Journal of Cellular Biochemistry 75:481–491 (1999)

Analysis of the Kinase-Related Protein Gene Found at Human Chromosome 3q21 in a Multi-Gene Cluster: Organization, Expression, Alternative Splicing, and Polymorphic Marker

D. Martin Watterson,[1,3]* James P. Schavocky,[1,3] Ling Guo,[1,3] Curtis Weiss,[1,3] Alexandre Chlenski,[1] Vladimir P. Shirinsky,[4] Linda J. Van Eldik,[2,3] and Jacques Haiech[5]

[1]Department of Molecular Pharmacology and Biological Chemistry, Northwestern University, Chicago, Illinois
[2]Department of Cell and Molecular Biology, Northwestern University, Chicago, Illinois
[3]Drug Discovery Program, Northwestern University, Chicago, Illinois
[4]Laboratory of Cell Motility, Institute of Experimental Cardiology, Russian Cardiology Research Center, Moscow, Russia
[5]Pharmacologie et Physico-Chimie des Interactions Cellulaires et Moléculaires, Faculté de Pharmacie, Université Louis Pasteur, Strasbourg, France

Abstract We report the amino acid sequence, genomic organization, tissue expression, and alternative splice patterns for the human kinase related protein (KRP) gene, as well as the discovery of a new CA repeat sequence polymorphic marker in an upstream intron of the myosin light chain kinase (MLCK) gene. The KRP/MLCK genetic locus is a prototype for a recently discovered paradigm in which an independently regulated gene for a non-enzymic protein is embedded within a larger gene for a signal transduction enzyme, and both classes of proteins are involved in the regulation of the same cellular structure. The MLCK/KRP gene cluster has been found only in higher vertebrates and is localized to human chromosome 3q21. The determination of the human KRP amino acid sequence through cDNA sequence analysis and its comparison to the exon/intron organization of the human KRP gene revealed an alternative splice pattern at the start of KRP exon 2, resulting in the insertion of a single glutamic acid in the middle of the protein. Examination of tissue distribution using Northern blot analysis revealed that the human expression pattern is more similar to the well-characterized chicken KRP gene expression pattern than to rodent or rabbit. Unexpected differences of the human gene from other species is the apparent expression of the human gene products in adult cardiac muscle, an observation that was pursued further by the production of a site-directed antiserum and immunohistochemistry analysis. The results reported here provide insight into the conserved and variable features of this late evolving genetic paradigm, raise new questions about the molecular aspects of cardiac muscle regulation, and provide tools needed for future clinical studies. The comparative analysis of the MLCK/KRP locus, combined with the recent discovery of a similar genomic relationship among other signal transduction proteins, suggest a diverse distribution of this theme among signal transduction systems in higher vertebrate genomes and indicate the utility of comparative genomics in revealing late evolving genetic paradigms. J. Cell. Biochem. 75:481–491, 1999.

Key words: genome; protein kinase; myosin; calmodulin; signal transduction; polymorphism; alternative splicing; smooth muscle; cardiac muscle; calcium

The kinase related protein (KRP) is a stabilizer and organizer of unphosphorylated vertebrate myosin II minifilaments that are characteristic of the unstimulated physiological state in vertebrate smooth muscle and nonmuscle tissues. KRP is encoded in higher vertebrate species by a gene embedded within the gene for myosin light chain kinase (MLCK), a calmodulin (CaM) regulated protein kinase that phosphorylates myosin II in response to intracellular Ca^{2+} signals in vertebrate tissues [for a recent review see Lukas et al., 1998]. Gallus gallus (chicken) is the only species for which a

Grant sponsor: NIH; Grant number: GM30861; Grant sponsor: Howard Hughes Medical Institute; Grant number: 75195.

*Correspondence to: D. Martin Watterson, Ward 8–196, 303 E. Chicago Avenue, Chicago, IL 60611. E-mail: m-watterson@nwu.edu

Received 11 March 1999; Accepted 30 April 1999

complete coding exon map is available for the MLCK/KRP genes and their respective products [Collinge et al., 1992; Watterson et al., 1995; Birukov et al., 1998] . This standard of comparison revealed that KRP and two isoforms of MLCK are encoded by differential use of 31 coding exons that have one variation in exon structure, the presence of exon 29 of the protein kinases within the first exon of KRP and the presence of the KRP gene promoter and transcription initiation site within the intron preceding exon 29 of the MLCK gene. The start of coding for KRP (encoded by a 2.7 Kb mRNA) is an internal methionine codon for MLCKs (two isoforms encoded by either a 5.5 or 9 Kb mRNA). This results in the same amino acid sequence being present in two entirely different classes of proteins. Although these two divergent classes of proteins have in common the KRP amino acid sequence and the myosin molecular motor system as a subcellular target, in vivo production of KRP mRNA and protein is regulated independently of that for the MLCKs [Collinge et al., 1992; Watterson et al., 1995; Birukov et al., 1998].

The gene-within-a-gene arrangement for the KRP/MLCK locus has only been found in higher vertebrate species, suggestive of a late evolving gene cluster. The recently completed yeast genome project has demonstrated that this early phylogenetic species lacks an MLCK [Hunter and Plowman, 1997]. Although there is an MLCK in the mold, *Dictyostelium discoidium*, it is not regulated by CaM and lacks a CaM binding motif [Tan and Spudich, 1991]. A CaM regulated MLCK has been found in the fruit fly, *Drosophila melanogaster*, but it lacks a KRP domain and no evidence was found for an embedded KRP gene or the presence of a KRP protein [Kojima et al., 1996; Tohtong et al., 1997]. The earliest phylogenetic species in which a KRP protein and an embedded KRP gene have been found is the chicken [Collinge et al., 1992; Birukov et al., 1998; Watterson et al., 1995]. In contrast to the rapidly increasing knowledge about this functionally linked gene cluster in phylogenetically earlier species, there is a lack of information about the MLCK/KRP locus in primates, except for a single report [Potier et al., 1995] on the localization of MLCK to 3qcen-21q. Therefore, we have used the recently available knowledge about the genomic organization of the chicken MLCK/KRP locus [Birukov et al., 1998] as a standard of comparison in the characterization of the human KRP gene organization and its relationship to the human MLCK gene.

