

Minireview

Phosphoproteomics toolbox: Computational biology, protein chemistry and mass spectrometry

Majbrit Hjerrild*, Steen Gammeltoft

Department of Clinical Biochemistry, Glostrup Hospital, Nordre Ringvej, DK-2600 Glostrup, Denmark

Received 1 June 2006; revised 21 July 2006; accepted 25 July 2006

Available online 4 August 2006

Edited by Francesc Posas

Abstract Protein phosphorylation is important for regulation of most biological functions and up to 50% of all proteins are thought to be modified by protein kinases. Increased knowledge about potential phosphorylation of a protein may increase our understanding of the molecular processes in which it takes part. Despite the importance of protein phosphorylation, identification of phosphoproteins and localization of phosphorylation sites is still a major challenge in proteomics. However, high-throughput methods for identification of phosphoproteins are being developed, in particular within the fields of bioinformatics and mass spectrometry. In this review, we present a toolbox of current technology applied in phosphoproteomics including computational prediction, chemical approaches and mass spectrometry-based analysis, and propose an integrated strategy for experimental phosphoproteomics.

© 2006 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

Keywords: Phosphoproteomics; Phosphorylation; Bioinformatics; Mass spectrometry

1. Phosphoproteomics of cellular regulation

The cellular proteome is highly dynamic because the expressed proteins, their abundance, and their post-translational modifications depend on the physiological state of the cell. Phosphorylation is one of the most common and best characterized post-translational modifications (PTM) of cellular proteins. Activation of protein kinases and phosphatases provides a powerful control of the phosphorylation state, and thus the subsequent biological process e.g. through alteration of protein activity, subcellular localization, degradation, conforma-

tion, or interaction with other proteins. Regulation of the cell cycle, membrane transport and permeability, cell adhesion, neurotransmission, and metabolism are examples of biological functions that are modulated through protein phosphorylation [1]. The human genome contains 518 different protein kinases and identification of their biological targets is an active research area [2]. Even though the number of identified phosphoproteins is rapidly increasing especially due to development of high-throughput methods for the identification of phosphoproteins, in particular within the fields of bioinformatics and mass spectrometry, it is believed that only a small fraction of physiological phosphorylation sites has been assigned. In this review a toolbox of techniques currently available for analysis of the phosphoproteome is presented (Fig. 1). Based on this overview we propose an integrated strategy for high-throughput analysis of phosphorylated proteins (Fig. 2).

2. Computational phosphoproteomics

The specificity of protein kinases is determined by e.g. acidic, basic, or hydrophobic amino acids adjacent to the phosphoacceptor site often referred to as the consensus sequence of the kinase. Alternatively, a key determinant for MAP kinase and CDK specificity is a proline in the +1 position. A wide range of computational approaches have been developed for predictions of phosphorylation sites ranging from simple motif searches to more complex methods like the artificial neural networks (ANN) where sequence correlations can be taken into account (Table 1).

Definition of sequence motifs for specific kinases have, for example, been included in the PROSITE database [3]. The drawback of this simple motif search is that the consensus sequence often is based on limited data and that the sensitivity of this approach tend to be quite low [4]. Another simple approach is the group-based phosphorylation scoring (GPS) method, which is based on comparison of the sequence surrounding the reported phosphorylation sites (three residues on each side of the site) to a given heptapeptide in the candidate substrate. This approach has recently been used for prediction of phosphorylation sites for 71 protein kinase subfamilies [5].

A more complex method for prediction of phosphorylation sites is based on weight matrices which defines more diverse patterns and makes it possible to rank the predicted phosphorylation sites. Scansite is a weight matrix-based kinase-specific phosphorylation sites prediction server containing more than

*Corresponding author. Fax: +45 43 23 39 29.
E-mail address: mh@dcb-glostrup.dk (M. Hjerrild).

