

# Deciphering Protein Kinase Specificity Through Large-Scale Analysis of Yeast Phosphorylation Site Motifs

Janine Mok,<sup>1\*</sup> Philip M. Kim,<sup>2†</sup> Hugo Y. K. Lam,<sup>3</sup> Stacy Piccirillo,<sup>1</sup> Xiuqiong Zhou,<sup>1</sup> Grace R. Jeschke,<sup>4</sup> Douglas L. Sheridan,<sup>4‡</sup> Sirlester A. Parker,<sup>4</sup> Ved Desai,<sup>4</sup> Miri Jwa,<sup>5</sup> Elisabetta Cameroni,<sup>6§</sup> Hengyao Niu,<sup>7</sup> Matthew Good,<sup>8</sup> Attila Remenyi,<sup>8||</sup> Jia-Lin Nianhan Ma,<sup>9</sup> Yi-Jun Sheu,<sup>10</sup> Holly E. Sassi,<sup>11</sup> Richelle Sopko,<sup>11</sup> Clarence S. M. Chan,<sup>5</sup> Claudio De Virgilio,<sup>6</sup> Nancy M. Hollingsworth,<sup>7</sup> Wendell A. Lim,<sup>12</sup> David F. Stern,<sup>9</sup> Bruce Stillman,<sup>10</sup> Brenda J. Andrews,<sup>11</sup> Mark B. Gerstein,<sup>2,3</sup> Michael Snyder,<sup>1,2||#</sup> Benjamin E. Turk<sup>4#</sup>

Phosphorylation is a universal mechanism for regulating cell behavior in eukaryotes. Although protein kinases target short linear sequence motifs on their substrates, the rules for kinase substrate recognition are not completely understood. We used a rapid peptide screening approach to determine consensus phosphorylation site motifs targeted by 61 of the 122 kinases in *Saccharomyces cerevisiae*. By correlating these motifs with kinase primary sequence, we uncovered previously unappreciated rules for determining specificity within the kinase family, including a residue determining P-3 arginine specificity among members of the CMGC [CDK (cyclin-dependent kinase), MAPK (mitogen-activated protein kinase), GSK (glycogen synthase kinase), and CDK-like] group of kinases. Furthermore, computational scanning of the yeast proteome enabled the prediction of thousands of new kinase-substrate relationships. We experimentally verified several candidate substrates of the Prk1 family of kinases in vitro and in vivo and identified a protein substrate of the kinase Vhs1. Together, these results elucidate how kinase catalytic domains recognize their phosphorylation targets and suggest general avenues for the identification of previously unknown kinase substrates across eukaryotes.

## INTRODUCTION

As one of the most widespread posttranslational modifications, protein phosphorylation is involved in virtually every basic cellular process, including DNA replication, gene transcription, protein translation, cell growth and metabolism, differentiation, and intercellular communication. With the advent of whole genome sequencing, the entire complement of kinases, or "kinome," for multiple organisms has been cataloged, revealing that most eukaryotes devote ~2% of their protein-coding capacity to these enzymes (1). Unraveling the function of each member of such a large family remains a challenge. Advances in phosphoproteomic methodologies, such as large-scale mass spectrometry (MS)-based phosphorylation site discovery, targeted small interfering RNA screens, the use of analog-sensitive kinase alleles that are engineered to accept specific inhibitors and adenosine triphosphate (ATP) analogs, and protein microarray analyses, have shed considerable light on the scope and complexity of phosphorylation-based signal transduction pathways in eukaryotes (2–5).

However, one aspect of protein kinase biology that remains poorly understood is how kinases achieve specificity for their target substrates. Understanding rules for substrate recognition by kinases has important applications in the mapping of phosphorylation sites in protein substrates, discovery of previously unknown substrates, and production of model substrates for small-molecule inhibitor screening (6). In addition, a detailed understanding of how kinases interact with their substrates enables both deciphering and genetic rewiring of kinase specificity, thereby uncovering fundamental ways in which signaling pathways are organized and propagated (7, 8).

In a typical eukaryotic cell, there are hundreds of thousands of Ser, Thr, and Tyr residues among the thousands of proteins. To ensure signaling fi-

delity, kinases must somehow discriminate among these vast numbers of potential phosphorylation sites. Mechanisms that influence substrate selection by a protein kinase include subcellular localization, substrate docking

<sup>1</sup>Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06520, USA. <sup>2</sup>Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA. <sup>3</sup>Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT 06520, USA. <sup>4</sup>Department of Pharmacology, Yale University School of Medicine, New Haven, CT 06520, USA. <sup>5</sup>Institute for Cellular and Molecular Biology, University of Texas, Austin, TX 78712, USA. <sup>6</sup>Department of Medicine, Division of Biochemistry, University of Fribourg, CH-1700 Fribourg, Switzerland. <sup>7</sup>Department of Biochemistry and Cell Biology, Stony Brook University, Stony Brook, NY 11794, USA. <sup>8</sup>Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA 94143, USA. <sup>9</sup>Department of Pathology, School of Medicine, Yale University, New Haven, CT 06520, USA. <sup>10</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA. <sup>11</sup>Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada M5S 3E1. <sup>12</sup>Eötvös Loránd University, Department of Biochemistry and Howard Hughes Medical Institute, Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary. \*Present address: Stanford Genome Technology Center, Department of Biochemistry, Stanford University, Palo Alto, CA 94304, USA.

†Present address: Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada M5S 3E1.

‡Present address: Alexion Pharmaceuticals Inc., 352 Knotter Drive, Cheshire, CT 06410, USA.

§Present address: Institute for Research in Biomedicine, CH-6500 Bellinzona, Switzerland.

||Present address: Eötvös Loránd University, Department of Biochemistry, Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary.

||Present address: Department of Genetics, Stanford University, Palo Alto, CA 94305, USA.

#To whom correspondence should be addressed. E-mail: ben.turk@yale.edu (B.E.T.); mpsnyder@stanford.edu (M.S.).

interactions, and binding to scaffold proteins (9). An important aspect of substrate recognition, however, is that the phosphorylation site on the substrate falls within a consensus amino acid sequence that is complementary to the active site of the kinase.

Consensus phosphorylation site motifs for protein kinases have been previously established on an individual basis through either the inspection of known phosphorylation sites, the systematic mutagenesis of protein and peptide substrates, or the screening of peptide libraries (10, 11). Although these studies have provided valuable insight into substrate recognition, such data are only available for a subset of known protein kinases. NetPhorest, which is the most comprehensive repository for kinase phosphorylation site motifs reported to date, includes motifs for only 35% of all human kinases (12). The incompleteness of available data and the heterogeneity of data collection methods limits their application to elucidating cellular signaling pathways and modeling larger phosphorylation networks. For example, the use of motif scanning approaches to link specific kinases to the thousands of *in vivo* phosphorylation sites discovered through MS-based phosphoproteomics has proven difficult in targeted kinase studies because multiple kinases can potentially target the same or similar motifs.

We thus set out to catalog consensus phosphorylation site motifs for the kinome of the model organism *Saccharomyces cerevisiae*. We adapted a peptide library screening approach (13) to a miniaturized format that would enable rapid analysis of large numbers of kinases. With this method, we determined consensus phosphorylation motifs targeted by 61 of the 122 yeast kinases. This large collection of phosphorylation site motifs provided new insight into the structural basis for substrate recognition by protein kinases as a family in a manner not possible through analyses of individual kinases. Furthermore, we used our motif collection to predict new kinase-substrate relationships through database scanning and integration with other yeast proteomic and genomic data sets.

## RESULTS

### A rapid peptide-based approach for the high-throughput determination of kinase consensus phosphorylation site motifs

To determine phosphorylation motifs for yeast protein kinases, we developed a high-throughput approach using our previously reported positional scanning peptide library (13). This library consisted of 200 distinct peptide mixtures in which each 16-mer peptide contained a central fixed phosphorylation acceptor (phosphoacceptor) site (an equimolar mixture of Ser and Thr) flanked by degenerate positions consisting of equimolar mixtures of the 20 amino acids excluding Ser, Thr, and Cys, and a C-terminal biotin tag (Fig. 1A). For each of the nine positions surrounding the phosphoacceptor site, there were 22 peptide mixtures in which each of the 20 unmodified amino acids, as well as phosphothreonine (pT) and phosphotyrosine (pY), were fixed. In addition to these 198 ( $9 \times 22$ ) peptide mixtures, two control peptide mixtures bearing either Ser or Thr alone as the fixed phosphoacceptor residue in the context of a fully degenerate sequence were also included. These control mixtures served as indicators of any preference the kinase had for either Ser or Thr residues at the phosphoacceptor site. Peptides were incubated with the kinase of interest in the presence of radiolabeled ATP. At the end of the incubation period, aliquots of each reaction were spotted simultaneously with a capillary pin-based liquid transfer device onto a streptavidin-coated membrane that captured the peptide substrates through their C-terminal biotin tags. After extensive washing, the membrane was dried and exposed to a phos-

phor screen, allowing the extent of radiolabel incorporation for each peptide to be visualized and quantified. To enable high-throughput analysis, all steps were performed in a 1536-well format, thereby reducing the amount of kinase and peptide required and enabling simultaneous analysis of four kinases.

Three yeast kinases (Tpk1, Tpk2, and Ste20) were assayed with both the miniaturized and the large-volume formats, and we performed multiple replicates with one of these kinases, Tpk1. Identical results were observed with the two formats and in replicate assays with the 1536-well format (data for Tpk1 are shown in fig. S1). These kinases also recapitulated preferences of their mammalian orthologs for basic residues upstream of the phosphorylation site (13, 14). These results confirm that the miniaturized peptide library screening system is reproducible and provides data that are quantitatively equivalent to lower-throughput approaches.

### Screening yeast kinases for their consensus phosphorylation site motifs

With our peptide array method, we screened 111 of the 122 yeast kinases. Kinases were initially purified from yeast strains that harbor galactose-inducible expression plasmids bearing either a C-terminal tandem affinity purification tag or an N-terminal glutathione *S*-transferase (GST) tag (15, 16). In a number of instances, it was necessary to perform the assay in the presence of known activating subunits [(for example, cyclins for cyclin-dependent kinases (CDKs)], phosphorylate the kinase *in vitro* or coexpress it with an activating kinase, or purify the kinase from yeast grown under activating conditions. For kinases with which poor yields were obtained from yeast, we used alternative bacterial and mammalian cell expression systems. Each kinase was assayed on the peptide substrates in duplicate on separate days. In total, we generated reproducible phosphorylation motifs for 61 of the 111 yeast kinases screened (Fig. 1B and table S1). Three distinct motifs were generated for the CDK Pho85 by analyzing its activity separately in complex with different cyclin subunits (Pho80, Pcl1, and Pcl2). The remaining kinases were not sufficiently active to phosphorylate the peptides above background levels. These kinases may be highly specific for particular protein substrates and thus do not phosphorylate peptides efficiently. For example, in keeping with previous observations for their mammalian orthologs (17), we did not observe activity on our peptide substrates for the eight kinases in the mitogen-activated protein kinase (MAPK) kinase (MAPKK) and MAPK kinase kinase (MAPKKK) families. Other kinases were likely simply inactive under exponential growth conditions or when assayed in the absence of obligate binding partners and may be suitable for analysis once their activation mechanisms are more completely understood.

