MOLECULAR AND CELLULAR BIOLOGY, Nov. 1996, p. 6486–6493 Vol. 16, No. 11
0270-7306/96/$04.00+0

A Structural Basis for Substrate Specificities of Protein Ser/Thr Kinases: Primary Sequence Preference of Casein Kinases I and II, NIMA, Phosphorylase Kinase, Calmodulin-Dependent Kinase II, CDK5, and Erk1

Z. SONGYANG,[1,2] KUN PING LU,[3] YOUNG T. KWON,[4] LI-HUEI TSAI,[4] ODILE FILHOL,[5] CLAUDE COCHET,[5] DEBRA A. BRICKEY,[6] THOMAS R. SODERLING,[6] CHERYL BARTLESON,[7] DONALD J. GRAVES,[7] ANTHONY J. DEMAGGIO,[8] MERL F. HOEKSTRA,[8] JOHN BLENIS,[2] TONY HUNTER,[3] AND LEWIS C. CANTLEY[1,2]*

Division of Signal Transduction, Beth Israel Hospital, Boston, Massachusetts 02215[1]; Molecular Biology and Virology Laboratory, The Salk Institute, La Jolla, California 92037[3]; Department of Cell Biology[2] and Department of Pathology,[4] Harvard Medical School, Boston, Massachusetts 02115; INSERM U244, DBMS/BRCE, CEN, Grenoble, France[5]; Vollum Institute for Advanced Biomedical Research, Oregon Health Sciences University, Portland, Oregon 97201[6]; Department of Biochemistry and Biophysics, Iowa State University, Ames, Iowa 50011[7]; and ICOS Corporation, Bothell, Washington 98021[8]

Received 11 April 1996/Returned for modification 16 May 1996/Accepted 15 July 1996

We have developed a method to study the primary sequence specificities of protein kinases by using an oriented degenerate peptide library. We report here the substrate specificities of eight protein Ser/Thr kinases. All of the kinases studied selected distinct optimal substrates. The identified substrate specificities of these kinases, together with known crystal structures of protein kinase A, CDK2, Erk2, twitchin, and casein kinase I, provide a structural basis for the substrate recognition of protein Ser/Thr kinases. In particular, the specific selection of amino acids at the +1 and −3 positions to the substrate serine/threonine can be rationalized on the basis of sequences of protein kinases. The identification of optimal peptide substrates of CDK5, casein kinases I and II, NIMA, calmodulin-dependent kinases, Erk1, and phosphorylase kinase makes it possible to predict the potential in vivo targets of these kinases.

The essential role of protein kinases in regulating signal transduction was established with the discovery of cyclic AMP-dependent protein kinase (PKA) (12). To respond to different extracellular stimuli, distinct groups of protein kinases have evolved. Each protein kinase is thought to phosphorylate a unique set of targets in the cell. The substrate specificities of protein kinases are therefore crucial for the fidelity of signaling events.

The classical approach for studying the specificity of a protein kinase is to compare the phosphorylation kinetics of synthetic peptides on the basis of known sequences phosphorylated by the kinase. This procedure is helpful in identifying the amino acids critical for efficient phosphorylation. However, it is not practical to synthesize and study each of the billions of possible variations of sequences that must be considered. Moreover, it is extremely difficult to apply this approach to study the specificity of a protein kinase with no known substrates. To overcome these problems, we developed a method for determining the primary sequence specificities of protein kinases by using an oriented degenerate peptide library (21). Optimal peptide substrates of a given protein kinase are identified by phosphorylation of a pool of degenerate peptides containing billions of different species. The specificities determined for PKA, CDC2, and CDK2 by using this technique were consistent with known substrates of these kinases. The results also allowed the prediction of in vivo kinase substrates. Synthetic peptides based on predicted optimal motifs were shown to act as low-K_m substrates for the kinases studied. Therefore, this method is a useful tool for studying substrate specificities of protein kinases.

We present here the specificities of eight additional protein Ser/Thr kinases: CDK5, casein kinase I (CKI) δ and γ, casein kinase II (CKII), NIMA, calmodulin-dependent (Cam) kinase II, Erk1, and phosphorylase kinase. Our findings demonstrate that each of these protein kinases has a distinct optimal peptide substrate. Critical determinants for recognition by the protein kinases were found at residues both N and C terminal to the phosphorylation site. The selectivities of these kinases were also rationalized on a structural basis. Finally, the optimal substrate sequences identified can be used to predict in vivo targets of these protein kinases.

MATERIALS AND METHODS

Peptide libraries. The following Ser-oriented degenerate peptide library was used (21) for CKI and -II, Cam kinases I and II, NIMA, and phosphorylase kinase: Met-Ala-Xxx-Xxx-Xxx-Xxx-Ser-Xxx-Xxx-Xxx-Xxx-Ala-Lys-Lys-Lys, where Xxx indicates all amino acids except Trp, Cys, Tyr, Ser, and Thr. For the proline-directed protein kinase CDK5 and Erk1, the following Ser-Pro-oriented library was used: Met-Ala-Xaa-Xaa-Xaa-Xaa-Ser-Pro-Xaa-Xaa-Xaa-Ala-Lys-Lys-Lys, where Xaa stands for all amino acids except Cys and Trp. For CKI δ and γ, a Ser-Ile-oriented library (Met-Ala-Xaa-Xaa-Xaa-Xaa-Xaa-Ser-Ile-Xaa-Xaa-Xaa-Xaa-Ala-Lys-Lys-Lys) was also used.

