Characterization and comparison of four serine- and arginine-rich (SR) protein kinases

Oliver NAYLER*[1], Stefan STAMM† and Axel ULLRICH*

*Max-Planck-Institute for Biochemistry, Am Klopferspitz 18A, D-82152 Martinsried, Germany, and †Max-Planck-Institute for Psychiatry, Am Klopferspitz 18A, D-82152 Martinsried, Germany

Phosphorylated serine- and arginine-rich (SR) proteins are components of the spliceosomal complex, and have been implicated in the control of alternative splicing. Kinases that regulate the phosphorylation and possibly the intranuclear distribution of SR proteins may therefore contribute to changes in choice of splice site. We have cloned three mouse cDNAs with high sequence identity to the family of LAMMER kinases (i.e. kinases carrying the conserved signature EHLAMMERILG in the catalytic domain). A comparison of their amino acid sequences revealed two related subfamilies with high evolutionary conservation. We have compared the expression patterns of these proteins in mouse tissues and transformed cell lines with that of a previously cloned family member (mCLK1/STY), and detected various transcripts for each gene. This underlines previous findings of alternative splicing of *mclk1/STY*. Our results suggest that the proportions of products for each gene are regulated independently. We further demonstrate that all variants encode autophosphorylating proteins that can phosphorylate several biochemically purified SR proteins *in vitro*, leading to hyperphosphorylation of at least one SR protein *in vivo*. The observed tissue distributions and substrate specificities suggest that these kinases may all be constituents of a network of regulatory mechanisms that enable SR proteins to control RNA splicing.

INTRODUCTION

The phosphorylation of proteins on serine/threonine and/or tyrosine residues is now well established as a principal mechanism in the control of many cellular functions, such as the cell cycle, differentiation and signal transduction. Recently, a family of highly phosphorylated proteins, the so-called serine- and arginine-rich (SR) proteins, have become a focus of interest due to their role in the regulation of splicing [1]. SR proteins [2] were originally characterized by a shared epitope that cross-reacts with the monoclonal antibody mAb104, at least one N-terminal RNA recognition motif and a basic C-terminal domain rich in serine and arginine residues, often arranged in tandem repeats [3]. Since that time, many more potential SR proteins or SR-like proteins have been identified using different monoclonal antibodies, so that the number of such proteins may well reach 80 or more [4]. There is evidence that the SR domain is involved in protein–protein interactions [5] as well as protein–RNA interactions [6], and may serve as a localization signal directing proteins to nuclear speckles [7].

Experiments suggesting that spliceosome assembly requires phosphorylation [8], whereas splicing catalysis is inhibited by phosphorylation [9], led to a search for kinases involved in the regulation of the splicing machinery, especially of SR proteins. Two enzymes (SRPK1 and topoisomerase I) were identified biochemically as kinases that phosphorylate SR proteins [10,11]. Recently, a CDC2-like kinase from mouse, mCLK1 (or STY), was shown by a yeast two-hybrid screen to interact with the SR proteins SRp20 and SRp30a (ASF/SF2) and with the RNA binding proteins heteronuclear ribonucleoprotein G and ribonucleoprotein S1. Furthermore, mCLK1 was shown to phosphorylate ASF/SF2 *in vitro* and to influence the localization of SR proteins *in vivo* [12]. mCLK1 or STY is a dual-specificity kinase that was originally isolated in an anti-phosphotyrosine antibody screen of mouse expression libraries [13,14], and human (hCLK1, hCLK2 and hCLK3), plant (AFC1, AFC2 and AFC3) and fly [DOA ('darkener of apricot')] homologues have since been found [15–18]. All previously identified family members carry the conserved signature EHLAMMERILG in the catalytic domain, and were hence classified as LAMMER kinases [18]. The N-terminal domain is rich in serine and arginine residues, whereas the catalytic domain contains amino acids that are conserved in CLKs [13].

Both mCLK1 and the *Drosophila* homologue DOA display extensive regulation at the pre-mRNA processing level, each having two alternatively spliced products coding for either the full-length catalytically active protein or a truncated protein lacking the catalytic domain [18,19]. Identical splice forms were also found for human homologues [16]. The ratio of these splice products appears to be regulated developmentally in *Drosophila* [18], and in a tissue- and cell type-specific manner in mammals [16,19]. In addition, the expression of several other, larger, transcripts was observed to be differentially regulated, and these were shown to represent partially spliced products [19]. The biological role of CLKs was unclear due to the absence of known endogenous substrates. However, this may have changed following the identification of SR proteins and RNA binding proteins as interacting molecules [12].

In the present study, we provide the sequences of three novel mouse homologues of mCLK1 and compare four CLKs from one species with regard to sequence, expression pattern, catalytic activity and subcellular localization. We also show that each mCLK can phosphorylate biochemically purified SR proteins *in vitro* and ASF/SF2 *in vivo*. We suggest that they may have

Abbreviations used: SR protein, serine- and arginine-rich protein; CLK, CDC2-like kinase (the prefixes m and h denote human and mouse respectively); GST, glutathione S-transferase; DOA, 'darkener of apricot'; RT-PCR, reverse transcription–PCR; LAMMER kinase, kinase carrying the conserved signature EHLAMMERILG in the catalytic domain; F-MEL cells, Friend murine erythroleukaemia cells; GFP, green fluorescent protein.

[1] To whom correspondence should be addressed.

different or partially overlapping substrate specificities, which, in addition to the regulation of their expression patterns, could provide a further level of control of alternative splicing events.

EXPERIMENTAL

cDNA cloning of *mclk1*, *mclk2*, *mclk3* and *mclk4*

Reverse transcription–PCR (RT-PCR) reactions were performed with 2 μg of total RNA prepared from confluent or differentiated (day 7) mouse C2C12 myoblasts [20] using degenerate oligonucleotide primers as described by Ciossek et al. [21]. Briefly, 2 μg of RNA was reverse-transcribed in the presence of 1 μM degenerate antisense primer, 250 μM of each nucleotide and 75 units of Stratascript reverse transcriptase (Stratagene) in a total volume of 20 μl for 30 min at 42 °C. Portions of 2 μl were used in a PCR reaction with degenerate sense and antisense oligonucleotides (1 μM each), 25 μM of each nucleotide and 2.5 units of *Taq* polymerase (Boehringer). A total of 30 cycles were performed at 94 °C for 1 min, 50 °C for 1 min and 72 °C for 1 min. Fragments of approx. 250 bp were gel-purified, cloned in Bluescript (Stratagene) and sequenced.

The genes *mclk2*, *mclk3* and *mclk4* were cloned from a mouse embryo 11.5 days *post coitum* λZAP cDNA library [21] using the isolated PCR fragment as a probe according to manufacturer's instructions (final wash in 0.5 × SSC/0.1 % SDS at 42 °C) (Stratagene). *mclk1* was cloned by RT-PCR from 1 μg of brain poly(A)$^+$ RNA using the specific primers *clk*1s–*Bam* (CGGGATCCCTTCGCCTTGCAAGCTTTGTC) and *clk1*as–*Eco*RI (CGGAATTCCTAGACTGATACAGTCTGTAAG) (where s and as denote sense and antisense respectively), and Pwo polymerase (Boehringer).

DNA sequencing was performed by the dideoxynucleotide chain-termination method [22] using Sequenase, reagents and protocols supplied by United States Biochemical Corp. Comparisons of the deduced protein sequences were carried out using MacDNASIS PRO software (Hitachi). Amino acid alignments were carried out with a Waterman algorithm. The phylogenetic tree was calculated using the Higgins–Sharp algorithm (CLUSTAL4).

