Phosphorylation by LAMMER Protein Kinases: Determination of a Consensus Site, Identification of In Vitro Substrates, and Implications for Substrate Preferences†

Eleni Nikolakaki,‡ Cheng Du,§ Jack Lai,|| Thomas Giannakouros,‡ Lew Cantley,|| and Leonard Rabinow*,⊥

Laboratory of Biochemistry, School of Chemistry, Aristotle University of Thessaloniki, 54 006 Thessaloniki, Greece, Department of Pathology and Microbiology, University Nebraska Medical Center, Omaha, Nebraska 68198-6495, and Division of Signal Transduction, Beth Israel Hospital and Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, and Laboratoire de Signalisation, Développement et Cancer, UPRES-A8080, Bâtiment 445, Université de Paris XI, 91400 Orsay, France

Received July 20, 2001; Revised Manuscript Received October 10, 2001

ABSTRACT: LAMMER protein kinases are ubiquitous throughout eukaryotes, including multiple paralogues in mammals. Members are characterized by similar overall structure and highly identical amino acid sequence motifs in catalytic subdomains essential for phosphotransfer and interaction with substrates. LAMMER kinases phosphorylate and regulate the activity of the SR protein class of pre-mRNA splicing components, both in vitro and in vivo. In this study, we define an optimum in vitro consensus phosphorylation site for three family members using an oriented degenerate peptide library approach. We also examine the substrate specificity and interactions of several LAMMER protein kinases from widely diverged species with potential substrates, including their own N-termini, predicted to be substrates by the peptide-based approach. Although the optimal in vitro consensus phosphorylation site for these kinases is remarkably similar for short peptides, distinct substrate preferences are revealed by in vitro phosphorylation of intact proteins. This finding suggests that these kinases may possess varied substrates in vivo, and thus the multiple LAMMER kinases present in higher eukaryotes may perform differentiable functions. These results further demonstrate that these kinases can phosphorylate a number of substrates in addition to SR proteins, suggesting that they may regulate multiple cellular processes, in addition to the alternative splicing of pre-mRNAs.

Members of the ubiquitous LAMMER protein kinase family phosphorylate SR proteins, both in vitro (*1*−*4*), and in vivo (*5*), affecting the efficiency and specificity of pre-mRNA splicing. SR proteins themselves constitute a conserved protein family, intimately involved in the regulation of pre-mRNA splicing and other steps of RNA metabolism (for reviews, see refs *6, 7*). Interactions between SR proteins in splicing complexes depend largely on arginine- and serine-rich (RS), domains that are highly phosphorylated in vivo. It is known moreover that phosphorylation of RS domains affects the specificity of both protein−protein and protein−RNA interactions (*8, 9*). Several studies have shown that changes in the phosphorylation state of SR proteins are accompanied by their spatial redistributions within the nucleus (*10*). Protein kinases from several different families possess the ability to phosphorylate SR proteins in vitro (*2*, *11*−*16*), although to date only LAMMER kinases have been demonstrated to affect SR protein activity and splicing in vivo (*5*).

LAMMER kinases display similar overall structure, with noncatalytic N-terminal extensions and similar spacing of catalytic subdomains (*17*). Furthermore, catalytic subdomains essential for phosphotransfer and substrate interaction are nearly 100% identical among family members. Although catalytic subdomain X is usually relatively less constrained among kinases, virtual 100% identity of the motif "EHLAMMERILG" is found in this subdomain in all LAMMER family members, leading to their name. Alignments of LAMMER kinase sequence with those of crystallized kinases suggest that this motif lies in a position where it might contact substrates or effectors.

LAMMER kinases possess the motifs characteristic of serine/threonine protein kinases. However, murine CLK1 (formerly known as CLK/STY), the first LAMMER kinase isolated, was identified in screens of erythroleukemia and P19 embryonal carcinoma cell cDNA expression libraries with anti-phosphotyrosine antibodies (*18, 19*). Although LAMMER kinases autophosphorylate with dual-specificity in vitro (*20, 21*), and may also do so in vivo (*22*), some doubt about the in vivo autophosphorylation on Tyr residues has been raised (*4*). Moreover, the biological significance,

† Supported by grants from the U.S. National Science Foundation, Association pour la Recherche Contre le Cancer, and Fondation pour la Recherche Medicale to L.R., by a grant from the Greek Secretariat of Research and Technology (ΠΕΝΕΔ 1999) to T.G., and by NIH award #GM56203 to L.C. The collaboration between the groups in Thessaloniki and Paris was supported by grants from the University of Paris XI (Program Epicure), a grant from the Greek Secretariat of Research and Technology (French−Greek exchange grant, ΕΠΑΝ−M.43) to T.G., and a grant from the French Ministère des Affaires Etrangères (Program Platon, EGIDE) to L.R.

* Corresponding author. Tel: 33-01-69157465. Fax: 33-01-6915-68-02. E-mail: Leonard.Rabinow@emex.u-psud.fr.

‡ Aristotle University of Thessaloniki.

§ University Nebraska Medical Center.

|| Harvard Medical School.

⊥ Université de Paris XI.

10.1021/bi011521h CCC: $22.00
Published on Web 01/18/2002

if any, of autophosphorylation on Tyr residues remains to be demonstrated.

Analysis of LAMMER kinase function in vivo has benefited from the characterization of mutations in the *Drosophila* LAMMER kinase locus *Darkener of apricot* (*Doa*). *Doa* alleles were initially identified as dominant suppressors of mutations induced by insertions of the retrotransposon copia, and they increase transcription of this retroelement 2−4-fold. They are almost invariably recessive lethal (*23, 24*). Adults recovered from specific combinations of hypomorphic alleles escaping lethality display pleiotropic phenotypes affecting many adult structures, suggesting that *Doa* plays an essential role in the differentiation and maintenance of a wide variety of cell types (*17*). *Doa* mutations affect the sex-specific splicing of *dsx* pre-mRNA, a key regulator of somatic sexual differentiation in *Drosophila*, yielding sexual transformations (*5*). It was also demonstrated that aberrant splicing of *dsx* pre-mRNA was accompanied by the altered intranuclear localization and hypophosphorylation of at least two SR proteins in *Doa* mutant animals.

Further support for the hypothesis that LAMMER kinases function in cellular differentiation was obtained with the finding that the tobacco family member PK12 is induced both transcriptionally and posttranscriptionally by the maturation hormone ethylene (*21*).

The high degree of structural similarity among LAMMER kinases, particularly in catalytic subdomains essential for peptide interaction and phosphotransfer, suggests that they possess similar substrates in widely diverged organisms, from the yeast *Saccharomyces cerevisiae* to humans. However, the lack both of authentic SR proteins and of alternative mRNA splicing in the yeast *S. cerevisiae* (*25*), as well as the dispensability of the LAMMER kinase KNS1 for vegetative growth (*26*, Rabinow et al., manuscript in preparation), suggests that these kinases may play roles in the regulation of cellular processes in addition to that of alternative splicing. Recent evidence demonstrates that *Doa* plays an important role in activating *hedgehog* signaling in *Drosophila* (Du and Rabinow, submitted), further supporting this idea.

Here we further characterize the activity of several LAMMER kinases in vitro, including the definition of a consensus phosphorylation site for three family members from widely diverged species, phosphorylation studies with several substrates, further characterization of interactions with SR proteins, and an examination of their ability to autophosphorylate on their N-termini. Despite high amino acid sequence identity in catalytic subdomains and the similarity of preferred peptide substrate phosphorylation sites, differences in protein substrate preferences exist among the kinases assayed.

EXPERIMENTAL PROCEDURES

Determination of a Consensus Phosphorylation Site for Three LAMMER Kinases, Clk2, DOA, and PK12, Using a Degenerate Peptide Library: Kinase Reactions and Phosphopeptide Separation. For small scale kinase assay, each LAMMER kinase was added to 30 μL of solution containing 100 μg of degenerate peptide mixture from each peptide library, 100 μM ATP, 5 μCi [γ-^{32}P]ATP, 10 mM $MgCl_2$, 10 mM $MnCl_2$, 20 mg/mL BSA, and 50 mM Tris-HCl pH 7.4. The reaction mixture was incubated at 37 °C for 2 h. Reaction mixtures were spotted onto P81 phosphocellulose paper and washed five times in 500 mL of 1.5% phosphoric acid. Incorporation of ^{32}P was determined by liquid scintillation counting. For each condition, values for control reactions lacking substrate peptide were subtracted as blanks. For large scale kinase assays with the RS-peptide library, the volume of each component except [γ-^{32}P]ATP and enzyme was scaled up 10-fold from small scale. The amount of [γ-^{32}P]ATP in the large-scale assay was reduced from 5 to 0.33 μCi. On the basis of the results of the small scale assay, the amount of enzyme added was varied to allow phosphorylation of about 1% of the total peptide mixture. A control phosphorylation experiment, in which the peptides were subjected to a mock phosphorylation without the presence of enzyme, was also performed. The reaction was stopped by adding 300 μL of 30% acetic acid. The phosphopeptides were separated from ATP and enzyme by using a DEAE-Sephacel column This mixture was added to a 0.8−1.2 mL DEAE ion-exchange column previously equilibrated with 30% acetic acid and eluted using the same solvent. After the first 600 μL void volume, the eluent was collected in 0.5 mL fractions. Liquid scintillation counting for each fraction was determined with and without a P81 phosphocellulose paper wash, to evaluate the efficiency of ATP removal at this step. Normally, the first two fractions that were free of [γ-^{32}P]ATP were pooled and concentrated by Speed-Vac at 4 °C. Further purification from unphosphorylated peptides was performed on an iron-chelating column. Briefly, a 0.3−0.4 mL column of ferric iminodiacetic acid−agarose (IDA, Sigma) was loaded with 2 mL of 20 mM $FeCl_3$ solution at a flow rate of 0.2 mL/min controlled by an Econo Pump (Bio-Rad). The column was then sequentially washed by 2 mL of H_2O, 3 mL of buffer C (500 mM NH_4HCO_3), pH 8.0, 2 mL of H_2O, and 3 mL of buffer A (50 mM MES pH 5.5, 1 M NaCl) at a flow rate of 0.4 mL/min. The dried sample was dissolved in 200 mL of buffer A and loaded onto the IDA column while the output flow rate was controlled at 0.2 mL/min. The column was then sequentially eluted (1 mL per fraction) with 3 mL of each buffer A, H_2O, and buffer C at flow rate 0.4 mL/min. Judging by the ^{32}P counting, the desired fractions were combined and concentrated to give 1−3 nmol of phosphopeptide mixture.