Kinase reaction. CKII, Cam kinase II and the γ subunit of phosphorylase kinase were purified by conventional liquid chromatography. Glutathione *S*-transferase (GST)-CKI, GST-Erk1, and GST-NIMA were purified by using glutathione-agarose beads. p35/CDK5 complex was purified by immunoprecipitation using anti-CDK5 antibodies. The peptide library was phosphorylated by the protein kinase of interest in a buffer containing 100 μM ATP with a trace of [γ-^{32}P] ATP (roughly 6×10^5 cpm), 1 mM dithiothreitol, 10 mM $MgCl_2$, and 50 mM Tris (pH 7.5) as previously described (21). For Cam kinase II, Ca^{2+} (1 mM) and calmodulin (1 μM) were also added during the kinase reaction. In a typical reaction, the protein kinases (GST-CKI, GST-Erk1, and CKII [~10 μg], NIMA [~1 μg],

* Corresponding author. Mailing address: Division of Signal Transduction, Beth Israel Hospital, 330 Brookline Ave., Boston, MA 02115. Phone: (617) 667-0947. Fax: (617) 667-0957.

TABLE 1. Substrate specificities of protein kinases

Ser kinase	Selectivity value										
	−5	−4	−3	−2	−1	0	+1	+2	+3	+4	+5
Arg/Lys directed											
Phosphorylase kinase (FRMMSFFLF)		F (1.7) M (1.5) K (1.4)	**R** (5.7) **K** (2.5)	**M** (2.0) R (1.8) Q (1.7) F (1.6)	**M** (2.2) F (2.0) L (1.8) I (1.5)	**S**	**F** (3.9) **I** (2.4) M (1.7) L (1.6)	**F** (2.5) **R** (2.0) K (2.0)	**L** (2.1) **I** (2.0)	**F** (2.3) L (1.8) I (1.7)	
Cam kinase II (KRQQSFDLF)		**K** (2.3) F (1.5)	**R** (3.9) **K** (2.7)	**Q** (4.5) M (1.8)	**Q** (2.1) **M** (2.0) K (1.9) L (1.8) F (1.5)	**S**	**F** (4.8) I (1.9) M (1.7) L (1.6) V (1.3)	**D** (2.5) E (1.7) I (1.5)	L (1.7) M (1.7) K (1.6) I (1.5)	F (1.7) K (1.4)	
Negative charge directed											
CKI δ (EFDTGSIIIFF)	E (1.8)	F (1.7) E (1.6)	**D** (5.7)	T (1.9) A (1.6) G (1.8)	G (1.8)	**S**	**I**	**I** (2.5) F (1.8) Y (1.8) G (1.7)	I (1.9) G (1.7) F (1.6)	**F** (2.1) G (1.6)	**F** (2.2) P (1.7) L (1.5)
CKI γ (YYDAASIIIFF)	Y (1.5)	Y (1.6) E (1.5)	**D** (2.7) **Y** (2.3)	A (3.0) D (2.0)	A (1.6) G (1.6)	**S**	**I**	**I** (2.8) Y (2.2) F (2.0)	I (1.9) F (1.6) Y (1.5)	**F** (2.6) Y (1.8)	**F** (2.1)
CKII (EDEESEDEE)		**E** (2.2) D (1.8) A (1.8)	**D** (3.2) **E** (2.2)	**E** (2.8) **D** (2.6)	**E** (2.7) **D** (2.0)	**S**	**E** (3.0) **D** (2.9) A (1.5)	**D** (3.4) **E** (2.3) A (1.6)	**E** (3.4) **D** (3.0)	**E** (4.1) **D** (2.1)	
Ser-Pro directed											
Erk1 (TGPLSPGPF)		**T** (2.0) P (1.5) S (1.4)	**G** (3.2) P (1.5) E (1.5) Y (1.5)	**P** (6.5) **L** (2.1) I (1.3)	**L** (2.0) M (1.9) P (1.7)	**S**	**P**	**G** (2.1) P (1.9) F (1.3)	**P** (2.4) F (1.9) G (1.5) Y (1.5)	**F** (2.3) Y (1.7) I (1.2)	
CDK5 (KHHKSPKHR)		**K** (1.8) H (1.6) G (1.5)	H (1.6)	**H** (2.5) **P** (2.1)	**K** (2.0) G (1.7) H (1.5)	**S**	**P**	**K** (2.9) **R** (2.9)	**H** (4.5) **R** (4.0) **K** (3.4)	**R** (3.2) **H** (3.2) **K** (2.5)	
Hydrophobic amino acid directed											
NIMA (RFRRSRRMI)		**R** (2.0) N (1.5)	**F** (7.3) **L** (2.3) M (1.7)	**R** (4.5) K (1.2)	**R** (4.6) K (1.6)	**S**	**R** (2.6) I (1.7) V (1.7) M (1.7)	**R** (2.0) **I** (2.0) M (1.9) V (1.6)	M (1.9) I (1.8) F (1.7) V (1.5)	**I** (2.1) F (1.7) M (1.7)	

[a] Amino acids with selectivity values higher than 1.5 or the best two at every degenerate position are shown. 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; S, Ser; T, Thr; V, Val; Y, Tyr. Boldface letters indicate amino acids which were most preferentially selected (value of >2.0).

Cam kinases I and II, and phosphorylase kinase γ subunit [~20 μg]) were added to 300 μl of solution containing 1 mg of degenerate peptide mixture. GST-Erk1 was activated by MEK in vitro before phosphorylation of the peptide library.

Phosphopeptide purification, sequencing, and data analysis. The phosphopeptides were separated from ATP by using a DEAE-Sephacel column and further purified from unphosphorylated peptides on an iron-chelating column (21). In this study, a relatively smaller iron column (300 μl instead of 1 ml as used previously [21]) was used such that unphosphorylated peptide contamination was significantly reduced. Peptide sequencing and data analysis were previously described (21).