Abbreviations: ANN, artificial neural network; ESI, electrospray ionization; GPS, group-based phosphorylation site predicting and scoring platform; HPLC, high performance liquid chromatography; HPRD, human protein reference database; IMAC, immobilized metal affinity chromatography; LC-MS/MS, liquid chromatography tandem mass spectrometry; MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight; PTM, post-translational modification; QTOF, quadrupole time-of-flight; SDS-PAGE, sodium dodecyl sulphate-polyacrylamide gel electrophoresis; SILAC, stable isotope labelling by amino acid in cell culture; siRNA, small interfering RNA; SVM, support vector machine

PHOSPHOPROTEOMICS TOOLBOX

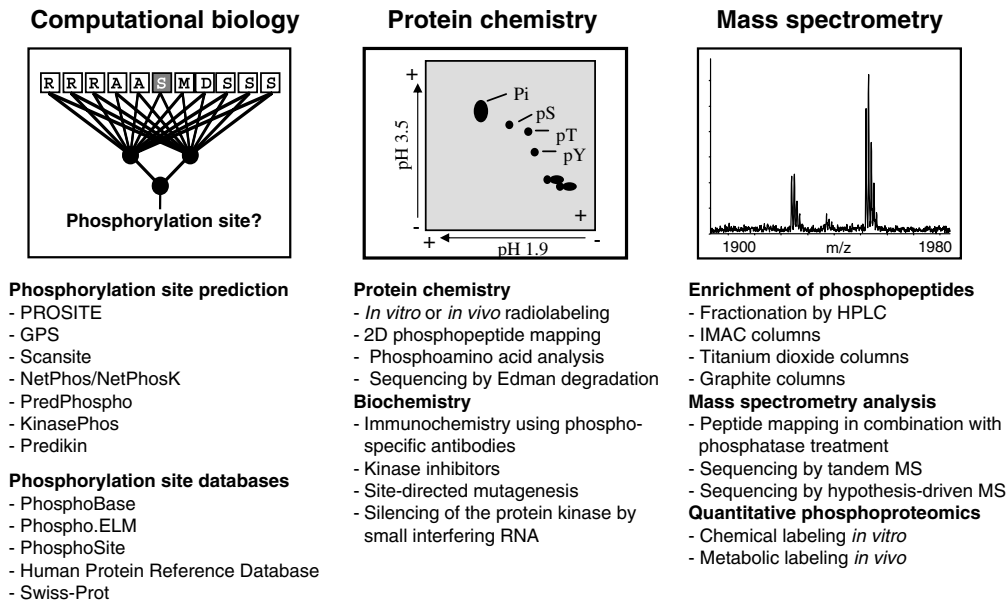


Fig. 1. Techniques currently used for phosphoproteome analysis including computational biology, protein chemistry, and mass spectrometry-based analysis.