In addition to reporting here the genomic organization of the KRP gene and the amino acid sequences of its expression products, we demonstrate that the human KRP gene, like the chicken gene, has a diverse pattern of expression. Some differences were seen between the pattern in humans vs. chicken and other vertebrate species, most notably the presence of comparatively high levels of KRP in adult human cardiac muscle. During the course of the studies reported here, we also found a CA repeat in an intron sequence approximately 11 Kb upstream of the human KRP gene, and demonstrated its potential use as a new human polymorphic marker. In addition, we found that there is an alternative use of a 5′splice acceptor site in KRP exon 2 that results in alternatively spliced forms of KRP and MLCK that differ by the presence or absence of a single glutamic acid residue. Altogether, the results demonstrate the highly conserved nature of the KRP amino acid sequence and exon organization, while raising the possibility of species differences in regulation for this late evolving gene. Sequence data from this article have been deposited with the GenBank/EMBL databases and can be found under accession numbers AF096766, AF096767, AF096768, AF096769, AF096770, AF096771, AF096772, AF096773, AF096774, and AF096775.

MATERIALS AND METHODS

Isolation and Characterization of Human KRP Gene and cDNA

Restriction enzyme digests, library screening, PCR, hybridizations, primer extensions, and DNA sequence analyses were done by standard methods, as previously described [Collinge et al., 1992; Watterson et al., 1995; Birukov et al., 1998]. To isolate KRP genomic clones, a human genomic DNA library in a P1 phage vector (pAD10SacBII; Genome Systems Inc.) was screened by PCR with two primers (5′-GATGTGTCCCAAGCTTT-3′ and 5′-TGCAGTCAAATCTAGCAG-3′) based on a human placenta cDNA sequence for the KRP domain of MLCK that we had previously isolated (accession number X90870). Two genomic clones, P1–209 and P1–454, were isolated. The P1–209 genomic clone was subjected to EcoRI digestion and the resulting fragments divided into eight

groups based on molecular weight. Twenty-five independent clones from each group were screened by hybridization with cDNA-based probes in order to locate clones containing exon sequences. Two subclones, E2–10F (6.8Kb) and E-14A (7.7Kb), were detected, isolated, and characterized. Human KRP cDNA clones were isolated by RT-PCR from human lung and placenta. The PCR primers used corresponded to nucleotides 706–727 in accession number AF096769 (includes KRP exon 1) and 799–820 in accession number AF096771 (includes KRP exon 3). Clones were screened by DNA sequence analysis. Primer extension analyses were done as previously described for chicken KRP [Collinge et al., 1992]. Briefly, commercially available (Clontech, Palo Alto, CA) poly(A)$^+$ RNAs (1.5 μg each) from adult human brain, kidney, lung, placenta, and heart tissue were used with four different primers corresponding to nucleotides 808–837, 1017–1046, 933–962, and 1137–1166 of the human KRP exon 1 sequence (accession number AF096769). A 30-base oligonucleotide primer (5′-TTCTGCATTGAGCGGGCTGGTTGGTGACCC-3′) that corresponds to a portion of the coding sequence of KRP exon 1 consistently produced a product with all of the mRNAs.

Microsatellite Isolation, Identification of CA Repeats, Heterozygosity Analysis, and Northern Blots

To isolate a microsatellite sequence close to the KRP gene, a sublibrary of the P1–209 clone was constructed from a HaeIII digest, and a subclone (P1-Hae-CA6) was identified by hybridization with a [γ-^{32}P]-ATP end-labeled $(CA)_9$ oligonucleotide. The sequence (accession number AF096772) of P1-Hae-CA6 contains the repetitive motif $(CA)_{19}(TA)_4$. The microsatellite was mapped to the KRP/MLCK gene locus by XL-PCR using a commercially available kit (Perkin-Elmer, Foster City, CA, Cat No. N808–0193). Primers used for PCR were located in exons and introns of both the MLCK and KRP gene. To investigate the polymorphic nature of the CA-repeat, two of the primers used in the mapping and based in the flanking MLCK intron sequences were used to amplify the $(CA)_{19}(TA)_4$ repeat from 23 unrelated Caucasian individuals from CEPH reference families (Centre d′Etude du Polymorphisme Humain, Paris, France). PCR was performed in a 50 ml reaction containing 0.1 μg of genomic DNA, 0.33 μM of each primer corresponding to nucleotides 299–317 and 437–461 of accession number AF096772, 1.88 pmol of the primer corresponding to nucleotides 299–317 that was end-labeled with [γ-^{32}P]-ATP and T4 polynucleotide kinase, 200 mM dNTPs, 0.5 mM $MgCl_2$, and 1.25 U of Ampli-Taq DNA polymerase. DNA was amplified by one cycle at 95°C for 5 min, then 10 cycles consisting of 30 s at 95°C, 1 min at 55°C, and 1 min at 72°C. For cycles 11–20, the extension time was 4.5 min and for cycles 21–30 the extension time was 8 min. PCR products were run on a 6% denaturing polyacrylamide gel, vacuum dried, and exposed to X-ray film. Sizes of PCR products were determined by migration next to a sequencing ladder. Northern blot analyses were done as previously described [Collinge et al., 1992; Watterson et al., 1995; Birukov et al., 1998] except commercially available (Clontech) multiple tissue mRNA blots were used and the KRP mRNA specific probe was made by PCR using primers corresponding to nucleotides 952–971 and 1029–1050 of KRP exon 1 (accession number AF096769).

