

## REVIEW

# Experimental and computational tools useful for (re)construction of dynamic kinase–substrate networks

Chris Soon Heng Tan<sup>1,2,3</sup> and Rune Linding<sup>4</sup>

<sup>1</sup>Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada

<sup>2</sup>Department of Molecular Genetics, University of Toronto, Toronto, Canada

<sup>3</sup>Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Canada

<sup>4</sup>Cellular and Molecular Logic Team, Section of Cell and Molecular Biology, The Institute of Cancer Research (ICR), London, UK

The explosion of site- and context-specific *in vivo* phosphorylation events presents a potentially rich source of biological knowledge and calls for novel data analysis and modeling paradigms. Perhaps the most immediate challenge is delineating detected phosphorylation sites to their effector kinases. This is important for (re)constructing transient kinase–substrate interaction networks that are essential for mechanistic understanding of cellular behaviors and therapeutic intervention, but has largely eluded high-throughput protein–interaction studies due to their transient nature and strong dependencies on cellular context. Here, we surveyed some of the computational approaches developed to dissect phosphorylation data detected in systematic proteomic experiments and reviewed some experimental and computational approaches used to map phosphorylation sites to their effector kinases in efforts aimed at reconstructing biological signaling networks.

Received: April 30, 2009  
Revised: August 11, 2009  
Accepted: August 22, 2009

**Keywords:**

Bioinformatics / Databases / Protein phosphorylation / Protein–protein interaction / Proteome analysis / Systems biology

## 1 Introduction

Extra- and intracellular cues are constantly assimilated and processed by the cell to elicit appropriate cellular behaviors, such as proliferation, differentiation, migration and apoptosis, needed for survival and morphological development. These cues are propagated, processed and integrated through complex cellular protein signaling networks, which are constantly being modulated by mechanisms such as reversible post-translational modifications [1], differential protein expression and degradation, that enable a cell to

respond differently to identical stimulus depending on the state of the cellular networks [2, 3].

Protein phosphorylation is probably the most prevalent reversible post-translational modification in cells, estimated to affect one-third of cellular proteins [4], and is known to influence proteins' enzymatic activities, 3-D structure, subcellular localization and interactions with other biomolecules [5]. The dynamic balance between protein phosphorylation (by protein kinases) and their dephosphorylation (*e.g.* by phosphatases or protein degradation) in response to combinatorial intra- and inter-cellular cues impose dynamic structures in protein–interaction networks which function as molecular switches or logic gates to regulate cellular activities [1, 6]. Regulation errors in protein phosphorylation or dephosphorylation often result in dysfunctional cellular processes leading to cancer and complex regulatory diseases [7, 8]. Hyper-phosphorylated retinoblastoma protein, for instance, is linked to multiple cancers [9–11] while hypo-phosphorylated retinoblastoma protein is implicated in adult T-cell lymphoma progression [12]. Hence, deciphering the modulation of cellular protein–interaction networks by

**Correspondence:** Dr. Rune Linding, Cellular & Molecular Logic Team, Section of Cell and Molecular Biology, The Institute of Cancer Research (ICR), London, SW3 6JB, UK

**E-mail:** linding@icr.ac.uk

**Fax:** +44-207-153-5078

**Abbreviations:** AS, analogue-sensitive; EGF, epidermal growth factor; FCM, fuzzy c-means; PLSR, partial least-square regression; SOM, self-organizing map

protein phosphorylation events is essential for understanding and modeling cellular behaviors including disease formation.

The enzymatic phosphorylation of proteins was probably first described by Burnett and Kennedy in 1954 [13, 14]. By the 1980s, the formations of some cancers have been identified as the repercussions of aberrant phosphorylation events [15]. Through the 1980s and 1990s, phosphoproteins and their phosphorylated sites continued to be discovered with their functions deciphered in a hypothesis-driven manner through targeted biochemistry and genetic experiments [13]. Despite this, our understanding of protein phosphorylation in the context of cell signaling and tumorigenesis remains relatively fragmented. A probable reason is the abstraction or reductionist's approach whereby complex and multi-factored diseases like cancer are attributed to one or a few genes when cellular processes, like the transformation of a normal cell to a cancerous cell, arise from complex interplay between cellular proteins through protein–protein interactions and post-translational modifications such as protein phosphorylation [1, 5, 8].

Phosphoproteins were traditionally detected using inorganic phosphate isotopes like  $P^{32}$  or using antibodies that target phosphor residues. Advances in instrumentation, phosphopeptide enrichment techniques and spectra searching algorithms have recently propelled MS as the method of choice for phosphorylation analysis because the approach allows rapid identification of numerous cellular phosphoproteins and their phosphorylated sites. An advantage of MS-based assays is that numerous phosphorylation sites can be detected under relevant physiological and cellular conditions (see, *e.g.* [16]). As phosphorylation events often occur in low stoichiometry (especially for tyrosine phosphorylation), enrichment of trypsin-derived phosphopeptides using affinity reagents like antibodies, IMAC, [17, 18], strong cation exchange chromatography (SCX) [19],  $TiO_2$  (titanium dioxide chromatography) [20, 21] often precede phosphopeptide identification by MS. With the development of quantitative approaches like stable isotope labeling by amino acids in cell culture (SILAC) [22], isotope-tagged amine-reactive reagents (iTRAQ) [23] and isotope-coded affinity tags (ICATs) [24] to differentially label proteins from different samples, the relative quantitative level of phosphorylation of many sites under different biological conditions can be unraveled.

The explosion of site- and context-specific *in vivo* phosphorylation events presents a potentially rich source of biological knowledge and calls for novel data analysis and modeling paradigms. The most immediate challenge is likely delineating detected phosphorylation sites to their effector kinases. This is important for (re)constructing the transient kinase–substrate interaction networks that are essential for mechanistic understanding of cellular behaviors and therapeutic intervention [1, 8] but have largely eluded high-throughput protein-interaction studies [25, 26] due to their transient nature and strong dependencies on cellular context. The rapid detection of thousand of phos-

phorylation sites and their dynamics under different cellular conditions, cell lines and tissues has also necessitated computational analysis of the underlying data. Here, we surveyed some of the computational and experimental approaches that have been adopted to analyze large-scale phosphorylation data sets and to (re)construct transient kinase–substrate interaction networks.