SIX STEPS IN EXPERIMENTAL PHOSPHOPROTEOMICS

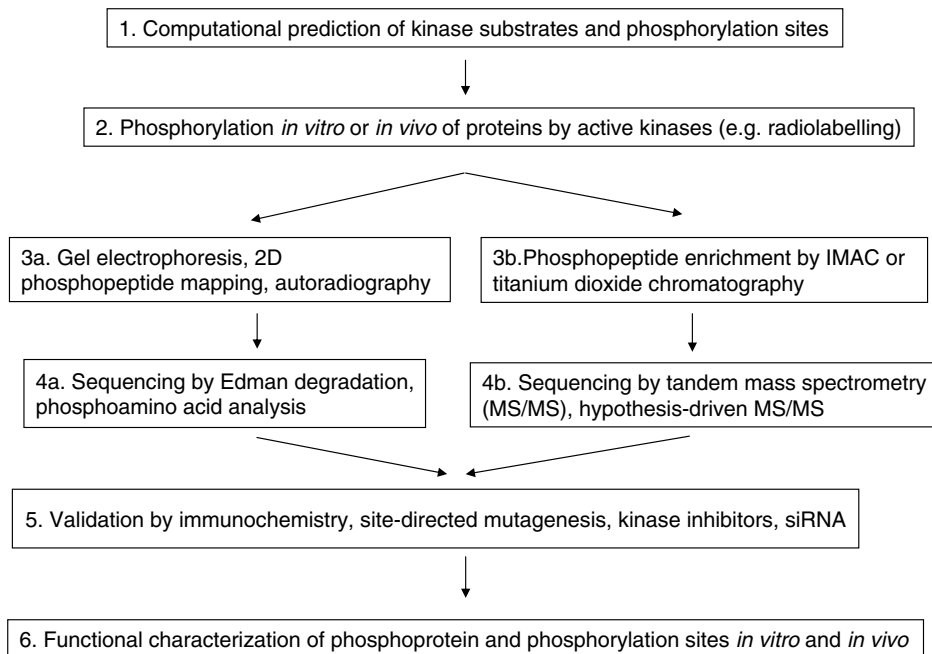


Fig. 2. Strategies for integrated analysis of phosphoproteins using computational biology, protein chemistry, and mass spectrometry.

60 motifs characterizing binding or substrate specificities of many families of Ser/Thr- or Tyr-kinases, SH2, SH3, PDZ, 14-3-3 and PTB domains [6]. The prediction of phosphorylation sites by Scansite is based on *in vitro* phosphorylation of an oriented peptide library by specific protein kinases [7]. A major advantage of this approach is that the sequence motifs are determined in unbiased experiments and that no prior knowledge of substrates is required. One limitation is that the optimal sequence might not be determined, since five

amino acids have been omitted during synthesis of the peptide library. Secondly, the synthetic peptide includes only eight amino acid residues around the site of phosphorylation and neglects the role of more distant residues for specificity. Finally, synergy between amino acids in two or more positions in the motif will be underestimated by this approach.

Sequence motifs are complex in the sense that positional correlation between several residues is significant for the specificity. Hence, complex machine learning approaches such as

Table 1
WWW-accessible phosphorylation site prediction servers and databases

| | WWW-accessibility |
|--|---|
| <i>Phosphorylation site prediction servers</i> | |
| GPS | http://973-proteinweb.ustc.edu.cn/gps/gps_web/ |
| KinasePhos | http://KinasePhos.mbc.nctu.edu.tw/ |
| NetPhos | www.cbs.dtu.dk/services/NetPhos/ |
| NetPhosK | www.cbs.dtu.dk/services/NetPhosK/ |
| PredPhospho | www.ngri.re.kr/proteo/PredPhospho.htm/ |
| PREDIKIN | http://florey.biosci.uq.edu.au/kinsub/home.htm |
| Prosite | http://us.expasy.org/prosite/ |
| Scansite | http://scansite.mit.edu |
| <i>Phosphorylation site databases</i> | |
| PhosphoBase | http://www.cbs.dtu.dk/databases/PhosphoBase/ |
| Phospho.ELM | http://phospho.elm.eu.org/ |
| Human protein reference database (HPRD) | http://www.hprd.org/ |
| PhosphoSite | http://www.phosphosite.org/ |
| Swiss-Prot | http://us.expasy.org/sprot/ |

ANN have been used for sequence analysis including prediction of phosphorylation sites. ANN is capable of classifying highly complex and nonlinear biological sequence patterns, where correlations between positions are important [8]. ANN has been extensively used in biological sequence analysis, because they do not only recognize patterns seen during training but also retain the ability to generalize and recognize similar patterns [9]. Originally, the ANN-based algorithm NetPhos, was developed for prediction of the general phosphorylation status using experimentally validated Ser, Thr, and Tyr phosphorylation sites in the curated database PhosphoBase [4,10]. Recently, a kinase-specific version, termed NetPhosK, was developed for prediction of phosphorylation sites for 17 protein kinases [11,12]. The false-negative and false-positive predictions of the PKA-specific neural network was determined experimentally to 0% and 59%, corresponding to a sensitivity and specificity of 100% and 41%, respectively [12].