Approximately half of the phosphorylation site motifs that we determined for yeast kinases were identical to known motifs, because they corresponded to yeast homologs of mammalian kinases that have been previously characterized (11, 12). In contrast, the remaining kinases and their mammalian homologs have either not been previously characterized (table S2 lists mammalian homologs and indicates which kinases have previously known motifs) or in one instance (Tos3) yielded a motif different from that reported. Representative spot arrays produced by four kinases for which phosphorylation motifs were not previously known (Atg1, Gin4, Mps1, and Prk1) are shown in Fig. 1B. Spot intensities from the peptide arrays were quantified, background-corrected, and normalized to provide the selectivity values shown in Table 1. We verified the consensus phosphorylation motifs for these kinases by performing kinase assays with optimized peptide substrates (named ATGtide, GINTide, MPSTide, and PRKTide, respectively) consisting of those residues that were most highly selected at each position. As shown in Fig. 1C, each kinase was highly specific for its corresponding peptide substrate, thus provid-

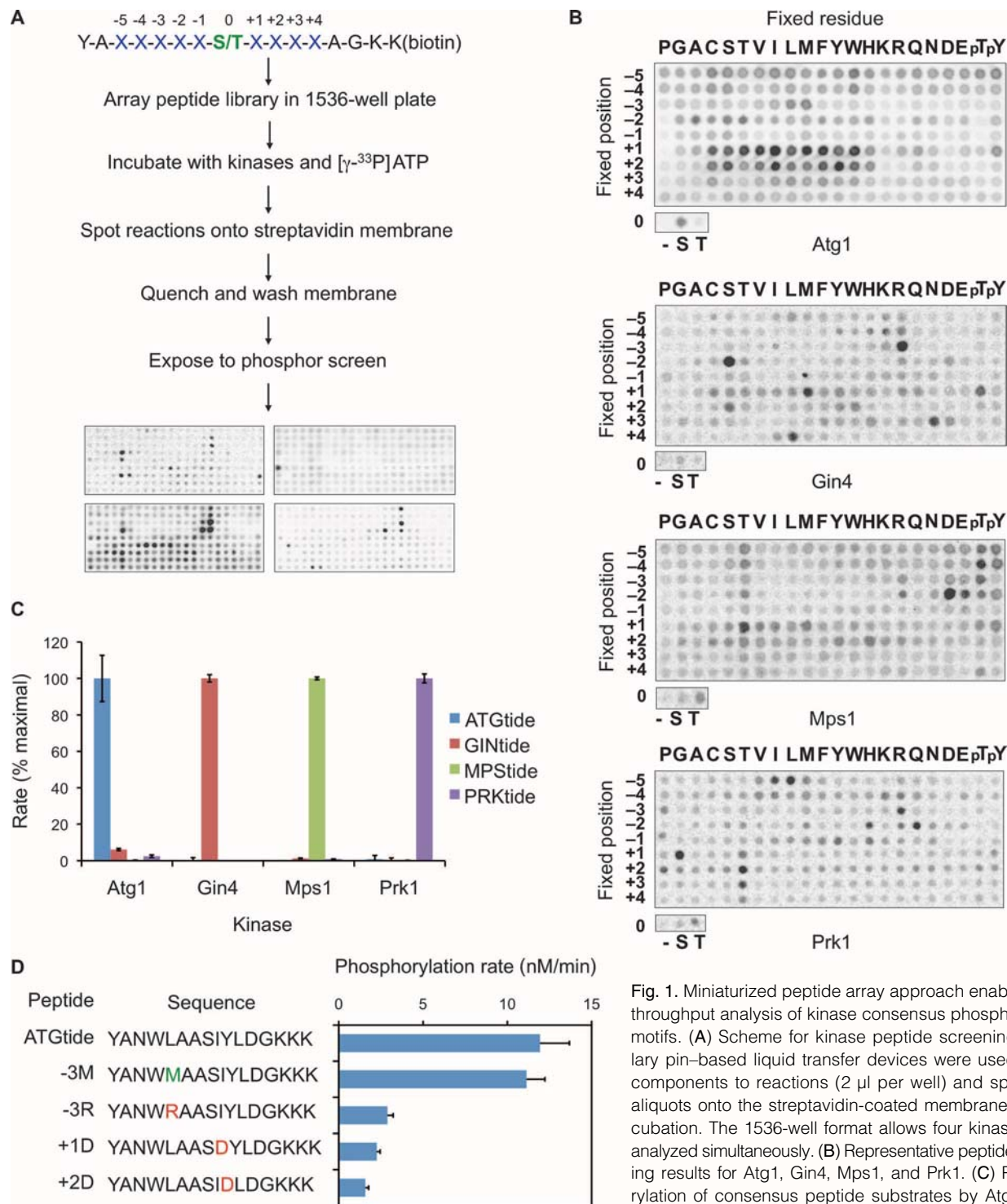


Fig. 1. Miniaturized peptide array approach enables high-throughput analysis of kinase consensus phosphorylation motifs. (A) Scheme for kinase peptide screening. Capillary pin-based liquid transfer devices were used to add components to reactions (2  $\mu$ l per well) and spot 0.2- $\mu$ l aliquots onto the streptavidin-coated membrane after incubation. The 1536-well format allows four kinases to be analyzed simultaneously. (B) Representative peptide screening results for Atg1, Gin4, Mps1, and Prk1. (C) Phosphorylation of consensus peptide substrates by Atg1, Gin4, Mps1, and Prk1. The sequence of each peptide is as follows:

ATGtide, YANWLAASIYLDGKKK; GINTide, YALRRSRSMWNLGKKK; MPStide, YADHDDDTMHFRGKKK; and PRKtide, YALKPQYTGPRGKKK. Abbreviations for the amino acids are as follows: A, Ala; C, Cys; 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; W, Trp; and Y, Tyr. Peptide phosphorylation was assayed at 10  $\mu$ M concentration by radiolabel kinase assay. Incorporation of radiolabeled phosphate into peptides was determined by phosphocellulose filter binding assay. Maximal rates for each kinase in these assays were as follows: Atg1, 250 nM/min, Gin4, 510 nM/min, Mps1, 130 nM/min, Prk1, 330 nM/min. (D) Rates of Atg1 phosphorylation of ATGtide variants with individual point substitutions. Peptide phosphorylation was assayed as for (C).

ing independent validation of our mixture-based peptide library screening approach.

Notably, the autophagy-linked kinase Atg1 had an atypical motif exhibiting selections for hydrophobic residues at multiple positions. We verified this motif by making targeted substitutions to the ATGtide substrate. As anticipated, substituting a different favorable hydrophobic residue (Met) at the most selective position (P-3) had no significant effect on the rate of ATGtide phosphorylation. Moreover, substituting unfavorable charged residues at any of three most strongly selective positions dramatically reduced the reaction rate (Fig. 1D).

### Overall features of kinase phosphorylation signatures

Normalized, background-corrected phosphorylation signals for each kinase were assembled into position weight matrices (PWMs), which are quantitative representations of the phosphorylation motif. We scored each position for its total selectivity, and a specificity heat map of all kinases and positions revealed the wide range of selectivity exhibited by kinases (Fig. 2). At one extreme, Yck1 and Cka1 (yeast casein kinase 1 and casein kinase 2 homologs) were highly sequence-specific, with requirements for particular amino acids at multiple positions. At the other extreme, Cak1

and Rad53 were the least selective in that, although the extent of substrate phosphorylation by these kinases is clearly dependent on peptide sequence, there were no residues that were required at any position surrounding the phosphoacceptor. Most kinases fell between these extremes, with a combination of required residues and more subtle propensities that influence the overall efficiency of phosphorylation. Furthermore, although selectivity was noted at each position surrounding the phosphorylation site, kinases were most frequently selective at the P-3 position, followed by the P-2 and P+1 positions. By contrast, few kinases were selective at the P-1 position.

The 61 yeast kinases were clustered into groups on the basis of phosphorylation site selectivity (Fig. 3). Thirty-five kinases were observed to target “basophilic” motifs. Thirty-one of these showed a classic basophilic signature (10), with a strong selectivity primarily for an Arg residue at the P-3 position. This was the single most common feature found among all motifs (Fig. 3 and table S1). Four other basophilic kinases, Ipl1, Skm1, Ste20, and Cla4, were selective for Arg at the P-2 position but did not show strong selectivity for Arg at the P-3 position (Fig. 3 and table S1). The basophilic kinases, however, diverged with respect to the residues selected at other positions. For example, basophilic kinases are

**Table 1.** Quantified selectivity values for protein kinases discussed in the text. Peptide array data were quantified and normalized to an average value of 1 within a position. Positively selected residues with

values greater than 1.5 are shown. Complete quantified PWMs for all kinases are provided as data set S1.

Kinase	Position								
	-5	-4	-3	-2	-1	+1	+2	+3	+4
Atg1		W (1.8)	L (2.8) M (2.4)	A (1.6)		V (2.4) I (2.3) F (1.9) M (1.9)	Y (2.8) I (2.3) F (2.1) M (1.8) W (1.6) V (1.6)	L (1.7)	D (1.8)
Gin4	L (2.0) I (1.8) M (1.7) R (1.6) K (1.6)	R (2.7) K (2.4) H (1.8)	R (9.9) K (1.6)	S (7.9) T (1.8) R (1.6)	M (2.9) K (1.8) P (1.8)	M (2.6)	W (2.4) Y (2.2)	N (2.8) D (1.8) F (1.6)	L (3.9) I (2.3)
Mps1			D (1.9) R (1.9)	D (4.5) E (2.4) R (2.2)	D (1.6)	M (1.9) V (1.7) I (1.7)	H (1.6)		
Prk1	L (4.5) I (2.7) M (1.6)		R (4.0) P (1.9)	Q (3.4) H (2.0) R (1.9)	Y (2.2) H (1.9)	G (6.8)			
Snf1	L (2.5) I (2.0)	R (2.8) K (2.5) H (1.7)	R (7.5) K (3.0)	R (3.0) T (2.4) S (2.2) C (1.7) V (1.6)	R (2.2) K (1.8) H (1.7) Y (1.7)	M (1.9) F (1.9) C (1.8) Y (1.6)		N (2.1) F (1.7)	L (4.5) I (2.1) M (1.7)
Kin1	K (1.7)	H (2.0) W (2.0)	H (2.2) W (2.0) F (1.9) Y (1.6)	N (4.0) S (2.3)	R (2.0)	Q (1.6) R (1.6)		D (2.2) N (1.7) C (1.6)	L (5.5) I (2.4) V (1.9)
Kss1		H (1.8)		P (6.2)	D (2.0) A (1.9) R (1.9)	P (8.5)		P (1.9)	K (1.7)
Yak1	R (1.7)	R (2.4)	R (8.0)	R (3.0) C (2.5) P (2.1) A (1.6)	R (2.5) K (2.0) H (1.6)	P (5.5) R (2.3)	C (1.8)	P (1.7)	
Vhs1	M (6.4)	R (2.1) K (1.6) V (1.6)	R (11)	S (3.1) T (2.7) R (2.0)	K (1.8) H (1.7)		S (2.7) T (2.0)		F (1.6)

often reported to be selective primarily for either Leu or Arg at the P-5 position, as well as selective for Arg at P-3 (13, 18–20). Among the various kinases that selected Arg at the P-3 position, we observed a spectrum of residues selected at the P-5 position, including not only Leu (Cmk1 and Cmk2) and Arg (Ypk1), but also Met (Vhs1), Val or Ile (Prr1), and His (Psk2) (Fig. 3 and table S1). The seven proline-directed kinases, which primarily selected for Pro at the P+1 position, were also distinguishable on the basis of selectivity at other positions. For example, Kss1, Hog1, and Fus3 all showed a secondary selectivity for proline at the P-2 position that was not observed in Pho85 or Cdc28. Other motifs were less common, and include multiple distinct “acidophilic” motifs in which the strongest selectivity was for Asp, Glu, or pThr. Such acidophilic motifs have been previously seen for various mammalian kinases, including GSK3 (glycogen synthase kinase 3; selectivity for acidic amino acids at the P+4 position), CK1 (casein kinase I; P-5 through P-3), PLK (Polo-like kinase; P-2), and CK2 (P+1 through P+3) (21–23). All yeast orthologs of these kinases recapitulated the motif found in their mammalian orthologs (table S2), but we also found additional yeast acidophilic kinases that were not anticipated (Mps1, Gcn2, and Cdc7). In addition, three kinases, Atg1, Kin1, and Kin3, exhibited their strongest selectivities for hydrophobic residues. The remaining kinases exhibited multiple strong selectivities and could not easily be categorized.