Sequencing and Data Analysis. Typically, 1−2 nmol of the phosphopeptide mixture was added to the sequencer. To calculate the relative preference of amino acids at each degenerate position and to correct the background, the initial unassayed peptide library (starting) and the control peptide mixture (background) were also sequenced. The abundance of amino acid present, calculated as a molar percentage ratio at each cycle, and its corresponding position relative to the phosphorylation site were mapped into a 2D data matrix. The background matrix was subtracted from the experimental matrix with a numerical adjusting factor equally applied to all the background data. The adjusting factor was chosen in order not to overly subtract the background from the experimental matrix, which would result in a negative value. The corrected data were then compared with the data from the starting mixture to calculate the ratio of amino acid abundance. The sum of abundance of each amino acid at a given cycle was normalized to 16 (the number of amino acids

present at the degenerate positions) so that each amino acid would have a value of 1 in the absence of selectivity at a particular position.

Yeast Two-Hybrid Constructs and Assays. The GAL4-based two-hybrid system was utilized, using plasmids pAS1, pACT, and pACTII, developed by Elledge and collaborators (*27*). The host yeast strain PJ69-4A (*28*) allowed for stringent growth selection on His^- media. Standard lithium acetate transformants were selected on Leu^--$Tryp^-$ medium, and single colonies were selected and streaked onto selective (Leu^--$Tryp^-$-His^-) medium for assay. At least three independent transformants were tested for each construction. Controls on the two-hybrid experiments included testing with the isolated transcriptional activating-domain plasmids without inserts as well as a number of potential false interactors. These included fusions with the GAL4 DNA binding domain in pAS1 and p53, lamin, and the CDK2 and SNF1 kinases in the activating-domain plasmid. False positive and murine Clk1 fusion plasmids were kind gifts of Karen Colwill (*1*). Tests for expression of the recombinant proteins expressed from the DNA-binding domain plasmid pAS1 and pACTII via immunoblots utilizing the HA epitope-tag carried in the constructs were used to verify protein expression.

Plasmid constructions were as follows.

pAS1:DOA full-length: An *Eco*RI-*Pst*I fragment (2.5 kb) of pCD1+2 (*17*) was blunted with the Klenow fragment of DNA polymerase I, *Nco*I linkers were added, and the near full-length kinase cDNA was inserted into the *Nco*I site of pAS1 (*5*).

pAS1:DOA-catalytic domain: The *Sty*I fragment (1.6 kb) containing the DOA-catalytic domain from pCD1+2 was blunted, *Nco*I linkers were added at both ends, and the new construct was inserted into the *Nco*I site of pAS1 (*5*).

pAS1:DOA-N-terminal domain: The pAS1:DOA full-length construct was digested with *Bam*HI, to remove the catalytic domain, and religated on itself (*5*).

pAS1:CLK2 full-length: A 1.7 kb *Kpn*I fragment containing a full-length human CLK2 cDNA (*29*) was cloned into pUC18 and redigested with *Bam*HI-*Eco*RV, the 1.5 kb fragment inserted into Bluescript SK+, and a *Bam*HI-*Sal*I fragment (1.5 kb) inserted into the corresponding sites on pAS1.

pAS1:CLK2 N-terminal domain: A *Kpn*I-*Eco*RI fragment of the CLK2 N-terminal clone SK-I3 (generous gift of S. Hanks) was cloned into pUC18, redigested with *Bam*HI-*Eco*RI, and inserted into Bluescript SK^+. A 740 bp *Bam*HI-*Sal*I fragment was ligated into pAS1.

pAS1:KNS1 N-terminal domain: A 440 bp *Ssp*I fragment from KNS1 (*26*), encoding the N-terminus of the kinase, was ligated to *Bam*HI linkers and inserted into the corresponding site on pAS1.

pAS1:KNS1 catalytic domain: The 1.3 kb *Not*I and *Bam*HI fragment that contains the KNS1 catalytic domain was blunted, ligated to *Bam*HI linkers, and inserted into the pAS1 *Bam*HI site.

pAS1:CLK1, pACTI ASF, and pACT X16 plasmids were gifts of Karen Colwill and Tony Pawson (*1*).

pACTII:TRA: tra coding sequences were amplified by the polymerase chain reaction (PCR) with primers 5′GCGGATCCCGATGAAAATGGATGCC3′ and 5′GCGGATCCTGTTCAATATGGGGGACG3′ from plasmid pKStra-B, which encodes a full-length female tra cDNA. The 630 bp reaction product encoding full length tra was digested with *Bam*HI and inserted into the *Bam*HI site of pACTII (*5*).

pACTII:TRA2: TRA2 coding sequences from plasmid pSKC13-B, containing a cDNA encoding a full-length tra-2 in its longest form (264 amino acid residues), were amplified by PCR, with the primers 5′CGGGATCCGCAAAAAGCCAATGGATCG3′ and 5′GCGGATCCGCTTTAAAATAGCGCGATG3′. The 830 bp amplification product encoding full-length TRA2 was digested with *Bam*HI and inserted into the corresponding site of pACTII.

pACTII:RBP1: A 543 bp *Bam*HI-*Eco*RI fragment from a RBP1 full length cDNA was directly cloned into the corresponding sites on pACTII (*5*).

pACTII:RBP1ΔRS: RBP1 cDNA sequences were amplified using a primer for the T3 RNA polymerase promoter present in the vector and the internal RBP1 primer 5′GCGACCCGAATTC ATCTCTAC3′. The PCR amplification product was digested with *Bam*HI and *Eco*RI, yielding a 300 bp fragment, which was ligated to pACTII (*5*).

Protein Expression, In Vitro Phosphorylation, Determination of K_m Values, and Phospho-Amino Acid Analysis.

The LAMMER protein kinases DOA, KNS1, CLK2, and PK12 were expressed as catalytic domain fusion proteins in *Escherichia coli* and purified as previously described (*17*, *20*, *21*). The pMalC2 bacterial expression vector was used to construct plasmids encoding the N-terminal domain of DOA and CLK2 fused with maltose binding protein.

pMalC2:Doa-N: pCD1+2 was cut with *Eco*RI and *Bam*HI, to yield a 640 bp fragment, which was ligated to pMalC2, generating a fusion protein including the initial 213 amino acid residues of DOA from the start of our longest cDNA clone (amino acid residues 14−227 of the kinase (*17*).

pMalC2:CLK2-N: A 730 bp *Sma*I-*Eco*RI fragment from pUC18:I3N, encoding the majority of the N-terminal domain of CLK2, was ligated to pBluescript SK^+. The resulting plasmid was digested with *Bam*HI-*Sal*I, yielding a 750 bp fragment, which was ligated to the corresponding sites of pMalC2. This construct produces a fusion protein including amino acid residues 1−133 of CLK2 kinase.

In vitro phosphorylation assays were performed in a total volume of 25 μL containing 25 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM DTT, 50 μM [γ-^{32}P]ATP (6000 Ci/mmol), 1.5−5 μg of the appropriate substrate, and 0.3−0.8 μg/mL purified LAMMER kinase fusion proteins. Samples were incubated for 30 min at 30 °C, and the reaction was stopped by adding 5 μL of 5 X Laemmli buffer and heating at 95 °C for 3 min. Following gel separation of the phosphorylated proteins, radiolabeled proteins were excised from the gel, acid hydrolyzed, and phospho-amino acid analysis by two-dimensional thin-layer electrophoresis was performed on cellulose thin-layer plates (*30*).

For the determination of K_m, the amount of substrate in the reaction mixture was varied between 0.1 and 15 μg, and the concentration of ATP was raised to 100 μM. Incorporation of radioactivity was measured by excising the respective radioactive bands from an SDS−PAGE gel and scintillation counting. The K_m values were calculated using the program MicroCAl Origin (version 2.94).

The nonphosphorylated form of human protamine P1 was a generous gift of Philippe Chevaillier (Laboratoire de Biologie Cellulaire, Université Paris-Val de Marne, 94010

Créteil, France) and purified from sperm nuclear basic proteins by chromatography on carboxymethylcellulose. It was characterized by its amino acid composition and partial amino acid sequence, as previously described (*31*, *32*). Myelin basic protein was purchased from Sigma (Sigma Chemical Co., Deisenhofen, Germany) or Boehringer (Boeringer Mannheim GMbH, Germany). Histones H1, H2A, H2B, H3 and H4 were purchased from Boehringer. Ribosomal proteins were extracted from mouse liver ribosomes using two volumes of CH_3COOH, as previously described (*33*). RBP1, B52, TRA, and TRA2 proteins were prepared from SF9 cells and used without further treatment or were phosphatase-treated, as previously described (*5*). The recombinant proteins prepared from SF9 cells were a generous gift from William Mattox. TRA2 was also prepared from *E. coli*. Wild-type and specifically altered forms of LBR protein were prepared as GST fusion proteins as previously described (*16*, *34*). Briefly, GST-wtNt contains the N-terminal domain of LBR (amino acids 1−205); GST-DRSNt contains the N-terminal domain of LBR but is lacking the RS motifs (amino acids 75−84); GST-wtNtG^{78}, GST-wtNtA^{80}, GST-wtNtA^{82}, and GST-wtNtA^{84} fusion proteins are identical to GST-wtNt, except that in each case Ser^{78}, Ser^{80}, Ser^{82}, or Ser^{84} was mutated to Gly or Ala. The GST domain itself was not an efficient substrate for any of the kinases tested under our conditions.