Production of KRP Antibody and Immunohistochemical Localization

A peptide corresponding to residues 39–53 of human KRP (VAEEKPHVKPYFSKT) was synthesized by the solid-phase method as previously described [Van Eldik et al., 1983]. The peptide was coupled to keyhole limpet hemocyanin (KLH) using glutaraldehyde followed by reduction with $NaBH_4$, and then dialyzed against PBS before injection into rabbits for antibody production. Antisera from rabbit no. 6091 reacted in Western blots with KRP and with MLCK that contains a KRP domain. Immunohistochemistry on formalin-fixed, paraffin-embedded sections of fetal and adult human heart tissues (obtained from Dr. Bruce Quinn and the Northwestern University Medical School tissue bank) was performed as previously described for chicken KRP [Birukov et al., 1998], except using a 1:100 dilution of no. 6091 anti-human KRP IgG.

RESULTS

Genomic clones containing the entire human KRP gene and three of the upstream exons encoding MLCK sequences were isolated as described in Materials and Methods. A summary of the exon and intron organization for the human KRP gene and the contiguous MLCK exons is

shown in Figure 1A. The human KRP gene encompasses approximately 6.0 Kb of the genome. The DNA sequence was determined for each of the human KRP exons, the contiguous MLCK exons 26–28, the immediate flanking intron sequences for each exon, and the intron region between MLCK exons 26 and 27 that contains the CA repeat described below. The DNA sequences have been deposited and are available as successive accession numbers, AF096766 through AF096772 from GenBank. The DNA sequence data was coupled with PCR analysis of intron size to develop the physical map of the human KRP gene and contiguous MLCK exons 26–28.

The DNA sequences of the exon/intron boundaries for the human KRP and MLCK exons 26–28 are given in Table I. The coding exons have canonical 3′-donor and 5′-acceptor splice consensus sequences, except at the 5′end (exon 1) and 3′ end (exon 3), where UTR sequences are present. There is potential for alternative splicing at the 5′ end of exon 2. This possibility was directly addressed by DNA sequence analysis of cDNA clones, summarized in a following section of results. The relative location of exon/intron boundaries for the human KRP gene and its contiguous MLCK exons is conserved between chicken and human.

The DNA sequence identity between human and chicken KRP exons 1, 2, and 3 is 61, 74, and 61%, respectively. If comparison is restricted to the coding DNA sequences in KRP exons 1– 3, then the DNA sequence identity between chicken and human KRP exons increases to about 75% for each exon. In contrast to the exons, intron sizes and sequences are not as well conserved between chicken and human. The intron sizes between MLCK exons 26, 27, 28, and KRP exons 1, 2, and 3 are, respectively, 8.0, 2.5, 5.5, 1.0, and 3.5 Kb in the human gene

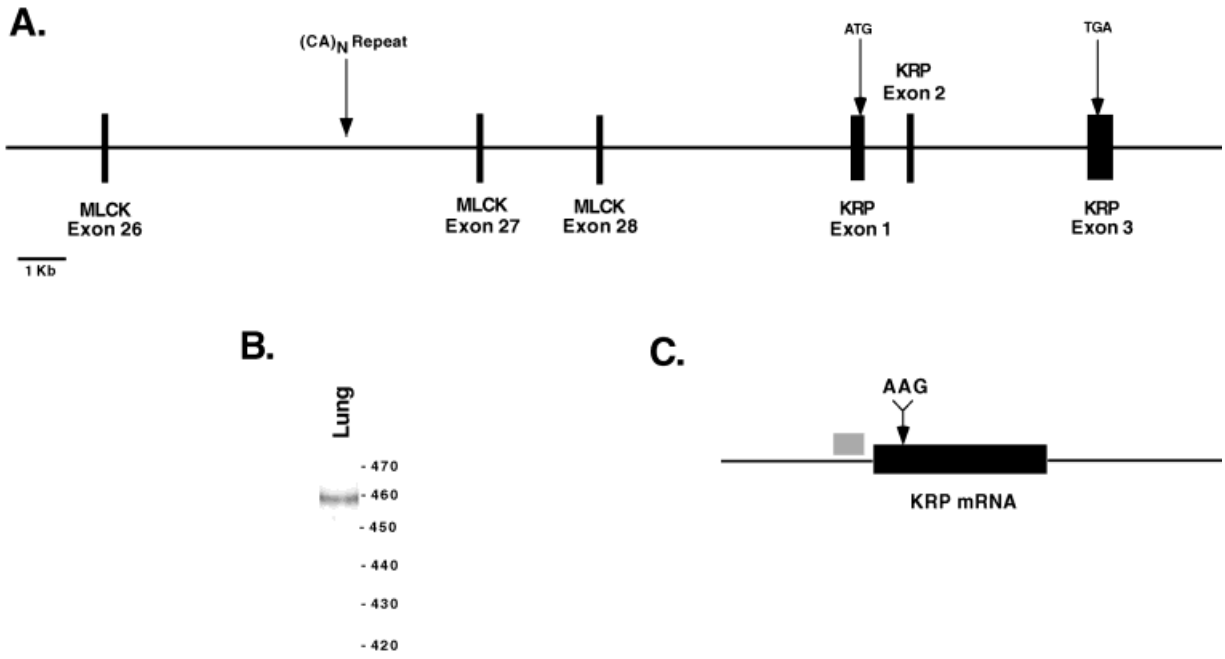

Fig. 1. Genomic organization of the human KRP gene. **A:** The organization of exons and introns for the KRP gene and the contiguous region of the MLCK gene are shown, as well as the relative location of a polymorphic marker ($(CA)_N$ repeat). Exons are presented as vertical lines. The distance from the CA repeat to KRP exon 1 is approximately 11 Kb. All DNA sequences are available from GenBank as a series of accession numbers from AF096766 through AF096775. **B:** Primer extension analysis of human lung mRNA. Primer extension was performed on 1.5 μg poly(A)+ RNA. The primer corresponds to nucleotides 432–461 of the human lung KRP cDNA (accession no. AF096773). The relative bp sizes indicated are based on the products of a sequencing reaction subjected to electrophoresis on the same 5% polyacrylamide gel. **C:** Diagrammatic summary of the relationship of KRP exons to the human KRP mRNA (see Fig. 2), and their relationship to the unique KRP probe used in Northern blot analyses (see Fig. 3). The coding region is represented by a solid black box. The arrow indicates the relative location in the mRNA of the alternative splice variation found for KRP exon 2. The gray shaded box shows the location of the KRP unique probe used in the Northern analysis in Figure 3.