RESULTS

Using the soluble oriented peptide library approach, we determined the optimal nonapeptide motifs for eight different protein Ser/Thr kinases. These enzymes were selected for study since they cover a broad spectrum of the known protein Ser/Thr kinases. The results are presented in Table 1. The selectivity values in parentheses indicate how strongly a particular amino acid is selected at a given position. A comparison of optimal substrate sequences of the protein Ser/Thr kinases studied (Table 1 and reference 21a) indicated that the −3 and +1 positions of the substrate play a major role in substrate recognition by most of the protein kinases. On the basis of these observations and previous studies by us and others, the protein Ser/Thr kinases studied can be classified in regard to their substrate specificity. A large number of protein Ser/Thr kinases (e.g., the Cam kinase family, PKC family, phosphorylase kinase γ, PKA, and SLK1) strongly prefer an Arg/Lys residue at the −3 position and are thus named Arg/Lys-directed protein kinases. In contrast, protein Ser/Thr kinases such as CKI and CKII select for substrates with acidic residues or phosphorylated amino acids at the −3 and other positions. Some protein Ser/Thr kinases are proline-directed kinases in that they select for substrates with Pro at the +1 position. These include mitogen-activated protein kinases and cyclin-dependent kinases. Finally, the protein kinase NIMA belongs to a unique family which preferentially selects a hydrophobic residue at the −3 position.

Arg/Lys-directed protein kinases. (i) Optimal substrates for phosphorylase kinase. Phosphorylase kinase regulates glycogen metabolism by phosphorylating phosphorylase. Although very few substrates have been reported for phosphorylase kinase, this kinase is an interesting enzyme which has the ability to phosphorylate serine, threonine, and, under certain conditions, tyrosine residues in vitro (24). We examined its specificity in phosphorylating serine-containing sequences by using the γ subunit (kinase domain) of this kinase. As shown in Table 1, this kinase preferred peptides with the motif Phe-Arg-Met-Met-Ser-Phe-Phe-Leu-Phe. This optimal peptide sequence is consistent with known sites phosphorylated by this kinase (Table 2).

TABLE 2. Comparison of optimal sequences determined by the peptide library with sequences at the phosphorylation sites of known protein substrates

Peptide or protein	Sequence[a]	Reference
Cam kinase II		
Consensus from library	K R Q Q S F **D** L F	
	F **K** M M I E M K	
Known sites		
Autophosphorylation	H **R** **Q** E T V **D** C L	17
Autophosphorylation	H **R** **Q** E T V E C L	17
Synapsin I	T **R** **Q** A S Q S G Q	17
Synapsin I	T **R** **Q** T S V S G Q	17
Tyrosine hydroxylase	G **R** R Q S L I E D	17
Calcineurin	A **R** V F S V L R E	17
Phospholamban	R **R** A S T I E M P	17
ATP-citrate lyase	S **R** T A S **F** S E S	17
Phosphorylase kinase		
Optimal sequence from library	F R M M S **F** **F** L F	
	M **K** R F I R I L	
Known substrate or site		
Phosphorylase	R **K** Q I S V R G L	17
Autophosphorylation	K **R** S G S I Y E P	17
CKI		
Optimal motif from library	E F D T G S I I I F F	
	Y E Y A A Y G Y P	
Known substrates		
αs2-Casein	S*S***S***E E S I I S Q E	17
β-Casein	S*S***S***E E S I T R I N	17
Glycogen synthase	T L **S*** V S S L P G L E	17
Phosphorylase kinase β	T K R S **G** S I Y E P L	17
CKII		
Optimal motif from library	E **D** **E** **E** S **E** **D** **E** E	
	D E **D** D **D** E **D** D	
Some known sites		
Nucleolar protein B23	E **D** A **E** S **E** D **E** D	17
Human a-Hsp90	E **D** V G S D E **E** **E**	17
PKA subunit R_{II}	A **D** S E S E D E E	17
Phosphatase inhibitor 2	E Q **E** S S G E **E** D	17
p35/CDK5		
Consensus from library	K H **H** **K** S P **K** H **R**	
	H **P** G **R** R **H**	
	K **K**	
Known sites		
Rat neurofilament	K E T **K** S P V **K** E	19
	V E V **K** S P E **K** A	19
Simian virus 40 large T antigen	A G **H** A T P P **K** K	19
Erk1		
Consensus from library	T **G** P L S P G P F	
	P P L M P F Y	
Known sites (Erk1 or Erk2)		
Myelin basic protein	V T **P** R T P P P S	17
TAL1	M V Q L S P P A L	3
Tyrosine hydrolase	E A V T S P R F I	9
Epidermal growth factor receptor	V E **P** L T P S G E	22
Myc	T P **P** L S P S R R	6
Jun	T P **P** L S P I D M	6
Oncoprotein 18	E L I L S P R S K	2
Caldesmon	D K V T S P T K V	1
Caldesmon	D G N K S P A P K	1
NIMA		
Optimal sequence from library	R F **R** **R** S **R** R M I	
	N K K I I I F	
Known site (phospholemman)	T **F** **R** S S I R R L	15

[a] Residues in boldface are strongly selected, and underlined residues are more strongly selected. S* indicates phosphoserine.

(ii) Substrate specificity of Cam kinase II. Like phosphorylase kinase, PKA, and PKC, Cam kinase II preferred Arg at the −3 position and hydrophobic amino acids at the +1 position (Table 1). Cam kinase I had a substrate specificity similar but not identical to that of Cam kinase II (data not shown). Both Cam kinase I and Cam kinase II had unusually strong selections for Phe at the +1 position compared with other kinases studied. However, Cam kinase I differed from Cam kinase II in that it did not strongly select for substrates with Gln at the −2 position or for acidic amino acids (Asp and Glu) at the +2 position (data not shown). Some known phosphorylation sites of Cam kinase II are presented in Table 2. They are in good agreement with the peptide library results.

Negative charge-directed protein kinases. (i) Substrate specificities of CKI γ and δ. Two different peptide libraries were used to investigate the specificities of CKI γ and δ. Using a Ser-oriented library with degenerate amino acids at four positions on either site of the fixed Ser, an apparent selection for Ile at +1 was observed for both of these enzymes (data not shown). To further examine substrate preference of CKI isomers, a Ser-Ile fixed peptide library was used. As shown in Table 1, the optimal substrates predicted for CKI γ and δ are nearly identical. Both kinases strongly select peptides with a negatively charged residue (Asp) at the −3 position. It has been reported that CKIs prefer to phosphorylate Ser or Thr residues in regions of proteins that have been previously phosphorylated (5). Although phosphorylated amino acids are excluded from libraries used in these studies, the strong selection for Asp at the −3 position suggests a preference for substrates with phosphorylated amino acids at this position. Interestingly, there was no selection for peptides with acidic residues C terminal of the fixed Ser, suggesting that these kinases are likely to phosphorylate residues C terminal rather than N terminal of previously phosphorylated sites. Some known phosphorylation sites of CKI are listed in Table 2.