## 2 Computational analysis of large-scale or dynamic phospho-proteomic data

### 2.1 Clustering

Receptor kinases are a class of kinases that assimilate extracellular signals to initiate appropriate cellular behaviors. Genome-wide studies of phosphorylation events initiated by one such receptor kinase, epidermal growth factor (EGF) receptor, have been conducted where sites differentially phosphorylated at time points 1, 5, 10 and 20 min after EGF activation in HeLa cells were obtained using quantitative phospho-proteomic techniques [27, 28]. These studies revealed the dynamic temporal nature of protein phosphorylation where different sites are preferentially up-, unchanged or down-phosphorylated at different times after EGF stimulation. This time-series data permits the grouping of sites with similar temporal profile to facilitate biological analysis. Fuzzy c-means (FCM) clustering was adopted for the purpose. Unlike other hard-partition methods like k-means and self-organizing map (SOM), FCM allows an instance to belong to different clusters with different score that add up to 1. Hence, FCM allows confident identification of members of a cluster based on low-resolution data.

In work from Forest White's lab [29], the temporal dynamic of a large number of tyrosine phosphorylation sites were analyzed under four time points (0, 5, 10 and 30 min) after 5, 10 and 30 min incubation of human mammary epithelial cell line with 25 nM EGF with untreated cells as control. SOM with U-matrix method of visualizing was used to identify co-regulated phosphorylation sites. In essence, SOM is a technique for mapping high-dimensional data (in this case, each site has 16 features corresponding to the four time points under the four different treatments) to lower dimension, often 2-D, that facilitate manual grouping of clusters by visual inspection. An advantage of SOM is it allows an overview of each cluster's similarities with other clusters.

### 2.2 Partial least-square regression

Partial least-square regression (PLSR) is a class of regression technique that combined data compression technique (through principal component analysis) with regression to predict dependent variables using input variables from limited samples. In situations where the number of input variables exceeds the number of observations, or when the

input variables exhibit multi co-linearity (meaning some variables are highly correlated), or when there are missing data, PLSR is a preferred choice for regression analysis over other conventional regression techniques. In a landmark Systems Biology paper, Yaffe and Laufenburger labs used PLSR to capture the quantitative physiological status of 11 signaling proteins at multiple time points after stimulation by combination of three cytokines (tumor necrosis factor, EGF and insulin) in HT-29 cells to predict cellular responses with high accuracy [3]. The work highlighted the importance of including the dynamic response of signaling molecules upon cytokine stimulation in regression analysis as using only the input cytokine concentration failed to predict the observed cellular responses. PLSR was subsequently used, by a similar group of researchers, to correlate the temporal dynamics of tyrosine phosphorylation sites with migration and proliferation cellular behaviors mediated by ErbB2 family of tyrosine receptors [30, 31]. Other works had also used PLSR to correlate multiple signaling events to observed cellular behaviors [2, 32]. A groundbreaking revelation from these studies is that apparently genetically identical cells can react differently to the same stimulus depending on the current signaling network state [1].

### 2.3 Motif-based analysis

Another common data analysis strategy is identifying over-represented sequence patterns among detected phosphorylation sites that may correspond to phosphorylation motifs of kinases. This strategy can detect novel phosphorylation motifs of uncharacterized kinases [33] or novel binding motifs of phosphoresidue binding domains like SH2 and PTB [5]. Although phosphorylation or binding motifs could be determined by *in vitro* methods like protein microarray and degenerate oriented peptide libraries (see below), the experiments were conducted in *in vitro* context and thus may not fully reflect the motifs *in vivo*. Motif discovery tools such as Gibbs motif samplers [34], MEME [35], PRATT [36], TEIRESIAS [37] and D-STAR [38] can be used to discover motifs [39] from sets of phosphorylation sites detected in phospho-proteomic experiments. For example, PRATT was previously used to extract phosphorylation motifs in kinases' substrates detected from protein chip experiments [40].

A shortcoming of many generic motif discovery algorithms is they do not explicitly correct for high occurrence of some amino acids in proteins that are not necessarily related to phosphorylation, leading to false-positive motifs. MotifX [41] is a recent motif extraction program specially developed to extract phosphorylation motifs from phospho-proteomic data in which background frequencies of amino acids are incorporated to reduce false positives. MotifX was applied in various studies to find known and novel motifs in mammalian species [42–44] and in *Arabidopsis* [45]. Motifs extracted by MotifX are restricted with either all amino acids or a single amino acid is allowed at each position, therefore

exclude motifs with degenerate positions like the pY-x-x-[LIV] motif of JAK2 kinase [46]. Moreover, the greedy iterative nature of the algorithm could potentially omit the discovery of some motifs. MoDL is a motif discovery algorithm created to extract degenerated motifs found in phosphorylation data [47] using the principle of minimum description length.

Motif extraction from phospho-proteomic data coupled with downstream experimental validation could lead to discovery of novel *in vivo* motifs of kinases and phosphoresidue binding domains. This is exemplified in Miller *et al.* [48] where a novel binding motif for a SH2 domain in inositol 5-phosphatase 2 SHIP2 was discovered. In the work, 481 unique phosphotyrosine peptides detected using tandem MS experiments on mammalian cell lines were grouped into 20 clusters followed by motif extraction using TEIRESIAS. A novel N-terminal hydrophobic motif [DE]-x-x-x-[ILV]-[ILV]-pY was extracted from one of the cluster, in which three out of four peptides expressing the motifs were validated to bind SHIP2 in pull-down assays. Mutational analysis on two amino acid positions immediate N-terminal to the phosphotyrosine confirmed the generality of the motif. Interestingly, proteins expressing the motif are enriched with “cell surface receptor linked signal transduction” GO term, in line with known association of SH2-containing protein with receptor-linked signaling. The work is probably the first system-wide approach that combined both bioinformatics analyses and experimental validation to discover novel motif. As our knowledge of specificity motifs of kinases and phosphoresidue binding domains increases, proteomics researchers can rely more and more on motif atlases; this is one possible usage of the NetPhorest algorithm (see later) [49].

## 3 Experimental mapping of kinase–substrate networks

Kinases typically interact transiently with their substrates to transfer a phosphate group to phospho-acceptor residues on substrates [1]. The transient nature of these kinase–substrate interactions renders their detection by conventional protein–interaction assays challenging. Here, we highlight an “orthogonal” set of experimental methods that can be useful for identifying substrates of different kinases and, hence, facilitate the construction of kinase–substrate networks.