Fusion protein constructs

Full-length *mclk1*, *mclk2*, *mclk3* and *mclk4* cDNAs were first subcloned by PCR in pGEX vectors (Pharmacia), creating in-frame glutathione S-transferase (GST) fusion constructs using the following primers for PCR: *clk1*s–*Bam* (as above); *clk1*as–*Not*I,TATAGCGGCCGCTAGACTGATACAGTCTGT; *clk2*s–*Sma*I, TCCCCCGGGATGCCCCATCCCCGAAGGTACCA; *clk2*as–*Not*I, TATAGCGGCCGCTCACCGACTGATATCCCGACTGGAGTC; *clk3*s–*Sma*I, TCCCCCGGGGAGACGATGCATCACTGTAAG; *clk3*as–*Not*I, TATAGCGGCCGCGCTGGCCTGCACCTGTCATCTGCTGGG; *clk4*s–*Eco*RI, CGGAATTCATGCGGCATTCCAAACGAACTC, *clk4*as–*Not*I, TATAGCGGCCGCCCTGACTCCCACTCATTTCCTTTTTAA. The cDNAs encoding the fusion constructs were then recloned in pcDNA3 (Invitrogen) by PCR using the following GST upstream primers: GST–*Eco*RI, CGGAATTCCGCCACCATGGCCCCTATACTAGGTTAT (for *clk1*); GST–*Hin*dIII,GCCAAGCTTGCCACCATGGCCCCTATACTAGGTTAT (for *clk2*, *clk3* and *clk4*).

To clone the 6 × His-tagged constructs, we produced a Bluescript vector containing an in-frame 6 × His stretch six amino acids downstream of the initiating ATG of the *lacZ* gene. The *clk* genes were then cloned in-frame into this vector by PCR, using the following primers: *clk1–Sac*II, TCCCCGCGGCAATGAGACATTCAAAGAGAACT; *clk2–Sac*I, GCGGAGCTCCATGCCCCATCCCCGAAGGTAC; *clk3–Sac*II, TCCCCGCGGCAATGCATCACTGTAAGCGATAC; *clk4–Sac*II, TCCCCGCGGCAATGCGGCATTCCAAACGAACT; and the antisense *Not*I primers described above. The 6 × His-tagged fusion constructs were cloned in pcDNA3 (Invitrogen) by PCR using the following primers: gal–*Bam*, GCGGATCCGCCACCATGGCCGTGATTACGCCA (for *clk1* and *clk3*); gal–*Hin*dIII, GCCAAGCTTGCCACCATGGCCGTGATTACGCCA (for *clk2* and *clk4*).

All PCR reactions were performed using Pwo polymerase (Boehringer). The integrity of the clones was checked by sequencing and by a coupled transcription–translation assay using T7 RNA polymerase and rabbit reticulocyte lysate, according to the manufacturer's protocol (Promega).

Preparation of mCLK1–mCLK4 mutants

mCLK1–CLK4 mutants containing a lysine (K) to arginine (R) substitution at positions 190 (mCLK1), 192 (mCLK2), 186 (mCLK3) and 189 (mCLK4) were generated using a site-directed mutagenesis protocol [23]. Oligonucleotide primers were as follows: mCLK1-K190R, GTAGCAGTAAGAATAGTTAAA; mCLK2-K192R, GTTGCCCTGAGGATCATTAAGAAT; mCLK3-K186R, GTTGCCCTGAGGATCATCCGGAAT; mCLK4-K189R, TACAATTCTCACTGCTACATGTAAGC CATC.

RNA extraction and Northern analysis

RNA was extracted from frozen adult mice tissues or tissue culture cells [24]. A 10 μg portion of total RNA was then electrophoresed in 1.2 % (w/v) agarose/formaldehyde gels and transferred to Hybond N membranes (Amersham). Hybridization was performed overnight in 50 % formamide, 5 × SSC (750 mM NaCl, 75 mM sodium citrate), 5 × Denhardt's (0.1 % Ficoll 400, 0.1 % polyvinylpyrrolidone, 0.1 % BSA), 0.2 % SDS and 100 μg/ml salmon sperm DNA. For labelling, (1–3) × 10^6 c.p.m./ml ^{32}P-labelled random-primed DNA probe (Amersham Megaprime kit) was used, followed by washes at 0.2 × SSC/0.1 % SDS at 42 °C. Blots were incubated with Hyperfilm-MP (Amersham) at −80 °C for 2 weeks. Membranes were stripped for re-use with boiling 0.1 % SDS/water.

In vitro kinase reaction

[^{35}S]Methionine-labelled fusion proteins GST–mCLK1 to GST–mCLK4 were produced in a 50 μl coupled *in vitro* transcription–translation reaction using the manufacturer's instructions (Promega).

Portions of 2 μl of each lysate were checked and quantified for the amount of protein produced by SDS/PAGE and autoradiography. Equal amounts (usually 20–30 μl of lysate) were added to 500 μl of PBS containing 1 mM PMSF, 10 μg/ml aprotinin and 30 μl of GSH–Sepharose beads (Pharmacia) and incubated on a rotating wheel for 2 h at 4 °C. This resulted in quantitative binding of the fusion proteins. The beads were then washed three times in 500 μl of PBS and once in 500 μl of kinase assay buffer (20 mM Hepes, 10 mM $MgCl_2$, 1 mM dithiothreitol, 200 μM sodium orthovanadate, 1 mM EGTA, pH 7.5). The assay was carried out for 30 min at room temperature in 30 μl of kinase assay buffer with 20 μM ATP, 4 μCi of [γ-^{32}P]ATP (10 mCi/ml; Amersham), 1 μM Microcystin LR (Sigma) and ~ 2.5 μg of dephosphorylated SR proteins. The reaction was stopped by adding 30 μl of 2 × SDS sample buffer. The samples were boiled for 5 min, and then 15 μl portions were loaded on an

SDS/15 %-polyacrylamide gel. Following electrophoresis, the gels were stained with Coomassie Blue, dried and exposed to Hyperfilm-MP (Amersham) for 24 h. The [^{35}S]methionine signal was suppressed with 3M Whatman paper placed between the film and the gel.

SR protein purification and dephosphorylation

SR proteins were purified from 5×10^9 Friend murine erythroleukaemia (F-MEL) cells according to the protocol described previously [3], and resuspended in buffer D [25]. The identity of the proteins was checked by Western blot using mAb104 (results not shown). Portions of 30 μl of SR proteins (~ 0.5 μg/μl) were incubated on ice for 10 min in 0.7 mM $MnCl_2$ and 5 m-units of protein phosphatase 1γ catalytic subunit (Boehringer), followed by 60 min at 30 °C [8]. Portions of 5 μl of dephosphorylated SR proteins were used per assay.

To create an ASF/SF2-enhanced green fluorescent protein (GFP) fusion protein, ASF/SF2 was cloned into pC2-EGFP (Clontech) using *Eco*RI and *Bam*HI sites.