TABLE I. Exon-Intron Boundaries and Exon Sizes for KRP/MLCK Gene[a]

Exon ID	Exon size (bp)	5′ Splice	3′ Splice
MLCK 26	124	ctggtttcagAGAATGCGGG	GCTACATCCTgtgagtcctg
MLCK 27	153	gtctctccagAGTCAGTGGC	AAGATATGAAgtaatgagtt
MLCK28	124	cctctgtcagAAACCGCCTG	GAAATGGCAGgtaatgagag
MLCK 29[b]	130	tattttccagAAAACGGGCA	GAATCTGAAGgtaaggagga
KRP 1	477	[c]	GAATCTGAAGgtaaggagga
KRP 2	129	cttctagaagATGTGTCCCA	AAGATTGAAGgtaagttgta
KRP 2A[d]	132	tgtcttctagAAGATGTGTC	AAGATTGAAGgtaagttgta
KRP 3	716	tgtgatgcagGATACCCAGA	[e]

[a]Exon-derived nucleotides are in capital letters, intron-derived nucleotides are in lowercase letters. The complete exon and adjacent intron flanking sequences have been deposited with GenBank (accession numbers AF096766 through AF096771).
[b]KRP exon 1 includes all of MLCK exon 29 plus additional 5′ sequences.
[c]No splice site because this is 5′ end.
[d]Exon 2A is the result of an alternative 5′ splice acceptor site.
[e]No splice site because this is 3′ end.

(Fig.1) but 4.0, 1.5, 1.1, 1.0, and 1.8, respectively, in the chicken gene [Birukov et al., 1998].

The 5′ end of the KRP mRNA (start of exon 1) was estimated by primer extension analysis of poly(A)$^+$ RNA from human brain, kidney, lung, placenta, and heart tissue as described in Materials and Methods. In all these tissues, a major extension product of approximately 455 nucleotides in length was detected. Figure 1B shows a representative result with mRNA from adult human lung tissue. This would place the 5′ end of the mRNA over 384 nucleotides upstream of the ATG codon (at or before nucleotide 706 of accession number AF096769), a longer length than the 151 nucleotides of the chicken KRP mRNA [Collinge et al., 1992]. The human 5′ flanking sequences (accession number AF096769) were analyzed for potential promoters and start of transcription by use of the neural network prediction algorithms [Reese et al., 1996]. The results suggested potential TATA boxes with start of transcription at nucleotide 490 or 764 of accession number AF096769, with the TATA box and start site between 724 and 774 giving a slightly better score (0.99 vs. 0.87). A variety of results (not shown), including DNA sequence analysis of RT/PCR products and primer extension data, are consistent with more than one TATA box and start of transcription being possible. Clearly, there was not a strong TATA box and initiation sequence motif as was found for the chicken KRP [Collinge et al., 1992], and there was not a single major class of products from primer extension and nuclease protection experiments as was found with the chicken KRP. Altogether, the results suggest more than one potential promoter and start site. This aspect of the human gene was not pursued further as part of this report. However, the results are consistent with differences among species in expression and regulation, as discussed in a following section.

In order to further verify the human KRP sequence predicted from the exons and to test whether the possible alternative splice site at the 5′ end of exon 2 is utilized, we isolated and characterized KRP cDNA clones from human lung and placenta as described in Materials and Methods. The three human KRP cDNA sequences are available from GenBank as accession numbers AF096773, AF096774, and AF096775. A composite cDNA sequence and translated amino acid sequence of the human KRP and a comparison to chicken KRP amino acid sequence are shown in Figure 2. The amino acid sequence identity between chicken and human KRP is 80%. The major differences are the presence of a short insertion of additional amino acids in the chicken protein near the carboxy terminus (see polyglutamic acid sequence and alignment gap in the human KRP in Fig. 2), and the presence of an extra tripeptide sequence in the human KRP just after the site of potential alternative splicing (indicated by a gap in the chicken sequence in Fig. 2). The cDNA sequences agree with the exon sequences obtained from genomic clone analyses.

The cDNAs from lung (accession number AF096773) and placenta (accession numbers AF096774 and AF096775) varied in the presence