RESULTS

Identification of Consensus Phosphorylation Sites for Three LAMMER Kinases from Widely Diverged Species.

To determine the optimum consensus and relative amino acid residue preferences in the phosphorylation site for the LAMMER kinases DOA (*Drosophila*), CLK2 (human), and PK12 (tobacco), we tested four degenerate peptide libraries. These were (1) a "RS" peptide library, with a Ser residue at the phosphorylation site and an Arg fixed at position −3; (2) a Ser-oriented degenerate peptide library (Met-Ala-Xxx-Xxx-Xxx-Xxx-Xxx-Ser-Xxx-Xxx-Xxx-Xxx-Ala-Lys-Lys-Lys); (3) a Ser-Pro-oriented library (Met-Ala-Xxx-Xxx-Xxx-Xxx-Ser-Pro-Xxx-Xxx-Xxx-Ala-Lys-Lys-Lys); and, finally, (4) a Tyr-oriented library. The three latter libraries were phosphorylated to a markedly lesser extent than the RS peptide libarary, demonstrating the importance of Arg at the P−3 position. This finding is consistent with previous studies that have suggested the importance of Arg at the P−3 position for substrates of the LAMMER kinase CLK1 (*35*) and SRPK2 (*13*), a close relative of LAMMER kinases.

The proline (P+1) directed library was phosphorylated by PK12, DOA, and Clk2, respectively, to 59%, 49%, and 42% of the values relative to the RS peptide library. The totally degenerate serine-oriented library was only phosphorylated to 25%, 22%, and 0%, respectively, of the values of the RS library. The phosphorylation level for the tyrosine-oriented library was 11% relative to RS for PK12 and was barely phosphorylated by the other two LAMMER kinases. Further characterization of LAMMER kinase phosphorylation specificity was thus performed only on the RS library. We also tested the LAMMER kinase KNS1 of *S. cerevisiae* on the RS peptide library. However, consistent with previous results (*20*), no significant phosphorylation was detected, due to its low activity on exogenous peptides mixtures or proteins.

The RS peptide library comprises peptides of the sequence Met-Ala-Xxx-Xxx-Xxx-Xxx-Arg-Xxx-Xxx-Ser-Xxx-Xxx-Xxx-Xxx-Xxx-Ala-Lys-Lys-Lys, where Xxx indicates all amino acids except Ser, Thr, Cys, and Trp; Trp and Cys were omitted to avoid problems with sequencing and oxidation, whereas Ser and Thr were omitted to ensure that the only potential site of phosphorylation was the Ser at position 10. The Met-Ala dipeptide tag at the amino terminus was included to verify that peptides from this mixture are being sequenced and to facilitate the initial coupling reaction during solid-phase synthesis. The embedded Ala residue at the C-terminal region provided an estimate of how much peptide loss occurred during sequencing. The poly-lysine tail at the end of each library sequence prevents wash-out during sequencing and improves the solubility of the mixture.

For sequencing, the RS peptide library was incubated with each LAMMER kinase under conditions in which approximately 1% of the total peptide mixture was phosphorylated. The phosphopeptide products were separated from nonphosphorylated peptides using a ferric−iminodiacetic acid column, and the mixture was sequenced. In Figure 1, the relative abundance of amino acids at each of the 11 degenerate positions is presented from experiments using CLK2, whereas the results obtained for the two other LAMMER kinases are summarized in Table 1. The selectivity values in parentheses indicate how strongly a particular amino acid is selected at a given position.

As shown in Figure 1 and Table 1, the three LAMMER kinases tested generally prefer Arg and to a lesser extent His near the phosphorylation site, consistent with the fact that these enzymes phosphorylate SR splicing factors. In contrast to SRPKs, which show a strict requirement for Arg on both sides of the phosphorylated serine, LAMMER kinases strongly selected for peptides with acidic amino acids (Glu and Asp) at the P−4, P−2, P+2, P+3, and P+4 positions. At the P−4 position, DOA preferred an acidic amino acid (Glu) rather than Arg, whereas Asp at the P+3 position was optimal for all three LAMMER kinases. There was some differential selectivity among the three kinases in that PK-12 had a rather strict requirement for a charged amino acid at the P−2 and P−1 positions (preferentially Arg and to a lesser extent Glu, Asp, and His), while both CLK2 and DOA also selected for peptide substrates with a hydrophobic (Ala being preferred but also Leu, Ile, Phe, Val) or a polar but not charged amino acid (Gln, Gly, and Tyr) at the P−2 position and to a lesser extent at the P-1 position. A hydrophobic amino acid (Phe being preferred but also Pro, Leu, and Val) was also moderately selected at the P+1 position by all three kinases. The finding of Pro as a favored amino acid at the P+1 position is consistent with previous observations on murine CLK1 kinase (*2*), as well as the sequenced phosphorylation site of DOA, human CLK2, and KNS1 LAMMER kinases on myelin basic protein (*20*). It might also be recalled that the phosphorylation of the peptide library with Pro at +1 (the Ser-Pro-oriented library) was noticeably higher in our initial experiments than the completely degenerate Ser library (see above). It was also previously suggested that Pro might be preferentially selected at the P+1 position of LAMMER substrates (*13*), due to the similarity of the P+1 pocket of the LAMMER family members to that of MAP kinases (*17*, *18*).

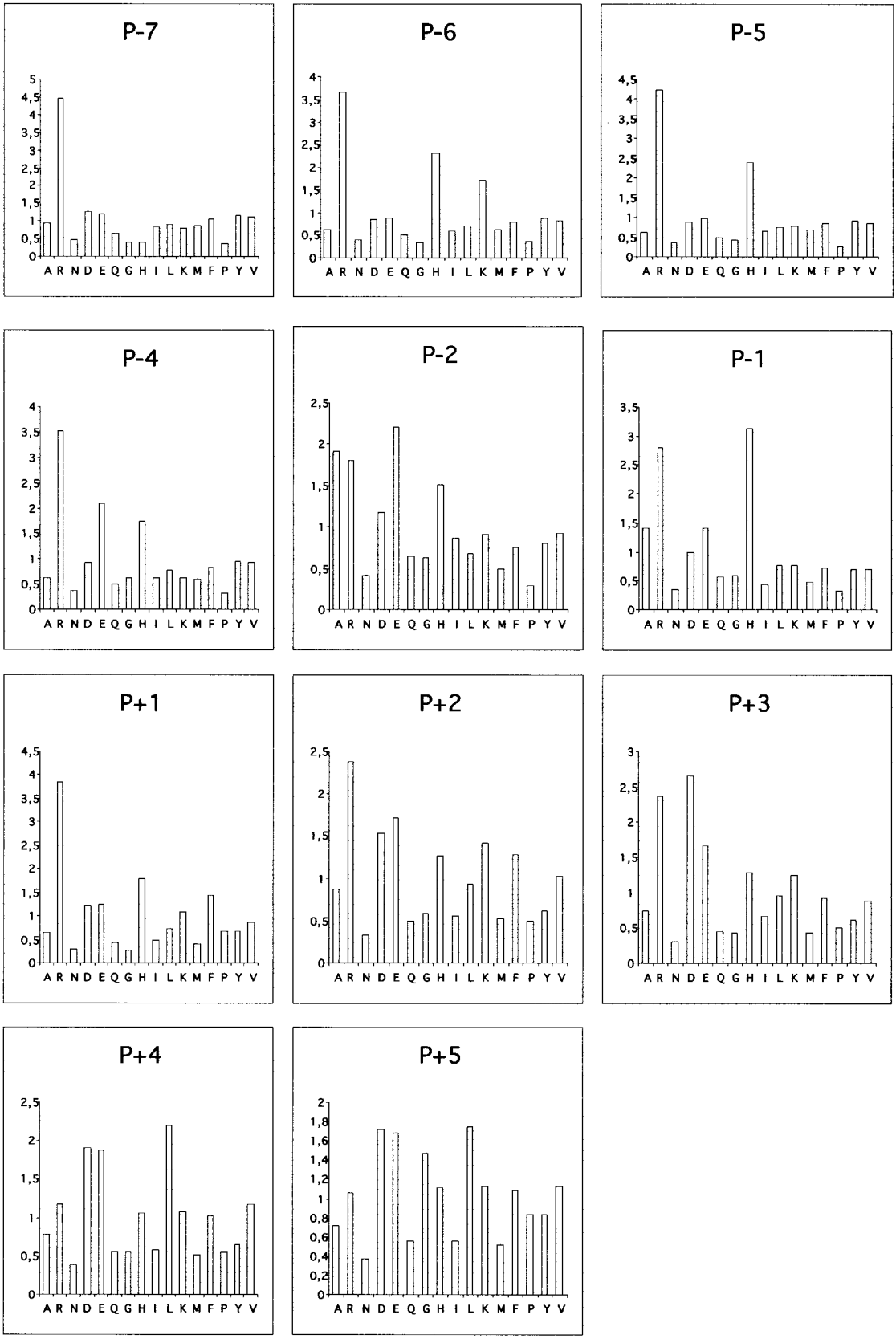

FIGURE 1: Substrate specificity of protein kinase CLK2. An Arg-Ser-oriented degenerate peptide library with the sequence Met-Ala-Xxx-Xxx-Xxx-Xxx-Arg-Xxx-Xxx-Ser-Xxx-Xxx-Xxx-Xxx-Xxx-Ala-Lys-Lys-Lys, (where Xxx indicates all amino acids except Ser, Thr, Cys, Trp), was incubated with CLK2 as described under Experimental Procedures. Phosphorylated peptides were separated on DEAE-Sephacel and ferric chelation columns and sequenced. Each panel indicates the relative abundance of the 16 amino acids at a given cycle of sequencing. Abbreviations for amino acid residues: A, Ala; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; V, Val; Y, Tyr.