Cell culture, transient expression and lysis procedure

Human 293 cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10 % (v/v) foetal calf serum. A total of 3×10^5 cells were seeded per 6 cm dish and were transfected 24 h later with 0.25–1 μg of DNA (co-transfection with 0.5 μg of each plasmid) using the calcium precipitation method [26]. After 16 h, the medium was changed and the cells were incubated for another 6–48 h (with or without 50 μM sodium orthovanadate) before lysis. Cells were lysed on ice for 30 min in 200 μl of buffer containing 50 mM Hepes, pH 7.5, 150 mM NaCl, 1 % Triton X-100, 10 % glycerol, 1 mM EDTA, 20 mM sodium pyrophosphate, 2 mM sodium orthovanadate, 100 mM NaF, 5 mM β-glycerolphosphate, 1 mM PMSF and 1 μg/ml aprotinin (HNTG buffer). The cell lysates were centrifuged at 13000 ***g*** for 10 min at 4 °C; then an equal volume of 2 × SDS sample buffer was added to the supernatant and 400 μl of 1 × SDS sample buffer was added to the pellet. Samples were boiled for 5 min and then vortexed vigorously, and 20 μl portions were loaded on SDS/10 %-polyacrylamide gels.

Western blotting and antibodies

Following SDS/PAGE, the proteins were transferred to nitrocellulose membranes (Schleicher & Schuell) and immunoblotted. Specific polyclonal antibodies were raised against the C-terminal 17 amino acids of mCLK1–mCLK4 fused to keyhole-limpet haemocyanin. Anti-phosphotyrosine antibody 4G10 was from Santa Cruz. For immunoblot detection, the ECL system from Amersham was utilized.

Immunofluorescence

NIH 3T3 cells were grown on glass coverslips and transfected with 1 μg of each of constructs *mCLK1–mCLK4*. The cells were fixed 24 h later in 2 % paraformaldehyde for 15 min on ice, permeabilized, blocked with PBS/0.1 % Triton X-100/1 % BSA (Sigma) for 1 h and washed three times with PBS. Incubation with the corresponding antibodies against mCLK1–mCLK4 (diluted 1:250 in PBS) was performed overnight. After three washes with PBS, the cells were incubated for 1 h with a Texas Red-linked secondary antibody (Amersham; diluted 1:1000 in PBS) and 0.5 μM TO-PRO iodine (Molecular Probes). After washing three times with PBS, the cells were mounted using Gelmount (Biomeda) and analysed by confocal laser scanning microscopy (Leica).

RESULTS

Cloning and analysis of three novel mouse *clk* cDNAs

The catalytic domains of protein kinases contain highly conserved regions, which have been used successfully to PCR-amplify and clone novel family members from a variety of species and tissues. In our experiments we used the signature sequence HRDLAAR in catalytic subdomain VI and D(V/M)WS(Y/F)G in subdomain IX to create degenerate oligonucleotides [21]. These primers were then used to search for hitherto unknown protein kinases involved in muscle cell differentiation by using RT-PCR of total RNA isolated from various *in vitro* differentiated stages of the mouse myoblast cell line C2C12 [20]. Of the approx. 300 fragments that were sequenced, one was novel. It was derived from a member of the LAMMER family of dual-specificity kinases [18] (also known as CLK [13] or STY [14]), and shared high identity with a part of the human cDNA *hclk2*. To obtain the full-length clone and to search for other, closely related, sequences, we screened a mouse 11.5 p.c. embryonic library at low stringency using the original 250 bp PCR fragment as a probe. We isolated three highly related full-length cDNA sequences defining different members of the CLK family. Sequence analysis revealed that two clones were mouse homologues of the human *hclk2* and *hclk3* genes [16], and hence were designated *mclk2* and *mclk3* respectively. The remaining clone appeared to be a novel *mclk* gene, which we named *mclk4*. The predicted molecular masses of the encoded proteins are 59.9 kDa (mCLK2), 58.5 kDa (mCLK3) and 57.2 kDa (mCLK4).

As illustrated in Figure 1(A), all of the proteins share essential features identifying them as LAMMER kinases [18]. They contain a nuclear localization signal [27] and an unusually basic N-terminal region composed of many serine and arginine residues, which probably contains a signal sequence to localize the protein to nuclear speckles [7,12]. The N-terminal domain is the most divergent part, suggesting that this area could contain information specific to each protein. The catalytic domain is highly conserved among all family members, with only few amino acid changes. Furthermore, all amino acids known to define the CLK subfamily are present in all four proteins [13]. The sequences of several CLKs from species as distantly related as *Drosophila* [18] and *Arabidopsis* [17] have been reported, underlining the evolutionary conservation of these proteins. We therefore compared the full-length amino acid sequences of several CLK proteins described in the literature, and prepared a phylogenetic tree (Figure 1B). It is evident that the *Arabidopsis* CLKs, as well as the *Drosophila* DOA kinase, are the most distant members of this family, whereas the mouse and human CLKs share a high degree of identity. hCLK1 and mCLK1 are 87 % identical, whereas hCLK2 and mCLK2 are 95 % identical. The highest degree of identity is seen between hCLK3 and mCLK3 (98 %). Interestingly, the mammalian kinases are divided into two subfamilies, with mCLK1/hCLK1 and mCLK4 being more closely related to each other (69 % identity) than to either mCLK2/hCLK2 or mCLK3/hCLK3, with whom they have 43 % identity. On the other hand, mCLK2/hCLK2 and hCLK3/mCLK3 are 51 % identical with each other.

Northern blot analysis

We analysed and compared the expression patterns of all *clk* genes (including the previously cloned *mclk1*) by Northern blot hybridization of total RNA from selected mouse tissues, as well