### 3.1 Protein and peptide microarray

Multiple putative protein substrates can be immobilized on a solid support, such as glass slide or streptavidin membrane, as miniature protein array or proteome chip [50]. They can then be overlaid with isotope-labeled ATP and purified kinase of interest [51, 52]. Phosphorylated substrates can subsequently be detected using high-resolution phosphorimaging

[53]. This approach allows rapid screening for putative substrates of a specific kinase using small amount of reagents and can be scaled up for genome-wide assay [50]. For example, using microarrays containing 4400 unique yeast proteins [52], Ptacek and colleagues tested 82 yeast protein kinases and identified 4200 phosphorylation events on 1325 yeast proteins [40]. In another study, potential substrates of Abl and Abl-related gene (Arg) tyrosine kinases were assessed using a microarray containing 2400 different human proteins [54].

Kinase–substrate interactions detected using protein microarray described above may not occur physiologically due to lack of biological context, such as cellular co-localization and/or protein co-expression between kinases and their detected substrates. Moreover, high kinase concentration that does not reflect physiological or cellular level is often used to increase sensitivity of these assays. Many physiological kinase–substrate interactions can be missed because contextual factors like adaptor proteins or co-activators (*e.g.* cyclins) and priming phosphorylation sites or kinases are often not present in the arrays. Nevertheless, protein microarray assays serve to identify potential kinase–substrate relations that can either be validated by downstream biochemical and genetic experiments or corroborated with biological data from existing studies.

Peptides with phospho-acceptor residue at a fixed position can be immobilized on a chip just like full-length proteins, and subsequently incubated with isotope-labeled ATP and kinase of interest, followed by phosphorimaging to identify phosphorylated peptides. Peptides spotted on microarrays can be random sequences [55], from degenerate oriented peptide libraries [56] or are subsequences found in proteins [32, 57, 58]. If peptides corresponding to subsequences in proteins are used, the precise sites of phosphorylation on substrates can be determined. However, both false positive and false negative phosphorylation sites can be detected by this approach as the 3-D structural context of the phospho-acceptor residues which can affect phosphorylation are not represented in the assays. On the other hand, if random peptides or degenerate oriented peptide libraries are used, the observed phosphorylated peptides can be used to derive a position-specific scoring matrix, as a statistical model, to quantify the phosphorylation propensity of a phospho-acceptor residue based on amino acids flanking the residue [49, 59]. The position-specific scoring matrix can then be used to scan a proteome to identify putative substrates for the kinase assayed [49, 59]. Exact sites of phosphorylation are predicted using this bioinformatics approach although the approach is known to have high false positive rates presumably because non-naturally occurring peptides devoid from the content/context of whole protein sequences are used.

A challenge for peptide chips based on random peptides is the huge possible peptide sequence space that can be assessed. An alternate approach is to incubate the kinase of interest and isotope labeled ATP in solution with a large set of possible peptide sequences of a fixed length. To facilitate

identification of sequence pattern needed for phosphorylation by a kinase of interest, peptides can be divided into individual pools where peptides in each pool match a consensus pattern. Such approach was taken by Cantley, Yaffe and Turk and coworkers to determine the sequence specificities of serine/threonine kinases [59, 60]. Unique peptide pools were generated such that in each pool, all peptides have a common amino acid at one of residue position whereas amino acids in other positions are degenerated. In Turk's approaches, a total of 198 unique peptide pools were generated as phospho-threonine, phosphotyrosine and the 20 naturally occurring unmodified amino acids were individually fixed at each of the nine residue positions flanking a central phospho-acceptor residue. Each peptide in the pools is biotin-tagged, allowing the peptides to be spotted onto a streptavidin membrane. Phosphorylated peptides can subsequently be detected by autoradiography or phosphorimaging. Amino acids preferred by the kinase of interest at each position flanking the phospho-acceptor residue can then be determined. In addition, the fixation of a phosphorylated residue at one of the position flanking a phosphorylated phospho-acceptor residue allows the detection of phosphorylation that requires priming phosphorylation sites [61].

### 3.2 Perturbation-based assays

Although protein or peptide chips can identify putative substrates of kinases, they are *in vitro* experimental techniques that do not necessarily capture the physiological or cellular concentration and co-localization factors of protein kinases and their substrates [62]. Smolka *et al.* [63] combined quantitative MS with perturbation studies to identify cellular phosphorylation sites and substrates of yeast DNA damage checkpoint kinases Mec1/Tel1 and Rad53 upon induction of DNA damage. Quantitative MS can detect sites that are differentially phosphorylated across two or more cellular conditions/perturbations but do not directly identify their effector kinases. However, by detecting phosphorylation sites that specifically altered between kinase-null (Mec1/Tel1 and Rad53) and wild-type *Saccharomyces cerevisiae*, Smolka *et al.* identified 62 putative target sites of Mec1/Tel1 and Rad53 in 55 proteins. These differentially phosphorylated sites were enriched in the known phosphorylation motifs (linear motifs) of Mec1/Tel1 and Rad53, which further support that many of the identified targets are cellular substrates of the kinases. Quantitative MS approach was also adopted to identify proteins in zebrafish Fyn/Yes morpholino knockdown embryos that were differentially phosphorylated compared to those in wild-type embryos [64]. Using similar approaches, Matsuoka *et al.* identified putative ATR and ATM phosphorylation sites that were altered upon DNA damage in human embryonic kidney 293T cells, and as such should correspond to physiological targets of the two kinases. Putative phosphorylation sites of ATR and ATM were identified using antibodies against pSQ

or pTQ sites that are known to target by ATM and ATR kinases. Among the phosphorylation sites detected, 905 sites among 700 proteins were found up-regulated fourfold after induction of DNA damage, of which 55 sites were found on 31 ATR and ATM substrates known to be implicated in DNA damage signaling. Hence, the 700 proteins are possible physiological substrates of ATM and ATR. The accuracy and scalability of this approach depend on the availability of suitable antibodies and their qualities. As many kinases are known to have similar specificities, the specificities of the antibodies are an important consideration in such approach.

A potential pitfall in above-mentioned approaches is that not all perturbed sites identified are direct targets of the deleted kinase. Instead, they could be targets of other kinases that are activated “downstream” of the deleted kinases in signaling cascades. For example, in Smolka *et al.* [63], about half of the sites down-phosphorylated in Mec1/Tel1 mutant are also down-phosphorylated in Rad53 mutant. As these site expressed Rad53 phosphorylation motifs and given that Rad53 acted downstream of Mec1/Tel1, they are likely targets of Rad53. The issue of direct *versus* indirect action by cellular kinases was addressed in Linding *et al.* [62] where we discussed and demonstrated that the combination of quantitative perturbations (in contrast to gene deletion) with MS and computational algorithms is a powerful approach to accurately map phosphorylation sites to their effector kinases [1, 62, 65].