Connecting phosphorylation site motifs to kinase specificity-determining residues

Yeast kinases have been classified into five groups on the basis of sequence homology: AGC [adenosine 3',5'-monophosphate-dependent protein kinase (PKA)/guanosine 3',5'-monophosphate-dependent protein kinase (PKG)/protein kinase C (PKC)], CAMK (calcium/calmodulin-regulated and structurally similar kinases), CMGC (CDK, MAPK, GSK, and CDK-like kinases), STE11/STE20, and STE7/MEK (MAPKK) (24). These groups have then been classified further into families that share a high degree of sequence similarity within their catalytic domains. Although related kinases generally recognized similar phosphorylation site motifs, kinases within the same family occasionally exhibited differences, both subtle and striking. One family that illustrates striking differences is the Snf1 kinase family, which belongs to the CAMK group. In yeast, the Snf1 [also known as the AMPK, adenosine monophosphate (AMP)-activated protein kinase] family has six family members—Gin4, Hsl1, Kcc4, Kin1, Kin2, and Snf1. We identified consensus phosphorylation site motifs for each of these kinases with the exception of Kin2 (Table 1 and table S1). All five kinases had common features in their motifs, which are also shared with mammalian AMPKs

(25, 26). For example, each one had preferences for a Ser residue as the phosphoacceptor site, a Ser residue at the P-2 position, an Asn residue at the P+3 position, and hydrophobic residues at the P+4 position (Gin4, Snf1, and Kin1 are summarized in Table 1; see data set S1 for quantitative data for Hsl1 and Kcc4). Strikingly, however, only four of the five Snf1 family kinases exhibited the hallmark basophilic P-3 Arg selectivity of the CAMK group, with Kin1 lacking this conserved feature. Instead, Kin1 had an additional preference for an Asn residue at the P-2 position.

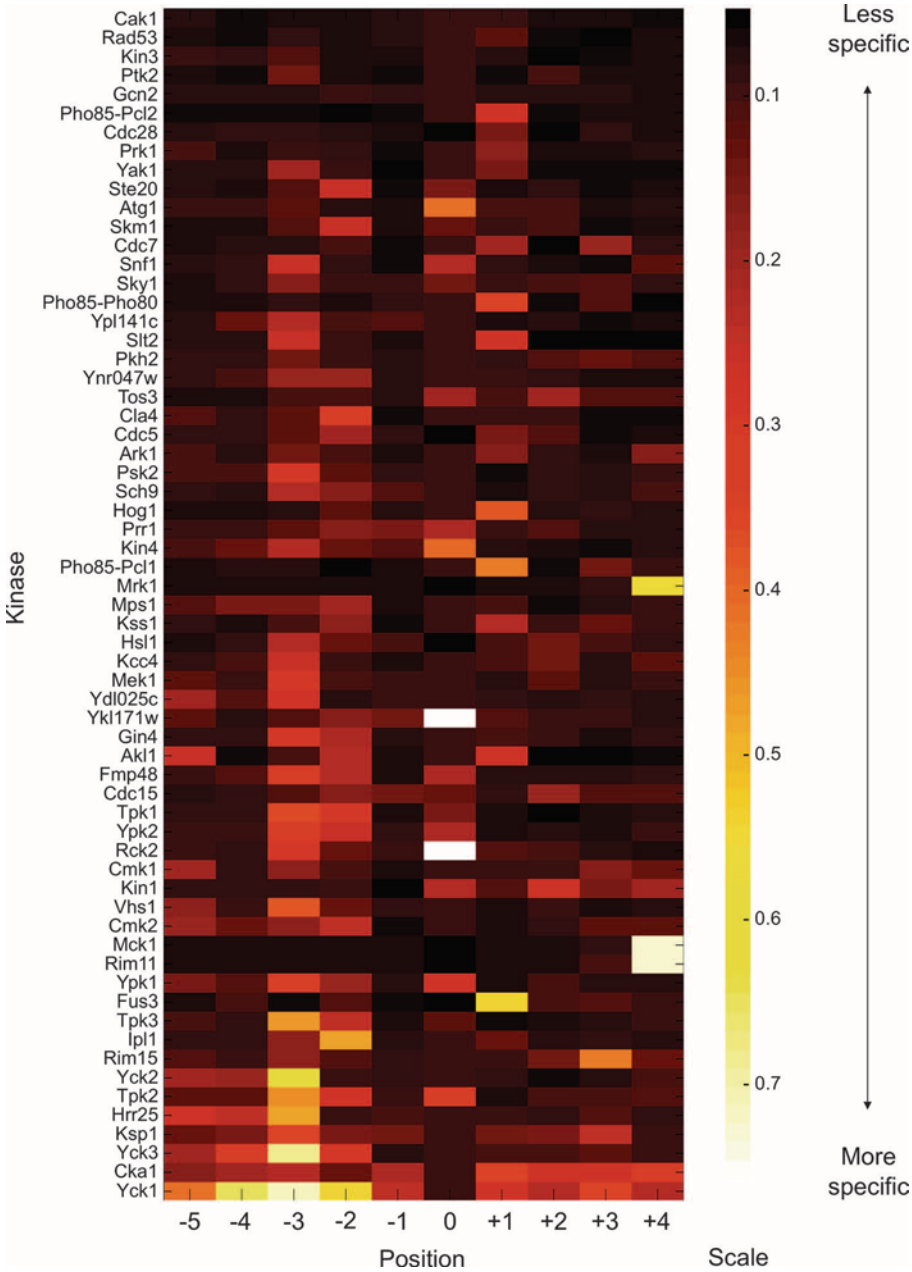
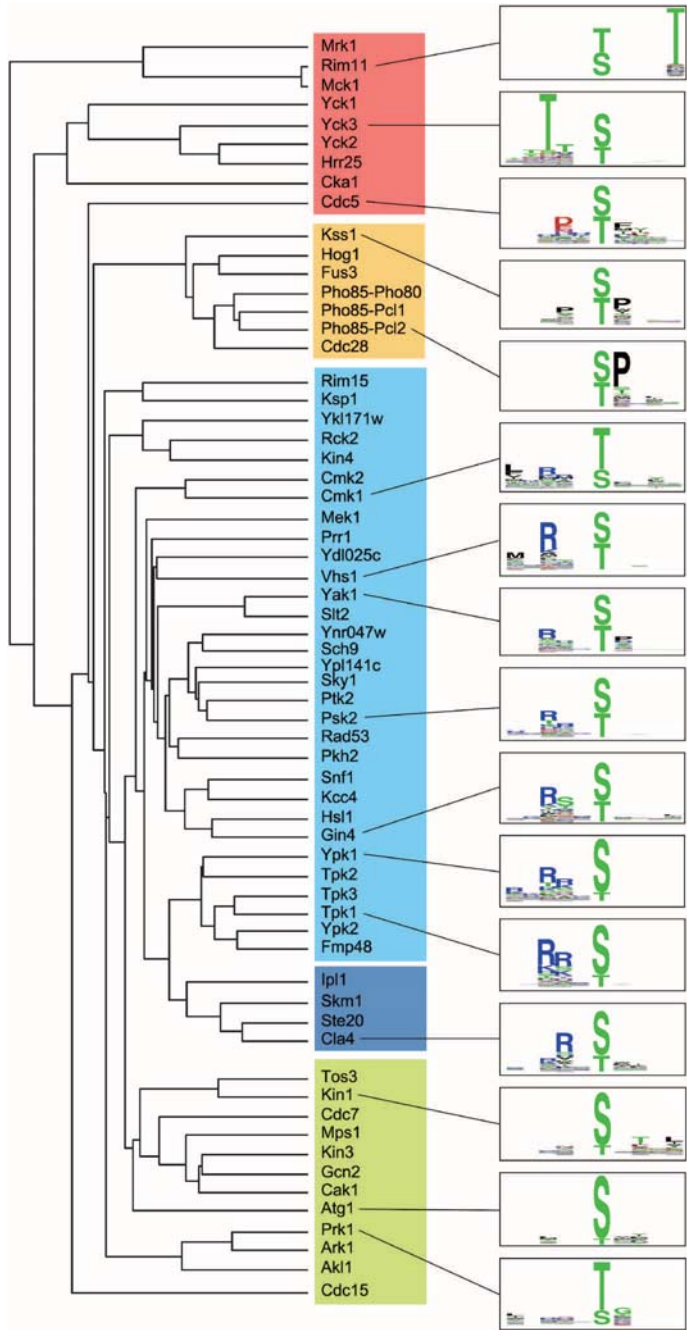
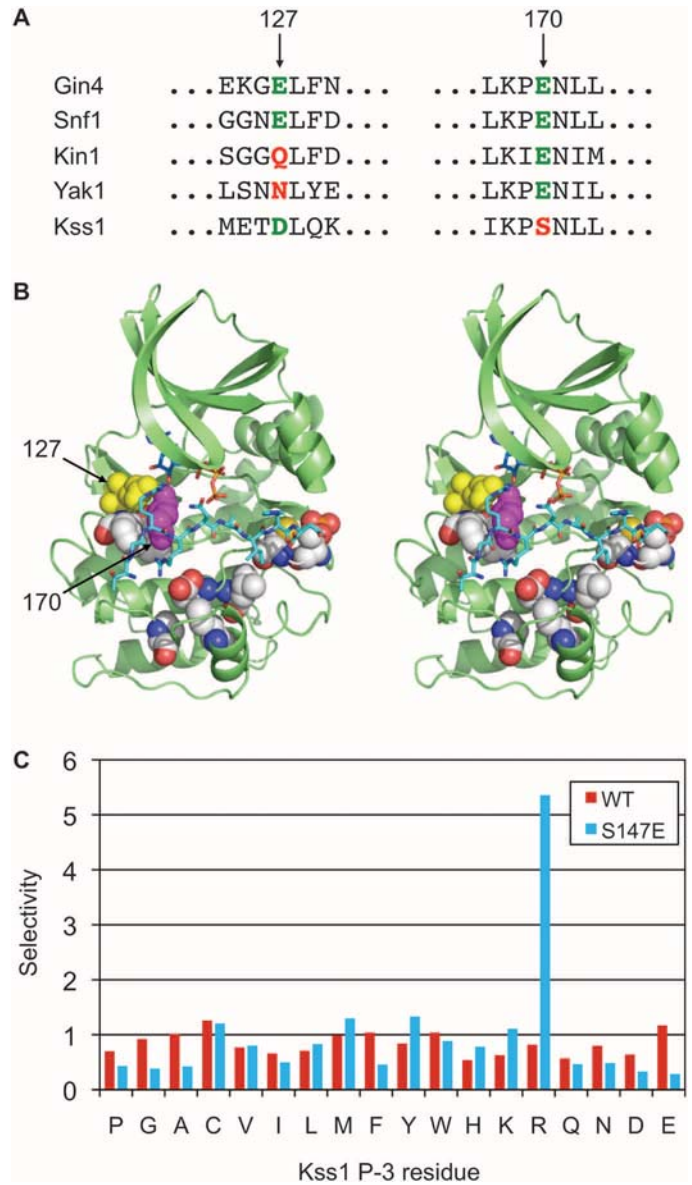


Fig. 2. Heat map ranking kinases by their specificity quotients as calculated from their average PWMs. Kinases are ranked from least specific (top) to most specific (bottom). The specificity in each position is defined as the information content in each position, equivalent to the total height of the sequence logo (see table S1 for logos).