A

```
MCLK1   MRHSKRTYC-------PDWDERDWDYGTWRSSSSHKRKKRSHSSAREQKR      43
MCLK2   .P.PR.YHSSERGSRGSYHEHYQSRKHKRRR.R.WSSSSDRTRRR.REDS      50
MCLK3   .H.C..YRSPEPDPYLTYRWK.RRS.SREHEGRLRYPSR.EPPPR.S---      47
MCLK4   .......H.-------....S.ESWGHESY.G-...R.....TQ.NRH        42

MCLK1   CRYDHSKTTDSYYLESRSINEKAYHSRRYVDEY--RNDYMGYEPGHPYGE      91
MCLK2   YHVRSRSSY.DHSSDR.LY-------D.RYCGSYR....SRDRGEAY.DT     93
MCLK3   ---RSRSHDRIP.QRRYREHRDSDTY.CEERSPSFGE.CYGSSRSRHRRR     94
MCLK4   .KPH.QFKDSDCHYLEARCLNERDYRD.RYIDEY-....CEGYVPRH.HR     91

MCLK1   PGSRYQMHS-SKSSGRSGRSSYKSKHRSRHHTSQHHSHGKSHRRKRSRSV     140
MCLK2   DFRQSYEYHREN..Y..Q...RRKHR.R.RRSRTFSRSSSHSS.RAK-..     142
MCLK3   SRE.APYRTRKHAHHCHK.RTRSCSSASSRSQQSSKRSSR--------..     136
MCLK4   DVESTYRIHC....V..R...P.R.RNRPCASH.S..--..........I     139

MCLK1   EDDEEGHLICQSGDVLSARYEIVDTLGEGAFGKVVECIDHKVGGRRVAVK     190
MCLK2   ...A.....YHV..W.QE.....S.....TS.R..Q.V..RR..T...L.     192
MCLK3   ...K....V.RI.SW.QE.....GN....T.......L..ARGKSQ..L.     186
MCLK4   ..............R.....................GMD.LH....        189

MCLK1   IVKNVDRYCEAAQSEIQVLEHLNTTDPHSTFRCVQMLEWFEHRGHICIVF     240
MCLK2   .I...EK.K...RL..N...KI.EK..DNKNL....FD..DYH..M..S.     242
MCLK3   .IR..GK.R...RL..N..KKIKEK.KENK.L..L.SD..NFH..M..A.     236
MCLK4   .....GR.R...R..........S...N.V..........D.H..V....     239

MCLK1   ELLGLSTYDFIKENSFLPFRMDHIRKMAYQICKSVNFLHSNKLTHTDLKP     290
MCLK2   .......F..L.D.NY..YPIHQV.H..F.L.QA.K...D..........     292
MCLK3   ....KN.FE.L...N.Q.YPLP.V.H....L.HALR...E.Q........     286
MCLK4   ..................QI....Q......Q.I....H..........     289

MCLK1   ENILFVKSDYTEAYNPKMKRDERTIVNPDIKVVDFGSATYDDEHHSTLVS     340
MCLK2   ......N...ELT..LEK.....SVKSTAVR........F.H.....I..     342
MCLK3   ......N.EFETL..EHKSCE.KSVK.TSIR.A......F.H...T.I.A     336
MCLK4   ..........VVK..S........LK.T.......................     339
                                                      ****
MCLK1   TRHYRAPEVILALGWSQPCDVWSIGCILIEYYLGFTVFPTHDSREHLAMM     390
MCLK2   ...........E..............IF...V...L.Q...N.......     392
MCLK3   .....P.....E...A............F...R...L.Q..EN....V..     386
MCLK4   .....................................Q....K......     389
        **
MCLK1   ERILGPLPKHMIQKTRKRRYFHHDRLDWDEHSSAGRYVSRRCKPLKEFML     440
MCLK2   ......V.SR..R....QK..YRG......NT......REN....RRYLT     442
MCLK3   .K....I.S...HR...QK..YKGG.V...N..D....KEN.....SY..     436
MCLK4   ......I.A.........K....NQ.............R...........     439

MCLK1   SQDAEHELLFDLIGKMLEYDPAKRITLKEALKHPFFYPLKKHT          483
MCLK2   .EAED.HQ.....ENM...E....L..G...Q....AC.RTEPPNTKLWD     492
MCLK3   QDSL..VQ....MRRM..F...Q....A...L....AG.TPEERSFHSSR     486
MCLK4   CHDE...K....VRRM......R....D...Q....DL..RK            489

MCLK1
MCLK2   SSRDISR                                              499
MCLK3   NPSR                                                 496
MCLK4
```

B

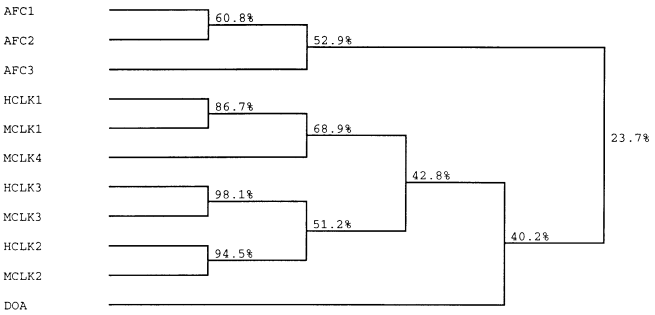

Figure 1 Amino acid sequences of mCLK2, mCLK3 and mCLK4 in comparison with that of mCLK1

(**A**) Residues that differ from the mCLK1 sequence are shown. Dots indicate identical amino acids, and hyphens indicate gaps introduced for optimal alignment. The predicted nuclear localization signals [27] are underlined. Invariant residues characteristic of CLKs are printed bold. The catalytic domain is indicated by arrows, and the LAMMER signature is indicated by asterisks. (**B**) Phylogenetic tree of mCLK1–mCLK4, hCLK1–hCLK3, *Arabidopsis* AFC1–AFC3 and *Drosophila* DOA proteins calculated using the Higgins–Sharp (CLUSTAL4) algorithm. The sequence identity between the proteins is indicated (%).

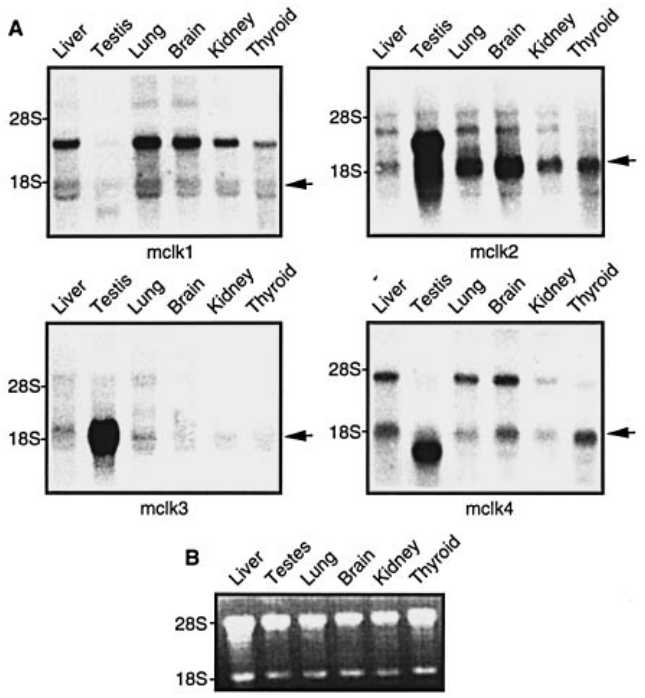

Figure 2 Northern blot analysis of *mclk1*, *mclk2*, mclk3 and *mclk4* expression in adult mouse tissues

(**A**) Portions of 10 μg of total RNA from the indicated tissues were resolved on a 1.2% (w/v) agarose/formaldehyde gel, blotted on Hybond N and hybridized to full-length cDNA probes. The positions of 18 S and 28 S rRNA are indicated. The mRNA ($\sim$ 1.8 kb) reflecting the length of the cDNA clones is marked by an arrow. (**B**) Ethidium bromide staining of blotted 18 S and 28 S rRNA indicates equal loading of RNA.

as from different mouse tumour cell lines. Figure 2 illustrates that, for each *clk* gene, we detected several mRNAs in all tissues investigated, although the expression patterns were different for each gene. Based on our cDNA clones, the expected size of the corresponding mRNA would be $\sim$ 1.8 kb for all *clk* genes, and such a band was detected in all tissues and cell lines examined, albeit at different levels. In testes, there were major differences between the four genes, with very low *mclk1* expression levels when compared with *mclk2*, *mclk3* and *mclk4*. However, whereas almost all of the *mclk3* signal represented an mRNA expected to encode the catalytically active form, *mclk4* was expressed predominantly as a smaller mRNA. We presume this to be an alternatively spliced form that encodes a catalytically inactive, truncated protein analogous to those described for *mclk1*, *hclk2* and *hclk3* [16,19]. *mclk2* is also highly expressed in this tissue, but as a larger transcript. Similar large transcripts, which did not correspond to the expected mRNA size, were detected for all *clk* genes. We suspect them to represent unspliced or partially spliced messages, by analogy with *hclk1* [19], as we found several partially spliced cDNA clones in our library screens (results not shown). However, we cannot completely rule out the possibility that these larger transcripts encode alternatively spliced mRNAs, although no such molecules appeared in our screens. The proportions of these larger RNA species, when compared with the coding mRNA, varied among the *clk* genes.