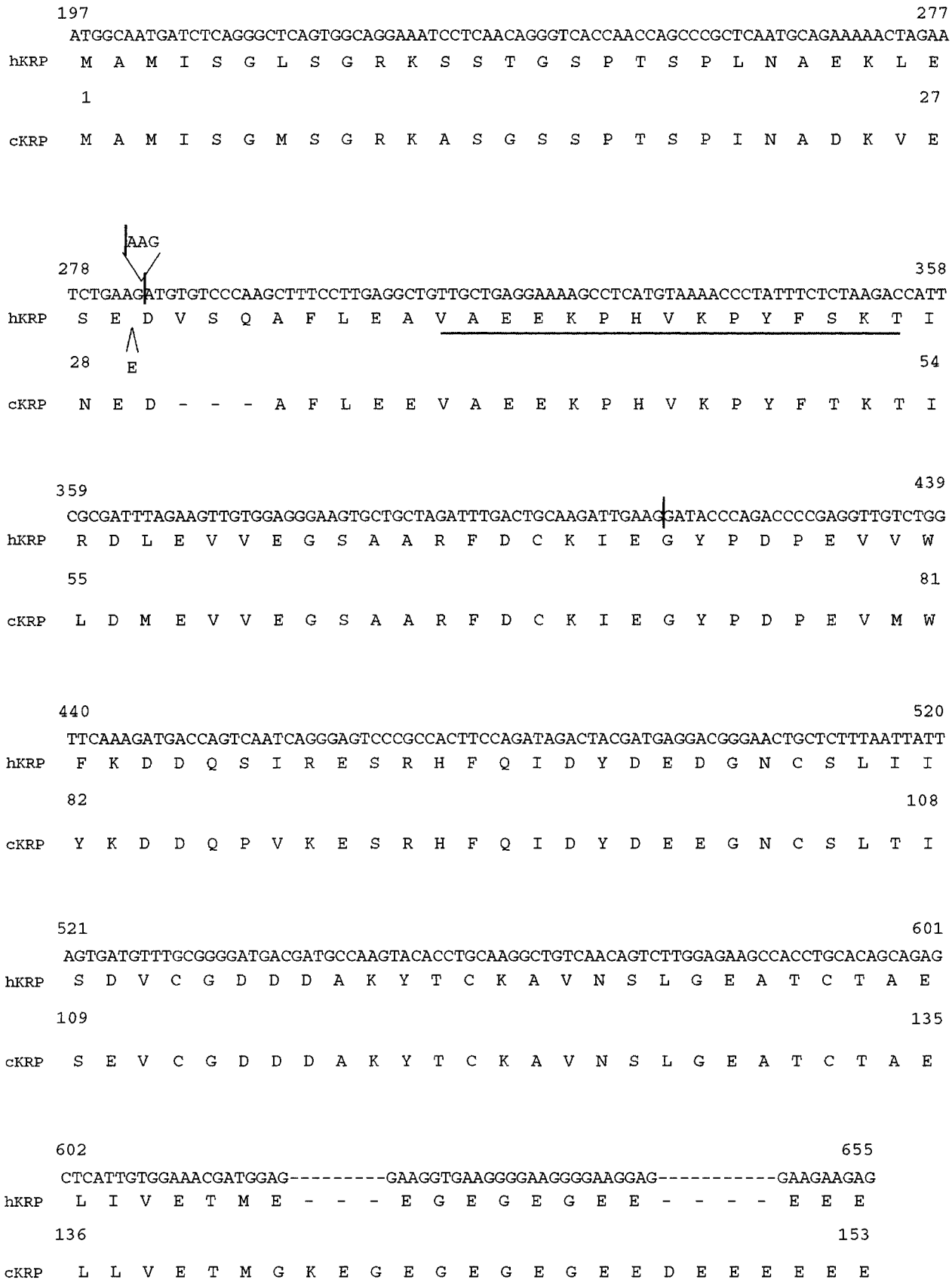

Fig. 2. Amino acid sequence comparison of human and chicken KRPs. The DNA sequence and translated amino acid sequence for human KRP (hKRP) are those determined as part of this investigation and deposited with GenBank (accession numbers AF096773, AF096774, and AF096775). The relative location of splice sites between exons (see Fig. 1) are shown by vertical lines. The relative positions of the trinucleotide insert encoding an extra glutamic acid that results from alternative splicing at human KRP exon 2 are indicated. The sequence of the synthetic peptide immunogen for the KRP antibody used in this study is delineated by the underlining. The chicken amino acid sequence [Collinge et al., 1992] is shown (cKRP) for comparison. Dashes indicate gaps inserted for the purpose of alignment.

(AF096773, AF096775) or absence (AF096774) of an AAG sequence, reflecting differential use of two potential splice acceptor sites on the 5′ end of KRP exon 2. The alternative use of these splice sites in exon 2 maintains the reading frame but leads to the presence or absence of a glutamic acid in the middle of the protein. The resultant amino acid sequence difference between . . .KLESEDVSQAF. . . and . . .KLESEEDVSQAF. . . would not be expected to alter drastically the secondary structure potential of this region of the protein. It is interesting to note that the various cDNA sequences available for human MLCK (accession numbers X85337, X90870, U48959) differ in terms of the presence or absence of this glutamic acid codon in the KRP domain of MLCK. This indicates that human MLCKs have variants in their carboxyl terminal KRP domains due to alternative splicing at this same exon.

We discovered a CA repeat sequence in an intron approximately 11 Kb upstream of the human KRP gene. Isolation of microsatellite sequences and mapping to the MLCK/KRP gene locus were done as described in Materials and Methods, and the relative location of this CA repeat sequence is indicated in Figure 1A. To investigate the polymorphic nature of the CA repeat, two of the primers used in the mapping, and based in the flanking MLCK intron sequences were used to amplify the $(CA)_{19}(TA)_4$ repeat from 23 unrelated individuals. Table II summarizes the results. The heterozygosity (0.85) indicates the potential utility of this microsatellite sequence as a new polymorphic marker for the human chromosome 3q21 region in future genomic analyses.

TABLE II. Sizes and Frequency of CA Repeat Alleles[a]

Size (bp)	Frequency
171	0.022
169	0.022
167	0.239
165	0.109
163	0.174
161	0.152
159	0.043
157	0.043
151	0.065
149	0.130

[a]Sizes and frequencies of 10 alleles detected by PCR of 23 unrelated Caucasian individuals from CEPH reference families (Centre d'Etude du Polymorphisme Humain, Paris). Using the allele frequencies shown, the calculated heterozygosity was 0.85. As expected for a dinucleotide repeat polymorphism, all allelic fragments differed by an even number of nucleotides.