Table 1: Substrate Specificities of LAMMER Protein Kinases

LAMMER kinase	−7	−6	−5	−4	−3	−2	−1	0	+1	+2	+3	+4	+5
CLK2	**R**(4.5)	**R**(3.7)	**R**(4.2)	**R**(3.5)	**R**	**E**(2.2)	**H**(3.1)	**S**	**R**(3.8)	**R**(2.4)	**D**(2.7)	**L**(2.2)	L(1.8)
		H(2.3)	**H**(2.4)	**E**(2.1)		A(1.9)	**R**(2.8)		H(1.8)	E(1.7)	**R**(2.4)	D/E(1.9)	D/E(1.7)
		K(1.7)		H(1.7)		R(1.8)				D(1.5)	E(1.7)		G(1.5)
DOA	**R**(3.6)	**R**(3.6)	**R**(3.5)	**E**(3.3)	**R**	**E**(2.6)	**H**(3.1)	**S**	**R**(4.1)	**R**(2.7)	**D**(3.3)	**L**(2.8)	G/D(1.8)
	A(1.6)	**K**(2.9)	E(1.5)	**R**(2.3)		**A**(2.5)	**R**(2.3)		F(1.8)	E(1.9)	**R**(2.7)	**E**(2.0)	L/E(1.6)
	D(1.5)					R(1.7)	E(1.7)			D(1.6)	E(1.7)	D(1.8)	R(1.5)
PK-12	**H**(9.3)	**R**(9.8)	**E**(2.6)	**E**(4.3)	**R**	**R**(4.2)	**R**(4.6)	**S**	**R**(4.6)	**R**(3.1)	**D**(3.9)	**D**(2.3)	**D**(2.2)
	R(7.4)	**H**(4.4)	**D**(2.5)	**R**(4.2)		**E**(3.3)	**H**(2.7)		**D**(2.1)	**E**(2.5)	**R**(2.6)	**E**(2.2)	G/E(1.9)
	D(3.6)	**E**(4.3)	**H**(2.4)			**D**(2.2)	**E**(2.0)		E(1.9)	**D**(2.4)	**E**(2.2)	**L**(2.0)	R(1.7)
	E(3.3)	**D**(4.2)					D(1.7)			K(1.5)		R(1.8)	
		K(4.0)											
		Y(1.7)											

[a] Values in parentheses indicate the relative selectivities for the amino acids; amino acids with values less than 1.5 are omitted. The one-letter amino acid code is used. Bold letters indicate amino acids which are most preferentially selected (value >2.0).

Table 2: Potential CLK2 Substrates Identified in a Database Search on Scansite (http://cansite.bidmc.harvard.edu)[a]

score	ID	accession	protein	position(s)	sequence
0.06879	HSP1_DIDMA	P35305	sperm protamine P1	24	RRRRRSRSRRRRSRR
0.08323, 0.12360	SFR6_HUMAN	Q13247	arginine/serine-rich 6	208, 206	RRRSRSRSRRSSRSR
			pre-mRNA splicing factor, SRp55		
0.09196, 0.09991	SFR2_HUMAN	Q01130	arginine/serine-rich 2	130, 132	RRRRRSRSRSRSRSR
0.13155			pre-mRNA splicing factor, SC35	138	RSRSRSRSRSRSRYS
0.10837	SFR5_RAT	Q09167	arginine/serine-rich 5	189	RHRSRSRSRSRTRSS
			pre-mRNA splicing factor, SRp40	205	RSRSRSRSRRSKSYS
0.11518	DDX8_HUMAN	Q14562	probable ATP-dependent	208, 210	RSRSRSRSRTRERNK
			RNA helicase HRH1 (DEAH box protein)		
0.12136	U2R2_HUMAN	Q15696	ribonucleoprotein	447	RSRSRSRSRRSRRSR
			auxiliary factor, 35 KDA		
			splicing factor U2AF35		
0.12175	SFR7_HUMAN	Q16629	arginine/serine-rich 7	132	RRRSRSRSRSHSRSR
			pre-mRNA splicing factor, 9G8		
0.12220	SFRB_HUMAN	Q05519	arginine/serine-rich 11	253	RRHSRSRSRSRRRRT
			arginine-rich 54 kD U1		
0.12268	RU17_HUMAN	P08621	ribonucleoprotein 70 kD	445	DRRRRSRSRDKEERR
			U1 snRNP 70 kD splicing factor		
0.12747	XE7_HUMAN	Q02040	protein XE7	660	DRHRRERSRERRGSA
			B-lymphocyte antigen precursor		
0.13155	SFR1_HUMAN	Q07955	arginine/serine-rich 1	212	RSRSRSRSRSRSRSN
			pre-mRNA splicing factor ASF/SF2		

[a] Only the first occurrence of a probable CLK2 substrate is listed, without respect to the species. However, multiple sites are indicated in the several cases where a single protein contained multiple probable phosphorylation sites within the first 20 "hits" obtained.

LAMMER Kinases from a Variety of Organisms Physically Interact with and Phosphorylate SR Proteins and SR-Related Splicing Regulators.

On the basis of our selection data we scanned the entire protein database for potential substrates of LAMMER kinases. In agreement with previous in vitro and in vivo phosphorylation studies (*1*−*3*, *5*, *36*), many SR proteins were identifed in the top 0.5 percentile hits using a matrix based on the data in Figure 1 and Table 1 applied to the SCANSITE website (http://cansite.bidmc.harvard.edu) (*37*) (Table 2). The phosphorylation sites predicted lie in the RS domains of SR proteins, including snRNP-associated (e.g., U1 70K) and non-snRNP associated (e.g., U2AF) splicing factors and other spliceosomal components. ASF/SF2, the most commonly used in vitro substrate for CLK1, was also identified in the top 0.009 percentile of SCANSITE hits.

It had been previously shown that the LAMMER kinases DOA from *Drosophila*, murine CLK1, and tobacco PK12 interact with and phosphorylate SR proteins and SR-related splicing regulators. Using the yeast two-hybrid assay (*38*), we additionally tested for interactions between the human CLK2 and yeast KNS1 LAMMER kinases with the murine SR proteins ASF/SF2 and SRp20 (also known as X16) and *Drosophila* RBP1, as well as the SR-related splicing regulators from *Drosophila* TRA and TRA-2. TRA, TRA2, and RBP1 proteins assemble into splicing complexes to regulate alternative splicing of sex-specific transcripts in *Drosophila* and bind an exonic splicing-enhancer in the *dsx* transcript (*39*−*41*). Murine CLK1 and *Drosophila* DOA kinases were included as positive controls, since their interactions with SR proteins in the two-hybrid system were previously documented (*1*, *5*).

Interactions of full-length human CLK2 (formerly SK-G1 (*42*)) and the *S. cerevisiae* LAMMER kinase KNS1p were observed with each of the SR and related proteins tested (Figure 2, Table 3). The lack of interactions between these kinases and a truncated version of RBP1 protein lacking its RS domain demonstrates that this domain is absolutely required for interaction with LAMMER kinases (Figure 2). In combination with similar interactions previously documented with *Drosophila* DOA and murine CLK1 (*1*, *5*), these results further demonstrate that the LAMMER kinases generally interact with SR-related proteins through the RS domains of the latter protein family.