Because it has been reported [13] that *mclk1* is overexpressed in certain cancer cell lines, we wished to extend these studies to all four genes (Figure 3). Although we were able to detect messages in all cell lines tested, albeit sometimes in very low quantities, we found significant differences in expression levels between the cell lines for each individual gene. However, we could not detect an overall increase in *clk* mRNA in transformed cells, even though in some cell lines we clearly detected higher

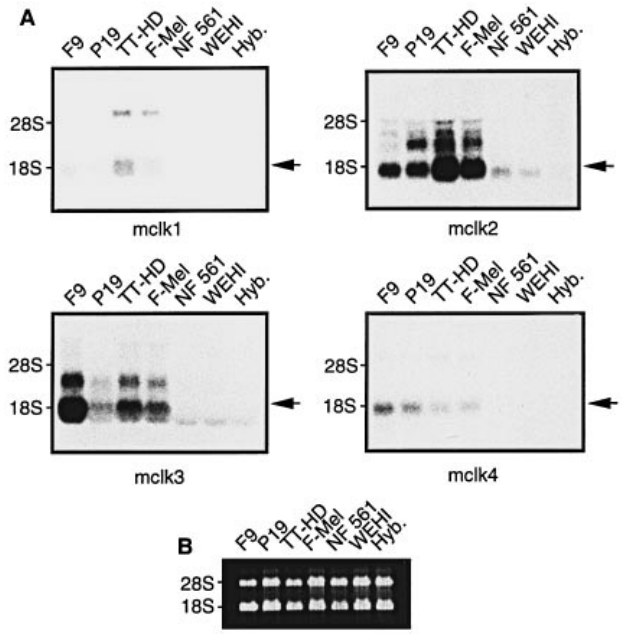

Figure 3 Northern blot analysis of *mclk1*, *mclk2*, *mclk3* and *mclk4* expression in different mouse cell lines

(**A**) Portions of 10 μg of total RNA were isolated from various cell lines and resolved on a 1.2% (w/v) agarose/formaldehyde gel, blotted on Hybond N and hybridized to full-length cDNA probes. The cell lines were: F9, embryonic carcinoma cell line (uninduced); P19, embryonic carcinoma cell line (uninduced); TT-HD, ovary teratoma cell line; F-Mel, Friend murine erythroleukaemic cell line; NF 561, myeloid leukaemic cell line; WEHI-3B, myelomonocyte cell line; Hybridoma (Hyb.), IgG1-producing B-cell line. The positions of 18 S and 28 S rRNA are indicated. The mRNA ($\sim$ 1.8 kb) reflecting the length of the cDNA clones is marked by an arrow. (**B**) Ethidium bromide staining of blotted 18 S and 28 S rRNA indicates equal loading of RNA.

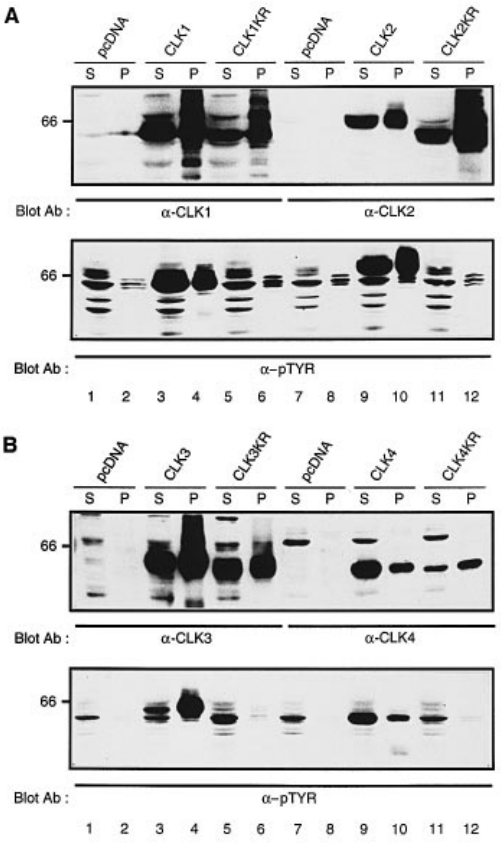

Figure 4 Autophosphorylation of mCLK1–mCLK4 on tyrosine residues *in vivo*

293 cells were transiently transfected with *mclk1*, *mclk2* or their catalytically inactive Lys $\rightarrow$ Arg (KR) mutant constructs (**A**), or with *mclk3*, *mclk4* or their catalytically inactive Lys $\rightarrow$ Arg (KR) mutant constructs (**B**). The cells were grown in 50 μM sodium orthovanadate and lysed 24 h later. The lysate was divided into supernatant (S) and pellet (P) fractions. Equal cell equivalents were loaded in each lane, blotted and stained with polyclonal antibodies against mCLK1–mCLK4 (α-CLK1 etc.) (**A** and **B**, upper panels) or with anti-phosphotyrosine antibody 4G10 (α-pTYR) (**A** and **B**, lower panels).

levels of a particular *clk* message. It is noteworthy that we found very low expression levels in Hybridoma, WEHI and NF 561 cell lines.

We also investigated the mRNA expression levels of all four genes in the C2C12 cell line and in L1 adipocytes during differentiation, but failed to detect any noticeable changes in expression (results not shown).

mCLK1–mCLK4 are autophosphorylated on tyrosine residues *in vivo*

mCLK1 was originally isolated from an expression library screen using anti-phosphotyrosine antibodies [13,14] and was subsequently shown to be autophosphorylated on tyrosine, serine and threonine residues *in vitro*, as well as on tyrosine residues *in vivo* [13,14]. Recently DOA, mCLK1, hCLK2 and the yeast homologue KNS1 were also shown to be autophosphorylated on tyrosine, serine and threonine residues *in vitro* [28]. We wished to examine the *in vivo* tyrosine phosphorylation status of all four family members in transiently overexpressing 293 cells. In Figures 4(A) and 4(B) we show that mCLK1–mCLK4 partitioned into Triton X-100-soluble and -insoluble fractions, and that the catalytically active kinases were tyrosine-phosphorylated (lanes 3, 4, 9 and 10), whereas the catalytic mutants displayed no obvious tyrosine phosphorylation signal (lanes 5, 6, 11 and 12). In some of the pellet fractions higher-molecular-mass bands were observed, which are likely to represent partially disrupted protein aggregates, as the number of bands decreased after excessive boiling in SDS buffer (results not shown). A molecular mass shift due to phosphorylation was most evident for mCLK2 (Figure 4A, lanes 9/10 and 11/12) and mCLK3 (Figure 4B, lanes 3/4 and 5/6), suggesting that these proteins were more heavily phosphorylated than either mCLK1 or mCLK4. These results strongly suggest that each mCLK is not only autophosphorylated *in vitro* (see Figure 5B; and results not shown), but also autophosphorylated on tyrosine residues *in vivo*.

mCLK1–mCLK4 phosphorylate SR proteins *in vitro*

To investigate the catalytic activity of our proteins and to test potential substrates, we initially used GST–mCLK fusion constructs, which were cloned in pcDNA3 and expressed *in vitro*. A recent report successfully demonstrated that mCLK1 could phosphorylate ASF/SF2 *in vitro*, suggesting that SR proteins are the natural substrates [12]. We therefore wished to examine the phosphorylation activity of mCLK1–mCLK4 using biochemically purified and dephosphorylated SR proteins as substrates. SR proteins were purified from 5×10^9 exponential-phase F-MEL cells in suspension according to standard procedures, and an aliquot of the purified proteins was analysed by SDS/PAGE