(ii) Substrate specificity of CKII. In agreement with previous peptide phosphorylation studies, CKII was found to preferentially phosphorylate peptides enriched in Glu and Asp (Table 1). In fact, Glu and Asp were preferred over other amino acids at all eight positions of degeneracy investigated. There was some position specificity in that Asp was strongly selected over Glu at the −3 and +2 positions whereas Glu was preferred over Asp at the +3 and +4 positions. In addition, acidic amino acids were more strongly selected at positions C terminal of Ser than at positions N terminal. Like CKIs, there is evidence that CKII phosphorylates regions of proteins that have been previously phosphorylated (14). Our results predict that while CKIs are more likely to phosphorylate C terminal of a previously phosphorylated region, CKII is more likely to phosphorylate N terminal of a previously phosphorylated site and CKII is less specific about the position of the negatively charged residue. Some known sites phosphorylated by CKII are presented in Table 2.

Pro-directed kinases. (i) Substrate specificity of p35/CDK5. CDK5, a member of the cyclin-dependent kinases, preferentially phosphorylated a library containing a fixed Ser-Pro motif, indicating a strong preference for Pro at the +1 position. The optimal substrate predicted for p35/CDK5 (Lys-His-His-Lys-Ser-Pro-Lys-His-Arg) is in close agreement with the sites that it phosphorylates on neurofilament proteins (19) (Table 2). This motif is similar to those that we previously obtained for cyclin A/CDK2 and cyclin B/CDC2 (21). These results indicate that the three kinases may share some common in vivo targets and that discrimination among potential in vivo targets is largely controlled by contacts outside the catalytic cleft. Our attempts to model peptide substrates into the crystal structure

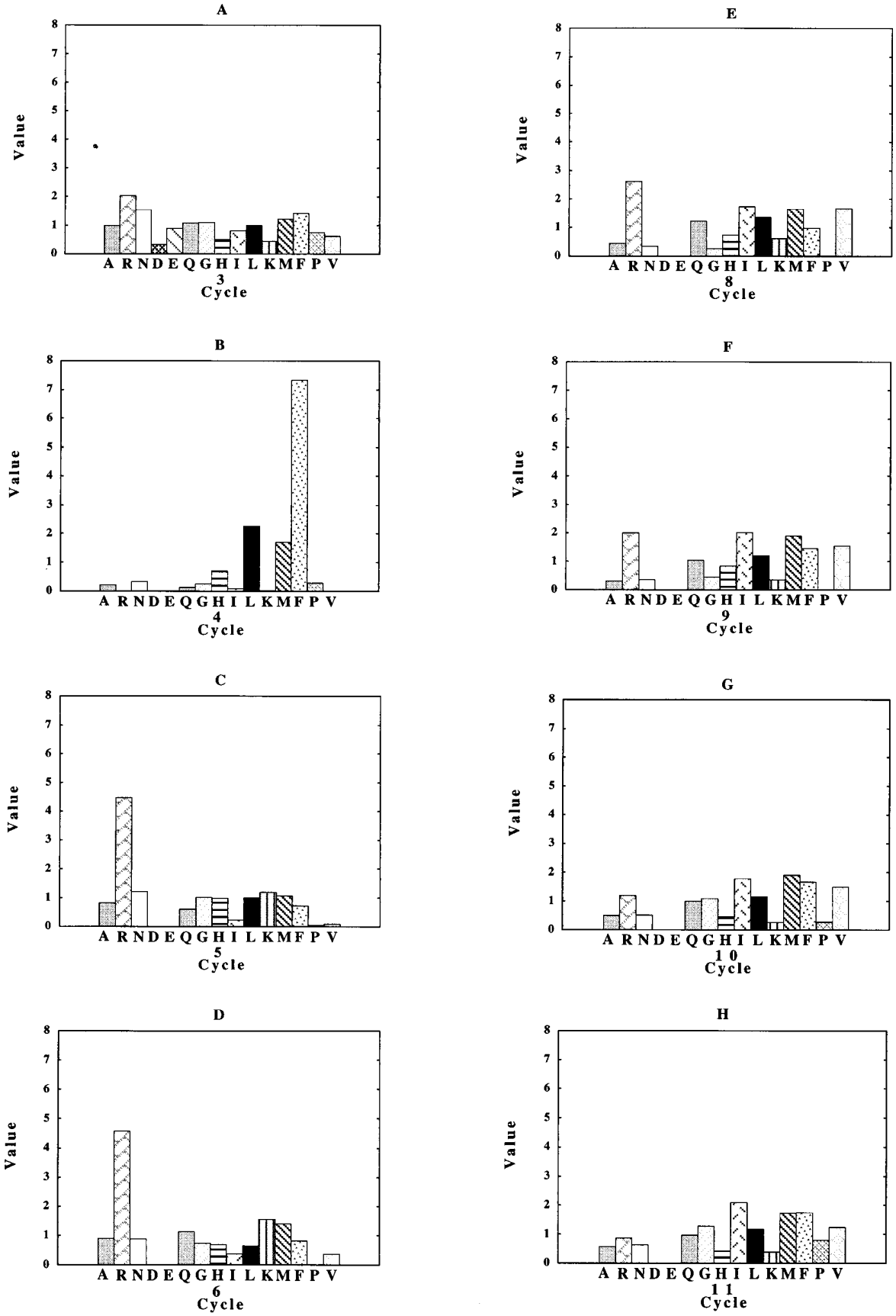

FIG. 1. Substrate specificity of protein kinase NIMA. Each panel indicates the relative abundance of the 15 amino acids at a given cycle of sequencing. Panels A to D indicate amino acid preferences at positions −4, −3, −2, and −1 N terminal of the phosphorylation site, and panels E to H indicate preferences at positions +1, +2, +3, and +4 C terminal of the phosphorylation site. 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; S, Ser; T, Thr; V, Val; Y, Tyr.

