right. The initiator methionine residues for MLCK-108 and KRP are indicated above the corresponding sites on the sequence.

2 4 were harvested for protein analysis. Extracts were made in SDS-PAGE sample buffer and processed for Western blot analysis as previously described [4], using the anti-210K antiserum at a 1:700 dilution and the anti-KRP antiserum at a 1:5000 dilution.

3. Results and discussion

As shown in Fig. 1, adult chicken tissue and cultured chicken cells contain at least three size classes of mRNA at 2.7 kb, 5.5 kb and 9.0 kb that hybridize with probes selective for KRP and MLCK mRNAs. The selective hybridization of the 9.0 kb mRNA (Fig. 1A) with probes based on the 5′ end of the DNA sequence (EMBL accession number X52876) demonstrates that the 9.0 kb mRNA has unique regions not found in the 5.5 kb and 2.7 kb size classes. The hybridization of both the 5.5 and 9 0 kb mRNAs (Fig. 1B) with probes based on the middle of the DNA sequence demonstrates that they share extended regions of sequence that are not found in the 2.7 kb size class of mRNA. Collinge et al. [2] showed previously that the 2.7 kb mRNA has approximately 109 nucleotides of unique sequence in its 5′-nontranslated region that allow selective detection of 2.7 kb mRNA. All three size classes of mRNA hybridize (Fig. 1C) with a probe based on the previously described [2] common 3′-nontranslated region of the mRNAs for MLCK (5.5 kb) and KRP (2.7 kb), indicating that all three RNA size classes utilize the comparatively large 3′ terminal exon that is common to the KRP and MLCK genes [2]. Overall, Northern blot results are consistent with the deposited DNA sequence (EMBL accession number X52876) and with the ORF that encodes a larger MLCK amino acid sequence extended at the amino terminus (Fig. 2). The amino acid sequence shown in Fig. 2 is that of the extended ORF. This extended ORF includes all of the amino acid sequences found in the ORFs previously reported by Shoemaker et al. [3] and Olson et al. [7] for chicken MLCK, plus additional unique amino terminal sequence.

A diagrammatic summary of the relationship among the multiple size classes of gene products is given in Fig. 3. The additional amino terminal sequence (see Fig. 2) found in the high molecular weight MLCK (MLCK-210) contains repeats of the sequence motifs described previously for the MLCK and KRP sequences [3], including a short segment (amino acids 74–119) that has similarity to a sequence segment in the KRP domain (amino acids 1790–1885).

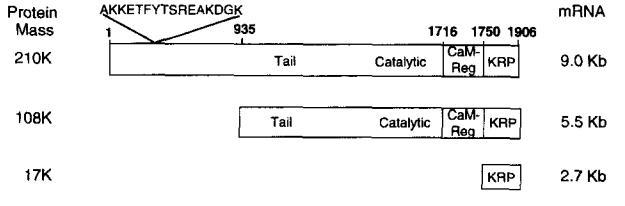

Fig. 3. Diagrammatic summary of the three size classes of gene products derived from the MLCK/KRP transcription unit. The mRNA size classes are indicated on the right, and the calculated mass of the protein products is indicated on the left. The segmental organization of the proteins into tail, catalytic, calmodulin-regulatory (CaM-Reg) and kinase-related protein (KRP) regions is as previously described [3]. The highlighted sequence above the MLCK-210 protein is the synthetic peptide sequence used to make selective anti-210K antibodies. The amino acid numbers delineate, respectively, the start of the MLCK-210 protein (residue 1), the MLCK-108 (residue 935 of MLCK-210), the CaM-Reg domain (residue 1716), the KRP domain (residue 1750), and the end of the protein (residue 1906).

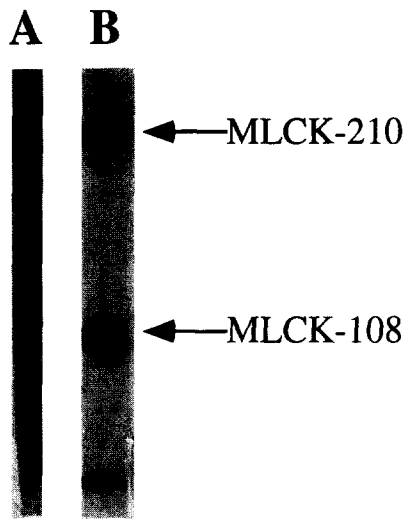

Fig. 4. Detection of MLCK-210 protein in cell extracts by Western blot analysis with site-directed antibodies. An extract from cultured adult chicken aorta-derived cells was subjected to SDS-gel electrophoresis on a 5% acrylamide/0.065% bisacrylamide gel, and Western blot analysis was performed using two different antibodies. Lane A: Selective detection of MLCK-210 with antiserum (1:700 dilution) generated against a synthetic peptide immunogen unique for the MLCK-210 protein (see section 2 and Fig 3). Lane B: Detection of both MLCK-108 and MLCK-210 with antisera made against purified KRP protein.