The tissue distribution of KRP mRNA was analyzed by probing human tissue Northern blots with a KRP-specific probe based on the 5′ nontranslated region of the KRP mRNA that corresponds to MLCK intron sequence in the genome. Previous analysis of the chicken gene demonstrated [Collinge et al., 1992; Birukov et al., 1998] that the promoter for KRP and the start of KRP exon 1 are in a region that is intron for MLCK, providing unique sequences at the 5′-end of the 2.7 Kb KRP mRNA that allow selective detection in Northern blots and RT-PCR assays. The use of an equivalent human KRP mRNA probe in Northern blots resulted in detection of exclusively the 2.7 Kb mRNA (Fig. 3). KRP mRNA is detected in a variety of tissues, with differing levels of expression among tissues. In general, this expression pattern is similar to that found with the chicken KRP [Birukov et al., 1998], with only a few potential differences detected. A notable difference in KRP gene expression between human and chicken tissues is that observed in adult heart tissue between the two species. Specifically, embryonic chicken heart muscle expresses KRP, but the adult chicken heart contains KRP only in associated blood vessels but not in the cardiac muscle [Birukov et al., 1998]. Similar to the chicken, the rabbit and the mouse also do not have high levels of KRP in the adult heart [Smith et al., 1998]. In contrast to these earlier phylogenetic species, human KRP appears to be present at significant levels in adult human heart. It is possible that this might reflect some variation among patient samples in terms of pathological state or mRNA isolation procedures. In this regard, another lot of mRNA from a different patient and the same vendor was analyzed and the same results were obtained as shown in Figure 3. However, a multi-tissue blot from another vendor showed a lack of detectable KRP mRNA in adult heart (data not shown). Because of the variability in RNA integrity that is characteristic of human pathology samples, we decided to analyze the expression of human KRP in heart at the level of protein expression, as described in the following sections, because of the close correlation between mRNA and protein levels with KRP [Collinge et

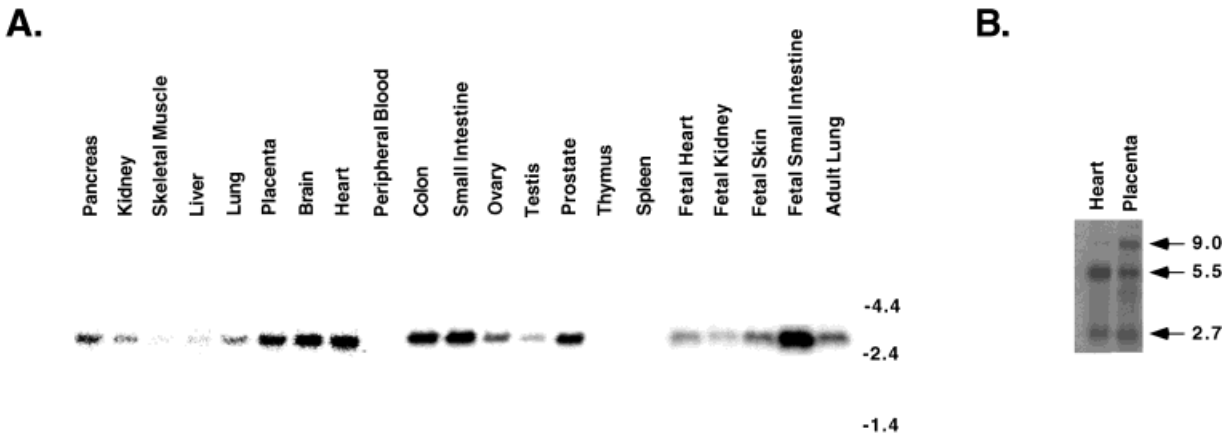

Fig. 3. Tissue distribution of KRP expression. **A:** Northern blot analysis of KRP mRNAs from human tissues using a unique KRP probe reveals a diverse tissue expression profile with varying levels of expression. The location of this KRP unique probe relative to the coding region of the KRP mRNA and common sequences in MLCK and KRP mRNAs is indicated in Figure 1. Commercially available (Clontech) human tissue blots containing 2 μg of poly(A)$^+$ RNA per lane were probed using a 98 bp [α-^{32}P] dCTP-labelled fragment specific to the 5′ UTR of KRP. **B:** Northern blot analysis of mRNAs from human placenta and human heart using a probe common to KRP (2.7 Kb) and MLCKs (5.5 and 9 Kb) reveals the presence of both MLCK108 mRNA (5.5 Kb) as well as KRP. The probe used in this blot includes the exon 2 region that encodes the amino acid sequence of the antigen used for the production of the antibodies used in Figure 4.

al., 1992; Watterson, et al., 1995; Birukov et al., 1998].

The unexpected finding that a protein such as KRP, which is considered to be a smooth muscle/nonmuscle regulatory protein, can be present in adult human cardiac tissue, at levels that cannot be explained by the smooth muscle content of blood vessels, prompted us to examine the localization of KRP and the smooth muscle/nonmuscle MLCK within the tissue by immunohistochemistry. We, therefore, engineered a site directed antibody against KRP, and analyzed heart sections from the human tissue bank at Northwestern University Medical School. The antisera were made as described in Materials and Methods with a synthetic peptide antigen corresponding to amino acids 39–53 of human KRP. This amino acid sequence encoded by exon 2 is conserved among phylogenetic species and is found in both human KRP splice variants. However, this amino acid sequence is unique in the GenBank database to KRP or the KRP domain of MLCK. The resultant antisera show reactivity in tissue Western blots with KRP and MLCKs that contain a KRP domain (data not shown), similar to previously reported results with other KRP antisera [Watterson et al., 1995]. However, an inherent limitation of studies of the MLCK/KRP system at the protein level is the presence of the same amino acid sequence in at least three distinct size classes of proteins: KRP, MLCK108, and MLCK210. In order to address what proteins we might be studying in cardiac tissue sections, we did Northern blot analysis of human heart tissue extracts using a labeled oligodeoxynucleotide probe that contains KRP coding sequences. Such probes allow the detection of the KRP mRNA at 2.7 Kb and the two MLCK mRNA size classes at 5.5 Kb (MLCK108) and 9 Kb (MLCK210). As shown in Figure 3B, the 5.5 Kb mRNA that encodes MLCK108 was detected as well as the expected KRP mRNA hybridization at 2.7 Kb. Because the amino acid sequence used in the production of the antisera is encoded by these mRNA sequences (KRP exon 2), the immunohistochemistry will be detecting KRP, MLCK108, or a combination of KRP and MLCK108.