### 3.3 Chemical genetics approaches

As mentioned above, a challenge in perturbation approaches is that one cannot always be certain which kinase(s) phosphorylated the observed altered sites as many kinases could share similar consensus motifs or targets. By structural alteration (through mutagenesis) of a kinase such that it can only incorporate a specific modified form of ATP, a detected phosphorylated protein containing the modified ATP is most likely targeted by the mutant kinase [66–68]. Cells (NIH 3T3) with such analogue-sensitive (AS) mutant of v-Src kinase were generated, lysed and incubated with analog ATP to identify putative substrates of v-Src under more realistic concentration of proteins found *in vivo* [69]. An AS mutant of Pho85 kinase was generated similarly and assayed for putative substrates in whole-cell extracts of yeast [70]. A similar approach was combined with bioinformatics search by Ubersax *et al.* [71] to identify *in vivo* substrates of Cdk1. The phosphorylation detection was performed on potential substrates expressing a Cdk1 consensus phosphorylation motif (S/T-P-x-K/R) using cell lysates incubated with purified AS mutant Cdk1 with analog ATP to identify 181 proteins that are efficiently phosphorylated by the AS mutant Cdk1. To further validate some of the identified substrates *in vivo*, small molecules that inhibit the AS mutant Cdk1 were added to cultures of yeast AS mutant Cdk1 strains to detect proteins with decreased phosphor-

ylation and hence likely to be targeted by Cdk1 *in vivo*. A total of 12 high confidence *in vivo* substrates of Cdk1 were identified in this manner. An issue with these approaches is that it can not be ruled out that ATP analog can be picked up and utilized by other distantly related kinases or that structural changes in the AS kinase has altered its *in vivo* specificity. Another potential concern is the scalability of the assays for genome-wide studies as it is unclear presently whether AS mutants can be created for most kinases.

## 4 Determining dynamics in kinase–substrate networks

### 4.1 Phospho-specific flow cytometry

Besides conventional immunoblotting using phosphorylation specific antibodies, only a few methods have the throughput and sensitivity to analyse the phosphorylation state of multiple signaling proteins simultaneously. One of these, phospho-specific flow cytometry, has provided insight to dysfunctional signaling networks at single cell level [72, 73]. Cells are first labeled by antibodies recognising phosphorylated proteins. Subsequently, cells are sorted and measured by FACS thereby providing quantitative measurement of cell specific phosphorylation events. This technique allows cells from heterogeneous samples to be sorted and assayed. This method is a promising way to examine the state of the signaling network in tumour samples for deciding optimum therapeutic strategies.

### 4.2 Real time imaging and integration with systems genetics

Development in microscopic techniques has recently allowed real-time measurement of signaling events and analysis of sub-cellular localisation of proteins regulated by phosphorylation. For example, a combination of FRET with phosphatase–substrate interactions analysis has recently provided insight into the function of PTP1B in signal-termination of RTKs [74, 75]. Real-time imaging of cellular morphology combined with RNAi disruption of signaling proteins constitutes a strategy to directly link cellular signaling events with phenotypic measurements with machine-learning approaches [76]. Recently, we in collaboration with Chris Bakal and Norbert Perrimon took this further by using combinatorial RNAi knockdown and a FRET-reporter to read-out the activation state of JNK kinase. Subsequently, the lists of enhancers and suppressors of JNK in different genetic backgrounds were combined with phospho-proteomics data to model a JNK phosphorylation network using NetworKIN and NetPhorest [65]. This resulted in the first systems-level model of a signaling system based on integration of systems-genetics and phospho-proteomics data.

## 5 Computational approaches for (re)construction of phosphorylation networks

Bioinformatics approaches using simple consensus motif searching can be used to associate identified phosphorylated sites or proteins to their effector kinases [63, 71, 77]. However, relying on simple regular expressions/patterns can be highly unspecific [49]. Thus, several methods have been developed to better identify potential phosphorylation sites of specific kinases or kinase groups. Here, we surveyed the methods and tools available (summarized in Table 1) and highlighted that their applications coupled with additional biological data can potentially increase the accuracy of kinase–substrate detection.

### 5.1 Machine-learning approaches

A subset of these tools deployed machine-learning algorithms to predict novel phosphorylation sites. The basic methodology involves training machine-learning models with positive examples and negative examples of sites phosphorylated by kinases of interest, and then validating the models on separate data set consisting of both positive and negative samples. The resulting computational models can subsequently be applied to new data to predict potential phosphorylation sites of specific kinases. Support vector machines, a statistical machine learning method, had been deployed in KinasePhos [78] and PredPhospho [79] for

predicting kinase-specific phosphorylation sites. Similarly, artificial neural networks and Bayesian Decision theory were employed in NetPhosK [80, 81], GANNPhos [82] and PPSP [83] to predict kinase-specific phosphorylation sites. Meta-PredPS [84], a meta-predictor, combined predictions from GPS [85], KinasePhos, PPSP, PredPhospho and Scansite [59] through a generalized weighted voting strategy to achieve better prediction for phosphorylation sites targeted by four protein kinase families (CDK, CK2, PKA and PKC).

### 5.2 Similarity based approaches

Other alternative approaches have been adopted to predict kinase-specific phosphorylation sites: GPS 2.0 [85] predicts kinase-specific phosphorylation sites in a query sequence based on the sites' similarity to known sites of kinases. To improve prediction performance, a derivative of the BLOSUM 62 substitution matrix was derived for each kinase group to optimize similarity comparisons between the sites. Predikin is a prediction tool that infers phosphorylation motifs using sequence-structure analysis for uncharacterized serine/threonine kinase sequences submitted by user [86]. The pkaPS [87] method uses a simplified analytical model to score physical and chemical requirements at amino acid positions from –18 to +23 of phospho-acceptor residues to predict putative phosphorylation sites of protein kinase A (cAMP-dependent kinase, PKA).