This difference correlated with a single amino acid substitution within the kinase catalytic domain (Fig. 4A). Gin4, Hsl1, Kcc4, and Snf1 each have a conserved Glu residue (corresponding to Glu<sup>127</sup> in PKA; Fig. 4B). Crystal



**Fig. 3.** Dendrogram of yeast kinases clustered by specificity. Specificity categories are indicated by shading: red, acidophilic; orange, Pro-directed; cyan, P-3 Arg selecting; blue, P-2 Arg selecting; green, other. Because there were multiple distinct acidophilic motifs in which selectivity varies by position, some kinases selecting primarily acidic residues clustered in the “other” category. Sequence logos (74) are shown for selected kinases from each group.



**Fig. 4.** Comparison of kinase consensus phosphorylation site motifs to primary sequence reveals specificity-determining residues. (A) Sequence alignment of the regions surrounding residues 127 and 170 (human PKA numbering) in the catalytic domain of representative Snf1 family kinases (Gin4, Snf1, Kin1), and the CMGC kinases Yak1 and Kss1. The presence of an acidic residue at position 127 correlates with Arg selectivity at the P-3 position for the Snf1 family, but not the CMGC group. Conversely, a Glu residue at position 170 correlates with Arg selectivity for CMGC group kinases, but not for the Snf1 family. (B) Stereo view of the crystal structure of PKA with bound pseudosubstrate peptide (shown in cyan in stick representation; for clarity, only the portion falling within the active-site cleft is shown) highlighting predicted specificity-determining residues (in sphere representation). Residues 127 and 170 are shown in yellow and magenta, respectively. The figure was generated with PyMOL from the coordinates in PDB code 1ATP. (C) Kss1 mutagenesis. Mutation Kss1 Ser<sup>147</sup> to Glu confers selectivity for Arg at P-3. The bar graph shows normalized spot intensities for the P-3 position taken from screens of the full peptide library (shown in fig. S2).

**Table 2.** Computationally predicted kinase specificity-determining residues. Correlation values and peptide-kinase distance measurements are defined in Materials and Methods.

Residue (PKA numbering)	Correlation	Distance from peptide (Å)	Representative kinase	Amino acid	Specificity	Reference
170	0.195	5.3	Tpk1	Glu	−2 Arg	(27)
			Sch9	Glu	−5 Arg	(29), (70)
			Yak1	Glu	−3 Arg	This study
169	0.186	7.2	Psk2	Asp	−5 Basic	(30)
			Cmk1	Pro	−5 Aliphatic	(18), (30)
197	0.176	9.5	Cdc28	(p)Thr	+3 Basic	(71)
129	0.172	6.4	Tpk1	Phe	−3 Arg	(27)
			Atg1	Thr	−3 Aliphatic	This study
127	0.157	7.4	Tpk1	Glu	−3 Arg	(27)
199	0.145	5.7	Tpk1	Cys	0 Ser	This study
246	0.134	8.4	Gcn2	Arg	−2 Acidic	This study
205	0.122	8.4	Kss1	Arg	+1 Pro	(54)
			lpl1	Leu	+1 Hydrophobic	(27), (29), (72)
237	0.120	9.9	Yak1	Leu	−2 Pro	This study
203	0.108	6.1	Yck1	Arg	−3 pSer/pThr	(73)
			Mps1	Asn	−2 Acidic	This study
			Gin4	His	−2 Ser	This study

structures of multiple basophilic kinases in complex with peptide substrates have shown that this residue forms a salt bridge with the guanidino group of the P−3 Arg residue of the bound substrate (27–30). Unlike the other family members, Kin1 has a Gln residue in place of this conserved Glu. These observations are thus consistent with a role for Glu<sup>127</sup> as the critical specificity-determining residue for Arg at the P−3 position in substrates, at least within the Snf1 family.

However, crystallographic insight into specificity determinants in protein kinases is limited to a handful of cases where structures of kinase-peptide complexes have been solved. Although computational approaches have offered additional insight into structural features that control specificity (31, 32), the existence of alternative binding modes, even between kinases with similar specificity (30), makes it difficult to make general conclusions regarding the relationship of kinase sequence to specificity. Indeed, multiple sequence alignment of the yeast kinome and comparison with our experimentally determined motifs indicated that the presence of an acidic residue at position 127 is neither necessary nor sufficient to direct selectivity for Arg at the P−3 position in substrates. For example, within the CMGC group, members of the MAPK and CDK families (Fus3, Kss1, Hog1, Cdc28, and Pho85), which are proline-directed kinases, have an Asp residue at that position, despite a lack of selectivity for Arg at the P−3 position. Conversely, Yak1 within the same group is basophilic, yet lacks an acidic residue at that position (Table 1 and Fig. 4A). Presumably, other residues within the catalytic domain are responsible for dictating a basophilic signature within this group of kinases.

With our large collection of kinase motifs, we identified previously unknown specificity-determining residues, including, but not restricted to, residues that might confer P−3 Arg selectivity for kinases that are not part of the Snf1 family. We used an approach based on the idea of covariation (33). We identified residues whose variation in the primary sequence of the catalytic domain significantly correlated with the variation in phosphorylation site specificity across kinases. To measure sequence variation, we used a simple pairwise similarity matrix, and to compare specificities, we calculated the Frobenius norm of the differences in PWMs (Table 2 and Fig. 4B). This approach reproduced several specificity-determining residues previously known from both structural and mutagenesis studies, including

Glu<sup>127</sup>. In addition, we uncovered many previously unknown candidate specificity-determining residues, seven of which were predicted to be within 10 Å to a bound protein substrate. Among these, an acidic Glu residue at position 170 (PKA numbering) correlated with P−3 Arg selectivity among CMGC kinases. This result contrasts with a previous prediction based on modeling of DYRK1A, the human homolog of Yak1 (34). To test our predictions, we examined the role of residue 170 in substrate selection. Indeed, a Ser to Glu mutation at the analogous position in the MAPK Kss1 (residue 147) conferred a basophilic signature (Fig. 4C and fig. S2). This result validates our ability to predict previously unknown specificity-determining residues on the basis of our large motif data set.

### Connecting kinases to substrates on the basis of phosphorylation site motifs

Because in vivo phosphorylation sites on protein substrates tend to fall within the context of the phosphorylation site motif for a particular kinase, database scanning has been used to predict previously unknown substrates and to pinpoint sites of phosphorylation (14, 26, 35–39). However, simple sequence-matching approaches are prone to false positives, because predicted sites may not be accessible for phosphorylation, and kinases can also depend on docking or scaffolding interactions for substrate recruitment. In addition, false negatives are frequent for kinases with low sequence specificity because their motifs occur in many proteins and are thus present with high frequency in databases (14, 18). To increase the accuracy of such predictions, we generated and used a motif analysis pipeline, MOTIPS (<http://motips.gersteinlab.org/>). MOTIPS scans sequence databases for sites that most closely match the PWM for a particular kinase with a modified algorithm based on the program Scansite (40). Predicted sites are then scored on the basis of a panel of features (evolutionary conservation, predicted surface accessibility, and disordered structure) that are characteristic of known phosphorylation sites (41–43).

We first analyzed established kinase substrates for the presence of their respective phosphorylation site motifs with MOTIPS. From a sampling of 174 in vivo kinase-substrate relationships curated from the literature, 99 of the substrates ranked among the top 0.5% of predicted sites for their respective kinase, with 27 substrates falling within the top 200 sites (Fig.

5A). We next analyzed predicted substrates for each of the 61 yeast kinases for their associated biological processes and respective localization according to Gene Ontology (GO) assignments in the *Saccharomyces* Genome Database (44) (Fig. 5B; the full list of predicted substrates for each kinase with associated GO terms and MOTIPS features is provided as data set S2). We found that predicted substrates were more likely to be associated with the same biological process and to localize to the same sub-

cellular compartment as their respective kinases than a randomly chosen set of proteins. Together, these observations suggest that motif scanning with our set of phosphorylation site motifs enriches for authentic kinase-substrate pairs.

To establish directly that our bioinformatics analysis had uncovered authentic substrates, we examined more closely the predicted substrates of the protein kinase Prk1. Prk1 is a member of a small family of kinases conserved throughout eukaryotes that mediates reorganization of the actin cytoskeleton during endocytosis (45). Our peptide array analysis revealed an unusual phosphorylation site motif that included strong preferences for aliphatic residues at the P-5 position, Gly at the P+1 position, and Thr as the phosphoacceptor (Fig. 1B and Table 1). We selected 107 Prk1 candidate substrates identified by MOTIPS for further analysis. These substrates contained sites of high, middle, and low rank among the top 2000 scoring sites. Because all five known Prk1 substrates undergo multisite phosphorylation (45–47), candidates were also chosen for having at least three predicted Prk1 phosphorylation sites. Of the 107 candidate substrates, we observed phosphorylation of 19 candidates in vitro with wild-type Prk1 but not with a Prk1 inactive mutant (fig. S3). To identify additional candidates, we used these 19 candidates as positive data points in a training set to educate MOTIPS by machine learning. Negative data points in the training set included 81 of the original Prk1 candidates that were unambiguously not substrates in vitro, as well as about 400 proteins identified in the yeast protein database as localizing solely to noncytosolic compartments (48).

