Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 325 (2004) 1443–1448
BBRC
www.elsevier.com/locate/ybbrc

GPS: a novel group-based phosphorylation predicting and scoring method

Feng-Feng Zhou[a,1], Yu Xue[b,1], Guo-Liang Chen[a], Xuebiao Yao[b,c,*]

[a] *Department of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui 230027, PR China*
[b] *School of Life Science, University of Science and Technology of China, Hefei 230027, PR China*
[c] *Department of Physiology, Morehouse School of Medicine, Atlanta, GA 30310, USA*

Received 22 October 2004
Available online 11 November 2004

Abstract

Protein phosphorylation is an important reversible post-translational modification of proteins, and it orchestrates a variety of cellular processes. Experimental identification of phosphorylation site is labor-intensive and often limited by the availability and optimization of enzymatic reaction. In silico prediction may facilitate the identification of potential phosphorylation sites with ease. Here we present a novel computational method named GPS: group-based phosphorylation site predicting and scoring platform. If two polypeptides differ by only two consecutive amino acids, in particular when the two different amino acids are a conserved pair, e.g., *isoleucine* (I) and *valine* (V), or *serine* (S) and *threonine* (T), we view these two polypeptides bearing similar 3D structures and biochemical properties. Based on this rationale, we formulated GPS that carries greater computational power with superior performance compared to two existing phosphorylation sites prediction systems, ScanSite 2.0 and PredPhospho. With database in public domain, GPS can predict substrate phosphorylation sites from 52 different protein kinase (PK) families while ScanSite 2.0 and PredPhospho offer at most 30 PK families. Using PKA as a model enzyme, we first compared prediction profiles from the GPS method with those from ScanSite 2.0 and PredPhospho. In addition, we chose an essential mitotic kinase Aurora-B as a model enzyme since ScanSite 2.0 and PredPhospho offer no prediction. However, GPS offers satisfactory sensitivity (94.44%) and specificity (97.14%). Finally, the accuracy of phosphorylation on MCAK predicted by GPS was validated by experimentation, in which six out of seven predicted potential phosphorylation sites on MCAK (Q91636) were experimentally verified. Taken together, we have generated a novel method to predict phosphorylation sites, which offers greater precision and computing power over ScanSite 2.0 and PredPhospho.

Keywords: GPS; Phosphorylation; Protein kinases; PKA; Aurora-B

In the eukaryotic cells, protein phosphorylation is one of the most ubiquitous post-translational modifications of proteins, orchestrating most of the cellular processes, including the cell cycle [1], transcriptional [2] and translational regulations [3], metabolic pathways [4], signal transductions [5], and the memory [6], etc. About 2% of the human and mouse proteomes encode protein kinases (PKs) with 518 and 540 distinct PKs determined in human [7] and mouse [8], respectively, among which 510 are the reciprocal orthology pairs. It was estimated that one-third of all the proteins could be phosphorylated, and about half of kinome were disease- or cancer-related by chromosomal mapping [7]. So it is in urgent need to identify the substrates accompanied with their phosphorylation sites in large-scale Phosphoproteome, which would help the drug design greatly. To date, several large-scale phosphoproteomics researches have been published on yeast [9], mouse [10], human [11,12] or plant [13], etc.

PKs' substrates and their sites can be investigated in vivo or in vitro [14], although they are time-consuming,

* Corresponding author. Fax: +86-551-3607141.
E-mail address: yaoxb@ustc.edu.cn (X. Yao).
[1] These authors contributed equally to this work.

0006-291X/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2004.11.001

labor-intensive, and expensive. Recently, there has been an intensive interest in developing novel technologies to identify phosphorylation sites in large scale, such as mass spectrometry (MS) [14], peptide microarray [15], and phosphospecific proteolysis [16].