**Table 1.** List of bioinformatics resources for phosphorylation studies

Name	URLs	References
<b>Site-Kinase Predictor</b>		
GPS	<a href="http://gps.biocuckoo.org">http://gps.biocuckoo.org</a>	[85]
KinasePhos	<a href="http://kinasephos2.mbc.nctu.edu.tw">http://kinasephos2.mbc.nctu.edu.tw</a>	[78]
MetaPredPS	<a href="http://metapred.umn.edu/MetaPredPS">http://metapred.umn.edu/MetaPredPS</a>	[84]
NetPhosK	<a href="http://www.cbs.dtu.dk/services/NetPhosK">http://www.cbs.dtu.dk/services/NetPhosK</a>	[80, 81]
NetPhorest	<a href="http://netphorest.info">http://netphorest.info</a>	[49]
NetworkKIN	<a href="http://networkin.info">http://networkin.info</a>	[62]
pkaPS	<a href="http://mendel.imp.ac.at/sat/pkaPS">http://mendel.imp.ac.at/sat/pkaPS</a>	[87]
Predikin	<a href="http://predikin.biosci.uq.edu.au/pkr">http://predikin.biosci.uq.edu.au/pkr</a>	[86]
PredPhospho	<a href="http://www.nih.go.kr/phosphovariant/html/predphospho.htm">http://www.nih.go.kr/phosphovariant/html/predphospho.htm</a>	[79]
Scansite	<a href="http://scansite.mit.edu">http://scansite.mit.edu</a>	[59]
<b>Phosphorylation Motif Resources</b>		
NetPhorest	<a href="http://netphorest.info">http://netphorest.info</a>	[49]
PhosphoMotif Finder	<a href="http://www.hprd.org/PhosphoMotif_finder">http://www.hprd.org/PhosphoMotif_finder</a>	[90]
Scansite	<a href="http://scansite.mit.edu">http://scansite.mit.edu</a>	[59]
<b>Phosphorylation Site Resources</b>		
Phosida	<a href="http://www.phosida.com">http://www.phosida.com</a>	[91]
Phospho.ELM	<a href="http://phospho.elm.eu.org">http://phospho.elm.eu.org</a>	[79]
PhosphoSitePlus	<a href="http://www.phosphosite.org">http://www.phosphosite.org</a>	[92]
PhosphoPep	<a href="http://www.phosphopep.org">http://www.phosphopep.org</a>	[93]

### 5.3 Contextual modeling of kinase specificity – NetworKIN

Computational and *in vitro* experimental detection of kinases' substrates and their phosphorylation sites often omitted contextual factors like sub-cellular compartmentalization and differentiated protein expression that can prevent phosphorylation. In addition, positive factors that co-regulate phosphorylation such as co-localization *via* anchoring proteins, scaffolds and substrate capture by non-catalytic interaction domains and docking motifs are often not captured in these experiments. These factors, in combination with the challenges of mapping transient and context-dependent kinase–substrate interactions using current protein–interaction assays, have in part led to a large gap between our understanding of *in vivo* phosphorylation sites and the kinases that modulate them. In the Phospho.ELM database [88], there are currently thousands of annotated *in vivo* phosphorylation sites, of which only  $\sim 1/4$  have been linked to at least one *in vivo* kinase [89].

To address this problem, we developed the NetworKIN algorithm to predict *in vivo* kinases for identified phosphorylation sites [62]. We demonstrated that the contextual information of kinases and substrates is crucial for accurate modeling of signaling networks, and demonstrated the capability of such approach in NetworKIN algorithm in reliable assembly of kinase–substrate networks [62]. The principle behind this algorithm is to model kinase specificity using contextual information for phosphoproteins and kinases in combination with sequence models of kinase consensus motifs [49]. By combining probabilistic modeling of network context with the linear motifs recognized by the catalytic kinase domain, NetworKIN can assign a specific kinase to an observed *in vivo* phosphorylation site with a 2.5-fold higher accuracy than previous methods such as Scansite and NetphosK. However, even given powerful predictions, it is still challenging to experimentally validate kinase substrate relationships, primarily due to the issue of direct *versus* indirect interactions, but also due to the challenges related to the use of phospho-specific antibodies [62]. Therefore, we proposed that a combination of quantitative perturbations (using chemical inhibitors, RNAi or stimulation's) with quantitative MS (specifically multiple reaction monitoring) and NetworKIN could allow systematic validation of cellular kinase–substrate interactions. As proof of principle, we showed that the BCLAF1 protein is likely a direct target of GSK3 since inclusion of a GSK3 inhibitor resulted in a decrease of the predicted target phosphorylation (S531). Although including contextual data markedly increases the accuracy of kinase–substrate relationship predictions, these methods are obviously still prone to errors in at least three sources: the experimentally determined phosphorylation sites, consensus sequence motifs and probabilistic association network [62]. As experimental approaches improve, these errors will continuously be eliminated.

### 5.4 The human kinome specificity atlas – NetPhorest

To meet the requirement of robust classification of sites (not prediction) according to their effector kinases, we recently extended the coverage of NetworKIN with the NetPhorest atlas of linear motifs involved in phosphorylation-dependent signaling [49]. This atlas includes kinase consensus motifs for 179 kinases and 104 phosphorylation-dependent binding domains including SH2 and PTB domains. The NetPhorest consists of a framework to automate data set construction and training of sequence models for linear motifs involved in phosphorylation mediated signaling. With the technologies described above, such as peptide arrays and MS, a rapid increase in the amount of phosphorylation and linear motif data are expected, underlining the importance of having a fully automated system for training, benchmarking, and selection of final sequence models. We envisage that the NetPhorest atlas will be useful for many different aspects of signaling- and proteomics-related research, for example, for generation of antibodies that recognize specific phosphorylated residues or sequences (phospho-specific antibodies). In particular, kinase “consensus motif antibodies” such as pS/T-Q are important tools for both targeted studies and systems biology studies [77]. The atlas can be used to predict peptide sequences for raising such antibodies. MS-based proteomics researchers frequently use simple sequence patterns for finding systematic biases in phosphopeptides identified in shotgun approaches. Such studies will benefit from the more robust and comprehensive collection of motifs in the NetPhorest atlas. NetPhorest will be used in future versions of resources like Scansite, NetPhosK, as well as NetworKIN. The ability of the NetPhorest framework to continuously incorporate the latest data will help accelerate systems-level modeling of cellular signaling [65]. We aim to develop NetPhorest into a global linear motif atlas by including other eukaryotic model organisms, such as yeast, as well as other types of post-translational modifications. Therefore, we continue to encourage the scientific community to help expand the coverage of this resource by submitting data and purified active enzymes to the NetPhorest consortium.

### 5.5 Outlook

Cell signaling networks are the foundation of cell fate and behavior. The aberrant activities of signal transduction networks are key mechanisms underlying the pathological behavior of cells during tumor development. However, signaling networks are remarkably complex, involving a myriad of dynamic interactions that flux in space and time. Thus, understanding how aberrant cell decisions arise requires a global view of cell signaling networks. Therefore, it will become increasingly important to use combinations of experimental and computational models to understand the principles of spatio-temporal assembly of mammalian

interaction networks at a systems level, and how they transmit and change information to mediate cell fate and behavior. We would argue that only through quantitative and contextually relevant assessments of these dynamic networks can we move towards grand goals such as the ability to simulate cell behavior and cell-to-cell communication in tissue boundaries. Essential to such efforts is the correlation of these networks to quantitative phenotype descriptions and genetic data, a process we recently initiated [65].