This set of positive and negative data points was used to retrain the Bayesian algorithm in MOTIPS to integrate the motif matching, conservation, surface accessibility, and disorder scores for each site, along with an additional score based on the number of predicted sites. The five known in vivo substrates of Prk1, which were excluded from the training set, all fell within the top seven targets (Fig. 6A). Five additional candidates taken from the top 15 putative substrates in the new Prk1 hit list were tested by an in vitro kinase assay that used the purified candidates as substrates. These in vitro assays revealed three additional substrates for Prk1: Gon7, a protein component of the EKC/KEOPS (endopeptidase-like kinase chromatin-associated/kinase, putative endopeptidase and other proteins of small size) complex involved in telomere regulation; Gph1, a protein involved in the mobilization of glycogen; and the key endocytic protein Las17. One of the five additional candidates tested was Ypl150w, which is a putative kinase that autophosphorylated in our assay and thus could not be confirmed or excluded as a substrate of Prk1. This second round of in vitro assays provides additional evidence that retraining our algorithm increased our success rate in predicting authentic kinase substrates. Furthermore, among the 22 in vitro confirmed Prk1 substrates, seven proteins (Bem2, Edl1, Las17, Sac3, Sla2, Syp1, and Yap1801) are reported to have roles in endocytosis or the regulation of the actin cytoskeleton, suggesting that they may be subject to regulation by Prk1 (Table 3).

We next investigated whether our predicted Prk1 candidate substrates represented bona fide substrates. Because a closely related kinase, Ark1, has an overlapping biological function and shares a nearly identical phosphorylation site motif with Prk1, we examined the phosphorylation state of candidate substrates in yeast strains deleted for both *PRK1* and *ARK1*. Changes in phosphorylation were monitored by electrophoretic mobility shifts in immunoblots of purified substrates, with phosphatase-treated samples serving as a control for the unphosphorylated species. We observed a change in mobility for two candidate substrates, Bem2 and Edl1, suggesting that they are in vivo targets of Prk1 or Ark1, or both (Fig. 6B). Although we did not observe gel shifts for other substrates, it is likely that some are authentic Prk1/Ark1 substrates as well but simply do not change

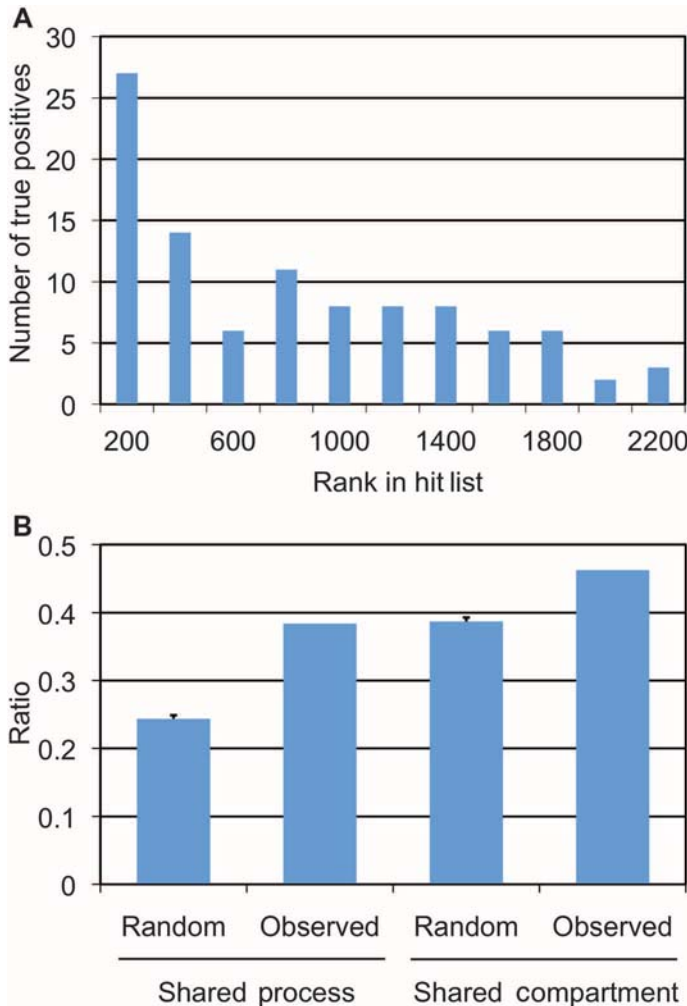


Fig. 5. MOTIPS ranking of known and predicted kinase-substrate pairs. (A) Bar graph showing the number of protein substrates reported in the literature (true positives) that have at least one phosphorylation site falling within the indicated rank value of predicted substrates for its respective kinase. Shown are the 99 sites of 174 known kinase-substrate pairs analyzed that fall within the top 0.5% predicted sites for that kinase among all Ser or Thr residues in the yeast proteome. (B) GO analysis of predicted kinase-substrate relationships that fall within the top 100 predicted substrates for all 61 kinases analyzed. The graph shows the ratio of predicted kinase-substrate pairs sharing either an annotated biological process (left bars) or a subcellular compartment (right bars) in comparison to pairs of proteins chosen at random. For both pairs, the probability that the observed value falls within the random distribution is extremely low ( $P < 10^{-35}$ ) based on the calculated area under the Gaussian curve corresponding to the random distribution.

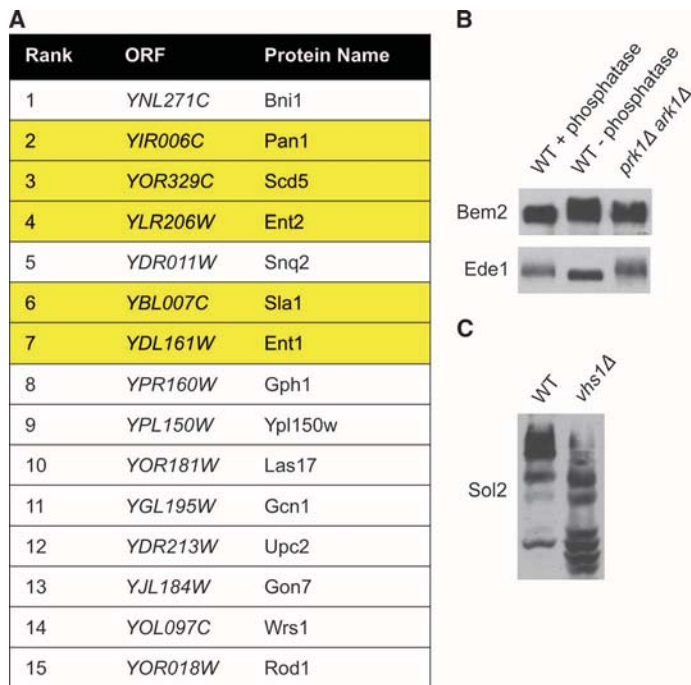


Fig. 6. Prediction and confirmation of kinase-substrate relationships. (A) Top 15 hits from the trained Prk1 MOTIPS output. The Prk1 hit list of candidate substrates was subjected to machine learning with a training set consisting of 19 true positives (experimentally derived) and ~480 true negatives (experimentally derived and supplemented with those proteins that are known to solely localize to noncytosolic compartments). Known *in vivo* substrates of Prk1 are highlighted in yellow. (B) Electrophoretic mobility shift analyses of Bem2 and Ede1. TAP (tandem affinity purification)-tagged Bem2 and Ede1 were purified from wild-type (WT) or *prk1Δ ark1Δ* strains by immobilized immunoglobulin G and then incubated in the presence or absence of phosphatases followed by immunoblotting against the TAP tag. (C) Mobility shift confirms Sol2 as an *in vivo* substrate of Vhs1. Lysates from wild-type (WT) or *vhs1Δ* strains expressing TAP-tagged Sol2 were fractionated on denaturing polyacrylamide gels impregnated with Phos-tag (57), which retards the mobility of phosphoproteins, followed by immunoblotting against the TAP tag.

mobility upon phosphorylation. Notably, previous MS phosphoproteomic analysis identified three of the *in vitro* Prk1 substrates (Ede1, Syp1, and Rpl5) as phosphorylated at Prk1 consensus sites *in vivo* (49–54) (the MOTIPS output for all kinases, which is available as data set S2, indicates which candidate phosphorylation sites have been identified by MS).

We also validated kinase-substrate pairs through integration with other proteomic data sets. We found that the kinase Vhs1, for which limited functional information is known, exhibited selectivity for the phosphorylation site motif MXRXXS (Table 1 and table S1). Fourteen *in vitro* substrates for the kinase Vhs1 (55) were previously identified by protein microarray analysis (4), and six of these, Mga1, Pfk26, Sef1, Sol1, Sol2, and Utr1, contain the Vhs1 consensus phosphorylation site motif. MS phosphoproteomic analysis (49) revealed that Sef1 was phosphorylated *in vivo* at a Vhs1 consensus phosphorylation site, and in an immunoprecipitation-MS analysis Sef1 and Vhs1 physically interacted (56). In addition, MS phosphoproteomic analysis identified Sol1 as phosphorylated at a Vhs1 consensus phosphorylation site *in vivo* (50),

**Table 3.** Proteins phosphorylated by Prk1 *in vitro*. Proteins functionally associated with actin rearrangement or endocytosis are Bem2, Ede1, Las17, Sac3, Sla2, Syp1, and Yap1801. Ub, ubiquitin; RhoGAP, Rho guanosine triphosphatase-activating protein; SCF, Skp1-Cullin-F-box; UPR, unfolded protein response; WASP, Wiscott-Aldrich syndrome protein.

ORF	Protein name	Function
YBL024W	Ncl1	m5C-methyltransferase
YBL047C	Ede1	Key endocytic protein; binds plasma membrane in a Ub-dependent manner
YCR030C	Syp1	Overexpression suppresses a <i>pfy1</i> (profilin) null mutation
YDL001W	Rmd1	Cytoplasmic protein required for sporulation
YDR159W	Sac3	Nuclear pore-associated protein; suppressor of actin mutations
YER155C	Bem2	RhoGAP involved in the control of cytoskeleton and morphogenesis
YGR094W	Vas1	Mitochondrial and cytoplasmic valyl-tRNA synthetase
YHL030W	Ecm29	Major component of the proteasome
YHR161C	Yap1801	Protein involved in clathrin cage assembly; binds clathrin and Pan1
YHR098C	Sfb3	Involved in sorting Pma1 into COPII vesicles
YJL090C	Dpb11	Subunit of DNA polymerase II epsilon complex
YJL129C	Trk1	Component of the Trk1-Trk2 potassium transport system
YJL184W	Gon7	Component of the EKC/KEOPS complex
YJR137C	Ecm17	Sulfite reductase beta subunit involved in amino acid biosynthesis
YML088W	Ufo1	F-box protein subunit of the SCF E3 ubiquitin ligase complex
YML103C	Nup188	Subunit of the nuclear pore complex
YNL243W	Sla2	Transmembrane protein that links actin to clathrin and endocytosis
YNL287W	Sec21	Gamma subunit of coatomer complex
YOR093C	Yor093c	Function unknown; deletion causes sensitivity to UPR-inducing agents
YOR181W	Las17	Actin assembly factor; homolog of human WASP
YPL131W	Rpl5	Component of the large 60S ribosomal subunit
YPR160W	Gph1	Glycogen phosphorylase

and its homolog Sol2 was the most highly phosphorylated Vhs1 *in vitro* substrate identified by protein microarray analysis (4). Mobility shift analysis of *VHS1* deletion strains using Phos-tag SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (57) was consistent with Sol2 as a substrate for Vhs1 *in vivo* (Fig. 6C). Although the presence of multiple Sol2 species in the presence and absence of Vhs1 indicates phosphorylation at multiple sites, likely by more than one kinase, the mobility shift indicates that in *vhs1* mutant cells, Sol2 is phosphorylated at fewer sites. Sol2, which promotes nucleocytoplasmic transfer RNA (tRNA) transport (58), is the first reported *in vivo* substrate for Vhs1 and suggests a role for this kinase in regulating this process. These results illustrate how integration of data from multiple proteomic approaches can shed light on the biology of poorly characterized molecules.