On the other hand, in silico prediction of phosphorylation sites based on primary protein sequences is much desirable and popular for its convenience and fast speed. There is a widely adopted rule that PKs' substrates could be phosphorylated at the specific sites with consensus sequences/motifs/functional patterns [17]. NetPhos was based on artificial neural networks (ANN) [18], which outperformed the consensus-sequence-based methods. However, it could not provide information about the corresponding kinases. The enhanced version, NetPhosK, incorporated the functionality of providing kinases' information, including ~17 PKs [19]. Another prediction system, Scansite [20], constructed the profiles of phosphorylation sites of ~20 kinases, and could predict potential phosphorylation sites accompanied with their kinases. This method only requires phosphopeptides even without full-length protein sequences for profile training. Another similar method, PredPhospho, is based on Support Vector Machines (SVM, classes of ANN) [21].

3D structure conservation/similarity was also regarded as an important characteristic to improve the substrate specificity [22,23]. These structural-based methods show excellent performance. But the 3D structure information of proteins is very limited compared to the huge number of proteins in the public databases. So these approaches still remain in their infancy.

Most of the current in silico prediction systems are based on ANNs, including SVMs. But they have very little biochemical meaning to biologists. In this article we design a more-meaningful method based on the substitution matrix, which includes many newly considered kinases.

Methods

Data collection. We get the data set of phosphorylation sites from Phospho.ELM [24] which also includes the data of PhosphoBase [25]. After removing the phosphorylation sites with ambiguous information of PKs, we get 1404 items. We also manually checked the recent publications and got 597 more items. After clustering some homology PKs with too few known phosphorylation sites into a unique group, we got 52 PK families/PK groups, including ABL, ALK, AMPK, ATM, AURORA-B, BTK, CAK, CAM-II, CDK, CHK, CK1, CK2, DAPK, DNA-PK, EGFR, EPHA, FAK, FGFR, FYN, GRK, GSK3, IGFR, IKK, IR, JAK, LCK, LYN, MAPKK, MAPK, MAPKAPK2, MET, MTOR, NEK, P34CDC2, PAK, PDGFR, PDK, PHK, PKA, PKB, PKC, PKG, PKR, PLK, ROCK, S6K, SGK, SRC, SYK, TRK, VEGFR, and ZAP70, etc. Each group should contain at least 10 experimental verified phosphorylation sites.

Scoring strategy. A phosphorylation site with m upstream and n downstream amino acids, respectively, is called a *phosphorylation site peptide* PSP (m,n). The biochemical characteristics of a phosphorylation site mainly depend on the neighboring amino acids [17,22]. So in this article, we only consider the heptapeptide sequence PSP$(3,3)$ [22].

For the PSP$(3,3)$ of a known phosphorylation site, and a given peptide sequence with length 7 AA, if all the amino acids except one are the same according to their positions, we may assume with confidence that the given peptide sequence can also be phosphorylated by the same kinase of the known phosphorylation site, especially when the pair of different amino acids has similar biochemical properties. For example, such pairs are *isoleucine* (I) and *valine* (V), or *serine* (S) and *threonine* (T). Actually, there are two examples in our data set of known phosphorylation sites for the above pairs respectively (see in Table 1). We use the amino acid substitution matrix BLOSUM62 [26] to evaluate the similarity between two peptide sequences with length 7 AA. Although other matrices could be used, the BLOSUM 62 matrix is chosen here.

For two amino acids a and b, let the substitution score between them in BLOSUM62 be Score(a,b). The *similarity* between two peptides A and B with length 7 AA is defined as:

$$S(\mathrm{A},\mathrm{B}) = \sum_{1\leqslant i\leqslant 7} \mathrm{Score}(\mathrm{A}[i],\mathrm{B}[i])$$

If $S(\mathrm{A},\mathrm{B}) < 0$, we redefine it as $S(\mathrm{A},\mathrm{B}) = 0$. And if $S(\mathrm{A},\mathrm{B}) > 0$, the *distance* between them is defined as: $D(\mathrm{A},\mathrm{B}) = 1/S(\mathrm{A},\mathrm{B})$. If $S(\mathrm{A},\mathrm{B}) = 0$, $D(\mathrm{A},\mathrm{B}) = \infty$.