In addition, Kim et al. have used support vector machines (SVM), which are classes of ANN, for prediction of phosphorylation sites at both the kinase family and group level. This prediction server, PredPhospho, is available for four kinase families (CDK, CK2, PKA and PKC) and four kinase groups (AGC, CAMK, CMGC and TK) [13]. Recently, SVMs were also applied for prediction of PKA, PKC, CK, CK2 and CDC2 kinase phosphorylation sites by Plewczynski et al. [14]. KinasePhos based on profile hidden Markov model is another tool for prediction of kinase-specific phosphorylation sites [15]. Finally, a different approach termed PREDIKIN has been developed for prediction of kinase specificity. PREDIKIN predicts optimal substrate heptapeptides by analysis of the primary sequence of the catalytic domain of the protein kinase [16]. One advantage of the PREDIKIN approach is that it can predict phosphorylation sites for uncharacterized kinases for which no information other than the amino acid sequence is available. Recently, an algorithm was designed to extract phosphorylation motifs from large mass spectrometry data sets of phosphorylated proteins [17]. The method relies on the intrinsic alignment of phospho-residues and the extraction of motifs through iterative comparison to a dynamic statistical background. The method was able to predict dozens of novel and known phosphorylation sites from published Ser, Thr and Tyr phosphorylation studies.

In conclusion, computational prediction methods have several advantages in phosphoproteomics. First, they are fast

and reproducible, and are publicly available at web servers. Second, they have proven to be sufficiently accurate for optimizing experimental design [18]. Examples are that Scansite predictions resulted in identification of tuberlin as a PKB substrate linking PI3K to tuberous sclerosis [19] as well as identification and characterization of novel PKA substrates [12,20]. On the other hand, computational approaches also have limitations. Computational prediction servers are based on experimental data and the accuracy of the predictions is strongly dependent on the quality of the data. The amount of data required for development of a high quality prediction method depends on the diversity of the acceptor motif of the kinase. In addition, it is not obvious how well a particular algorithm based on a limited data set can predict phosphorylation sites in the sequence space. Kinase specificity is also affected by other factors than the amino acid sequence of the substrates including co-localization of the kinase and its substrate, surface accessibility, spatial configuration and kinase-docking sites. It should always be borne in mind that computational predictions of phosphorylation sites are based solely on sequence comparison and predicted sites must be experimentally verified before considered valid. Another caveat is that the above mentioned bioinformatics programs are developed for prediction of mammalian phosphorylation sites. Hence, the predictions may fail or have higher false-negative and false-positive rates when employed in other organisms such as fungi and plants.

Because the number of experimentally characterized phosphorylation sites grows tremendously another important aspect of bioinformatics is creation of public phosphorylation site containing databases. PhosphoBase [10], Phospho.ELM [21], PhosphoSite [22], the human protein reference database [23], and Swiss-Prot [24] are examples where entries are compiled and curated from the literature as well as from different protein sequence databases (Table 1).

3. Classical protein chemistry and biochemistry approaches in phosphoprotein analysis

A range of techniques has been applied in chemical and biochemical analysis of phosphoproteins (Fig. 1). A classic strategy for detection of phosphoproteins is radioactive labelling in vivo with [³²P]orthophosphate or in vitro using purified ki-