Immunohistochemical analysis of adult and fetal human heart (Fig. 4) shows positive staining of both cardiac muscle as well as staining of the smooth muscle layer in nearby major blood vessels, consistent with the presence of KRP and MLCK containing a KRP domain. The result shown is representative of the staining seen with heart sections from two fetal and two adult samples. As a control for the immunohistochemistry staining, we examined the human cardiac tissue sections with an anti-MLCK antibody that recognizes MLCK but not KRP and found similar, although less intense staining (data not shown). The staining is selective within the tissue, as staining is not seen in

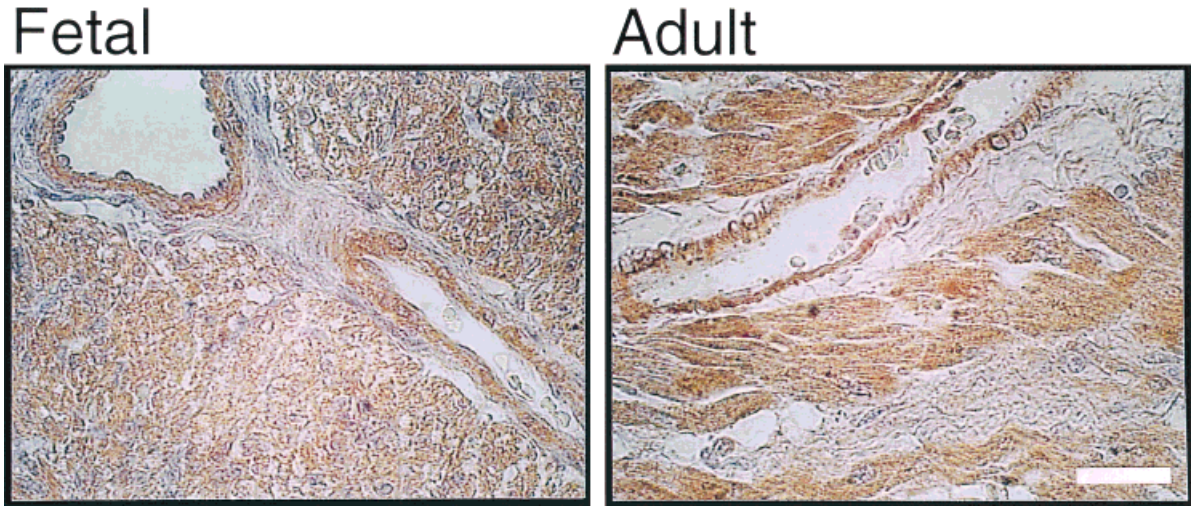

Fig. 4. Immunohistochemical localization of KRP in human heart. Sections from a 21-week fetus and a 51-year-old adult were stained with anti-KRP no. 6091 IgG (brown immunoreaction product) and counterstained with hematoxylin (blue). Note the KRP staining in both the walls of the blood vessels and in the cardiac muscle. Scale bar = 20 μm.

nonmuscle sections of the heart, and is consistent with the relatively high abundance of KRP and MLCK detected in whole tissue extracts. A more detailed localization within cardiac muscle was not done as part of this report. The salient feature of these results is that we found evidence consistent with authentic KRP and smooth muscle/nonmuscle MLCK in adult cardiac muscle.

DISCUSSION

The investigations summarized here and in previous reports [Collinge et al., 1992; Birukov et al., 1998] demonstrate the highly conserved nature of KRP's amino acid sequence, exon organization, and relationship to the MLCK gene once it became established in the genomes of higher vertebrates. The general pattern of KRP gene expression is also conserved, except in certain tissues. Because of its unexpected nature, the presence of KRP and MLCK mRNA and protein in adult human cardiac muscle was examined in more detail in order to confirm this indication of a difference between human gene expression and that for chicken, rabbit, and rodent. This suggests that the regulation of tissue selective expression of the KRP gene might differ in certain cases among higher vertebrate species, and is clearly more complex than previously assumed. The molecular basis of this differential gene expression and the potential biological role of KRP in adult human heart are intriguing subjects for future clinical investigations. However, defining the role of smooth muscle/nonmuscle MLCK/KRP in human cardiac muscle may be a nontrivial endeavor, especially if an animal or cell culture model for this expression is not found in future investigations, based on the continuing debate over the physiological role of skeletal muscle MLCK in a tissue with troponin-linked regulation. The results indicate that the chicken is closer to human in certain aspects of gene expression during development and among tissues. Clearly, the discovery of embedded genes in higher vertebrates that are under independent regulation and encode functionally linked proteins adds a new perspective to the interpretation of the genome projects, results from functional genetics, and attempts to selectively knock out the expression of physically clustered genes that affect the same cellular structure.

The initial report [Collinge et al., 1992] in which KRP was shown to be encoded by a small gene embedded within the larger MLCK gene was novel in regards to its organization and the production of the same amino acid sequence in the context of two entirely different proteins. The subsequent demonstration [Shirinsky et al., 1993] that KRP functions as a stabilizer of unphosphorylated myosin minifilaments indicated that KRP could play a complementary role to the catalytic activity of MLCK in the organization and regulation of the myosin II molecular motors in vertebrate cell function and homeostasis. This raised the interesting

hypothesis that vertebrates might have gene clusters that would encode functionally linked signal transduction proteins. More recently, identical paradigms have been found [Richardson and Parsons, 1996] in other signal transduction and structural protein systems. This suggests that the organization of such functionally linked gene clusters in vertebrate cells might be a more general genomic and regulatory paradigm than originally thought. The results presented here for the human KRP gene and neighboring MLCK exons are consistent with the phylogenetic conservation of this prototypical gene cluster once it appeared in evolution, but raise questions about conservation of gene promoter regulation.