As described in the above, two peptides may have similar biochemical characteristics if the score between them is high enough.

Clustering strategy. Taking all the PSP$(3,3)$ of a given kinase K as nodes, we connect them with edges whose weight is the distance between the pair of nodes. The nodes can be partitioned into several clusters according to the distances between them. If a peptide sequence with length 7 AA is close enough to one of the clusters, we may assume that this peptide can also be phosphorylated by kinase K. As in Fig. 1, we assume that the nodes in cluster 1 and 2 represent the PSP$(3,3)$ of kinase K. Since node P_1 is closer to one of the clusters than P_2, we have more confidence to predict the corresponding peptide of P_1 as a potential phosphorylation site of kinase K. If we take all the known sites as one cluster, it will be difficult to distinguish P_1 and P_2.

Since there are limited number of known phosphorylation sites, such a strategy may fail to identify a potential phosphorylation site only when this peptide belongs to a cluster whose nodes are all unknown peptides.

We adopt the Markov Cluster Algorithm (MCL for short) [27,28] to partition the above graph into several clusters. After this operation, we get a set of clusters of phosphorylation site peptides for each kinase, respectively.

Group-based phosphorylation scoring method (GPS). Based on the above clustering and scoring strategies, we designed the following algorithm to generate a score for a potential phosphorylation site P of kinase K.

Table 1
Four pairs of known phosphorylation sites with only one different amino acid

Substrate	Position	PK	PSP(3,3)	PMID
P08238	S225	CK2	KEISDDE	92065884
P07900	S230	CK2	KEVSDDE	89123325
P20700	S392	CDK	LSPSPSS	90263082
P11516	S392	CDK	LSPSPTS	93238752
P15127	T1376	PKC	RVLTLPR	92065884
P06213	T1375	PKC	RILTLPR	92065884
P19491	S717	PKC	VRKSKGK	96275134
P19490	T710	PKC	VRKTKGK	96275134

The data are from Phospho.ELM (integrated with PhosphoBase).

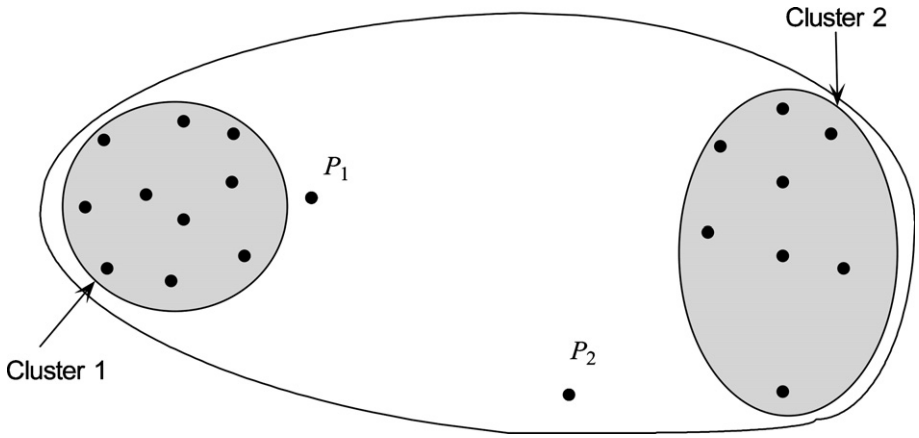


Fig. 1. Why we should partition the known phosphorylation sites into clusters. We assume that the nodes in clusters 1 and 2 represent the PSP(3,3) of kinase K. Since node P_1 is closer to one of the clusters than P_2, we have more confidence to predict the corresponding peptide of P_1 as a potential phosphorylation site of kinase K. If we take all the known sites as one cluster, it will be difficult to distinguish P_1 and P_2.