nase and [γ - ^{32}P]ATP followed by electrophoretic separation [25]. However, if the protein is phosphorylated *in vivo* at low stoichiometry or on multiple sites by different protein kinases it can be difficult to isolate sufficient phosphorylated protein. Hence, many protein phosphorylation studies are performed with proteins modified *in vitro*. ^{32}P -labeled phosphoproteins can be analyzed by proteolytic digestion and separation of peptides by two-dimensional electrophoresis. Phosphopeptides are detected by autoradiography and the relative stoichiometry of phosphorylation among the phosphopeptides can be determined. Phosphopeptides can be purified and phosphorylation sites identified by N-terminal peptide sequencing using Edman degradation. Recently, mass spectrometry has been introduced successfully for identification of phosphorylation sites (see below). The phosphoamino acid content of phosphoproteins can be determined by electrophoresis of ^{32}P -labeled phosphoamino acids together with phosphoamino acid standards followed by autoradiography and ninhydrin staining [26,27]. The localization of phosphorylation sites in peptides or proteins can be determined by N-terminal sequencing of phosphopeptides using Edman degradation. The phosphoprotein is first enzymatically digested and radioactive peptides are isolated by reverse phase high performance liquid chromatography (RP-HPLC) [28,29]. A disadvantage of RP-HPLC is that very hydrophilic phosphopeptides may elute in the flow-through and very hydrophobic peptides may not elute at all.

Biochemical approaches include mutation of the suspected phosphorylation residues to alanine or phenylalanine followed by phosphorylation assay and two-dimensional phosphopeptide mapping [27]. Immunochimistry using phospho-specific antibodies is also widely used for identification of phosphoproteins [30,31]. However, the specificity of some commercially available phospho-specific antibodies has not been adequately tested. Because phospho-specific antibodies frequently cross-react with other phosphorylation sites on a protein, it is essential to demonstrate that mutation of the phosphorylated residue abolishes recognition of the protein by the antibody. This type of cross-reactivity resulted for example in misidentification of Ser112 in Bad as a PKA phosphorylation site [32,33]. In addition, protein kinase inhibitors may be used to identify substrates of a specific kinase. However, many kinase inhibitors targets more than one kinase like for example the PKA inhibitor H89, which also is a potent inhibitor of MSK1, S6 kinase-1 and Rho-dependent kinase II [34]. Recently, small interfering RNA (siRNA) have been applied *in vivo* for validation of substrate phosphorylation by a selected protein kinase [35].

In conclusion, classical protein chemistry and biochemical analyses are fundamental and commonly used in phosphoproteome analysis. The disadvantage of these techniques is that radioactive labelling is often required and that high amounts of protein are needed. In addition, they are labour intensive and cannot be used for high-throughput analysis of the phosphoproteome.

4. Mass spectrometry analysis of phosphoproteome

Mass spectrometry-based methods, including matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry and electrospray ionization

quadrupole time-of-flight (ESI-QTOF) analysis, are well established techniques in phosphoproteomics. Characterization of phosphoproteins by mass spectrometry requires four basic steps including isolation of the phosphoproteins e.g. immunoprecipitation followed by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS–PAGE), enzymatic digestion of the protein with a protease such as trypsin, identification of phosphopeptides, and sequencing of the phosphopeptide to localize phosphorylation sites [36]. Although mass spectrometry-based analysis of phosphoproteins sounds simple, it is still a major challenge in proteomic research for a number of reasons. First, identification of phosphorylation sites requires isolation and analysis of the peptide that contains the modified residue. Furthermore, ion signals corresponding to phosphorylated peptides are significantly suppressed in the presence of non-phosphorylated peptides. In addition, phosphorylation can inhibit the tryptic digestion of the protein depending on the position of the phosphorylated residue relative to the tryptic cleavage site at arginine or lysine. Finally, if proteins are phosphorylated at low stoichiometry, the detection of the phosphopeptide is difficult.