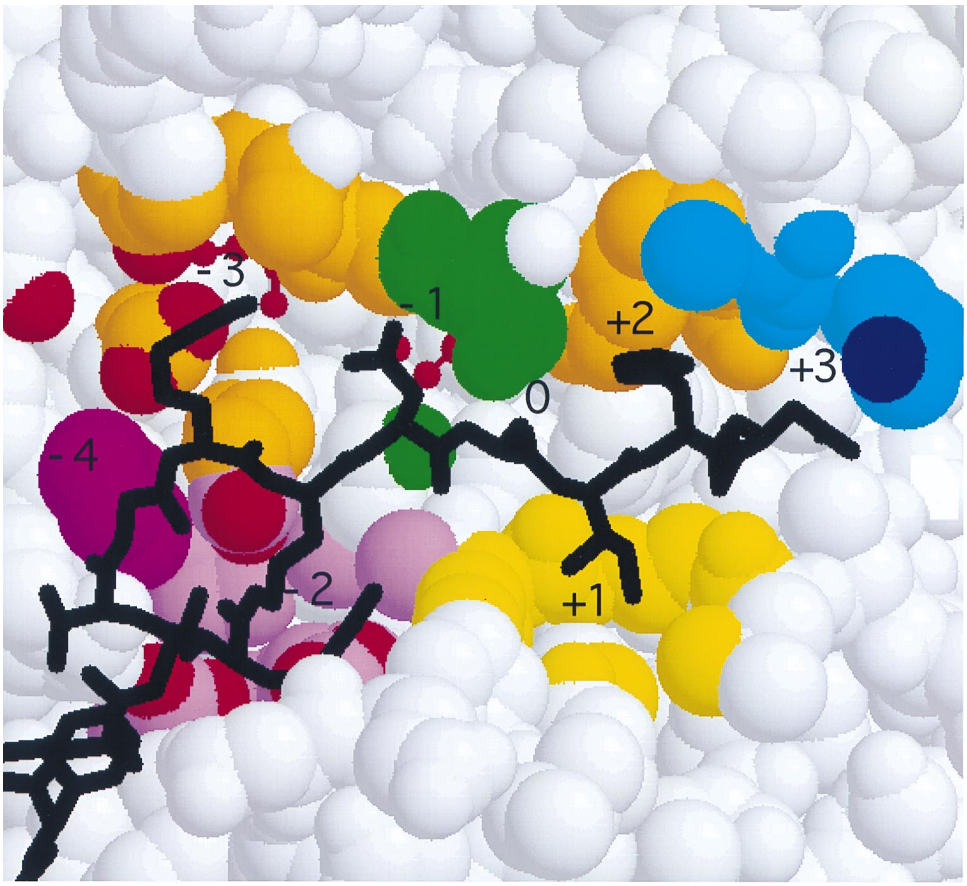

FIG. 2. Peptide substrate recognition by PKA. The catalytic cleft of PKA is presented with the pseudo-substrate, PKI (black) bound, on the basis of the PKI-PKA cocrystal (13). Those atoms in PKA that are proximal to the side chains of residues −4 to +3 of PKI [Gly-Arg-Arg-Asn-Ala(0)-Ile-His-Asn] are colored as follows: −4 pocket, violet; −3, orange; −2, pink; −1, green; +1 yellow; +2, orange; +3, cyan. The carboxylate oxygens of Glu and Asp residues in PKA that form the −4, −3, and −2 pockets are red, and an ε-nitrogen of Lys in PKA at the +3 pocket is blue.

of CDK2 (21) or the cyclin A/CDK2 (12) complex suggest that only amino acids that are more than four residues C terminal of the phosphorylated Ser are likely to directly contact the cyclin moiety. Libraries are being constructed to explore specificity at positions further from the catalytic cleft in order to determine the influence of the cyclin subunit in substrate selection.

(ii) Substrate specificity of Erk1. Like the cyclin-dependent kinases, Erk1 preferentially phosphorylated the Ser-Pro-oriented library. The optimal peptide substrate predicted for Erk1 was Thr-Gly-Pro-Leu-Ser-Pro-Gly-Pro-Phe (Table 1). Consistent with previous studies (6), peptides containing Pro at the −2 position were strongly preferred. Table 2 lists some sites known to be phosphorylated by Erk kinases. In close agreement with our prediction, five of the six known substrates have Leu at the −1 position. At the −2 position, hydrophobic amino acids, especially Pro, are frequently found. In fact, several known substrates of Erk1 have the motif PLS/TP (Table 2). Interestingly, the optimal motif is similar to consensus motifs that are known to bind SH3 domains in a proline-helix conformation.

Phe/hydrophobic amino acid-directed kinases. (i) Substrate specificity of NIMA. NIMA, a protein Ser/Thr kinase essential for G_2-M progression in *Aspergillus nidulans* (15, 16), was shown to have a unique optimal peptide substrate. Although genetic evidence exists for the importance of this kinase in cell growth regulation, in vivo substrates have not been identified. Using the peptide library, the optimal substrate is predicted to be Arg-Phe-Arg-Arg-Ser-Arg/hydrophobic-Arg/hydrophobic-hydrophobic-hydrophobic (Table 1). The human NIMA-like kinase (NLK1) had a similar substrate preference (not shown). The residue that dominates selectivity is the Phe at −3, which is distinct from the optimal motif of any previously characterized protein kinase. The strong selection for Phe at the −3 position is in agreement with in vitro substrates identified for NIMA (reference 15 and Table 2). The complete data for a selectivity experiment with this kinase are presented in Fig. 1 to elucidate the strong selection for Phe at the −3 position (cycle 4). To test the library data, we synthesized the optimal peptide (GRFRRSRRMI) for NIMA. This peptide is a good substrate for NIMA (K_m = 40.9 μM), with a higher V_{max}/K_m value than a peptide (GTFRSSIRRL) from previous best substrate, phospholemman (15).