Algorithm. ***G**roup-based **P**hosphorylation **S**coring method (GPS)*

***Input*:** Kinase K, and the set C of clusters of K got in the above section. A potential phosphorylation site peptide sequence *Seq*.

***Output*:** The calculated score for the possibility that *Seq* is a phosphorylation site peptide of kinase K.

***Operations*:**

For each cluster C_i in C

{

  For each peptide P_j in C_i

  {

    Calculate the score S(*Seq*, P_j)

  }

  $$S_i = \left\{ \sum_j S(Seq, P_j) \right\} \Big/ |C_i|$$

}

$$S(Seq) = \max_j \{S_i\}$$

Return S(*Seq*).

If the score of a peptide sequence by GPS for kinase K is high enough, we may assert that this peptide is a potential phosphorylation site of kinase K. Such a method is so simple that biologists may calculate the scores by hand. And we may also ignore the clustering strategy, and only calculate the score between a given peptide sequence and a known phosphorylation site to illustrate the possibility that the given peptide is a potential phosphorylation site.

Results and discussion

Performance on kinase PKA

We try to evaluate the performance of GPS method against two popular phosphorylation prediction systems, ScanSite 2.0 [20] and PredPhospho [21]. ScanSite 2.0 provides phosphorylation site prediction for 26 kinases, while PredPhospho only provides such functionality for four groups and four families of kinases. Besides most of the kinases in the above two systems, GPS method also includes several kinases which came into focus recently, e.g., Aurora-B. In the following, we will mainly evaluate the performance of the three systems on kinase PKA.

Sensitivity (S_n) and specificity (S_p) are used to evaluate the prediction system's performance. The known phosphorylation sites of PKA are regarded as the positive data, while all the other [S, T, and Y] sites in the known phosphorylation substrates of PKA are regarded as the negative data. For the data which are predicted as positive, the real positive ones are called *true positives* (TP), while the others are called *false positives* (FP). For the data which are predicted as negative, the real positive ones are called *false negatives* (FN), while the others are called *true negatives* (TN). The sensitivity and specificity are defined as follows:

$$S_n = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad \text{and} \quad S_p = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}.$$

As illustrated in Table 2, when a loose cut-off value 2 is applied, GPS (91.81%) can get a similar sensitivity to PredPhospho (93.60%), which outperforms that of ScanSite with any stringency (26.75%, 50.05%, and 70.72% for high, medium, and low stringency, respectively). At the same time, GPS can still get a satisfying specificity, 85.02% against 91.34% for PredPhospho, and 99.95%, 96.92%, and 92.86% for ScanSite with high, medium, and low stringency, respectively. To increase the specificity, we apply a more stringent cut-off value 3. From Table 2, the specificity of GPS method with the cut-off value 3 is increased to 92.15%, while the sensitivity is reduced by 9.35%. GPS method still outperforms ScanSite over the sensitivity, while keeping a satisfying specificity. Although ScanSite outperforms GPS method over specificity, the sensitivity is a bit too low, especially under the high and medium stringencies.

Table 2
Comparison of sensitivity and specificity of phosphorylation site predictions for PKA among GPS, ScanSite 2.0, and PredPhospho

PKA	ScanSite with stringency			PredPhospho (%)	GPS with cut-off value	
	High (%)	Medium (%)	Low (%)		2 (%)	3 (%)
S_n	26.75	50.05	70.72	93.60	91.81	82.46
S_p	99.95	96.92	92.86	91.34	85.02	92.15

Scansite has three stringencies: high, medium, and low, while the PredPhospho scores each predicted sites, and all predictions were adopted. The cut-off values of GPS are set as 2 and 3 for comparison, separately.

PredPhospho is a good prediction system for phosphorylation sites, but they only provide the prediction for four groups and four families of kinases. Our method can predict phosphorylation sites for 52 kinases. So GPS method may be a good complementary tool to ScanSite 2.0 and PredPhospho.