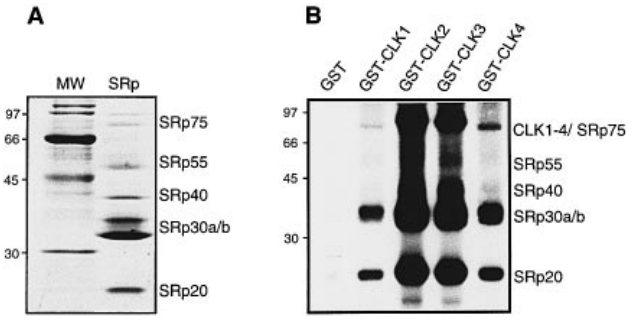

Figure 5 Phosphorylation of SR proteins *in vitro* by mCLK1–mCLK4

(**A**) SR proteins were purified from F-MEL cells using the protocol described by Zahler et al. [3]. Samples of ~ 3 μg of total SR proteins were analysed by SDS/15%-PAGE and stained with Coomassie Blue. The identity of the SR proteins was established by size [3]. Molecular masses of marker proteins (lane MW) are indicated in kDa. (**B**) Total SR proteins were dephosphorylated using the protein phosphatase 1γ catalytic subunit and then used as substrates in an *in vitro* kinase assay using equal amounts of *in vitro* produced and purified GST–mCLK fusion proteins. The assay was carried out in the presence of 1 μM of the phosphatase inhibitor Microcystin LR. The reaction was terminated with 2 × SDS sample buffer and the samples were loaded on a 15% polyacrylamide gel. Autoradiography was carried out for 24 h. The positions of the SR proteins and the autophosphorylated GST–mCLK fusion proteins are shown. Molecular masses are indicated in kDa.

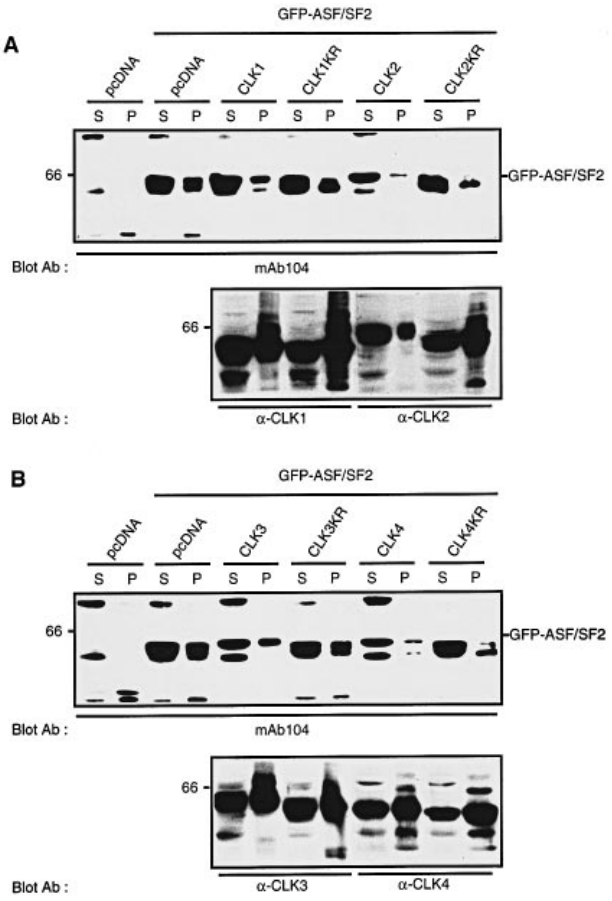

Figure 6 mCLK1–mCLK4 induce phosphorylation of ASF/SF2 *in vivo*

293 cells were transiently transfected with GFP–ASF/SF2 and various constructs as indicated [**A**, *mclk1*, *mclk2* and their catalytically inactive Lys → Arg (KR) mutant constructs; **B**, *mclk3*, *mclk4* and their catalytically inactive Lys → Arg (KR) mutant constructs], lysed 48 h after transfection and divided into supernatant (S) and pellet (P) fractions. Equal amounts of cells were loaded in each lane, blotted and stained with mAb104 to detect phosphorylated GFP–ASF/SF2 proteins (**A** and **B**, upper panels). Expression of mCLK1–mCLK4 was confirmed by reprobing with polyclonal antibodies against mCLK1–mCLK4 (α-CLK1 etc.) (**A** and **B**, lower panels).

(Figure 5A) to confirm that we had purified the expected proteins to near homogeneity. In addition, all bands were recognized by mAb104, identifying them as SR proteins (results not shown).

Following dephosphorylation by the protein phosphatase 1γ catalytic subunit, we used the SR proteins as substrates for the *in vitro* produced and purified fusion proteins GST–mCLK1 to GST–mCLK4 in an *in vitro* kinase assay. As shown in Figure 5(B), mCLK1–mCLK4 were able to phosphorylate SRp20, SRp30a/b and to a lesser extent SRp40 and SRp55. The lower signals of SRp40 and SRp55 relative to SRp20 and SRp30a/b most likely reflect the lower quantity of these proteins in the assay (Figure 5A). SRp75 was not seen in our experiments, because the autophosphorylated mCLK proteins migrate at the same position. When catalytically inactive Lys → Arg mutants were used, only background signals were observed (results not shown). We assume that the observed differences in catalytic activity between mCLK1 and mCLK4 compared with mCLK2 and mCLK3 are most likely caused by *in vitro* folding problems, as equal amounts of protein were used in each assay. To investigate the specificity of mCLK1–mCLK4, we used the recombinant human peptidyl-prolyl isomerase PIN1 as a substrate [29]. Although this protein also contains several N-terminal serine and arginine residues and is localized to nuclear speckles, none of the mCLKs was able to phosphorylate it *in vitro* (results not shown).

mCLK1–mCLK4 induce phosphorylation of the SR protein ASF/SF2 *in vivo*

It was shown previously that catalytically active mCLK1 could phosphorylate ASF/SF2 *in vitro* and cause a redistribution of mAb104 antigens *in vivo* when overexpressed in COS cells [12]. However, it was never directly demonstrated that the CLK could phosphorylate ASF/SF2 *in vivo*. We used an enhanced GFP–ASF/SF2 fusion protein, which localizes to nuclear speckles (O. Nayler and S. Stamm, unpublished work), as an *in vivo* substrate, and co-transfected the cells with the *mclk1–mclk4* constructs. GFP–ASF/SF2 was detected by mAb104, which recognizes the endogenous phosphoepitope on SR proteins [30]. In Figure 6 we show that GFP–ASF/SF2 was readily phosphorylated by endogenous kinases and partitioned into Triton X-100-soluble and -insoluble fractions, even in the absence of overexpressed mCLK1–mCLK4. However, the GFP–ASF/SF2 protein signal was sharpened and shifted to a higher molecular mass upon co-transfection with catalytically active mCLK1–mCLK4, indicating that phosphorylation had occurred. The effect was most clearly visible for mCLK2 and mCLK3 and was most easily observed for mCLK1 in the pellet fraction. GFP alone was not phosphorylated by mCLK1–mCLK4 (results not shown).