*Thanks to Claus Jørgensen for comments on the manuscript. We apologize to colleagues with works we did not include in this review due to space constraints.*

*The authors have declared no conflict of interest.*

## 6 References

- [1] Jørgensen, C., Linding, R., Directional and quantitative phosphorylation networks. *Brief Funct. Genomic Proteomic* 2008, 7, 17–26.
- [2] Miller-Jensen, K., Janes, K. A., Brugge, J. S., Lauffenburger, D. A., Common effector processing mediates cell-specific responses to stimuli. *Nature* 2007, 448, 604–608.
- [3] Janes, K. A., Albeck, J. G., Gaudet, S., Sorger, P. K. *et al.*, A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. *Science* 2005, 310, 1646–1653.
- [4] Cohen, P., The role of protein phosphorylation in human health and disease. The Sir Hans Krebs Medal Lecture. *Eur. J. Biochem.* 2001, 268, 5001–5010.
- [5] Seet, B. T., Dikic, I., Zhou, M. M., Pawson, T., Reading protein modifications with interaction domains. *Nat. Rev. Mol. Cell Biol.* 2006, 7, 473–483.
- [6] Dueber, J. E., Mirsky, E. A., Lim, W. A., Engineering synthetic signaling proteins with ultrasensitive input/output control. *Nat. Biotechnol.* 2007, 25, 660–662.
- [7] Zheng, B., Jeong, J. H., Asara, J. M., Yuan, Y. Y. *et al.*, Oncogenic B-RAF negatively regulates the tumor suppressor LKB1 to promote melanoma cell proliferation. *Mol. Cell* 2009, 33, 237–247.
- [8] Pawson, T., Linding, R., Network medicine. *FEBS Lett.* 2008, 582, 1266–1270.
- [9] Kitagawa, M., Higashi, H., Takahashi, I. S., Okabe, T. *et al.*, A cyclin-dependent kinase inhibitor, butyrolactone I, inhibits phosphorylation of RB protein and cell cycle progression. *Oncogene* 1994, 9, 2549–2557.
- [10] Chatterjee, S. J., George, B., Goebell, P. J., Alavi-Tafreshi, M. *et al.*, Hyperphosphorylation of pRb: a mechanism for RB tumour suppressor pathway inactivation in bladder cancer. *J. Pathol.* 2004, 203, 762–770.
- [11] Broceno, C., Wilkie, S., Mittnacht, S., RB activation defect in tumor cell lines. *Proc. Natl. Acad. Sci. USA* 2002, 99, 14200–14205.
- [12] Nakayama, K., Yamada, Y., Koji, T., Hayashi, T. *et al.*, Expression and phosphorylation status of retinoblastoma protein in adult T-cell leukemia/lymphoma. *Leuk. Res.* 2000, 24, 299–305.
- [13] Pawson, T., Scott, J. D., Protein phosphorylation in signaling – 50 years and counting. *Trends Biochem. Sci.* 2005, 30, 286–290.
- [14] Burnett, G., Kennedy, E. P., The enzymatic phosphorylation of proteins. *J. Biol. Chem.* 1954, 211, 969–980.
- [15] Hunter, T., Cooper, J. A., Epidermal growth factor induces rapid tyrosine phosphorylation of proteins in A431 human tumor cells. *Cell* 1981, 24, 741–752.
- [16] Huang, P. H., Mukasa, A., Bonavia, R., Flynn, R. A. *et al.*, Quantitative analysis of EGFRvIII cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma. *Proc. Natl. Acad. Sci. USA* 2007, 104, 12867–12872.
- [17] Posewitz, M. C., Tempst, P., Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal. Chem.* 1999, 71, 2883–2892.
- [18] Ficarro, S. B., McClelland, M. L., Stukenberg, P. T., Burke, D. J. *et al.*, Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat. Biotechnol.* 2002, 20, 301–305.
- [19] Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E. *et al.*, Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. USA* 2004, 101, 12130–12135.
- [20] Pinkse, M. W., Uitto, P. M., Hilhorst, M. J., Ooms, B., Heck, A. J., Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal. Chem.* 2004, 76, 3935–3943.
- [21] Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P., Jørgensen, T. J., Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell Proteomics* 2005, 4, 873–886.
- [22] Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B. *et al.*, Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell Proteomics* 2002, 1, 376–386.
- [23] Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B. *et al.*, Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell Proteomics* 2004, 3, 1154–1169.
- [24] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F. *et al.*, Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 1999, 17, 994–999.
- [25] Björklund, A. K., Light, S., Hedin, L., Elofsson, A., Quantitative assessment of the structural bias in protein-protein interaction assays. *Proteomics* 2008, 8, 4657–4667.
- [26] Jensen, L. J., Bork, P., Biochemistry. Not comparable, but complementary. *Science* 2008, 322, 56–57.
- [27] Olsen, J. V., Blagoev, B., Gnäd, F., Macek, B. *et al.*, Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* 2006, 127, 635–648.
- [28] Blagoev, B., Kratchmarova, I., Ong, S. E., Nielsen, M. *et al.*, A proteomics strategy to elucidate functional protein-