We also make a leave-one-out validation on kinase PKA. There are only 17 known phosphorylation sites with scores lower than the cut-off value 2 (90.56%, that is 173 out of 180). For a more stringent cut-off value 3, the value is increased to 3 (81.67%, that is 147 out of 180).

To illustrate the performance of GPS method on the random sequences, we randomly generated 1000 serine sites and threonine sites. The distributions of GPS scores of serine sites and threonine sites are as in Figs. 2A and B, respectively.

Performance of kinase Aurora-B

Protein kinase Aurora-B is a member of the Aurora/Ipl1 family, which is important in cell division [29], orchestrating chromosome segregation [30] and progression of cytokinesis [31]. Aurora/Ipl1 family is much

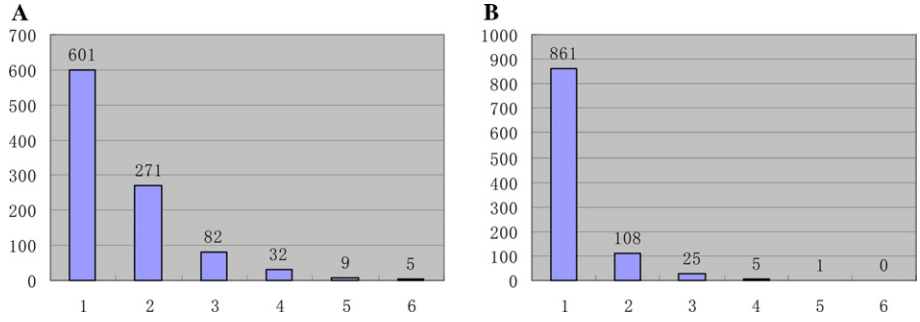
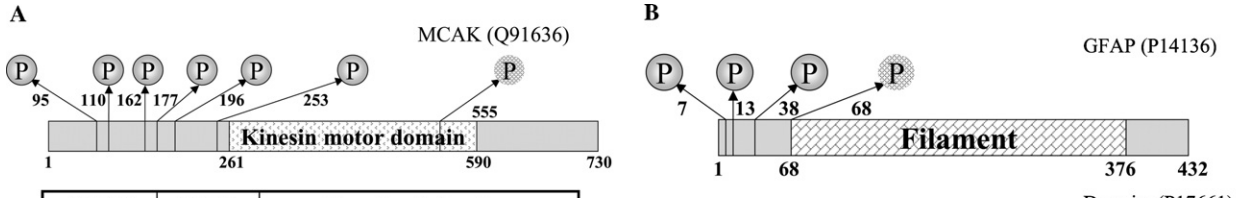


Fig. 2. The distributions of GPS scores of kinase PKA on randomly generated heptapeptides of serine (XXXSXXX) and threonine (XXXTXXX), which X represent any amino acid. With the cut-off value as 2, ~8% will be predicted as positive totally. And only 2.6% is predicted as positive with cut-off as 3, totally. (A) We randomly generate 1000 serine heptapeptides. If we follow the cut-off value as 2, then 12.8% will be predicted as positive, while only 4.6% is predicted as positive with cut-off as 3. (B) The distribution of GPS scores on 1000 threonine heptapeptides. Following the cut-off as 2, 3.1% will be predicted as positive, with 0.6% positive prediction with the cut-off as 3.



MCAK	PMID	Phosphorylation sites
Ref. 1 [37]	14972678	S161, T162, S177, S196, T229, and S253
Ref. 2 [38]	15064354	T95, S110, S161, S177, and S196
GPS	---	T95, S110, S161, S177, S196, S253, and S555

Fig. 3. Validation of GPS predicted results of kinase Aurora-B against experimental results. (A) The prediction of *Xenopus* MCAK (Swissprot Accession No.: Q91636). GPS method predicts seven sites as positive (T95, S110, S161, S177, S196, S253, and S555), of which six sites (T95, S110, S161, S177, S196, and S253) were experimentally verified as phosphorylation sites of PK Aurora-B [37,38]. (B) Aurora-B can phosphorylate type III intermediate filaments GFAP (Swissprot Accession No.: P14136) and Desmin (Swissprot Accession No.: P17661). Our GPS method predicts four and three sites on GFAP (T7, S13, S38, and S68) and Desmin (T16, S44, and S59), respectively, which the phosphorylation sites T7, S13, and S38 of GFAP with T16 and S59 were experimental verified.