## DISCUSSION

The elucidation of the mechanisms underlying kinase specificity remains an integral part of understanding phosphorylation-based signal transduc-

tion pathways. Previous methods for determining consensus phosphorylation site motifs have not been suitable for large-scale screening of a eukaryotic kinome. Here, we have described an approach for the high-throughput identification of consensus phosphorylation site motifs in which multiple kinases, with no previously known substrates, can be analyzed simultaneously. We have used this approach to provide comprehensive analysis of kinase specificity in a single eukaryotic organism, the yeast *S. cerevisiae*. Among other applications, this large data set has provided much broader insight into the structural basis for kinase selectivity than has been possible through individual analyses of single kinases.

With our data, we linked protein kinases to previously unknown substrates, thus elucidating mechanisms of phosphorylation-dependent signaling. A limitation to our approach, however, is that the peptide arrays treat each position in the substrate independently, and thus the potential interdependence between multiple positions is ignored. This approach is nonetheless a valuable first-pass screen for analyzing kinase specificity because it involves the systematic and exhaustive analysis of each amino acid residue at each position surrounding the phosphorylation site. Preferences observed with this approach can provide the basis for the design of kinase-specific peptide libraries to uncover positional interdependence. Furthermore, the presence of a consensus phosphorylation sequence alone is insufficient to direct phosphorylation of a protein by a particular kinase, and, accordingly, identification of previously unknown substrates on the basis of motif scanning is difficult. However, integration with other proteomic data sets provides a means of increasing confidence in predicted kinase-substrate relationships. In addition, specific kinase-substrate pairs can be inferred through computational methods that make use of non-sequence-based "contextual" features, such as subcellular localization and molecular function (38). For example, prediction of substrates targeted by relatively non-specific kinases with phosphorylation site motifs alone is unlikely to be successful because these sequences occur frequently in proteomes. In such cases, selection of authentic substrates is driven by docking or scaffolding interactions, and consensus sequences for substrate recruitment can be used in combination with phosphorylation site motifs to identify previously unknown substrates (59, 60).

For previously characterized kinases, we observed a high degree of conservation of phosphorylation site motifs between yeast and mammalian orthologs. These similarities suggest that the many previously unknown consensus motifs reported here are also conserved. Therefore, this data set will serve as a resource for studies of phosphorylation-dependent signaling in higher eukaryotes, as well as yeast.

## MATERIALS AND METHODS

Details regarding yeast strain information, kinase preparation, characterization of purified kinases, in vitro kinase assays, and electrophoretic mobility shift analyses are available in the Supplementary Materials.

### Peptide library screening

The peptide library (Anaspec Inc.) has been previously reported (13). For this study, fresh stock solutions were made from 5 mg of powder by dissolving peptides in dimethyl sulfoxide (DMSO), quantifying by absorbance at 280 nm, and adjusting to a stock concentration of 10 mM by adding the appropriate volume of DMSO. Stock solutions were stored at  $-20^{\circ}\text{C}$  in microcentrifuge tubes. Working 0.6 mM aqueous stocks were prepared by diluting the DMSO stock in 20 mM Hepes (pH 7.4) and arrayed into 1536-well stock plates containing 5- $\mu\text{l}$  aliquots in each well. Plates were sealed with adhesive foil and stored at  $-20^{\circ}\text{C}$ .

Peptides (0.2  $\mu\text{l}$  per well) were transferred to assay plates containing 2  $\mu\text{l}$  of kinase reaction buffer [generally 20 mM Hepes (pH 7.4), 10 mM

$\text{MgCl}_2$ , 1 mM dithiothreitol (DTT), 0.1% Tween 20] from stock plates manually with a  $48 \times 6$  slot pin replicator (VP Scientific). Reactions were initiated by adding a solution (0.2  $\mu\text{l}$  per well) containing purified kinase and  $[\gamma\text{-}^{33}\text{P}]\text{ATP}$  (0.55 mM, 0.3 to 0.4  $\mu\text{Ci}/\mu\text{l}$ ; Perkin Elmer) with a  $48 \times 1$  slot pin replicator (VP Scientific). Plates were sealed and incubated for 1 to 8 hours at  $30^{\circ}\text{C}$ . The final concentrations of the reaction components in each well were 50  $\mu\text{M}$  peptide and 50  $\mu\text{M}$  ATP at a specific activity of 0.55 to 0.73  $\text{mCi}/\mu\text{mol}$ . After incubation, 0.2  $\mu\text{l}$  from each well was spotted onto streptavidin-coated membrane (SAM2 Biotin Capture Membrane, Promega) simultaneously using the  $48 \times 6$  slot pin replicator. Membranes were washed three times with 10 mM tris-HCl (pH 7.5) with 140 mM NaCl and 0.1% SDS, twice with 2 M NaCl, twice with 2 M NaCl with 1%  $\text{H}_3\text{PO}_4$ , and twice with water, then dried and exposed to a phosphor storage screen. Processing of final images of the spot arrays consisted of copying the  $4 \times 22$  grid corresponding to the P+1, P+2, P+3, and P+4 peptide mixtures and pasting it below the  $5 \times 22$  grid corresponding to the P-5, P-4, P-3, P-2, and P-1 peptide mixtures with Adobe Photoshop to provide the  $9 \times 22$  spot grids shown in Fig. 1 and table S1.

### PWM generation

For each array, peptide phosphorylation signals were quantified with Genepix Pro 6.0 (Molecular Devices) by manually aligning a  $48 \times 8$  grid of circles onto each scanned phosphorimage to calculate the median intensity for each spot. These median intensity values were then background-corrected by subtracting the median intensity value corresponding to the negative control spot (reaction carried out in the absence of any peptide substrate). Signal scores for each amino acid at each position were then normalized by the following equation:

$$Z_{ca} = \frac{S_{ca}}{\sum_i S_{ci}} \times m$$

where  $Z_{ca}$  stands for the normalized score of amino acid  $a$  at position  $c$  having a signal score  $S_{ca}$ , and  $m$  stands for the total number of amino acids.  $S_{ci}$  is the signal score of amino acid  $i$  at position  $c$  where  $i$  is defined in the summation of all the  $m$  amino acids. The PWM is an  $N \times 20$  matrix of  $N$  positions with the normalized, background-corrected value given as the weight for each amino acid at each position. To account for spurious phosphorylation of Ser and Thr residues at other positions, the PWM entries in all Ser and Thr positions were set to 1 (equivalent to neutral selection at that position) with subsequent renormalization of the PWM.

### Proteome scanning

The entire yeast proteome was scanned to identify the best matches to each PWM. Our approach used a window-sliding method based on the normalized PWM similar to the method used in Scansite (40). Briefly, it extracted every possible 15-mer sequence from the yeast proteome and calculated the match score to the PWM, based on the formula:

$$S = \sum_i \sum_{a=r_i} \log(M_{ia})$$

where  $i$  stands for the position in the motif and  $r_i$  stands for the residue that is present at position  $i$  in the peptide in question.  $M_{ia}$  is the normalized PWM as described above. The resulting score was then normalized, such that zero stands for an optimal match to the motif and larger positive scores correspond to weaker matches. The top 10,000 potential phosphorylation sites for each kinase are reported in data set S2. This algorithm was implemented in a modular form in Java. All sequences

and features were loaded into an SQL database that is interactively queried by the Java search module.

### Feature collection

A number of different genomic features were gathered to supplement the initial match score. To compute the conservation score, we collected all orthologs for 13 proteomes of related yeast species (*Saccharomyces paradoxus* as the closest and *Schizosaccharomyces pombe* as the farthest) using the comparative genomics algorithm implemented in INPARANOID (61). We then aligned these orthologs using the automated alignment method MUSCLE (62) (the full set of alignments is available as data set S3). For each PWM hit, we calculated the conservation score by estimating the entropy at each position based on the aligned orthologs with the AL2CO program. The disorder score was based on the prediction program DISOPRED (63). DISOPRED was run for each protein in the yeast proteome. We used the DISOPRED probability score, corresponding to the likelihood of the residue in question being in a disordered region, as the measure of disorder. Finally, the surface accessibility score was calculated with the prediction program SABLE for each protein in the yeast proteome (64). The simple numerical surface score was used as the measure of surface accessibility.

### Feature integration

An integration algorithm based on the naïve Bayes framework was used to integrate the four features. We used a number of experimentally determined gold-standard kinase-substrate pairs, “positives,” to train the algorithm. For gold-standard negatives, we supplemented a set of experimentally determined negatives with a set of randomly chosen protein pairs. Each of these pairs is a pair of proteins that are annotated to always localize to two different compartments (for example, nucleus only and cytoplasm only). Thus, we biased the randomly chosen set of protein pairs further toward a set that was highly unlikely to contain any spurious positive interactions. The conditional probability was calculated from the four features according to the following formula:

$$p(I|D_1, D_2, D_3, D_4) = p(I)p(D_1|I)p(D_2|I)p(D_3|I)p(D_4|I)$$

where  $I$  denotes either interaction or noninteraction and  $D_1$  through  $D_4$  denote the four features. Data were thus integrated under the assumption that the four features are independent. To formally assess independence of the features, we calculated pairwise correlation coefficients. The results showed the pairwise correlation coefficients ranging from 0.01 to 0.57 (absolute values) have an average of 0.18, indicating the features are to a large extent independent (see table S3). Moreover, we performed principal components analysis (PCA) using the statistical software R to transform the possibly correlated values of the five features (hits per protein, match score, disorder score, accessibility score, and conservation score) of the Prk1 targets into uncorrelated values. The first three vectors were chosen to build a naïve Bayes model followed by a 10-fold stratified cross-validation. The area under the curve (AUC; 75.9%) of the receiver operating curve (ROC) resulting from the PCA validation was then compared to the AUC (78.6%) of Prk1 without the PCA transformation. The close performance of the two further indicated a certain level of independence of the features. Bayesian integration was implemented with the Java machine learning package *Weka* (65). The entire methodology is available as the modularized software packages MOTIPS (<http://motips.gersteinlab.org/>).