DISCUSSION

On the basis of similar crystal structures observed for PKA, CDK2, Erk1, CKI, twitchin, and insulin receptor (4, 10, 11, 13, 23, 25) and homologous primary sequences (7, 8), it is likely that most of the eukaryotic protein kinases have similar tertiary structures. The structure of the PKA-PKI cocrystal established the first model for how a substrate binds to the active site of a protein kinase (13). The basis for the amino acids selected at specific positions N terminal and C terminal of the phosphor-

POSITION/SUBSTR	POCKET			-4	POCKET			-3	POCKET						-2	POCKET		-1
SUBDOMAIN	V				V	VI			VI	VIII			IX			I		
RESIDUE NO.	127	133	328		127	170	330		170	201	203	204	230	236		52	53	
ARG/LYS-DIRECTED KINASES																		
PKA/	E	R	D	ARG	E	E	Y	ARG	E	T	E	Y	E	P	ARG	G	S	ARG/ASN
PKC ALPHA/	D	Q	D	ALA*	D	D	F	ARG*	D	T	D	Y	E	P	LYS*	G	S	GLY*
PKC EPSILON/	D	Q	D	LYS*	D	D	D	ARG*	D	T	D	Y	E	P	GLN*	G	S	GLY*
CAM KIN2 ALPHA/	E	V	I	LYS	E	E	T	ARG	E	T	G	Y	I	P	GLN	G	A	GLN
PHK GAMMA/	E	T	D	PHE	E	E	Y	ARG	E	T	S	Y	T	P	MET/ARG	G	V	MET
PHE-DIRECTED KINASE																		
NIMA/	D	K	R	ARG	D	E	W	PHE	E	T	F	Y	E	P	ARG	G	S	ARG
PRO-DIRECTED KINASES																		
CDC2/	D	D	-	HIS	D	Q	-	HIS	Q	T	W	Y	E	P	HIS/PRO	G	T	LYS
CDK2/	Q	E	-	HIS	Q	Q	-	HIS	Q	T	W	Y	E	A	HIS/PRO	G	T	ARG
CDK5/	D	S	-	LYS	D	Q	-	HIS	Q	T	W	Y	E	P	HIS/PRO	G	T	LYS
ERK1/	D	L	?	THR	D	S	?	GLY	S	T	W	Y	E	P	PRO	G	A	LEU
GLU/ASP-DIRECTED KINASES																		
CKI GAMMA/	L	L	Q	TYR	L	E	R	ASP	E	T	R	Y	Y	L	ALA	G	N	ALA
CKI DELTA/	L	F	S	PHE	L	D	R	ASP	D	T	R	Y	Y	L	THR	G	S	GLY
CK2 ALPHA/	T	Y	?	GLU	T	H	?	ASP	H	S	Y	F	S	P	GLU	G	K	GLU

POSITION/SUBSTRATE	POCKET					+1	POCKET				+2	POCKET			+3
SUBDOMAIN	VII	VIII			X		I	II-III				II-III			
RESIDUE NO.	187	198	202	205	247		53	54	82	84		81	82	83	
ARG/LYS-DIRECTED KINASES															
PKA/	F	L	P	L	Y	ILE	S	F	L	Q	ILE	K	L	K	PHE/ASP
PKC ALPHA/	M	F	P	I	F	LEU*	S	F	D	D	ARG*	Q	D	D	GLN*
PKC EPSILON/	M	F	P	L	F	VAL*	S	F	D	D	ARG*	Q	D	D	ARG*
CAM KIN2 ALPHA/	L	F	P	L	Y	PHE	A	F	R	H	ASP	A	R	D	LEU/LYS
PHK GAMMA/	F	V	P	L	L	PHE	V	S	L	E	PHE/ARG	E	L	R	LEU
PHE-DIRECTED KINASE															
NIMA/	L	Y	P	M	V	ARG/ILE	S	F	T	E	ARG/ILE	S	T		MET
PRO-DIRECTED KINASES															
CDC2/	L	E	L	R	F	PRO	T	Y	E	V	ARG	S	E	G	ARG
CDK2/	L	E	L	R	I	PRO	T	Y	E	V	ARG	T	E	G	LYS
CDK5/	L	E	L	R	F	PRO	T	Y	E	V	ARG	D	E	G	HIS
ERK1/	L	Y	R	R	L	PRO	A	Y	Q	Y	GLY	H	Q	T	PRO
GLU/ASP-DIRECTED KINASES															
CKI GAMMA/	L	L	A	M	Y	ILE	N	F	Q	H	ILE	P	Q	L	ILE
CKI DELTA/	L	L	A	A	Y	ILE	S	F	Q	H	ILE	P	Q	L	ILE
CK2 ALPHA/	L	R	R	K	I	GLU	K	Y	P	K	ASP	K	P	V	GLU

FIG. 3. Protein kinase residues predicted to contact side chains of peptide substrates. Single-letter codes are used to indicate the residues of various protein kinases that are predicted to contact the side chains of the −4 to +3 residues of bound peptide substrates. The top row indicates the residues of PKA proximal to the −4 to +3 residues of PKI (Fig. 2). The arabic numbers indicate the positions of the residues in the linear sequence of PKA. The roman numerals designate the subdomain locations of the indicated residues (using the numbering system of Hanks et al. [8]). The analogous residues in the other protein kinases are based on sequence alignments (8) or, in some cases, on known crystal structures (see text). Three-letter codes are used to indicate the amino acids preferentially selected at the −4 to +3 positions for each of the kinases, using the oriented peptide library approach. The results for CDC2, CDK2, and PKA are from reference 21. For PKCα and PKCε, the sequences of pseudosubstrates, rather than optimal substrates, are included. Basic residues are blue, acidic residues are red, hydrophilic, uncharged residues are green, and hydrophobic residues (including Tyr) are black. A dash indicates that the kinase terminates prior to this residue. A question mark indicates uncertainty about alignment. An asterisk indicates a residue from an intrinsic pseudosubstrate rather than an optimal substrate.

ylation site can be rationalized from this structure. The PKA-PKI cocrystal is presented in Fig. 2. The atoms from PKA that are within 6 Å (0.6 nm) of the β-carbons of residues −4 through +3 of PKI are indicated in colors with different colors indicating distinct pockets for side chains.