conserved in eukaryotic organisms, even between budding yeast and human. During mitosis, Aurora-B will localize on kinetochore in metaphase, forming a protein complex with INCENP, Survivin, and Borealin [32]. And it will move to midbody in cytokinesis [33]. Aurora-B is overexpressed in several cancer cells [34] and implicated in tumorigenesis [35]. So it is in urgent need to map the phosphorylation substrates with their sites of Aurora-B.

In the literature, we manually curated 10 well-known phosphorylation substrates with 19 sites of Aurora-B. By leave-one-out test, there is only one known phosphorylation site with score (3.22) lower than the cut-off value 3.7 (sensitivity: 94.44%, that is 17 out of 18). The specificity of Aurora-B with cut-off value 3.7 can even reach 97.14%.

It is reported that Aurora-B could phosphorylate the microtubule depolymerase MCAK (mitotic centromere-associated kinesin) during cell division and control its centromere/kinetochore localization and catalytic activity [36]. Two research groups mapped the phosphorylation sites on MCAK by Mass Spectrometry method [37,38]. Here we validate our predictions against the experimental results (see in Fig. 3A). Our GPS method predicts seven potential phosphorylation sites on *Xenopus* MCAK (Q91636), among which six predicted sites are in at least one of the experiments (T95, S110, S161, S177, S196, and S253). Another predicted phosphorylation site S555 may need to be verified in vivo or in vitro.

Another example is the phosphorylation of the type III intermediate filaments GFAP and Desmin by Aurora-B at cleavage furrow/midbody during cytokinesis, which will reduce their filament forming ability [31]. There are three and two verified phosphorylated sites of Aurora-B on GFAP (T7, S13, and S38) and Desmin (T16 and S59), respectively. And our GPS approach predicts four and three sites on GFAP (T7, S13, S38, and S68) and Desmin (T16, S44, and S59), respectively (see in Fig. 3B).

So GPS method on kinase Aurora-B may be a helpful tool to experimentalists, who focus on mitosis dynamics.

Acknowledgments

We thank Dr. T.J. Gibson and Dr. F. Diella for providing the data set of Phospho.ELM for this study. This work was supported by grants from Chinese Natural Science Foundation (39925018 and 30121001), Chinese Academy of Science (KSCX2-2-01), Chinese 973 project (2002CB713700), and American Cancer Society (RPG-99-173-01) to X. Yao. X. Yao is a GCC Distinguished Cancer Research Scholar.

References

[1] Y. Lou, J. Yao, A. Zereshki, Z. Dou, K. Ahmed, H. Wang, J. Hu, Y. Wang, X. Yao, NEK2A interacts with MAD1 and possibly functions as a novel integrator of the spindle checkpoint signaling, J. Biol. Chem. 279 (2004) 20049–20057.

[2] S. Uddin, F. Lekmine, A. Sassano, H. Rui, E.N. Fish, L.C. Platanias, Role of Stat5 in type I interferon-signaling and transcriptional regulation, Biochem. Biophys. Res. Commun. 308 (2003) 325–330.

[3] F. Yoshizawa, E. Watanabe, K. Sugahara, Y. Natori, Translational initiation regulators are hypophosphorylated in rat liver during ethionine-mediated ATP depletion, Biochem. Biophys. Res. Commun. 298 (2002) 235–239.

[4] A.J. Meijer, P.F. Dubbelhuis, Amino acid signalling and the integration of metabolism, Biochem. Biophys. Res. Commun. 313 (2004) 397–403.