In contrast, when the catalytically inactive mCLKs were overexpressed, no such shift was seen, and some GFP–ASF/SF2 showed a decrease in molecular mass, suggesting a dominant-negative effect of the catalytically inactive kinase.

mCLK1–mCLK4 are localized to the nucleus

mCLK1 was shown previously to be localized to the nucleus [12,19], suggesting that the putative nuclear localization signal

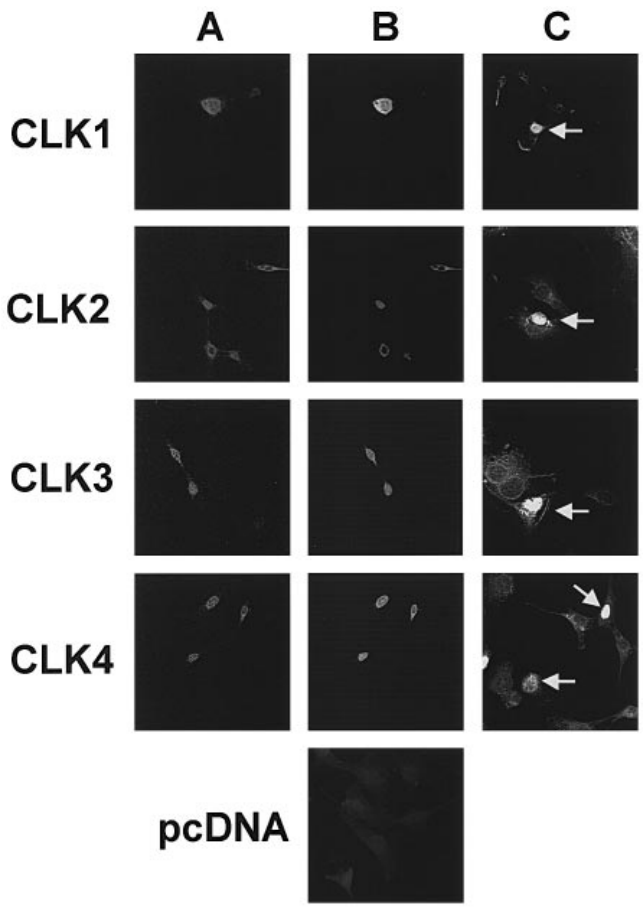

Figure 7 mCLK1–mCLK4 are localized in the nucleus

NIH 3T3 cells were transiently transfected with constructs *mclk1–mclk4*, fixed 24 h later and examined using confocal laser scanning microscopy. Cells were stained with TO-PRO iodine, a monomeric cyanine nucleic acid stain (**A**), and with polyclonal antibodies against mCLK1–mCLK4 (**B**), demonstrating co-localization of mCLK1–mCLK4 to the nucleus. In (**C**), a different set of cells stained with polyclonal antibodies against mCLK1–mCLK4 is shown using longer exposure times. Arrows pointing at transfected cells highlight the nuclear localization of mCLK1–mCLK4 (magnification $\times 31.5$).

found in the N-terminal part of the protein was indeed functional. We have found similar sequence motifs for mCLK2, mCLK3 and mCLK4, and wished to examine whether these proteins are also localized to the nucleus. We therefore transiently transfected NIH 3T3 cells with *mclk1–mclk4* constructs and used our polyclonal antibodies to immunolocalize the overexpressed proteins. As shown in Figure 7, mCLK1–mCLK4 were clearly nuclear (Figures 7B and 7C) and co-localized with a fluorescent nucleic acid dye (Figure 7A). Identical results were obtained with 293 and COS7 cells (not shown). These results confirm previous findings for mCLK1, and demonstrate that all mCLKs are indeed nuclear proteins.

DISCUSSION

We have isolated three novel mouse *clk* genes and found a high degree of conservation of amino acids within the C-terminal domain, with more diversity at the N-terminal end. All key amino acids required to identify them as kinases are present in all three proteins, and we classified them as LAMMER kinases [18] due to the EHLAMMERILGPLP signature in the catalytic domain, which is almost completely preserved. The N-terminal end is rich in serine and arginine residues, suggesting that the CLKs are SR-like proteins and possibly interact with other proteins of their kind [2]. When we compared all mCLK proteins (including the previously cloned mCLK1/STY), we found two not previously recognized subfamilies, formed by mCLK2/mCLK3 and mCLK1/mCLK4. It is possible that the mouse genome contains only four *clk* members, as our attempts to isolate additional genes by PCR and low-stringency cDNA library screening failed. However, we cannot exclude the possibility that there might be other, more distantly related, genes to be discovered.

In order to investigate and compare the expression patterns of all known mCLKs, we performed Northern blot analyses on selected adult mouse tissues and transformed cell lines. Based on published results describing *mclk1* expression [13,14], we anticipated the presence of several splice forms, especially larger molecules, which were shown to be partially spliced mRNAs [19].

We detected almost ubiquitous expression of all genes, although the expression levels differed slightly and in one tissue quite dramatically. In testis we detected significantly elevated expression levels for all but the *mclk1* gene, but only *mclk3* was represented predominantly by its 1.8 kb message. *mclk4* expression was also high, although the message size was somewhat smaller than that expected for the full-length coding mRNA. We suspect that this smaller mRNA represents an alternatively spliced molecule, as we found a clone that encodes a truncated protein, analogous to the truncated splice products reported for *mclk1*, *hclk2* and *hclk3* (results not shown). We can only speculate on the significance of these expression patterns seen in testis, although there is evidence that molecules involved in RNA processing may be critical in spermatogenesis. A mutation creating human azoospermia was mapped to a gene encoding a protein (azoospermia factor) that was presumed to be a component of the RNA processing machinery [31]. If CLKs are involved in splicing, as proposed [12], then mCLK3 could be an important component in the testis.

Although our data cannot provide proof of a regulatory function, we nevertheless observed tissue- and cell type-specific alternative splicing of the *clk* genes, which may indicate an involvement of this mechanism in the control of CLK function.

We also investigated *clk* expression levels in transformed cell types, but could not detect overexpression of *clk* genes as a common denominator in malignant cells; more thorough studies will be required. The same is true for the role of mCLK1–mCLK4 in differentiation [14,32], as we observed no significant changes in expression during L1 adipocyte and C2C12 cell differentiation (results not shown). As an explanation, there might be other family members still to be discovered (see above), or the appearance of novel transcripts in P19 cells may simply reflect a consequence rather than a cause of the differentiation process.

Although the sequences of the human homologues *hclk2* and *hclk3* have been reported [16], no further investigations were carried out. We show clearly that all mouse *clk* genes encode kinases that are autophosphorylated on tyrosine residues *in vivo*, although the significance of this is currently unclear. It was shown previously that overexpressed mCLK1 can partition into a Triton X-100-soluble fraction (catalytically active kinase) or an insoluble fraction (catalytically inactive kinase) [12]. We have attempted to repeat these results in 293 and NIH 3T3 cells, and detected similar but highly variable behaviour for mCLK2, mCLK3 and mCLK4 (results not shown). We have also found that the distribution depended on the amount of DNA used for transfection, as well as on the time before lysis. This indicated that cells regulate the distribution of these kinases using en-

dogenous factors and that localization may not depend exclusively on the autophosphorylation activity. We are currently in the process of investigating this further, and have some evidence suggesting cross-talk between mCLK1, mCLK2, mCLK3 and mCLK4 (O. Nayler, S. Stamm and A. Ullrich, unpublished work).