The apparent species variability in the regulation of KRP gene expression is consistent with the lack of a clear pattern in the DNA sequence of the 5′-flanking sequences for exon 1 that have been characterized to date. Comparison of the 5′ flanking sequences for the human KRP gene determined as part of this study with GenBank entries [Altschul et al., 1997] revealed some similarities with other KRP or telokin entries (the protein from rabbit has been termed telokin, but is the same protein as KRP). For example, both rabbit telokin exon 2 (accession number M76234) and the entry for rabbit telokin 5′ region (accession number U40712) showed similarity to the human gene 5′ flanking sequences. However, comparison of the rabbit telokin exon 2 entry (accession number M76234) with that for rabbit telokin cDNA (accession number M76181) suggests that the exon 2 entry (accession number M76234) may actually represent exon 1 and 5′ flanking sequences for rabbit telokin. With this assumption, one can find some sequence similarity among the various KRP and telokin sequences that are upstream of the initiator ATG codons for each species. However, the evidence for potential multiple promoters and start sites seen with rabbit telokin [Smith et al., 1998] and human KRP (this report) raise the logical possibility that these similarities might be a reflection of some conservation in the non-coding regions of exon 1 among species. Regardless, these interesting correlations are clearly a subject of future study in the context of promoter analyses in homologous systems (e.g., primate promoters in primate cells).

Another important outcome of our studies on the human MLCK/KRP gene locus was the discovery of a CA repeat sequence in an intron approximately 11 Kb upstream of the KRP gene, and the demonstration of its potential use as a new human polymorphic marker. Although more detailed studies with a larger number of individuals need to be done in the future, the heterozygosity (0.85) demonstrated in this report indicates the potential utility of this marker in future clinical investigations. Because there are relatively few polymorphic markers in this region of human chromosome 3, the availability of a marker within the gene cluster should provide a useful reagent for human genome studies and in attempts to link pathology observations with the genetic loci for proteins that regulate cardiovascular form and function.

ACKNOWLEDGMENTS

This work was supported in part by NIH grant GM30861 (D.M.W.) and Howard Hughes Medical Institute grant 75195 (V.P.S.). D.M.W. is the John G. Searle Professor of Molecular Biology and Biochemistry. We thank Drs. Thomas Lukas and Bruce Quinn for their generous advice and assistance.

REFERENCES

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.

Birukov KG, Schavocky JP, ShirinskyVP, Chibalina MV, Van Eldik LJ, Watterson DM. 1998. The organization of the genetic locus for chicken myosin light chain kinase is complex: multiple proteins are encoded and exhibit differential expression and localization. J Cell Biochem 70:402–413.

Collinge M, Matrisian PE, Zimmer WE, Shattuck RL, Lukas TJ, Van Eldik LJ, Watterson DM. 1992. Structure and expression of a calcium-binding protein gene contained within a calmodulin-regulated protein kinase gene. Mol Cell Biol 12: 2359–2371.

Hunter T, Plowman GD. 1997. The protein kinases of budding yeast: six score and more. TIBS 22:18–22.

Kojima S, Mishima M, Mabuchi I, Hotta Y. 1996. A single *Drosophila melanogaster* myosin light chain kinase gene produces multiple isoforms whose activities are differently regulated. Genes Cells 1:855–871.

Lukas TJ, Mirzoeva S, and Watterson DM. 1998. Calmodulin regulated protein kinases. In: Van Eldik LJ, Watterson DM, editors. Calmodulin and signal transduction. Orlando, FL: Academic Press. p 65–167.

Potier MC, Chelot E, Pekarsky Y, Gardiner K, Rossier J, Turnell WG. 1995. The human myosin light chain kinase (MLCK) from hippocampus: cloning, sequencing, expression, and localization to 3qcen-q21. Genomics 29:562–570.

Reese MG, Harris NL, Eeckman FH. 1996. Large scale sequencing specific neural networks for promoter and splice site recognition. In: Hunter L, Klein T, editors. Biocomputing: Proceedings of the 1996 Pacific Symposium. Singapore: World Scientific Publishing Co.

Richardson A, Parsons JT. 1996. A mechanism for regulation of the adhesion-associated protein tyrosine kinase $pp125^{FAK}$. Nature 380:538–540.

Shirinsky VP, Vorotnikov AV, Birukov KG, Nanaev AK, Collinge M, Lukas TJ, Sellers JR, Watterson DM. 1993. A kinase related protein stabilizes unphosphorylated smooth muscle myosin minifilaments in the presence of ATP. J Biol Chem 268:16578–16583.

Smith AF, Bigsby RM, Word RA, Herring PB. 1998. A 310-bp promoter mediates smooth muscle cell-specific expression of telokin. Am J Physiol 274(Cell Physiol 43):C1188–C1195.

Tan JL, Spudich JA. 1991. Characterization and bacterial expression of the *Dictyostelium* myosin light chain kinase cDNA. J Biol Chem 266:16044–16049.

Tohtong R, Rodriguez D, Maughan D, Simcox A. 1997. Analysis of cDNAs encoding Drosophila melanogaster myosin light chain kinase. J Muscle Res Cell Motil 18:43–56.

Van Eldik LJ, Fok K-F, Erickson BW, Watterson DM. 1983. Engineering of site-directed antisera against vertebrate calmodulin by using synthetic peptide immunogens containing an immunoreactive site. Proc Natl Acad Sci USA 80:6775–6779.

Watterson DM, Collinge M, Lukas TJ, Van Eldik LJ, Birukov KG, Stepanova OV, Shirinsky VS. 1995. Multiple gene products are produced from a novel protein kinase transcription region. FEBS Lett 373:217–220.