In order to obtain evidence that the ORF actually represented an expressed protein, a synthetic peptide antigen was made based on the unique region of the amino acid sequence (see Fig. 3) and antisera were produced in rabbits. A Western blot of a whole cell extract prepared from chicken aorta derived cells is shown in Fig. 4. A single protein band with an estimated polypeptide chain weight of 210,000 is the major immunoreactive species detected in the extract (Fig. 4A). The close similarity between computed mass and estimated polypeptide chain molecular weight based on SDS-PAGE for MLCK-210 is in contrast to the divergence between the computed mass of 107,534 and estimated chain weight of 135,000 for MLCK-108 [7]. It is not known if this is due to unusual SDS binding properties of MLCK-108 or if there are other protein isoforms. As further confirmation of the MLCK-210 ORF, antibodies were made against purified KRP and used in Western blot analyses of the cell extracts. As shown in Fig. 4B, MLCK-210 and MLCK-108 are both detected with this antiserum. In separate experiments (data not shown), the M_r = 210,000 protein was shown to be distinct from other cytoskeletal proteins in this size class, such as myosin and filamin, consistent with the fact that the amino acid sequence does not have regions of extended identity with either of these proteins.

In summary, the MLCK/KRP transcription region of genomic DNA encodes MLCK-210, MLCK-108 and KRP. In contrast to the detailed knowledge about the KRP gene and its mRNA and protein products, the genomic organization and mRNA characterizations for MLCK-108 and MLCK-210 remain to be completed. However, it is clear from the knowledge to date that the same amino acid sequence can be found as an apparent cytoskeletal targeting domain of a protein kinase sequence or as a distinct low molecular weight cytoskeletal protein that has no known enzyme activity. Although the function of MLCK-210 has not been demonstrated yet, logical possibil-

ities are suggested by an analysis of its sequence. First, all of the features previously shown [3] to encode a calmodulin-regulated MLCK activity are found in MLCK-210. Second, there are sequence motifs associated with subcellular targeting [3,4]. For example, MLCK-210 has nine type II motifs (IgG-C2-related), three of which are also found in MLCK-108. In addition, matrix analysis showed that a 96 residue segment in the unique region of MLCK-210 (residues 24–119) has a 40% identity with a corresponding segment in the myosin binding KRP domain (residues 1790–1885). Consistent with this possibility, our preliminary experiments indicate that MLCK-210 is, indeed, capable of interacting with smooth muscle myosin (data not shown). Therefore, a single transcriptionally active region in the vertebrate genome appears to encode and produce, by a mechanism that involves redundant use of DNA sequences, a series of proteins involved in the regulation of the myosin II molecular motor system.

Acknowledgements: This work was supported in part by NIH Grant GM 30861 and HHMI Award 75195-546901.

References

[1] Kilhoffer, M.-C., Lukas, T.J., Watterson, D.M. and Haiech, J. (1992) Biochim. Biophys. Acta 1160, 8–15.
[2] Collinge, M., Matrisian, P.E., Zimmer, W.E., Shattuck, R.L., Lukas, T.J., Van Eldik, L.J. and Watterson, D.M. (1992) Mol. Cell. Biol. 12, 2359–2371.
[3] Shoemaker, M.O., Lau, W., Shattuck, R.L., Kwiatkowski, A.P., Matrisian, P.E., Guerra-Santos, L., Wilson, E., Lukas, T.J., Van Eldik, L.J. and Watterson, D.M. (1990) J. Cell Biol. 111, 1107–1125.
[4] Shirinsky, V.P., Vorotnikov, A.V., Birukov, K.G., Nanaev, A.K., Collinge, M., Lukas, T.J., Sellers, J.R. and Watterson, D.M. (1993) J. Biol. Chem. 268, 16578–16583.
[5] Van Eldik, L.J., Fok, K.-F., Erickson, B.W. and Watterson, D.M. (1983) Proc. Natl. Acad. Sci. USA 80, 6775–6779.
[6] Birukov, K.G., Frid, M.G., Rogers, J.D., Shirinsky, V.P., Koteliansky, V.E., Campbell, J.H. and Campbell, G.R. (1993) Exp. Cell Res. 204, 46–53.
[7] Olson, N.J., Pearson, R.B., Needleman, D.S., Hurwitz, M.Y., Kemp, B.E. and Means, A.R. (1990) Proc. Natl. Acad. Sci. USA 87, 2284–2288.