- protein interactions applied to EGF signaling. *Nat. Biotechnol.* 2003, 21, 315–318.
- [29] Zhang, Y., Wolf-Yadlin, A., Ross, P. L., Pappin, D. J. *et al.*, Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol. Cell Proteomics* 2005, 4, 1240–1250.
- [30] Wolf-Yadlin, A., Kumar, N., Zhang, Y., Hautaniemi, S. *et al.*, Effects of HER2 overexpression on cell signaling networks governing proliferation and migration. *Mol. Syst. Biol.* 2006, 2, 54.
- [31] Kumar, N., Wolf-Yadlin, A., White, F. M., Lauffenburger, D. A., Modeling HER2 effects on cell behavior from mass spectrometry phosphotyrosine data. *PLoS Comput. Biol.* 2007, 3, e4.
- [32] Gordus, A., Krall, J. A., Beyer, E. M., Kaushansky, A. *et al.*, Linear combinations of docking affinities explain quantitative differences in RTK signaling. *Mol. Syst. Biol.* 2009, 5, 235.
- [33] Joughin, B. A., Naegle, K. M., Huang, P. H., Yaffe, M. B. *et al.*, An integrated comparative phosphoproteomic and bioinformatic approach reveals a novel class of MPM-2 motifs upregulated in EGFRvIII-expressing glioblastoma cells. *Mol. Biosyst.* 2009, 5, 59–67.
- [34] Lawrence, C. E., Altschul, S. F., Boguski, M. S., Liu, J. S. *et al.*, Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science* 1993, 262, 208–214.
- [35] Bailey, T. L., Elkan, C., Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 1994, 2, 28–36.
- [36] Jonassen, I., Efficient discovery of conserved patterns using a pattern graph. *Comput. Appl. Biosci.* 1997, 13, 509–522.
- [37] Rigoutsos, I., Floratos, A., Combinatorial pattern discovery in biological sequences: The TEIRESIAS algorithm. *Bioinformatics* 1998, 14, 55–67.
- [38] Tan, S. H., Hugo, W., Sung, W. K., Ng, S. K., A correlated motif approach for finding short linear motifs from protein interaction networks. *BMC Bioinformatics* 2006, 7, 502.
- [39] Neduva, V., Linding, R., Su-Angrand, I., Stark, A. *et al.*, Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS Biol.* 2005, 3, e405.
- [40] Ptacek, J., Devgan, G., Michaud, G., Zhu, H. *et al.*, Global analysis of protein phosphorylation in yeast. *Nature* 2005, 438, 679–684.
- [41] Schwartz, D., Gygi, S. P., An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat. Biotechnol.* 2005, 23, 1391–1398.
- [42] Smith, J. C., Duchesne, M. A., Tozzi, P., Ethier, M., Figeys, D., A differential phosphoproteomic analysis of retinoic acid-treated P19 cells. *J. Proteome Res.* 2007, 6, 3174–3186.
- [43] Pan, C., Gnad, F., Olsen, J. V., Mann, M., Quantitative phosphoproteome analysis of a mouse liver cell line reveals specificity of phosphatase inhibitors. *Proteomics* 2008, 8, 4534–4546.
- [44] Ballif, B. A., Carey, G. R., Sunyaev, S. R., Gygi, S. P., Large-scale identification and evolution indexing of tyrosine phosphorylation sites from murine brain. *J. Proteome Res.* 2008, 7, 311–318.
- [45] Sugiyama, N., Nakagami, H., Mochida, K., Daudi, A. *et al.*, Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in Arabidopsis. *Mol. Syst. Biol.* 2008, 4, 193.
- [46] Argetsinger, L. S., Kouadio, J. L., Steen, H., Stensballe, A. *et al.*, Autophosphorylation of JAK2 on tyrosines 221 and 570 regulates its activity. *Mol. Cell Biol.* 2004, 24, 4955–4967.
- [47] Ritz, A., Shakhnarovich, G., Salomon, A. R., Raphael, B. J., Discovery of phosphorylation motif mixtures in phosphoproteomics data. *Bioinformatics* 2009, 25, 14–21.
- [48] Miller, M. L., Hanke, S., Hinsby, A. M., Friis, C. *et al.*, Motif decomposition of the phosphotyrosine proteome reveals a new N-terminal binding motif for SHIP2. *Mol. Cell Proteomics* 2008, 7, 181–192.
- [49] Miller, M. L., Jensen, L. J., Diella, F., Jorgensen, C. *et al.*, Linear motif atlas for phosphorylation-dependent signaling. *Sci. Signal.* 2008, 1, ra2.
- [50] MacBeath, G., Schreiber, S. L., Printing proteins as microarrays for high-throughput function determination. *Science* 2000, 289, 1760–1763.
- [51] Zhu, H., Klemic, J. F., Chang, S., Bertone, P. *et al.*, Analysis of yeast protein kinases using protein chips. *Nat. Genet.* 2000, 26, 283–289.
- [52] Zhu, H., Bilgin, M., Bangham, R., Hall, D. *et al.*, Global analysis of protein activities using proteome chips. *Science* 2001, 293, 2101–2105.
- [53] Johnston, R. F., Pickett, S. C., Barker, D. L., Autoradiography using storage phosphor technology. *Electrophoresis* 1990, 11, 355–360.
- [54] Boyle, S. N., Michaud, G. A., Schweitzer, B., Predki, P. F., Koleske, A. J., A critical role for cortactin phosphorylation by Abl-family kinases in PDGF-induced dorsal-wave formation. *Curr. Biol.* 2007, 17, 445–451.
- [55] Rychlewski, L., Kschischo, M., Dong, L., Schutkowski, M., Reimer, U., Target specificity analysis of the Abl kinase using peptide microarray data. *J. Mol. Biol.* 2004, 336, 307–311.
- [56] Songyang, Z., Blechner, S., Hoagland, N., Hoekstra, M. F. *et al.*, Use of an oriented peptide library to determine the optimal substrates of protein kinases. *Curr. Biol.* 1994, 4, 973–982.
- [57] Lemeer, S., Jopling, C., Naji, F., Ruijtenbeek, R. *et al.*, Protein-tyrosine kinase activity profiling in knock down zebrafish embryos. *PLoS ONE* 2007, 2, e581.
- [58] Landgraf, C., Panni, S., Montecchi-Palazzi, L., Castagnoli, L. *et al.*, Protein interaction networks by proteome peptide scanning. *PLoS Biol.* 2004, 2, E14.
- [59] Yaffe, M. B., Leparo, G. G., Lai, J., Obata, T. *et al.*, A motif-based profile scanning approach for genome-wide prediction of signaling pathways. *Nat. Biotechnol.* 2001, 19, 348–353.
- [60] Huttli, J. E., Jarrell, E. T., Chang, J. D., Abbott, D. W. *et al.*, A rapid method for determining protein kinase phosphorylation specificity. *Nat. Methods* 2004, 1, 27–29.
- [61] Rena, G., Woods, Y. L., Prescott, A. R., Pegg, M. *et al.*, Two novel phosphorylation sites on FKHR that are critical for its nuclear exclusion. *EMBO J.* 2002, 21, 2263–2271.