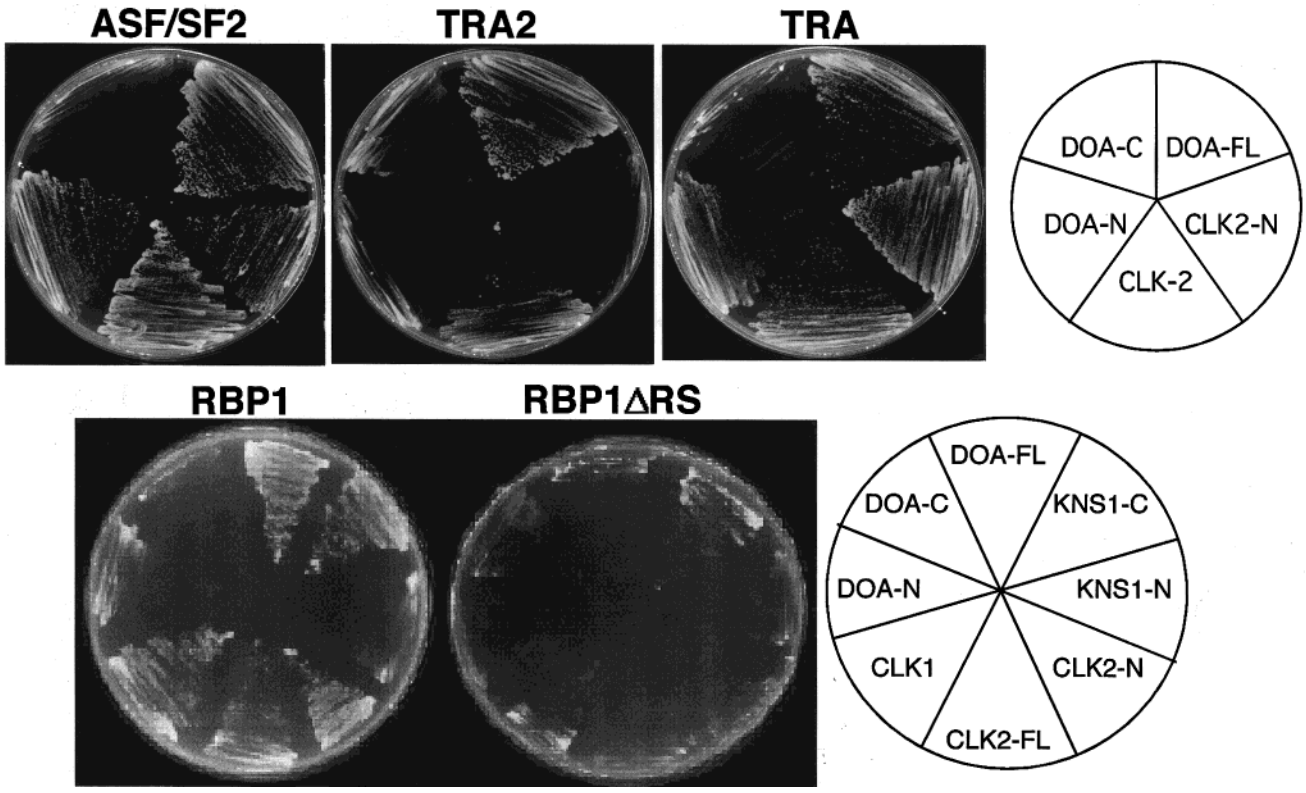

FIGURE 2: LAMMER protein kinases generally interact with SR proteins in vitro. Yeast two-hybrid assays were conducted between LAMMER protein kinases from widely diverged species and *Drosophila* and mammalian SR proteins. The growth of yeast on selective media demonstrating interaction between these protein classes indicates a preference, though not an absolute requirement, for the entire LAMMER kinase (both the noncatalytic N-terminal and the catalytic domain). Illustrated in the top set of panels are the murine SR-protein ASF/SF2 and the *Drosophila* SR-related proteins TRA and TRA2. In the lower panels, the *Drosophila* SR protein RBP1 was tested for interaction with LAMMER kinases from diverse species, with and without its RS domain. These results confirm and extend previous demonstrations that interaction between these protein families is dependent on an intact RS domain. FL= full-length kinase; C = C-terminal catalytic domain only; N = N-terminal noncatalytic domain only. LAMMER kinases are DOA, *Drosophila*, CLK2:human, CLK1:mouse, KNS1:*S. cerevisiae*. SR proteins were assayed as fusions with the GAL4 transcriptional activation domain of pACT, and the kinase genes were fused with the GAL4 DNA-binding domain of pAS1, transformed and expressed in yeast PJ69-4A. Selection for double transformants was on Leu^--Trp^- medium, following which individual colonies were streaked and assayed for interaction on Leu^--Trp^--His^- medium (shown).

Table 3: Summary of SR protein−DOA Interactions in the Yeast Two-Hybrid System

pACT	TRA	TRA2	ASF/SF2	SRp20/X16	RBP1	RBP1(ΔRS)
pAS1						
CLK2-N	+++	−	++	+	+++	−
CLK2-FL	++	++	+++	+++	+++	−
KNS1-N	−	−	−	−	−	−
KNS1-C	−	+	+++	+++	+++	+/−
DOA-FL	++	+++	+++	+++	+++	−
DOA-N	++	+/−	+++	−	++	−
DOA-C	+/−	+	−	−	−	−
Clk1	+++	+++	+++	+++	+++	−

[a] PACT, GAL4 activation domain fusion construct; pAS1, GAL4 DNA-binding domain fusion construct; FL = full-length protein; N = N-terminal noncatalytic domain of kinase; C = kinase catalytic domain; ΔRS= construct lacking RS domain. Kinases were CLK2 (human); KNS1 (S. cerevisiae); DOA (*Drosophila*); Clk1 (mouse). +++ = strong interaction/growth; ++ = moderate interaction/growth; + = weak interaction/growth; +/− = minimal to no interaction/growth; − = no interaction/growth.

Many of the LAMMER kinases themselves contain short Arg-Ser repeats within their amino terminal noncatalytic domains, although these RS repeats are limited to at most two repetitions of the motif. Because these regions could conceivably mediate interactions with SR proteins, we tested both the catalytic and noncatalytic regions of LAMMER kinases for their ability to interact with SR and related proteins. Previous results suggested that both the N-terminal noncatalytic, as well as the catalytic, domains of these kinases are required for efficient interaction with SR proteins. In contrast, using several SR proteins and the LAMMER kinases listed in Table 3, we found that interactions between SR proteins and the kinases lacking their catalytic domains does occur in some instances, as represented by weak growth on selective media (Figure 2, Table 3). These correspond to the cases where the interaction between the SR protein and the full-length LAMMER kinase is particularly strong in the yeast system. In agreement with previous results, the isolated catalytic domains of these kinases essentially failed to interact with the SR proteins (Figure 2; Table 3). These data generalize previous findings by our group as well as others that both catalytic and noncatalytic regions of LAMMER kinases are necessary for the most efficient interactions with the RS domains of SR proteins in the yeast two-hybrid system, but the data suggest that the Arg-Ser-rich N-termini of the LAMMER kinases are capable of independently mediating at least partial interactions with SR proteins.

We next tested whether the two-hybrid interactions demonstrated were indicative of in vitro phosphorylation. Four LAMMER kinases (DOA, CLK2, PK12, and KNS1) were expressed as kinase catalytic domain fusion proteins, since we were unable to obtain their expression as full-length proteins in *E. coli.* All four efficiently phosphorylated TRA,

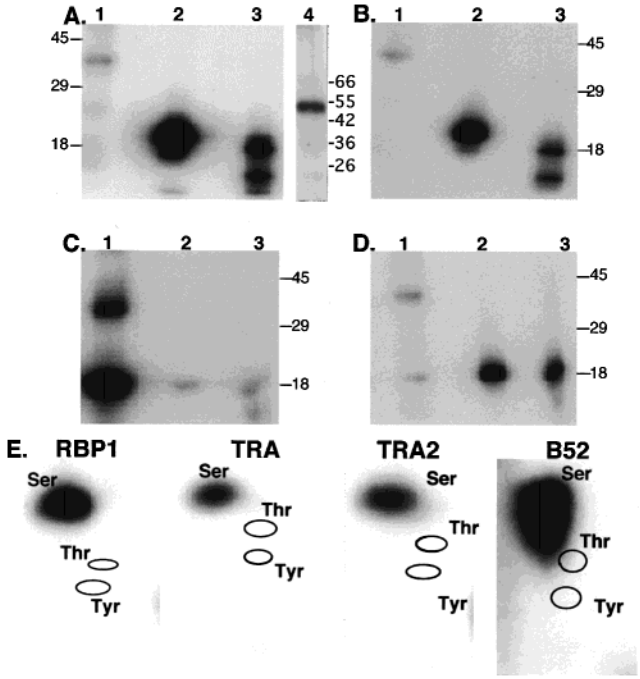

FIGURE 3: SR proteins are phosphorylated on serine residues in vitro by LAMMER kinases from evolutionarily diverged organisms. In vitro phosphorylation reactions were performed using the recombinant LAMMER protein kinases indicated, to assay their ability to phosphorylate the SR and related proteins shown. SR proteins were expressed and purified from insect Sf9 cells [TRA2, B52 and TRA, (not shown)], or bacteria (RBP1). Myelin basic protein (MBP), a known LAMMER protein kinase substrate, was included as a positive control. Lanes: (1) TRA2, (2) RBP1, (3) MBP, (4) B52 (DOA only). TRA protein was also an efficient substrate for all four kinases (not shown): (A) DOA: *Drosophila*; (B) CLK2: human; (C) KNS1: S. cerevisiae; (D) PK12: tobacco. Note the relative differences in activity of the various kinases toward the various substrates (e.g., compare panels A and B, lanes 1 and 2 with panel C, lanes 1 and 2). (E) Phosphorylation of SR proteins by LAMMER kinases occurs exclusively on Ser residues. Phosphoamino acid analysis of the proteins phosphorylated by CLK2 (panel B; except B52, phosphorylated by DOA, Panel A, lane 4) are shown. The migration of unlabeled phosphorylated Ser, Thr, and Tyr standards is indicated. Only Ser residues were phosphorylated on the SR protein substrates RBP1, TRA, TRA2, and B52 (which was assayed with DOA only) by all four LAMMER kinases assayed (CLK2, PK12, KNS1, and DOA).

TRA2, and RBP1 expressed in SF9 insect cells and/or in *E. coli* (Figure 3A−D and data not shown). DOA also efficiently phosphorylated the *Drosophila* SR protein B52 (Figure 3A).

Phospho-amino acid analysis of the SR proteins phosphorylated by these LAMMER kinases demonstrated that all detectable phosphorylation occurred on serine residues for these substrates (Figure 3E and not shown). Surprisingly, the four LAMMER kinases differed somewhat in their relative phosphorylation of these substrates (Figure 3A−D). This finding was unexpected, given the high identity among all LAMMER kinases in catalytic subdomains essential for substrate interaction, the similarities in phosphorylation sites of these kinases on peptide substrates, and that all those tested phosphorylate the same site in myelin basic protein (*20*).

LAMMER Kinases Do Not Interact in the Two-Hybrid System or Autophosphorylate Their N-Termini In Vitro.

Although the LAMMER kinases are very similar in amino acid sequence in their catalytic domains, their noncataytic N-terminal extensions are extremely diverged. However, as mentioned, the animal LAMMER kinases, including mammalian CLK 1-4 and *Drosophila* DOA as well as the single *C. elegans* orthologue encoded by transcript E02H4.3, possess two or more repeats of the RS motif in their N-terminal noncatalytic domain, similar to those found in their SR protein substrates (*1*, *3*, and unpublished observation of Rabinow). The presence of these repeats suggested that they might be sites of efficient autophosphorylation. Indeed, the N-termini of the mammalian CLKs and DOA were predicted to be substrates by the peptide-based approach at the CANSITE website. Furthermore, it has been previously reported that murine CLK1 dimerizes through its N-terminal domain or with an isolated N-terminal protein (*22*).