The identities of the residues from PKA that form the various side chain pockets are summarized in the top row of Fig. 3. Figure 3 also indicates the subdomain locations of the residues, using the nomenclature of Hanks et al. (8). Single-letter codes indicate the residues that form the pockets, and three-letter codes indicate the optimal amino acid selected at each pocket by the oriented peptide library approach. The residues predicted to make up the −4 to +3 pockets of the various protein kinases in Fig. 3 were assigned on the basis of sequence alignments with the analogous positions in PKA or, in the cases of CDK2 (4, 12), CKI (23), and phosphorylase kinase (12a), information from crystal structures. The protein kinases have been divided into Arg/Lys-directed, Pro-directed, Phe-directed, and Glu/Asp-directed on the basis of the selectivities observed at the −3 or +1 positions, as discussed in Results.

A quick scan of the colors in Fig. 3 reveals some general trends. Those pockets with multiple acidic residues (red) usually select for basic (blue) amino acids (Arg/Lys/His). This is particularly clear for the −4, −3, and −2 pockets of the Arg/Lys-directed kinases. The cyclin-dependent kinases are also acidic (although less so) in the −4, −3, and −2 pockets and weakly selected for substrates with His at these positions. In contrast, the casein kinases have basic residues in the −3 pocket and selected for Asp at this position.

The explanation for the strong selectivity of NIMA for substrates with Phe at −3 is not completely clear. The predicted −3 pocket, like that of PKA, has two acidic residues. A possible explanation for the selectivity is the presence of a Trp residue at the analogous position of Tyr-330 of PKA in this region (Fig. 3). The Trp may provide the opportunity for aromatic ring interactions to stabilize Phe binding to this region.

Similarities between NIMA and PKA in the regions that bind Arg at −2 have been previously discussed (15).

Most kinases investigated showed relatively low selectivity at the −1 position (Table 1). Side chains at the −1 position are predicted to interact with residues in subdomain I (e.g., residues 52 and 53 of PKA [Fig. 3]). Most kinases have small hydrophilic, uncharged residues at these positions. Interestingly, CKII has a Lys in this region and selects for Glu at −1, and phosphorylase kinase has a Val and selects for Met (Fig. 3). Some of the kinases selected for Arg/Lys at the −1 position, probably as a result of electrostatic attractions from residues in the nearby −3 pocket (Fig. 1).

As we previously discussed (21), the +1 pocket is very important for selectivity as well as for orienting the adjacent Ser/Thr residue for phosphorylation. This pocket is well separated from the other pockets and is entirely composed of residues between subdomains VII and VIII, near the C-terminal end of the T loop (yellow in Fig. 2). An examination of Fig. 3 reveals that most of the kinases that we have studied have very hydrophobic +1 pockets, and these kinases select for substrates with hydrophobic amino acids at +1. We previously discussed the prediction that the presence of an Arg residue at the base of this pocket (analogous to the Leu-205 position of PKA) results in selectivity for substrates with Pro at the +1 position (21). This prediction is based on modeling studies showing that the Arg side chain fills up the pocket and excludes residues with large side chains (or water molecules). Pro may also be selected at this position because it is the only naturally occurring amino acid that is a secondary amine. Thus, peptides with Pro at +1 can bind to this site without the necessity of dissociating a hydrogen-bonded water molecule from the amide nitrogen of the peptide (21). All Pro-directed protein kinases that we have investigated to date have Arg at this position. However, since several other residues are necessary to form the pocket, it is not clear that an Arg at this position alone is sufficient to predict selectivity for Pro at +1.

CKII is the only kinase that we have investigated that has a Lys residue at the base of the +1 pocket (Fig. 3). Consistent with this pocket being basic, CKII selects substrates with Glu at the +1 position.

A few of the kinases that we have investigated have strong selectivities at the +2 and/or +3 positions. For example, the cyclin-dependent kinases strongly select for substrates with basic residues at these positions (Table 1). Consistent with these selectivities, these kinases have acidic residues in the +2 and +3 pockets. Certain PKC family members have been shown to prefer substrates with basic residues at the +2 and +3 positions, and these kinases also have acidic pockets to account for these selectivities (Fig. 3). In contrast, CKII has basic residues in the +2 and +3 pockets and selects for substrates with Asp and Glu at these positions. Thus, while the presentation in Fig. 3 cannot explain every selectivity that we observed, it certainly provides explanations for most of the strong selectivities observed in Table 1.

In summary, we have determined the primary sequence specificities of eight protein Ser/Thr kinases by using an oriented peptide library technique. The results are in good agreement with known substrates of these kinases and should allow predictions of novel phosphorylation sites from primary sequences. On the basis of these results and alignments of these kinases with PKA, we have proposed a structural basis for substrate selectivities of these and other protein kinases. These predictions will guide future mutational studies to test the proposed models.

ACKNOWLEDGMENTS

We thank Michael Berne at Tufts University for peptide sequencing and Louise Johnson for providing crystal coordinates of phosphorylase kinase.

This work was supported by grants from American Cancer Society and Lucille P. Markey Charitable Trust.