- [62] Linding, R., Jensen, L. J., Ostheimer, G. J., van Vugt, M. A. *et al.*, Systematic discovery of *in vivo* phosphorylation networks. *Cell* 2007, 129, 1415–1426.
- [63] Smolka, M. B., Albuquerque, C. P., Chen, S. H., Zhou, H., Proteome-wide identification of *in vivo* targets of DNA damage checkpoint kinases. *Proc. Natl. Acad. Sci. USA* 2007, 104, 10364–10369.
- [64] Lemeer, S., Jopling, C., Gouw, J., Mohammed, S. *et al.*, Comparative phosphoproteomics of zebrafish Fyn/Yes morpholino knockdown embryos. *Mol. Cell Proteomics* 2008, 7, 2176–2187.
- [65] Bakal, C., Linding, R., Llense, F., Heffern, E. *et al.*, Phosphorylation networks regulating JNK activity in diverse genetic backgrounds. *Science* 2008, 322, 453–456.
- [66] Shah, K., Liu, Y., Deirmengian, C., Shokat, K. M., Engineering unnatural nucleotide specificity for Rous sarcoma virus tyrosine kinase to uniquely label its direct substrates. *Proc. Natl. Acad. Sci. USA* 1997, 94, 3565–3570.
- [67] Liu, Y., Shah, K., Yang, F., Witucki, L., Shokat, K. M., Engineering Src family protein kinases with unnatural nucleotide specificity. *Chem. Biol.* 1998, 5, 91–101.
- [68] Eblen, S. T., Kumar, N. V., Shah, K., Henderson, M. J. *et al.*, Identification of novel ERK2 substrates through use of an engineered kinase and ATP analogs. *J. Biol. Chem.* 2003, 278, 14926–14935.
- [69] Shah, K., Shokat, K. M., A chemical genetic screen for direct v-Src substrates reveals ordered assembly of a retrograde signaling pathway. *Chem. Biol.* 2002, 9, 35–47.
- [70] Dephoure, N., Howson, R. W., Blethrow, J. D., Shokat, K. M., O’Shea, E. K., Combining chemical genetics and proteomics to identify protein kinase substrates. *Proc. Natl. Acad. Sci. USA* 2005, 102, 17940–17945.
- [71] Ubersax, J. A., Woodbury, E. L., Quang, P. N., Paraz, M. *et al.*, Targets of the cyclin-dependent kinase Cdk1. *Nature* 2003, 425, 859–864.
- [72] Irish, J. M., Kotecha, N., Nolan, G. P., Mapping normal and cancer cell signalling networks: towards single-cell proteomics. *Nat. Rev. Cancer* 2006, 6, 146–155.
- [73] Irish, J. M., Hovland, R., Krutzik, P. O., Perez, O. D. *et al.*, Single cell profiling of potentiated phospho-protein networks in cancer cells. *Cell* 2004, 118, 217–228.
- [74] Verveer, P. J., Wouters, F. S., Reynolds, A. R., Bastiaens, P. I., Quantitative imaging of lateral ErbB1 receptor signal propagation in the plasma membrane. *Science* 2000, 290, 1567–1570.
- [75] Yudushkin, I. A., Schleifenbaum, A., Kinkhabwala, A., Neel, B. G. *et al.*, Live-cell imaging of enzyme-substrate interaction reveals spatial regulation of PTP1B. *Science* 2007, 315, 115–119.
- [76] Bakal, C., Aach, J., Church, G., Perrimon, N., Quantitative morphological signatures define local signaling networks regulating cell morphology. *Science* 2007, 316, 1753–1756.
- [77] Matsuoka, S., Ballif, B. A., Smogorzewska, A., McDonald, E. R., 3rd., *et al.*, ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* 2007, 316, 1160–1166.
- [78] Wong, Y. H., Lee, T.Y., Liang, H. K., Huang, C. M. *et al.*, KinasePhos 2.0: a web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns. *Nucleic Acids Res.* 2007, 35, W588–594.
- [79] Kim, J. H., Lee, J., Oh, B., Kimm, K., Koh, I., Prediction of phosphorylation sites using SVMs. *Bioinformatics* 2004, 20, 3179–3184.
- [80] Hjerrild, M., Stensballe, A., Rasmussen, T. E., Kofoed, C. B. *et al.*, Identification of phosphorylation sites in protein kinase A substrates using artificial neural networks and mass spectrometry. *J. Proteome Res.* 2004, 3, 426–433.
- [81] Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S., Brunak, S., Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* 2004, 4, 1633–1649.
- [82] Tang, Y. R., Chen, Y. Z., Canchaya, C. A., Zhang, Z., GANNPhos: a new phosphorylation site predictor based on a genetic algorithm integrated neural network. *Protein Eng. Des. Sel.* 2007, 20, 405–412.
- [83] Xue, Y., Li, A., Wang, L., Feng, H., Yao, X., PPSP: prediction of PK-specific phosphorylation site with Bayesian decision theory. *BMC Bioinformatics* 2006, 7, 163.
- [84] Wan, J., Kang, S., Tang, C., Yan, J. *et al.*, Meta-prediction of phosphorylation sites with weighted voting and restricted grid search parameter selection. *Nucleic Acids Res.* 2008, 36, e22.
- [85] Xue, Y., Ren, J., Gao, X., Jin, C. *et al.*, GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. *Mol. Cell Proteomics* 2008, 7, 1598–1608.
- [86] Brinkworth, R. I., Breinl, R. A., Kobe, B., Structural basis and prediction of substrate specificity in protein serine/threonine kinases. *Proc. Natl. Acad. Sci. USA* 2003, 100, 74–79.
- [87] Neuberger, G., Schneider, G., Eisenhaber, F., pkaPS: prediction of protein kinase A phosphorylation sites with the simplified kinase-substrate binding model. *Biol. Direct.* 2007, 2, 1.
- [88] Diella, F., Gould, C. M., Chica, C., Via, A., Gibson, T. J., Phospho.ELM: a database of phosphorylation sites – update 2008. *Nucleic Acids Res.* 2008, 36, D240–244.
- [89] Linding, R., Jensen, L. J., Pasculescu, A., Olhovskiy, M. *et al.*, NetworkKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res.* 2008, 36, D695–699.
- [90] Amanchy, R., Periaswamy, B., Mathivanan, S., Reddy, R. *et al.*, A curated compendium of phosphorylation motifs. *Nat. Biotechnol.* 2007, 25, 285–286.
- [91] Gnäd, F., Ren, S., Cox, J., Olsen, J. V. *et al.*, PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biol.* 2007, 8, R250.
- [92] Hornbeck, P. V., Chabra, I., Kornhauser, J. M., Skrzypek, E., Zhang, B., PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* 2004, 4, 1551–1561.
- [93] Bodenmiller, B., Campbell, D., Gerrits, B., Lam, H. *et al.*, PhosphoPep – a database of protein phosphorylation sites in model organisms. *Nat. Biotechnol.* 2008, 26, 1339–1340.