We therefore assessed the ability of DOA to interact with itself as a full-length protein or truncations of the C-terminal catalytic or the N-terminal noncatalytic domains, using the yeast two-hybrid system. The DOA N-terminal and catalytic domains, as well as full-length DOA protein, were assayed as fusions to both the GAL4 activation and binding domains, by themselves or with full-length murine CLK1 in these assays. CLK1 was similarly assayed by itself. In contrast to our expectations, little or no growth was observed in any combination (data not shown). These results suggest that no interactions occur, that any interactions are too transient to be detected by this assay, or that a secondary factor necessary for interaction is lacking in yeast.

We further tested the ability of two LAMMER kinase catalytic domains to autophosphorylate their N-termini in vitro. Because full-length kinases of any fusion constructs have been unrecoverable, we used maltose-binding domain fusion proteins of the kinase catalytic domains for DOA and CLK2 in in vitro assays, testing the ability of each of kinase to phosphoryate its own N-terminal noncatalytic domain as a separate protein, as for the exogenous substrates phosphorylated in Figures 3 and 4. In no case was phosphorylation observed (data not shown), although the N-termini of KNS1 and CLK2 were efficient in vitro substrates for PKA in control experiments, showing that the N-terminal protein was both intact and capable of being phosphorylated (not shown).

At first glance these results contradict a previous report of murine CLK2 autophosphorylation on Ser141 influencing its intranuclear localization (*4*). However, our human CLK2 N-terminal fusion protein terminated just N-terminal to that point, at residue 133. It might also be pointed out that although the mammalian LAMMER kinases possess a potentially autophosphorylatable Ser residue at this site, none of the LAMMER kinases from other organisms do. Thus, the presence of RS motifs is not a determinant sufficient for LAMMER kinase phosphorylation, and autophosphorylation in the noncatalytic N-terminal domains of these kinases is at best limited, despite the predictions made by the peptide library-based approach.

Differential In Vitro Specificity of LAMMER Kinases toward Potential Substrate Proteins.

Further exmination of the in vitro phosphorylation specificity of LAMMER kinases was performed using non-SR protein potential substrates identified in SCANSITE database searches. *L*amin *B r*eceptor (LBR), a ubiquitous component of the inner nuclear membrane (*43*), and P1 protamine, a highly basic protein that replaces histones during spermatogenesis (for review see ref *44*), were among the proteins

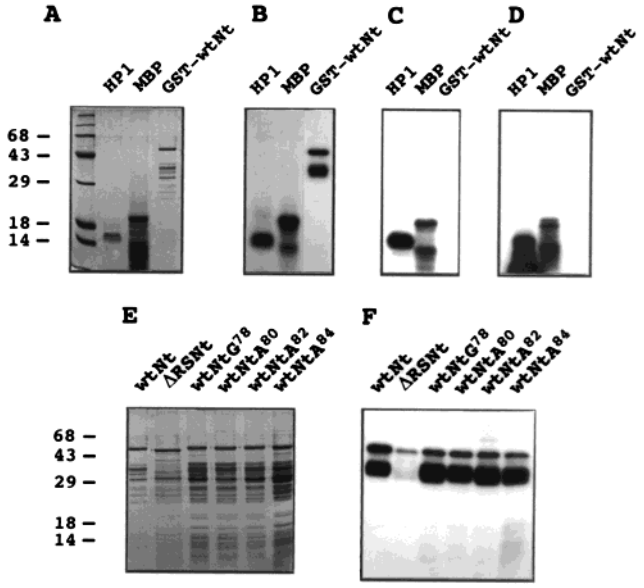

FIGURE 4: In vitro phosphorylation of LBR and P1 protamine by the LAMMER kinases. In vitro phosphorylation reactions were performed using the recombinant LAMMER protein kinases indicated, to assay their ability to phosphorylate the N-terminal domain of LBR protein (amino acids 1−205) expressed as a GST fusion protein (GST-wtNt) and purified human P1 protamine. Myelin basic protein (MBP) was included as a positive control. The samples were analyzed by SDS−PAGE on 14% gels and stained with Coomassie Blue (panel A) or autoradiographed (panels B, C, and D): (B) PK12; (C) CLK2; (D) DOA. The full-length GST-wtNt migrates with an apparent molecular mass of approximately 51 kDa. The lower bands are degradation products (*16, 57*). Coomassie blue staining (panel E) and phosphorylation (panel F) of wild-type and specifically altered forms of LBR protein by PK12. GST-wtNt contains the N-terminal domain of LBR; GST-ΔRSNt contains the N-terminal domain of LBR but is lacking the RS motifs (amino acids deleted 75−84); GST-wtNtG^{78}, GST-wtNtA^{80}, GST-wtNtA^{82}, and GST-wtNtA^{84}: fusion proteins are identical to GST-wtNt except that in each case Ser^{78}, Ser^{80}, Ser^{82}, or Ser^{84} was mutated to Gly or Ala. Molecular mass standards are shown at left (in kDa). The upper two molecular mass standards correspond to 200 and 97 kDa. Control experiments show that DOA and CLK2 do not detectably phosphorylate the GST domain when expressed as an isolated protein (not shown). Additionally, the drastic drop in phosphorylation of the GST-ΔRSNt (panel F, lane 2), compared with the LBR constructs containing RS repeats, suggests that the vast majority of phosphorylation events by this kinase occur in the RS domain and not on the GST fusion domain or the non-RS domain of LBR.

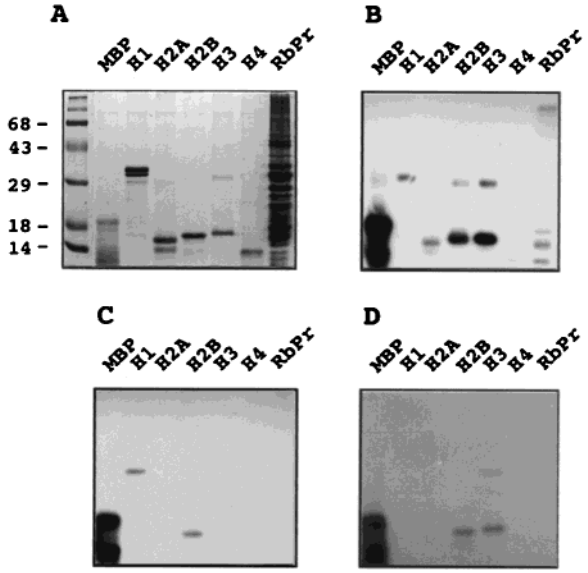

FIGURE 5: Phosphorylation of histones H1, H2A, H2B, H3, and H4 and ribosomal proteins (RbPr) by the LAMMER kinases. Myelin basic protein (MBP) was included as a positive control. The samples were analyzed by SDS−PAGE on 14% gels and stained with Coomassie Blue (panel A) or autoradiographed (panels B, C, and D): (B) PK12; (C) CLK2; (D) DOA. Molecular mass standards are shown at the left (in kDa). The upper two molecular mass standards correspond to 200 and 97 kDa. Phosphorylations were performed with equivalent levels of kinase activity, and exposures were for identical periods of time to demonstrate the relative specificty differences for the three kinases toward these substrates.

predicted to be efficient substrates of the LAMMER kinases based upon the peptide specificity data. Both of these proteins contain RS domains and are efficiently phosphorylated by human SRPK1 (*34, 45*). Accordingly, purified human P1 protamine and the N-terminal domain of LBR protein (amino acids 1−205), expressed as a GST fusion protein (construct termed GST-wtNt), were tested for their ability to be phosphorylated by the LAMMER kinases. As shown in Figure 4, CLK2, DOA, and PK12 all efficiently phosphorylated P1 protamine, whereas to our surprise only PK12 efficiently modified the N-terminal domain of LBR. The RS motifs of LBR included the major phosphorylation sites of PK12, since the phosphorylation of a fusion protein consisting of GST and LBR N-terminal domains, but missing the RS domain (residues 75−84; construct termed GST-DRSNt), was significantly reduced (Figure 4F). To determine more specifically the serine residues of the RS domain of LBR that were phosphorylated by PK12, we used as substrates fusion proteins identical to GST-wtNt except that in each case Ser^{78}, Ser^{80}, Ser^{82}, or Ser^{84} was mutated to Gly or Ala (constructs termed GST-wtNtG^{78}, GST-wtNtA^{80}, GST-wtNtA^{82}, and GST-wtNtA^{84}, respectively). As shown in Figure 4F, all four fusion proteins were phosphorylated similarly to wtNt, suggesting that any of the serines of the RS domain could be targeted by PK12.

PK12 displayed similar K_m values for MBP, P1 protamine, and GST-wtNt (6.3, 6.5, and 5.5 μM, respectively). The K_m values displayed by CLK2 and DOA for P1 protamine were 8 and 9.5 μM, respectively. However, both these enzymes displayed significantly higher K_m values for MBP (40 and 65 μM, respectively). This result is consistent with the ratio of P1 protamine/MBP phosphorylation observed with these two kinases. The respective K_m value displayed by SRPK1 for P1 protamine was 5 mM (*45*).

Previous work showed that histone H1 was moderately phosphorylated by murine CLK1 (*2*). However, it has also been reported that while DOA efficiently phosphorylated MBP, there was no detectable phosphorylation of either histone H1 or S6 ribosomal protein (*20*). To address these discrepancies, we tested the ability of PK12, CLK2, and DOA kinases to phosphorylate histones H1, H2A, H2B, H3, H4, as well as a mixture of ribosomal proteins. Consistent with our other data, PK-12 displayed the broadest specificity and phosphorylated histones H2B and H3, whereas histones H1 and H2A and some low molecular weight ribosomal proteins were poorly targeted (Figure 5B). On the other hand, CLK2 and DOA kinases displayed stricter specificity and weakly phosphorylated histones H1 and H2B and histones H2B and H3, respectively (Figure 5C,D).

Thus it is clear from Figures 3, 4, and 5 that the different LAMMER kinases selectively phosphorylate different protein substrates, despite similarities in their consensus phosphorylation sites based on peptide analysis.