REFERENCES

1. **Adam, L. P., and D. R. Hathaway.** 1993. Identification of mitogen-activated protein kinase phosphorylation sequences in mammalian h-caldesmon. FEBS Lett. **322:**56–60.
2. **Brattsand, G., U. Marklund, K. Nylander, G. Roos, and M. Gullberg.** 1994. Cell-cycle-regulated phosphorylation of oncoprotein 18 on Ser16, Ser25 and Ser38. Eur. J. Biochem. **220:**359–368.
3. **Cheng, J. T., M. H. Cobb, and R. Baer.** 1993. Phosphorylation of the TAL1 oncoprotein by the extracellular-signal-regulated protein kinase ERK1. Mol. Cell. Biol. **13:**801–808.
4. **De Bondt, H. L., J. Rosenblatt, J. Jancarik, H. D. Jones, D. O. Morgan, and S. H. Kim.** 1993. Crystal structure of cyclin-dependent kinase 2. Nature (London) **363:**595–602.
5. **Flotow, H., P. R. Graves, A. Q. Wang, C. J. Fiol, R. W. Roeske, and P. J. Roach.** 1990. Phosphate groups as substrate determinants for casein kinase I action. J. Biol. Chem. **265:**14264–14269.
6. **Gonzalez, F. A., D. L. Raden, and R. L. Davis.** 1991. Identification of substrate recognition determinants for human ERK1 and ERK2 protein kinases. J. Biol. Chem. **266:**22159–22163.
7. **Hanks, S. K., and A. M. Quinn.** 1991. Protein kinase catalytic domain sequence database: uses in identification of conserved features of primary structure and classification of novel family members. Methods Enzymol. **200:**38–62.
8. **Hanks, S. K., A. M. Quinn, and T. Hunter.** 1988. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science **241:**42–52.
9. **Haycock, J. W., N. G. Ahn, M. H. Cobb, and E. G. Krebs.** 1992. ERK1 and ERK2, two microtubule-associated protein 2 kinases, mediate the phosphorylation of tyrosine hydroxylase at serine-31 in situ. Proc. Natl. Acad. Sci. USA **89:**2365–2369.
10. **Hu, S. H., M. W. Parker, J. Y. Lei, M. C. Wilce, G. M. Benian, and B. E. Kemp.** 1994. Insights into autoregulation from the crystal structure of twitchin kinase. Nature (London) **369:**581–584.
11. **Hubbard, S. R., L. Wei, L. Ellis, and W. A. Hendrickson.** 1994. Crystal structure of the tyrosine kinase domain of the human insulin receptor. Nature (London) **372:**746–754.
12. **Jeffrey, P. D., A. A. Russo, K. Polyak, E. Gibbs, J. Hurwitz, J. Massague, and N. P. Pavletich.** 1995. Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. Nature (London) **376:**313–320.

12a. **Johnson, L.** Unpublished data.

13. **Knighton, D. R., J. H. Zheng, L. F. Ten Eyck, N. H. Xuong, S. S. Taylor, and J. M. Sowadski.** 1991. Structure of a peptide inhibitor bound to the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. Science **253:**414–420.
14. **Kuenzel, E. A., J. A. Mulligan, J. Sommercorn, and E. G. Krebs.** 1987. Substrate specificity determinants for casein kinase II as deduced from studies with synthetic peptides. J. Biol. Chem. **262:**9136–9140.
15. **Lu, K. P., B. E. Kemp, and A. R. Means.** 1994. Identification of substrate specificity determinants for the cell cycle-regulated NIMA protein kinase. J. Biol. Chem. **269:**6603–6607.
16. **Osmani, A. H., S. L. McGuire, and S. A. Osmani.** 1991. Parallel activation of the NIMA and p34cdc2 cell cycle-regulated protein kinases is required to initiate mitosis in A. nidulans. Cell **67:**283–291.
17. **Pearson, R. B., and B. E. Kemp.** 1991. Protein kinase phosphorylation site sequences and consensus specificity motifs: tabulations. Methods Enzymol. **200:**62–81.
18. **Resing, K. A., S. J. Mansour, A. S. Hermann, R. S. Johnson, J. M. Candia, K. Fukasawa, G. F. Vande Woude, and N. G. Ahn.** 1995. Determination of v-Mos-catalyzed phosphorylation sites and autophosphorylation sites on MAP kinase kinase by ESI/MS. Biochemistry **34:**2610–2620.
19. **Shetty, K. T., W. T. Link, and H. C. Pant.** 1993. cdc2-like kinase from rat spinal cord specifically phosphorylates KSPXK motifs in neurofilament proteins: isolation and characterization. Proc. Natl. Acad. Sci. USA **90:**6844–6848.
20. **Solomon, M. J., J. W. Harper, and J. Shuttleworth.** 1993. CAK, the p34cdc2 activating kinase, contains a protein identical or closely related to p40MO15. EMBO J. **12:**3133–3142.
21. **Songyang, Z., S. Blechner, N. Hoagland, M. F. Hoekstra, W. H. Piwnica, and L. C. Cantley.** 1994. Use of an oriented peptide library to determine the optimal substrates of protein kinases. Curr. Biol. **4:**973–982.

21a. **Songyang, Z., and L. C. Cantley.** Unpublished data.

22. **Williams, R., J. Sanghera, F. Wu, D. Carbonaro-Hall, D. L. Campbell, D. Warburton, S. Pelech, and F. Hall.** 1993. Identification of a human epider-

mal growth factor receptor-associated protein kinase as a new member of the mitogen-activated protein kinase/extracellular signal-regulated protein kinase family. J. Biol. Chem. **268:**18213–18217.

23. **Xu, R. M., G. Carmel, R. M. Sweet, J. Kuret, and X. Cheng.** 1995. Crystal structure of casein kinase-1, a phosphate-directed protein kinase. EMBO J. **14:**1015–1023.
24. **Yuan, C. J., C. Y. Huang, and D. J. Graves.** 1993. Phosphorylase kinase, a metal ion-dependent dual specificity kinase. J. Biol. Chem. **268:**17683–17686.
25. **Zhang, F. M., A. Strand, D. Robbins, M. H. Cobb, and E. J. Goldsmith.** 1994. Atomic structure of the MAP kinase ERK2 at 2.3 Å resolution. Nature (London) **367:**704–711.