### Covariation calculation to estimate specificity-determining residues

Sequences of the 61 yeast kinase catalytic domains (obtained from the kinase.com database) were initially aligned with ClustalW2 (66). A high-

quality sequence alignment was generated by manual editing of the initial alignment in Jalview (67) on the basis of multiple pairwise alignments with kinases of known three-dimensional (3D) structure and conserved catalytic residues (table S4). In addition, 89 orthologous kinases from *S. pombe*, *Dictyostelium discoideum*, and *Homo sapiens* were added and manually aligned. For these orthologs, the PWM was inferred to be identical to its yeast counterpart. A correlation-based methodology was implemented to identify specificity-determining residues:

For each  $\left(\begin{smallmatrix} n \\ m \end{smallmatrix}\right)$  pair of sequence positions ( $n$ ) and positions in the PWM ( $m$ ), 2D vectors were generated;  $k$  is the total number of kinases in the alignment and is equal to the number of PWMs. The first vector contained all pairwise similarities between the primary sequences of the kinases in that position, based on the McLachlan matrix (that is, the similarity of the amino acid in position X in kinase A to the similarity of the amino acid in the same position in kinase B) (68). The McLachlan matrix was chosen because it scores for residue substitutions based on chemical similarity (that is, physicochemical properties). The second vector contained the pairwise similarity of all PWMs to each other, based on the Frobenius norm (69):

$$\sqrt{\sum_i \sum_j (M_{1,ij} - M_{2,ij})^2}$$

Each position was then scored with the Pearson correlation coefficient of these two vectors (listed under “correlation” in Table 2). This method was implemented in the programming package MATLAB. Distances of the residue in question from bound peptide were estimated by mapping the residue onto the PKA-PKI structure [Protein Data Bank (PDB) ID 1ATP] with the program VMD. The peptide-kinase distances were measured as the closest distances between the geometric centers of the residue on the kinase, as mapped to the PKA structure, to the bound peptide, as in the PKA structure.

### SUPPLEMENTARY MATERIALS

[www.sciencesignaling.org/cgi/content/full/3/109/ra12/DC1](http://www.sciencesignaling.org/cgi/content/full/3/109/ra12/DC1)

Materials and Methods

Fig. S1. Assay reproducibility.

Fig. S2. Representative peptide array screening results for the Kss1 S147E mutant.

Fig. S3. Representative Prk1 in vitro assays.

Table S1. Representative peptide array screening results and sequence logos for each of the 61 kinases assayed (also available as a separate PDF with high-resolution images as 2000482tableS1.pdf)

Table S2. Protein kinases analyzed in this study.

Table S3. Pairwise correlation coefficients for each of four genomic features and the Scansite match score.

Table S4. Alignment of yeast kinases analyzed in this study.

#### References

Data set S1. Average PWMs for each of the 61 kinases assayed (plain text files).

Data set S2. MOTIPS output for each of the 61 kinases assayed (plain text files).

Data set S3. MUSCLE alignment of all predicted *S. cerevisiae* ORFs with orthologs from 12 other yeast species (clustal alignment files).

Data set S4. Alignment of yeast kinases analyzed in this study (clustal alignment file).

### REFERENCES AND NOTES

- G. Manning, G. D. Plowman, T. Hunter, S. Sudarsanam, Evolution of protein kinase signaling from yeast to man. *Trends Biochem. Sci.* **27**, 514–520 (2002).
- J. Moffat, D. M. Sabatini, Building mammalian signalling pathways with RNAi screens. *Nat. Rev. Mol. Cell Biol.* **7**, 177–187 (2006).
- K. Schmelzle, F. M. White, Phosphoproteomic approaches to elucidate cellular signaling networks. *Curr. Opin. Biotechnol.* **17**, 406–414 (2006).
- J. Ptacek, G. Devgan, G. Michaud, H. Zhu, X. Zhu, J. Fasolo, H. Guo, G. Jona, A. Breitkreutz, R. Sopko, R. R. McCartney, M. C. Schmidt, N. Rachidi, S. J. Lee, A. S. Mah, L. Meng, M. J. Stark, D. F. Stern, C. De Virgilio, M. Tyers, B. Andrews, M. Gerstein, B. Schweitzer, P. F. Predki, M. Snyder, Global analysis of protein phosphorylation in yeast. *Nature* **438**, 679–684 (2005).

5. L. M. Elphick, S. E. Lee, V. Gouverneur, D. J. Mann, Using chemical genetics and ATP analogues to dissect protein kinase function. *ACS Chem. Biol.* **2**, 299–314 (2007).
6. B. E. Turk, Understanding and exploiting substrate recognition by protein kinases. *Curr. Opin. Chem. Biol.* **12**, 4–10 (2008).
7. S. H. Park, A. Zarrinpar, W. A. Lim, Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms. *Science* **299**, 1061–1064 (2003).
8. J. M. Skerker, B. S. Perchuk, A. Siryaporn, E. A. Lubin, O. Ashenberg, M. Goulian, M. T. Laub, Rewiring the specificity of two-component signal transduction systems. *Cell* **133**, 1043–1054 (2008).
9. J. A. Ubersax, J. E. Ferrell Jr., Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell Biol.* **8**, 530–541 (2007).
10. L. A. Pinna, M. Ruzzene, How do protein kinases recognize their substrates? *Biochim. Biophys. Acta* **1314**, 191–225 (1996).
11. R. Amanchy, B. Periaswamy, S. Mathivanan, R. Reddy, S. G. Tattikota, A. Pandey, A curated compendium of phosphorylation motifs. *Nat. Biotechnol.* **25**, 285–286 (2007).
12. M. L. Miller, L. J. Jensen, F. Diella, C. Jorgensen, M. Tinti, L. Li, M. Hsiung, S. A. Parker, J. Bordeaux, T. Sicheritz-Ponten, M. Olhovskiy, A. Pasculescu, J. Alexander, S. Knapp, N. Blom, P. Bork, S. Li, G. Cesareni, T. Pawson, B. E. Turk, M. B. Yaffe, S. Brunak, R. Lindling, Linear motif atlas for phosphorylation-dependent signaling. *Sci. Signal.* **1**, ra2 (2008).
13. J. E. Huttli, E. T. Jarrell, J. D. Chang, D. W. Abbott, P. Storz, A. Toker, L. C. Cantley, B. E. Turk, A rapid method for determining protein kinase phosphorylation specificity. *Nat. Methods* **1**, 27–29 (2004).
14. U. E. Rennefahrt, S. W. Deacon, S. A. Parker, K. Devarajan, A. Beeser, J. Chernoff, S. Knapp, B. E. Turk, J. R. Peterson, Specificity profiling of Pak kinases allows identification of novel phosphorylation sites. *J. Biol. Chem.* **282**, 15667–15678 (2007).
15. D. M. Gelperin, M. A. White, M. L. Wilkinson, Y. Kon, L. A. Kung, K. J. Wise, N. Lopez-Hoyo, L. Jiang, S. Piccirillo, H. Yu, M. Gerstein, M. E. Dumont, E. M. Phizicky, M. Snyder, E. J. Grayhack, Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes Dev.* **19**, 2816–2826 (2005).
16. H. Zhu, M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R. A. Dean, M. Gerstein, M. Snyder, Global analysis of protein activities using proteome chips. *Science* **293**, 2101–2105 (2001).
17. P. Dent, Y. H. Chow, J. Wu, D. K. Morrison, R. Jove, T. W. Sturgill, Expression, purification and characterization of recombinant mitogen-activated protein kinase kinases. *Biochem. J.* **303**, 105–112 (1994).
18. I. A. Manke, A. Nguyen, D. Lim, M. Q. Stewart, A. E. Elia, M. B. Yaffe, MAPKAP kinase-2 is a cell cycle checkpoint kinase that regulates the G<sub>2</sub>/M transition and S phase progression in response to UV irradiation. *Mol. Cell* **17**, 37–48 (2005).
19. T. Obata, M. B. Yaffe, G. G. Lepar, E. T. Piro, H. Maegawa, A. Kashiwagi, R. Kikkawa, L. C. Cantley, Peptide and protein library screening defines optimal substrate motifs for AKT/PKB. *J. Biol. Chem.* **275**, 36108–36115 (2000).
20. D. R. Alessi, F. B. Caudwell, M. Andjelkovic, B. A. Hemmings, P. Cohen, Molecular basis for the substrate specificity of protein kinase B; comparison with MAPKAP kinase-1 and p70 S6 kinase. *FEBS Lett.* **399**, 333–338 (1996).
21. C. J. Fiol, J. H. Haseman, Y. H. Wang, P. J. Roach, R. W. Roeske, M. Kowalczyk, A. A. DePaoli-Roach, Phosphoserine as a recognition determinant for glycogen synthase kinase-3: Phosphorylation of a synthetic peptide based on the G-component of protein phosphatase-1. *Arch. Biochem. Biophys.* **267**, 797–802 (1988).
22. Z. Songyang, K. P. Lu, Y. T. Kwon, L. H. Tsai, O. Filhol, C. Cochet, D. A. Brickey, T. R. Soderling, C. Bartleson, D. J. Graves, A. J. DeMaggio, M. F. Hoekstra, J. Blenis, T. Hunter, L. C. Cantley, 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. *Mol. Cell Biol.* **16**, 6486–6493 (1996).
23. E. F. Johnson, K. D. Stewart, K. W. Woods, V. L. Giranda, Y. Luo, Pharmacological and functional comparison of the polo-like kinase family: Insight into inhibitor and substrate specificity. *Biochemistry* **46**, 9551–9563 (2007).
24. T. Hunter, G. D. Plowman, The protein kinases of budding yeast: Six score and more. *Trends Biochem. Sci.* **22**, 18–22 (1997).
25. S. Dale, W. A. Wilson, A. M. Edelman, D. G. Hardie, Similar substrate recognition motifs for mammalian AMP-activated protein kinase, higher plant HMG-CoA reductase kinase-A, yeast SNF1, and mammalian calmodulin-dependent protein kinase I. *FEBS Lett.* **361**, 191–195 (1995).
26. D. M. Gwinn, D. B. Shackelford, D. F. Egan, M. M. Mihaylova, A. Mery, D. S. Vasquez, B. E. Turk, R. J. Shaw, AMPK phosphorylation of raptor mediates a metabolic checkpoint. *Mol. Cell* **30**, 214–226 (2008).
27. D. R. Knighton, J. H. Zheng, L. F. Ten Eyck, N. H. Xuong, S. S. Taylor, J. M. Sowadski, Structure of a peptide inhibitor bound to the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* **253**, 414–420 (1991).
28. E. D. Lowe, M. E. Noble, V. T. Skamaki, N. G. Oikonomakos, D. J. Owen, L. N. Johnson, The crystal structure of a phosphorylase kinase peptide substrate complex: Kinase substrate recognition. *EMBO J.* **16**, 6646–6658 (1997).
29. J. Yang, P. Cron, V. M. Good, V. Thompson, B. A. Hemmings, D. Barford, Crystal structure of an activated Akt/protein kinase B ternary complex with GSK3-peptide and AMP-PNP. *Nat. Struct. Biol.* **9**, 940–944 (2002).
30. A. N. Bullock, J. Debreczeni, A. L. Amos, S. Knapp, B. E. Turk, Structure and substrate specificity of the Pim-1 kinase. *J. Biol. Chem.* **280**, 41675–41682 (2005).
31. R. I. Brinkworth, R. A. Breinl, B. Kobe, Structural basis and prediction of substrate specificity in protein serine/threonine kinases. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 74–79 (2003).
32. R. I. Brinkworth, A. L. Munn, B. Kobe, Protein kinases associated with the yeast phosphoproteome. *BMC Bioinformatics* **7**, 47 (2006).
33. L. Li, E. I. Shakhnovich, L. A. Mirny, Amino acids determining enzyme-substrate specificity in prokaryotic and eukaryotic protein kinases. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4463–4468 (2003).
34. S. Himpel, W. Tegge, R. Frank, S. Leder, H. G. Joost, W. Becker, Specificity determinants of substrate recognition by the protein kinase DYRK1A. *J. Biol. Chem.* **275**, 2431–2438 (2000).
35. M. B. Yaffe, G. G. Lepar, J. Lai, T. Obata, S. Volinia, L. C. Cantley, A motif-based profile scanning approach for genome-wide prediction of signaling pathways. *Nat. Biotechnol.* **19**, 348–353 (2001).
36. B. D. Manning, A. R. Tee, M. N. Logsdon, J. Blenis, L. C. Cantley, Identification of the tuberous sclerosis complex-2 tumor suppressor gene product tuberlin as a target of the phosphoinositide 3-kinase/Akt pathway. *Mol. Cell* **10**, 151–162 (2002).
37. L. J. Holt, J. E. Huttli, L. C. Cantley, D. O. Morgan, Evolution of Imed2 phosphorylation sites on Cdk1 substrates provides a mechanism to limit the effects of the phosphatase Cdc14 in meiosis. *Mol. Cell* **25**, 689–702 (2007).
38. R. Lindling, L. J. Jensen, G. J. Ostheimer, M. A. van Vugt, C. Jorgensen, I. M. Miron, F. Diella, K. Colwill, L. Taylor, K. Elder, P. Metelnikov, V. Nguyen, A. Pasculescu, J. Jin, J. G. Park, L. D. Samson, J. R. Woodgett, R. B. Russell, P. Bork, M. B. Yaffe, T. Pawson, Systematic discovery of in vivo phosphorylation networks. *Cell* **129**, 1415–1426 (2007).
39. J. E. Huttli, R. R. Shen, D. W. Abbott, A. Y. Zhou, K. M. Spott, J. M. Asara, W. C. Hahn, L. C. Cantley, Phosphorylation of the tumor suppressor CYLD by the breast cancer oncogene IKK $\epsilon$  promotes cell transformation. *Mol. Cell* **34**, 461–472 (2009).
40. J. C. Obenaus, L. C. Cantley, M. B. Yaffe, Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.* **31**, 3635–3641 (2003).
41. Y. V. Budovskaya, J. S. Stephan, S. J. Deminoff, P. K. Herman, An evolutionary proteomics approach identifies substrates of the cAMP-dependent protein kinase. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13933–13938 (2005).
42. L. M. Iakoucheva, P. Radivojac, C. J. Brown, T. R. O'Connor, J. G. Sikes, Z. Obradovic, A. K. Dunker, The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **32**, 1037–1049 (2004).
43. N. Blom, T. Sicheritz-Ponten, R. Gupta, S. Gammeltoft, S. Brunak, Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4**, 1633–1649 (2004).
44. E. L. Hong, R. Balakrishnan, Q. Dong, K. R. Christie, J. Park, G. Binkley, M. C. Costanzo, S. S. Dwight, S. R. Engel, D. G. Fisk, J. E. Hirschman, B. C. Hitz, C. J. Krieger, M. S. Livstone, S. R. Miyasato, R. S. Nash, R. Oughtred, M. S. Skrzypek, S. Weng, E. D. Wong, K. K. Zhu, K. Dolinski, D. Botstein, J. M. Cherry, Gene Ontology annotations at SGD: New data sources and annotation methods. *Nucleic Acids Res.* **36**, D577–D581 (2008).
45. J. Toshima, J. Y. Toshima, A. C. Martin, D. G. Drubin, Phosphoregulation of Arp2/3-dependent actin assembly during receptor-mediated endocytosis. *Nat. Cell Biol.* **7**, 246–254 (2005).
46. B. Huang, G. Zeng, A. Y. Ng, M. Cai, Identification of novel recognition motifs and regulatory targets for the yeast actin-regulating kinase Prk1p. *Mol. Biol. Cell* **14**, 4871–4884 (2003).
47. H. A. Watson, M. J. Cope, A. C. Groen, D. G. Drubin, B. Wendland, In vivo role for actin-regulating kinases in endocytosis and yeast epsin phosphorylation. *Mol. Biol. Cell* **12**, 3668–3679 (2001).
48. W. K. Huh, J. V. Falvo, L. C. Gerke, A. S. Carroll, R. W. Howson, J. S. Weissman, E. K. O'Shea, Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
49. C. P. Albuquerque, M. B. Smolka, S. H. Payne, V. Bafna, J. Eng, H. Zhou, A multi-dimensional chromatography technology for in-depth phosphoproteome analysis. *Mol. Cell. Proteomics* **7**, 1389–1396 (2008).
50. S. B. Ficarro, M. L. McClelland, P. T. Stukenberg, D. J. Burke, M. M. Ross, J. Shabanowitz, D. F. Hunt, F. M. White, Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **20**, 301–305 (2002).
51. A. Gruhler, J. V. Olsen, S. Mohammed, P. Mortensen, N. J. Faergeman, M. Mann, O. N. Jensen, Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol. Cell. Proteomics* **4**, 310–327 (2005).
52. X. Li, S. A. Gerber, A. D. Rudner, S. A. Beausoleil, W. Haas, J. Villen, J. E. Elias, S. P. Gygi, Large-scale phosphorylation analysis of  $\alpha$ -factor-arrested *Saccharomyces cerevisiae*. *J. Proteome Res.* **6**, 1190–1197 (2007).