DISCUSSION

We have determined the optimal peptide substrates of three LAMMER kinases from diverged organisms using an oriented peptide library approach. As in the case of the SRPK family kinases, DOA (*Drosophila*), CLK2 (human), and PK12 (tobacco) generally prefer Arg and to a lesser extent His at P−1 to the phosphorylation site. Although the optimal in vitro consensus phosphorylation site for these kinases was found to be remarkably similar for short peptides (Table 1), and members of this family display remarkably high sequence identity in catalytic subdomains, distinct substrate preferences are revealed by the in vitro phosphorylation of intact proteins.

PK12 possesses the broadest specificity of the kinases and phosphorylated all substrates tested (SR proteins, LBR, P1 protamine, and MBP), with more or less the same efficiency. Mutational analysis using a panel of LBR mutants demonstrated that PK12 phosphorylates the RS domain in a protein substrate and that this kinase can target any of the serines of the RS domain. While PK12 could thus conceivably play the roles of both LAMMER and SRPK kinases due to its broader specificity, the recently completed sequencing of the genome of *Arabidopsis thaliana* (*46*) demonstrates the existence of several putative SRPK-like kinases (Rabinow, unpublished observations), as well as three LAMMER kinases previously described (*47*). Thus the specificity of the plant LAMMER kinases will require further investigation in vivo, as is the case for their animal orthologues.

In contrast to PK12, both CLK2 and DOA only very weakly phosphorylated the RS domain of LBR. These two kinases also exhibit slightly different specificities, considering the ratio of their P1 protamine/MBP phosphorylation as well as the phosphorylation of histones. CLK2 modified H1 and H2B to a limited extent, while DOA barely phosphorylated H2B and H3. Differences in the specificity among members of the LAMMER family kinases have also been reported in the past (Table 4). The major site of phosphorylation by murine CLK1 on MBP, Ser109 (GRGLS̲LSRFS) (*35*), is different from that of the three other LAMMER family members tested, DOA, CLK2, and KNS1, whose major site of phosphorylation on this substrate is Ser164 (SRSGS̲PMAR) (*20*). The former authors proposed that basic residues (especially Arg) at positions flanking the phosphorylation site are critical determinants of CLK1 substrate specificity, with an Arg being absolutely required at positions P−3 and P+3. However, the same group had reported earlier that the major site of phosphorylation of tyrosine phosphatase PTB-1B by both CLK1 and CLK2 is Ser50 (YRDVS̲PFDH), where an acidic residue (Asp) is found at position P+3 (*48*). The latter finding is more in accordance with the data presented here characterizing consensus phosphorylation sites for the three LAMMER kinases. Also consistent with this report, Lee et al. (*20*) did not observe any phosphorylation of H1 histone by DOA, whereas Colwill et al. (*2*) reported that murine CLK1 phosphorylated histone H1 to the same extent as MBP. It is of course possible that CLK1 possesses slightly different substrate specificity compared with other family members.

Table 4: Sequences at the Phosphorylation Sites of Known Protein Substrates of the LAMMER Kinases

protein substrates	phosphorylation sites
myelin basic protein[a]	GRDS**R**SG**S**PMAR
myelin basic protein[b]	QGKG**R**GLS̲LSRF
PKC-zetazide[c]	MPRK**R**QGS̲VRRR
protein tyrosine phosphatase-1B[d]	RNRY**R**DVS̲PFDH
P1 protamine[e]	YRCC**R**SQS̲RSRY[14]
CLK2 autophosphorylation site	HRRK**R**TRS̲VEDD
ASF/SF2[f]	[198]RSPSYG(RS)$_8$NSRSRSYSPR-RSRGSPRYSPRHSRSRSRT[248]
consensus optimal CLK2 site[g]	RRRR**R**EHS̲RRDLL
consensus optimal DOA site[g]	RRRE**R**EHS̲RRDLG/D

[a] Site phosphorylated by DOA, CLK2, and KNS1 (*20*). [b] Site phosphorylated by CLK1 (*35*). [c] The polybasic PKC-zetazide peptide was the best peptide substrate tested out of a survey of 14 peptide substrates (*35*). [d] Site phosphorylated by CLK1 and CLK2 (*48*). [e] Ser10 and Ser8 of the RS domain were previously identified as the only in vivo phosphorylated sites of human P1 protamine (*32*). [f] ASF/SF2 is phosphorylated by CLK1 in its RS domain (amino acids 198−248; *1*). It is not clear which of the serines of the RS domain are targeted by the enzyme (see also Discussion). [g] The optimum consensus phosphorylation sites for DOA and CLK2 are based on the peptide library approach described here, presented for purposes of comparison with sequenced sites of known substrates.

On the basis of our phospho-peptide selection data, in combination with the in vitro phosphorylation results of our group and those of others, it appears that phosphorylation mediated by the LAMMER kinases requires five criteria (see Table 4): (1) they prefer to phosphorylate serine and not threonine residues; (2) they prefer a basic residue (Arg preferentially) at the P−3 position; (3) they prefer a basic environment around the **R**XX**S** motif; (4) they prefer an arginine or a hydrophobic residue at the P+1 position (in native protein substrates Pro seems to be preferred, somewhat in contrast to the peptide data, where it is not a particularly selected residue), and (5) they prefer polar (noncharged) or hydrophobic and to a lesser extent charged residues at the P−2 and P−1 positions. PK12 is an exception, since it modifies substrates with both basic and noncharged polar residues at these positions.

The validity of the fifth criterion is questionable when examining the phosphorylation of the SR proteins by the mammalian CLK kinases and DOA. The characterization of the specificity of CLK1 was based on the phosphorylation of a panel of ASF/SF2 mutants (*2*). From phosphopeptide mapping the sites of phosphorylation of ASF/SF2 are limited to the RS domain (amino acids 198−248 (*1*). However, it is not clear which of the serines in the RS domain are targeted by CLK1. Taking into consideration the phosphorylation of known peptide substrates (this study), as well as the phosphorylation of the GRSRSRSRSR peptide (*2*), most if not all the serines of the RS domain could be modified by CLK1. On the other hand, in line with the consensus sequence observed with other protein substrates (see Table 4), Ser221 (SRSNS̲RSRS), Ser227 (SRSYS̲PRRS), and to some extent Ser234 (RSRGS̲PRYS) might be preferentially targeted by CLK1. Interestingly, *Drosophila* RBP1 protein, which interacts with the LAMMER kinases via its RS domain and is phosphorylated both in vitro and in vivo (Table 3, (*5*)), also contains such a potential site, Ser122 (RSRSFS̲RDRRSRS). The observations that CLK2 and DOA fail

to phosphorylate either the RS domain of LBR (RRSRS-RSRSRSPGRP), or the short RS regions (RSRS) in their own N-termini suggest that the lack of serines in a basic environment with polar or hydrophobic amino acids at the P−2 and P−1 positions may contribute to the discrimination of potential substrates.

The interesting finding that the LAMMER kinases tested did not autophosphorylate on their N-termini conflicts with predictions made by SCANSITE using the consensus phosphorylation sites determined from the peptide-based approach. The N-termini of several LAMMER kinases were identified in these scans, yet our data clearly demonstrate that they are not efficient substrates for their own catalytic domains. One possibility is that phosphorylation of a single Ser residue, perhaps by a different kinase, creates an acidic environment at a P−2, P+2, or P+3 at a LAMMER kinase site, facilitating N-terminal autophosphorylation. Whether another kinase(s) can perform such an activating phosphorylation is unknown, but SRPK1 at least does not efficiently phosphorylate the N-termini of either human CLK2 or DOA in vitro (Giannakouros, unpublished). Another possible explanation that was not investigated here is that autophosphorylation of the noncatalytic domain requires its tethering to the peptide encoding the catalytic activity of the kinase.

It has been well-documented that the assembly of the splicing particle (the "spliceosome"), the splicing reaction, and the appropriate intranuclear localization of splicing proteins are strongly influenced by the state of phosphorylation of the RS domain of SR proteins (*49*−*51*; for a review, see ref *10*). Our results confirm and extend previous reports suggesting a central role for LAMMER kinases in pre-mRNA splicing through the phosphorylation of SR proteins. The initial description of the murine CLK1 LAMMER kinase interacting with SR proteins through their RS domains was based upon yeast two-hybrid and biochemical experiments (*1*, *2*). Subsequent reports confirmed the interaction of murine CLK1 with SR proteins in vitro and extended the observation to cultured cells of animals (*3*, *52*) and to PK12 in plants (*36*). In *Drosophila*, *Doa* mutants possess altered sexual differentiation due to aberrant pre-mRNA splicing of *dsx* (*5*), demonstrating similar functions for LAMMER kinases in vivo in developmentally regulated alternative splicing. Previous biochemical and genetic studies demonstrate that female-specific *dsx* splicing depends on the SR-related splicing regulators TRA and TRA2, as well as the authentic SR-protein RBP1, which form complexes on a splicing enhancer located within the *dsx* female-specific exon. These proteins are thought to function by recruiting U2AF and possibly other components of the general splicing machinery to *dsx* RNA. The activity and localization of these proteins are directly affected by mutations in DOA kinase, due to their hypophosphorylation.

Therefore, one major function of the LAMMER family of kinases overlaps with the SRPKs in the phosphorylation of SR proteins, although the functional requirements for the specificity of the two kinase families with respect to SR protein substrates remains unknown. The potential for overlapping substrate specificity could conceivably allow for the modulation of general and specific SR protein activities in constitutive or alternative splicing events, since the level of serine phosphorylation in RS domains is thought to be essential for SR protein specificity in protein−protein and protein−RNA interactions (*8*, *9*). Interestingly, it was noted that *Doa* mutations do not affect the alternative splicing of all transcripts (*5*). Although the basis for this specificity remains unknown, it is possible that SRPK phosphorylates SR proteins instead of DOA in specific tissues or developmental stages or on specific substrates. Another possible explanation for the specficity of DOA kinase in affecting alternative splicing of specific transcripts may depend on the nature of the SR-protein−RNA interactions themselves, since it was recently found that requirements for intact RS domains are substrate specific (*53*).