Complete characterization of the primary sequence and PTM of a protein requires much more material than mass spectrometry sequencing of a few peptides. At least Coomassie-stainable amounts (several picomoles $\sim 1\ \mu\text{g}$) should be purified to increase the chance of identification of phosphorylation sites. In addition, any purification step that enriches for phosphopeptides will provide better detection of phosphorylation sites. Phosphoproteins can be enriched from crude samples by affinity-based purification including immunoprecipitation with phospho-specific antibodies [31]. Anti-pTyr antibodies are used successfully for immunoprecipitation of phosphoproteins prior to mass spectrometry analysis [37,38], and recently also good anti-pSer and anti-pThr antibodies have been developed [31,39]. Furthermore, immobilized metal affinity chromatography (IMAC) using Fe^{3+} and Ga^{3+} is widely used for enrichment of phosphoproteins or -peptides prior to mass spectrometry analysis [40,41]. Some IMAC protocols result in problems with non-specific binding of non-phosphorylated acidic peptides. Hence, blocking the acidic residues by *O*-methyl esterification has been shown to enhance the specificity of the phosphopeptide binding [42]. Recently, titanium dioxide chromatography has also been applied for enrichment of phosphopeptides [43]. In addition, graphite columns has been shown to capture hydrophilic peptides such as phosphopeptides and glycopeptides [44]. Furthermore, chromatographic separation of phosphopeptides prior to mass spectrometry has been applied to facilitate identification of phosphorylation sites. Liquid chromatography (LC) using nano-columns with reversed-phase C_{18} material coupled directly to an ESI-QTOF tandem mass spectrometry (MS/MS) is a successful procedure for identification of PTM in a high-throughput fashion [45].

Detection of phosphopeptides is often based on fragmentation of the phosphoester bonds in p-Ser, p-Thr, or p-Tyr. Particularly beta-elimination of phosphoric acid from p-Ser or p-Thr, lead to identification of phosphopeptides in mass spectra [46,47]. In addition, p-Tyr-containing peptides can specifically be detected by the production of immonium ion from the p-Tyr residues [48]. Enzymatic dephosphorylation of the digested peptide by protein phosphatases has also been applied for identification of phosphopeptides [49]. Typically, the

masses of the dephosphorylated peptides after phosphatase treatment are compared to an untreated control by MALDI-TOF analysis. Furthermore, chemical techniques for conversion of phosphoamino acids into more tractable species have been developed. Affinity tags like biotin are specifically coupled to the phosphopeptide by beta-elimination of phosphoric acid from p-Ser or p-Thr followed by phosphopeptide enrichment using affinity chromatography [50]. Another chemical approach uses beta-elimination followed by addition of cysteamine to convert p-Ser and p-Thr to lysine analogues resulting in phosphoamino acid specific cleavage for trypsin [51]. However, the chemical techniques based on beta-elimination are only applicable for p-Ser and p-Thr analysis. Another chemical approach based on phosphoramidate chemistry enables analysis of p-Ser, p-Thr as well as p-Tyr [52]. The major disadvantage of the chemical techniques described above is that a large amount of sample is required.

In addition to identification of phosphoproteins and mapping of phosphorylation sites there is also a need for quantitative analysis of phosphorylated proteins. Recently, mass spectrometry-based techniques for relative quantification of PTMs has been developed using comparison with the intensity of chemically similar peptides. These techniques include differential stable isotope incorporation in proteins by metabolic labelling *in vivo* or chemical labelling *in vitro*. Ong et al. have designed a technique for *in vivo* labelling termed stable isotope labelling by amino acid in cell culture (SILAC), where they use amino acids containing a stable isotope [53]. *In vitro* labeling with iTRAQ reagent has also been applied successfully in quantitative phosphoproteomics [54]. An alternative to labelling of phosphopeptides is to run two samples separately with a chemically synthesized heavier version of the phosphopeptide as an internal standard [55].

5. Integrated strategy in experimental phosphoproteomics

No single high-throughput method is available for phosphoproteome analysis. Integration of analytical techniques including computational prediction, protein chemistry, and mass spectrometry can facilitate detailed characterization of phosphoproteins. Recently, high-throughput strategies for the identification of phosphoproteins have been developed, in particular within the fields of bioinformatics and mass spectrometry. We propose an integrated strategy for experimental analysis of the phosphoproteome. This includes six consecutive steps: prediction, phosphorylation, separation, identification, validation, and functional characterization of the phosphorylated protein (Fig. 2). The first step is computational analysis of the proteome/genome for prediction of protein kinase substrates and phosphorylation sites for a given kinase or group of kinases. Although the rate of false-positive and false-negative predictions is generally high, they may be useful to select potential kinase substrates for further analysis. The second step is phosphorylation *in vitro* or *in vivo* by activated kinases, which is followed by the third step of chromatographic separation and enrichment of phosphorylated proteins or peptides. In the fourth step phosphoproteins or peptides are sequenced either by classical protein chemistry or by tandem mass spectrometry. Today, mass spectrometry is often the analytical method of choice because of developments in different mass