53. A. Chi, C. Huttenhower, L. Y. Geer, J. J. Coon, J. E. Syka, D. L. Bai, J. Shabanowitz, D. J. Burke, O. G. Troyanskaya, D. F. Hunt, Analysis of phosphorylation sites on proteins from *Saccharomyces cerevisiae* by electron transfer dissociation (ETD) mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2193–2198 (2007).
54. M. B. Smolka, C. P. Albuquerque, S. H. Chen, H. Zhou, Proteome-wide identification of in vivo targets of DNA damage checkpoint kinases. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 10364–10369 (2007).
55. I. Muñoz, E. Simón, N. Casals, J. Clotet, J. Ariño, Identification of multicopy suppressors of cell cycle arrest at the G<sub>1</sub>-S transition in *Saccharomyces cerevisiae*. *Yeast* **20**, 157–169 (2003).
56. Y. Ho, A. Gruhler, A. Heilbut, G. D. Bader, L. Moore, S. L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutillier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreau, B. Musk, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A. R. Willems, H. Sassi, P. A. Nielsen, K. J. Rasmussen, J. R. Andersen, L. E. Johansen, L. H. Hansen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B. D. Sorensen, J. Matthiesen, R. C. Hendrickson, F. Gleeson, T. Pawson, M. F. Moran, D. Durocher, M. Mann, C. W. Hogue, D. Figeys, M. Tyers, Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002).
57. E. Kinoshita, E. Kinoshita-Kikuta, K. Takiyama, T. Koike, Phosphate-binding tag, a new tool to visualize phosphorylated proteins. *Mol. Cell. Proteomics* **5**, 749–757 (2006).
58. D. R. Stanford, M. L. Whitney, R. L. Hurto, D. M. Eisaman, W. C. Shen, A. K. Hopper, Division of labor among the yeast Sol proteins implicated in tRNA nuclear export and carbohydrate metabolism. *Genetics* **168**, 117–127 (2004).
59. J. L. Snead, M. Sullivan, D. M. Lowery, M. S. Cohen, C. Zhang, D. H. Randle, J. Taunton, M. B. Yaffe, D. O. Morgan, K. M. Shokat, A coupled chemical-genetic and bioinformatic approach to Polo-like kinase pathway exploration. *Chem. Biol.* **14**, 1261–1272 (2007).
60. D. L. Sheridan, Y. Kong, S. A. Parker, K. N. Dalby, B. E. Turk, Substrate discrimination among mitogen-activated protein kinases through distinct docking sequence motifs. *J. Biol. Chem.* **283**, 19511–19520 (2008).
61. M. Remm, C. E. Storm, E. L. Sonnhammer, Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* **314**, 1041–1052 (2001).
62. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
63. J. J. Ward, L. J. McGuffin, K. Bryson, B. F. Buxton, D. T. Jones, The DISOPRED server for the prediction of protein disorder. *Bioinformatics* **20**, 2138–2139 (2004).
64. R. Adamczak, A. Porollo, J. Meller, Accurate prediction of solvent accessibility using neural networks-based regression. *Proteins* **56**, 753–767 (2004).
65. E. Frank, M. Hall, L. Trigg, G. Holmes, I. H. Witten, Data mining in bioinformatics using Weka. *Bioinformatics* **20**, 2479–2481 (2004).
66. M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, D. G. Higgins, Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
67. M. Clamp, J. Cuff, S. M. Searle, G. J. Barton, The Jalview Java alignment editor. *Bioinformatics* **20**, 426–427 (2004).
68. A. D. McLachlan, Tests for comparing related amino-acid sequences. Cytochrome c and cytochrome C<sub>951</sub>. *J. Mol. Biol.* **61**, 409–424 (1971).
69. G. H. Golub, C. F. Van Loan, *Matrix Computations* (Johns Hopkins University Press, Baltimore, MD, ed. 3, 1996).
70. G. Zhu, K. Fujii, Y. Liu, V. Codrea, J. Herrero, S. Shaw, A single pair of acidic residues in the kinase major groove mediates strong substrate preference for P-2 or P-5 arginine in the AGC, CAMK, and STE kinase families. *J. Biol. Chem.* **280**, 36372–36379 (2005).
71. J. K. Holmes, M. J. Solomon, The role of Thr160 phosphorylation of Cdk2 in substrate recognition. *Eur. J. Biochem.* **268**, 4647–4652 (2001).
72. G. Zhu, K. Fujii, N. Belkina, Y. Liu, M. James, J. Herrero, S. Shaw, Exceptional disfavor for proline at the P + 1 position among AGC and CAMK kinases establishes reciprocal specificity between them and the proline-directed kinases. *J. Biol. Chem.* **280**, 10743–10748 (2005).
73. R. M. Xu, G. Carmel, R. M. Sweet, J. Kuret, X. Cheng, Crystal structure of casein kinase-1, a phosphate-directed protein kinase. *EMBO J.* **14**, 1015–1023 (1995).
74. G. E. Crooks, G. Hon, J. M. Chandonia, S. E. Brenner, WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
75. Funding: This study was supported by US NIH grants to M.S., B.E.T. (GM079498), N.M.H. (GM50717), W.A.L. (GM55040), and D.F.S. (CA82257), by Howard Hughes Medical Institute (W.A.L.), and by a Swiss National Science Foundation grant to C.D. Author contributions: J.M., S.P., G.R.J., D.L.S., S.A.P., V.D., and B.E.T. performed experiments; P.M.K., H.Y.K.L., and M.B.G. performed computational work; J.M., X.Z., G.R.J., S.A.P., V.D., M.J., E.C., H.N., M.G., A.R., J.N.M., Y.S., H.E.S., R.S., C.S.M.C., C.D., N.M.H., W.A.L., D.F.S., B.S., and B.J.A. prepared and characterized protein kinases and expression constructs; J.M., P.M.K., H.Y.K.L., M.B.G., M.S., and B.E.T. designed experiments, analyzed data, and wrote the paper. Competing interests: M.S. consults for Afformix, which has an interest in proteomics, including phosphoproteomics.