The rather broad specificity of the LAMMER kinases and the diversity of their in vitro substrates suggests that these kinases may play a more global regulatory role than the SRPKs. CLK1 has been previously suggested to activate components of growth-factor-stimulated signal transduction cascades, resulting in the activation of ERK-1, ERK-2, and pp90RSK (*54*), although genetic analyses in *Drosophila* failed to identify a role for LAMMER kinases in MAP kinase activation (*55*). CLK1 and CLK2 phosphorylate and activate the tyrosine phosphatase PTP-1B (*48*), implicating them in yet other signaling pathways. Furthermore, mCLK2 and mCLK4 are highly expressed in testis, while CLK3 is almost exclusively expressed in testis in both mice and rats (*3*, *56*). On the basis of this observation, P1 protamine is an intriguing candidate as an in vivo substrate for the LAMMER protein kinases. Besides being phosphorylated in vitro by all LAMMER kinases tested, the existence of phosphorylated P1 protamines in mammals is well-documented, while Ser10 and Ser8 of the RS domain (^{7}**RSQSR**SR^{13}) were previously identified as the phosphorylation sites of phosphorylated human P1 protamine (*32*). Furthermore, the alternating RS or SR dipeptides are highly conserved in all vertebrate P1 protamines examined to date and are even more so in several molluscan protamine-like proteins (*45*), indicating that they play a crucial role in the condensation of sperm chromatin. Members of the LAMMER protein kinases together with the SRPKs might be important regulators of this process.

ACKNOWLEDGMENT

We thank Karen Colwill and Tony Pawson for helpful discussions and clones; Steven Elledge for the yeast two-hybrid system reagents; K.H. Scheit and Steve Hanks for human CLK2 clones; Robert Fluhr for the PK12 expression clone; Bill Mattox for RBP1, TRA and TRA2 proteins; and Philippe Chevaillier for the gift of protamine P1.

REFERENCES

1. Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J. L., Bell, J. C., and Duncan, P. I. (1996) *EMBO J. 15*, 265−275.
2. Colwill, K., Feng, L. L., Yeakley, J. M., Gish, G. D., Caceres, J. F., Pawson, T., and Fu, X. D. (1996) *J. Biol. Chem. 271*, 24569−24575.
3. Nayler, O., Stamm, S., and Ullrich, A. (1997) *Biochem. J. 326*, 693−700.
4. Nayler, O., Schnorrer, F., Stamm, S., and Ullrich, A. (1998) *J. Biol. Chem. 273*, 34341−34348.
5. Du, C., McGuffin, M. E., Dauwalder, B., Rabinow, L., and Mattox, W. (1998) *Mol. Cell 2*, 741−750.
6. Manley, J. L., and Tacke, R. (1996) *Genes Dev. 10*, 1569−1579.
7. Valcarcel, J., and Green, M. R. (1996) *Trends Biochem. Sci. 21*, 296−301.

8. Tacke, R., Chen, Y., and Manley, J. L. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 1148−1153.
9. Xiao, S. H., and Manley, J. L. (1997) *Genes Dev. 11*, 334−344.
10. Misteli, T., and Spector, D. L. (1997) *Trends Cell Biol. 7*, 135−138.
11. Kuroyanagi, N., Onogi, H., Wakabayashi, T., and Hagiwara, M. (1998) *Biochem. Biophys. Res. Commun. 242*, 357−64.
12. Tang, Z., Yanagida, M., and Lin, R. J. (1998) *J. Biol. Chem. 273*, 5963−5969.
13. Wang, H. Y., Lin, W., Dyck, J. A., Yeakley, J. M., Songyang, Z., Cantley, L. C., and Fu, X. D. (1998) *J. Cell Biol. 140*, 737−50.
14. Gui, J. F., Lane, W. S., and Fu, X. D. (1994) *Nature 369*, 678−682.
15. Gui, J. F., Tronchere, H., Chandler, S. D., and Fu, X. D. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10824−10828.
16. Nikolakaki, E., Simos, G., Georgatos, S. D., and Giannakouros, T. (1996) *J. Biol. Chem. 271*, 8365−8372.
17. Yun, B., Farkas, R., Lee, K., and Rabinow, L. (1994) *Genes Dev. 8*, 1160−1173.
18. Ben-David, Y., Letwin, K., Tannock, L., Bernstein, A., and Pawson, T. (1991) *EMBO J. 10*, 317−325.
19. Howell, B. W., Afar, D. E. H., Lew, J., Douville, E. M. J., Iceley, P. L. E., Gray, D. A., and Bell, J. C. (1991) *Mol. Cell. Biol. 11*, 568−572.
20. Lee, K., Du, C., Horn, M., and Rabinow, L. (1996) *J. Biol. Chem. 271*, 27299−27303.
21. Sessa, G., Raz, V., Savaldi, S., and Fluhr, R. (1996) *Plant Cell 8*, 2223−2234.
22. 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.
23. Rabinow, L., and Birchler, J. A. (1989) *EMBO J. 8*, 879−889.
24. Rabinow, L., Chiang, S. L., and Birchler, J. A. (1993) *Genetics 134*, 1175−1185.
25. Yun, C. Y., and Fu, X. D. (2000) *J. Cell Biology 150*, 707−718.
26. Padmanabha, R., Gehrung, S., and Snyder, M. (1991) *Mol. Gen. Genet. 229*, 1−9.
27. Durfee, T., Becherer, K., Chen, P. L., Yeh, S. H., Yang, Y., Kilburn, A. E., Lee, W. H., and Elledge, S. J. (1993) *Genes Dev. 7*, 555−569.
28. James, P., Halladay, J., and Craig, E. A. (1996) *Genetics 144*, 1425−1436.
29. Hanes, J., von der Kammer, H., Klaudiny, J., and Scheit, K. H. (1994) *J. Mol. Biol. 244*, 665−672.
30. Boyle, W. J., Van Der Geer, P., and Hunter, T. (1991) in *Methods Enzymol.* (Hunter, T., and Sefton, B. M., Eds.) pp 110−149, Academic Press, New York.
31. Gusse, M., Sautiere, P., Belaiche, D., Martinage, A., Roux, C., Dadoune, J. P., and Chevaillier, P. (1986) *Biochim. Biophys. Acta 884*, 124−134.
32. Chirat, F., Arkhis, A., Martinage, A., Jaquinot, M., Chevaillier, P., and Sautière, P. (1993) *Biochim. Biophys. Acta 1203*, 109−114.
33. Hardy, S. J. S., Kurland, C. G., Voynow, P., and Mora, C. (1969) *Biochemistry 8*, 2897−2905.
34. Papoutsopoulou, S., Nikolakaki, E., and Giannakouros, T. (1999) *Biochem. Biophys. Res. Commun. 255*, 602−607.
35. Menegay, H. J., Myers, M. P., Moeslein, F. M., and Landreth, G. E. (2000) *J. Cell Sci. 113*, 3241−3253.
36. Savaldi-Goldstein, S., Sessa, G., and Fluhr, R. (2000) *Plant J. 21*, 91−96.
37. Yaffe, M. B., Leparc, G. G., Lai, J., Obata, T., Volinia, S., and Cantley, L. C. (2001) *Nature Biotech. 19*, 348−353.
38. Fields, S., and Song, O. (1989) *Nature 340*, 245−246.
39. Tian, M., and Maniatis, T. (1992) *Science 256*, 237−40.
40. Tian, M., and Maniatis, T. (1994) *Genes Dev. 8*, 1703−1712.
41. Heinrichs, V., and Baker, B. S. (1995) *EMBO J. 14*, 3987−4000.
42. Hanks, S. K., and Quinn, A. M. (1991) In *Methods in Enzymology* (Hunter, T., and Sefton, B. M., Eds.) pp 38−62, Academic Press, New York.
43. Worman, H. J., Evans, C. D., and Blobel, G. (1990) *J. Cell Biol. 111*, 1535−1542.
44. Oliva, R., and Dixon, G. H. (1991) *Prog. Nucleic Acid Res. Mol. Biol. 40*, 25−94.
45. Papoutsopoulou, S., Nikolakaki, E., Chalepakis, G., Kruft, V., Chevaillier, P., and Giannakouros, T. (1999) *Nuc. Acids Res. 27*, 2972−2980.
46. Initiative, A. G. (2000) *Nature 408*, 796−814.
47. Bender, J., and Fink, G. R. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 12105−12109.
48. Moeslein, F. M., Myers, M. P., and Landreth, G. E. (1999) *J. Biol. Chem. 274*, 26697−26704.
49. Tazi, J., Kornstadt, U., Rossi, F., Jeanteur, P., Cathala, G., Brunel, C., and Luhrmann, R. (1993) *Nature 363*, 283−286.
50. Mermoud, J. E., Cohen, P. T. W., and Lamond, A. I. (1994) *EMBO J. 13*, 5679−5688.
51. Yitzhaki, S., Miriami, E., Sperling, R., and Sperling, J. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 8830−8835.
52. Duncan, P. I., Stojdl, D. F., Marius, R. M., and Bell, J. C. (1997) *Mol. Cell. Biol. 17*, 5996−6001.
53. Zhu, J., and Krainer, A. R. (2000) *Genes Dev. 14*, 3166−3178.
54. Myers, M. P., Murphy, M. B., and Landreth, G. (1994) *Mol. Cell. Biol. 14*, 6954−6961.
55. Yun, B., Lee, K., Farkas, R., Hitte, C., and Rabinow, L. (2000) *Genetics 156*, 749−761.
56. Becker, W., Kentrup, H., Heukelbach, J., and Joost, H. G. (1996) *Biochim. Biophys. Acta 1312*, 63−67.
57. Ye, Q., and Worman, H. J. (1994) *J. Biol. Chem.* 11306−11311.

BI011521H