[5] S. Choudhary, A. Kumar, R.K. Kale, L.G. Raisz, C.C. Pilbeam, Extracellular calcium induces COX-2 in osteoblasts via a PKA pathway, Biochem. Biophys. Res. Commun. 322 (2004) 395–402.

[6] P.K. Dash, S.A. Orsi, M. Moody, A.N. Moore, A role for hippocampal Rho-ROCK pathway in long-term spatial memory, Biochem. Biophys. Res. Commun. 322 (2004) 893–898.

[7] G. Manning, D.B. Whyte, R. Martinez, T. Hunter, S. Sudarsanam, The protein kinase complement of the human genome, Science 298 (2002) 1912–1934.

[8] S. Caenepeel, G. Charydczak, S. Sudarsanam, T. Hunter, G. Manning, The mouse kinome: discovery and comparative genomics of all mouse protein kinases, Proc. Natl. Acad. Sci. USA 101 (2004) 11707–11712.

[9] S.B. Ficarro, M.L. McCleland, 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 (2002) 301–305.

[10] B.A. Ballif, J. Villen, S.A. Beausoleil, D. Schwartz, S.P. Gygi, Phosphoproteomic analysis of the developing mouse brain, Mol. Cell. Proteomics (2004) [Epub ahead of print].

[11] S.A. Beausoleil, M. Jedrychowski, D. Schwartz, J.E. Elias, J. Villen, J. Li, M.A. Cohn, L.C. Cantley, S.P. Gygi, Large-scale characterization of HeLa cell nuclear phosphoproteins, Proc. Natl. Acad. Sci. USA 101 (2004) 2130–2135.

[12] Y.P. Lim, L.S. Diong, R. Qi, B.J. Druker, R.J. Epstein, Phosphoproteomic fingerprinting of epidermal growth factor signaling and anticancer drug action in human tumor cells, Mol. Cancer Ther. 2 (2003) 1369–1377.

[13] T.S. Nuhse, A. Stensballe, O.N. Jensen, S.C. Peck, Phosphoproteomics of the *Arabidopsis* plasma membrane and a new phosphorylation site database, Plant Cell 16 (2004) 2394–2405.

[14] C. Kraft, F. Herzog, C. Gieffers, K. Mechtler, A. Hagting, J. Pines, J.M. Peters, Mitotic regulation of the human anaphase-promoting complex by phosphorylation, EMBO J. 22 (2003) 6598–6609.

[15] L. Rychlewski, M. Kschischo, L. Dong, M. Schutkowski, U. Reimer, Target specificity analysis of the Abl kinase using peptide microarray data, J. Mol. Biol. 336 (2004) 307–311.

[16] Z.A. Knight, B. Schilling, R.H. Row, D.M. Kenski, B.W. Gibson, K.M. Shokat, Phosphospecific proteolysis for mapping sites of protein phosphorylation, Nat. Biotechnol. 21 (2003) 1047–1054. Erratum in: Nat. Biotechnol. 21 (2003) 1396..

[17] A. Kreegipuu, N. Blom, S. Brunak, J. Jarv, Statistical analysis of protein kinase specificity determinants, FEBS Lett. 430 (1998) 45–50.

[18] N. Blom, S. Gammeltoft, S. Brunak, Sequence and structure-based prediction of eukaryotic protein phosphorylation sites, J. Mol. Biol. 294 (1999) 1351–1362.