Our immunofluorescence studies show that mCLK1, mCLK2, mCLK3 and mCLK4 are predominantly nuclear, and we assume that the nuclear localization signals found in the N-terminal part of all CLK proteins are responsible for the observed targeting. The work reported by Colwill et al. [12] provided strong evidence that some SR proteins are natural substrates for mCLK1 and that overexpressed mCLK1 caused a redistribution of mAb104 antigens from nuclear speckles. We have investigated the substrate specificity of the three other mCLK family members, and showed that all of them were able to phosphorylate certain SR proteins *in vitro*. It is apparent that all the SR proteins that we investigated were recognized by mCLK1–mCLK4, which might be explained by the fact that ASF/SF2 (as an example) is likely to be a target of several kinases such as SRPK1, topoisomerase I and an enzymic activity associated with U1 70K [10,11,33]. It should also be kept in mind that the small number of SR proteins that we tested may represent only a fraction of the total number found in a cell [4].

We have also investigated the *in vivo* phosphorylation of a GFP–ASF/SF2 construct that was co-expressed with mCLK1–mCLK4 or their respective catalytically inactive mutants. In these experiments we observed a molecular mass shift of GFP–ASF/SF2 upon co-transfection with mCLK1, mCLK2, mCLK3 or mCLK4, indicative of *in vivo* phosphorylation, which may be caused by the CLKs directly (as shown *in vitro*) or through activation of a downstream kinase. Furthermore, mCLK1–mCLK4 dissolved nuclear speckles, as judged from the observation that GFP–ASF/SF2 was no longer localized to these sites (results not shown). However, whether this effect is truly caused by hyperphosphorylation of GFP–ASF/SF2 alone is still under investigation. It will be of future interest to map and compare the phosphorylation sites on ASF/SF2 for mCLK1–mCLK4, as this may provide an explanation for the overlapping substrate specificity that we observed.

In view of our Northern blot data, it is conceivable that mCLKs are co-expressed in most tissues and are indeed part of a regulatory mechanism that recruits and activates SR proteins, which can then participate in RNA splicing. One could then imagine that mCLK1–mCLK4 (and others) phosphorylate SR proteins at different sites, so that the SR proteins are regulated through multiple, stepwise, phosphorylation events. Alternatively, each mCLK could serve as an end-point of diverse signals, leading to the phosphorylation of the same splice factor. It is therefore of considerable interest to have this set of mCLKs available, in order to compare the phosphorylation sites and to investigate their precise nuclear localization and possible upstream signalling events. Such studies will hopefully increase our understanding of the complicated mechanisms that regulate RNA processing.

We thank M. Stein-Gerlach, C. Wallasch and K. Martell for reading the manuscript and for many helpful discussions, and S. Woodford for technical assistance. The PIN1 protein was kindly provided by G. Niederfellner. O.N. was supported by a grant from the Max-Planck-Society and the Schweizerischer Nationalfond.

REFERENCES

1 Krämer, A. (1996) Annu. Rev. Biochem. **65**, 367–409
2 Fu, X.-D. (1995) RNA **1**, 663–680
3 Zahler, A. M., Lane, W. S., Stolk, J. A. and Roth, M. B. (1992) Genes Dev **6**, 837–847
4 Blencowe, B. J., Issner, R., Kim, J., McCaw, P. and Sharp, P. A. (1995) RNA **1**, 852–865
5 Kohtz, J. D., Jamison, S. F., Will, C. L., Zuo, P., Lührmann, R., Garcia-Blanco, M. A. and Manley, J. L. (1994) Nature (London) **368**, 119–124
6 Harada, K., Martin, S. S. and Frankel, A. D. (1996) Nature (London) **380**, 175–179
7 Hedley, M. L., Amrein, H. and Maniatis, T. (1995) Proc. Natl. Acad. Sci. U.S.A. **92**, 11524–11528
8 Mermoud, J. E., Cohen, P. T. W. and Lamond, A. I. (1994) EMBO J. **13**, 5679–5688
9 Tazi, J., Kornstädt, U., Rossi, F., Jeanteur, P., Cathala, G., Brunel, C. and Lührmann, R. (1993) Nature (London) **363**, 283–286
10 Gui, J.-F., Lane, W. S. and Fu, X.-D. (1994) Nature (London) **369**, 678–682
11 Rossi, F., Labourier, E., Forné, T., Divita, G., Derancourt, J., Riou, J. F., Antoine, E., Cathala, G., Brunel, C. and Tazi, J. (1996) Nature (London) **381**, 80–82
12 Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J. L., Bell, J. C. and Duncan, P. I. (1996) EMBO J. **15**, 265–275
13 Ben-David, Y., Letwin, K., Tannock, L., Bernstein, A. and Pawson, T. (1991) EMBO J. **10**, 317–325
14 Howell, B. W., Afar, D. E. H., Lew, J., Douville, E. M. J., Icely, P. L. E., Gray, D. A. and Bell, J. C. (1991) Mol. Cell. Biol. **11**, 568–572
15 Johnson, K. W. and Smith, K. A. (1991) J. Biol. Chem. **266**, 3402–3407
16 Hanes, J., von der Kammer, H., Klaudiny, J. and Scheit, K. H. (1994) J. Mol. Biol. **244**, 665–672
17 Bender, J. and Fink, G. R. (1994) Proc. Natl. Acad. Sci. U.S.A. **91**, 12105–12109
18 Yun, B., Farkas, R., Lee, K. and Rabinow, L. (1994) Genes Dev. **8**, 1160–1173
19 Duncan, P. I., Howell, B. W., Marius, R. M., Drmanic, S., Douville, E. M. J. and Bell, J. C. (1995) J. Biol. Chem. **270**, 21524–21531
20 Lechner, C., Zahalka, M. A., Giot, J.-F., Moller, N. P. H. and Ullrich, A. (1996) Proc. Natl. Acad. Sci. U.S.A. **93**, 4355–4359
21 Ciossek, T., Lerch, M. and Ullrich, A. (1995) Oncogene **11**, 2085–2095
22 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
23 Kunkel, T. A. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 488–492
24 Puissant, C. and Houdebine, L. M. (1990) BioTechniques **8**, 148–149
25 Dignam, J. D., Lebovitz, R. M. and Roeder, R. G. (1983) Nucleic Acids Res. **11**, 1475–1489
26 Chen, C. and Okayama, H. (1987) Mol. Cell. Biol. **7**, 2745–2752
27 Dingwall, C. and Laskey, R. A. (1991) Trends Biochem. Sci. **16**, 478–481
28 Lee, K., Du, C., Horn, M. and Rabinow, L. (1996) J. Biol. Chem. **271**, 27299–27303
29 Lu, K. P., Hanes, S. D. and Hunter, T. (1996) Nature (London) **380**, 544–547
30 Roth, M. B., Murphy, C. and Gall, J. G. (1990) J. Cell Biol. **111**, 2217–2223
31 Ma, K., Inglis, J. D., Sharkey, A., Bickmore, W., Hill, R. E., Prosser, E. J., Speed, R. M., Thomson, E. J., Jobling, M., Taylor, K., et al. (1993) Cell **75**, 1287–1295
32 Myers, M. P., Murphy, M. B. and Landreth, G. (1994) Mol. Cell. Biol. **14**, 6954–6961
33 Woppmann, A., Will, C. L., Kornstadt, U., Zuo, P., Manley, J. L. and Lührmann, R. (1993) Nucleic Acids Res. **21**, 2815–2822

Received 14 January 1997/23 April 1997; accepted 9 May 1997