[19] 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 (2004) 1633–1649.
[20] J.C. Obenauer, L.C. Cantley, M.B. Yaffe, Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs, Nucleic. Acids. Res. 31 (2003) 3635–3641.
[21] J.H. Kim, J. Lee, B. Oh, K. Kimm, I. Koh, Prediction of phosphorylation sites using SVMs, Bioinformatics (2004) [Epub ahead of print].
[22] R.I. Brinkworth, R.A. Breinl, B. Kobe, Structural basis and prediction of substrate specificity in protein serine/threonine kinases, Proc. Natl. Acad. Sci. USA 100 (2003) 74–79.
[23] L. Li, E.I. Shakhnovich, L.A. Mirny, Amino acids determining enzyme-substrate specificity in prokaryotic and eukaryotic protein kinases, Proc. Natl. Acad. Sci. USA 100 (2003) 4463–4468.
[24] F. Diella, S. Cameron, C. Gemund, R. Linding, A. Via, B. Kuster, T. Sicheritz-Ponten, N. Blom, T.J. Gibson, Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins, BMC Bioinformatics 5 (2004) 79.
[25] A. Kreegipuu, N. Blom, S. Brunak, PhosphoBase, a database of phosphorylation sites: release 2.0, Nucleic Acids Res. 27 (1999) 237–239.
[26] S. Henikoff, J.G. Henikoff, Amino acid substitution matrices from protein blocks, Proc. Natl. Acad. Sci. USA 89 (1992) 10915–10919.
[27] Stijn van Dongen, Graph Clustering by Flow Simulation, PhD thesis, University of Utrecht, May 2000.
[28] A.J. Enright, S. Van Dongen, C.A. Ouzounis, An efficient algorithm for large-scale detection of protein families, Nucleic Acids Res. 30 (2002) 1575–1584.
[29] Y.W. Ke, Z. Dou, J. Zhang, X.B. Yao, Function and regulation of Aurora/Ipl1p kinase family in cell division, Cell Res. 13 (2003) 69–81.
[30] P. Meraldi, R. Honda, E.A. Nigg, Aurora kinases link chromosome segregation and cell division to cancer susceptibility, Curr. Opin. Genet. Dev. 14 (2004) 29–36.
[31] A. Kawajiri, Y. Yasui, H. Goto, M. Tatsuka, M. Takahashi, K. Nagata, M. Inagaki, Functional significance of the specific sites phosphorylated in desmin at cleavage furrow: Aurora-B may phosphorylate and regulate type III intermediate filaments during cytokinesis coordinatedly with Rho-kinase, Mol. Biol. Cell 14 (2003) 1489–1500.
[32] R. Gassmann, A. Carvalho, A.J. Henzing, S. Ruchaud, D.F. Hudson, R. Honda, E.A. Nigg, D.L. Gerloff, W.C. Earnshaw, Borealin: a novel chromosomal passenger required for stability of the bipolar mitotic spindle, J. Cell Biol. 166 (2004) 179–191.
[33] A.R. Skop, H. Liu, J. Yates III, B.J. Meyer, R. Heald, Dissection of the mammalian midbody proteome reveals conserved cytokinesis mechanisms, Science 305 (2004) 61–66.
[34] R.R. Adams, D.M. Eckley, P. Vagnarelli, S.P. Wheatley, D.L. Gerloff, A.M. Mackay, P.A. Svingen, S.H. Kaufmann, W.C. Earnshaw, Human INCENP colocalizes with the Aurora-B/AIRK2 kinase on chromosomes and is overexpressed in tumour cells, Chromosoma 110 (2001) 65–74.
[35] H. Katayama, W.R. Brinkley, S. Sen, The Aurora kinases: role in cell transformation and tumorigenesis, Cancer Metastasis Rev. 22 (2003) 451–464.
[36] G.J. Gorbsky, Mitosis: MCAK under the aura of Aurora B, Curr. Biol. 14 (2004) R346–R348.
[37] W. Lan, X. Zhang, S.L. Kline-Smith, S.E. Rosasco, G.A. Barrett-Wilt, J. Shabanowitz, D.F. Hunt, C.E. Walczak, P.T. Stukenberg, Aurora B phosphorylates centromeric MCAK and regulates its localization and microtubule depolymerization activity, Curr. Biol. 14 (2004) 273–286.
[38] R. Ohi, T. Sapra, J. Howard, T.J. Mitchison, Differentiation of cytoplasmic and meiotic spindle assembly MCAK functions by Aurora B-dependent phosphorylation, Mol. Biol. Cell 15 (2004) 2895–2906.