spectrometry-based approaches for high-throughput qualitative and quantitative characterization of phosphoproteins. However, one major limitation of mass spectrometry is that the sequence coverage never is 100% especially when complex sample mixtures are analysed. One way to improve the mass spectrometry analysis can be affinity-based enrichment of phosphopeptides, e.g., by IMAC or titanium dioxide chromatography. Alternatively, it can be an advantage to use computational prediction of phosphopeptides for targeted selection of peptides for sequencing in tandem mass spectrometry. The fifth step involves validation of phosphorylation sites using phospho-specific antibodies, mutation of phosphorylated residues, kinase inhibitors, and siRNA. Finally, functional evaluation of the identified phosphorylation sites requires detailed studies *in vitro* and *in vivo*. These techniques have been applied for characterization of phosphorylation sites for decades, but the drawback is that high-throughput analysis of phosphoproteins is not possible.

This integrated strategy is limited to characterization of kinase-substrate pairs for well characterized mammalian protein kinases since the prediction programs are based on known phosphorylation motifs that are determined to yield a fairly robust definition of specificity. However, this level of knowledge exists for only about 10% of the 518 human protein kinases. Consequently, the prediction aspect of the first step of the integrated strategy does not function without a priori knowledge of a kinase target. Moreover, the recent technical development related to mass spectrometry has led to generation of vast amounts of empirical data in phosphoproteomics. Hence, an alternative strategy would be to begin with the large datasets and then use bioinformatics to sort the mass spectrometry data [17]. This means that empirically determined phosphorylation sites are used for the predicting kinases substrates *de novo* in contrast to predicting new substrates for known kinases as outlined above. For instance, taking a rough MAP kinase consensus of Ser/Thr-Pro presents a tremendous false-negative population. Further refinement to an 'improved' consensus of Pro-X-Ser/Thr-Pro decreases the false-negative rate, but still includes a great deal of noise. If one begins with empirically determined pSer/pThr-Pro sequences from mass spectrometry phosphoproteomics, this step would greatly decrease the amount of experimental validation required for each kinase.

References

- [1] Hunter, T. (1998) The Croonian lecture 1997. The phosphorylation of proteins on tyrosine: its role in cell growth and disease. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 353, 583–605.
- [2] Manning, G., Whyte, D.B., Martinez, R., Hunter, T. and Sudarsanam, S. (2002) The protein kinase complement of the human genome. *Science* 298, 1912–1934.
- [3] Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C.J., Hofmann, K. and Bairoch, A. (2002) The PROSITE database, its status in 2002. *Nucleic Acids Res.* 30, 235–238.
- [4] Blom, N., Gammeltoft, S. and Brunak, S. (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* 294, 1351–1362.
- [5] Xue, Y., Zhou, F., Zhu, M., Ahmed, K., Chen, G. and Yao, X. (2005) GPS: a comprehensive www server for phosphorylation sites prediction. *Nucleic Acids Res.* 33, W184–W187.
- [6] Obenaus, J.C., Cantley, L.C. and Yaffe, M.B. (2003) Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.* 31, 3635–3641.
- [7] Songyang, Z., Blechner, S., Hoagland, N., Hoekstra, M.F., Pivnicka-Worms, H. and Cantley, L.C. (1994) Use of an oriented

- peptide library to determine the optimal substrates of protein kinases. *Curr. Biol.* 4, 973–982.
- [8] Presnell, S.R. and Cohen, F.E. (1993) Artificial neural networks for pattern recognition in biochemical sequences. *Annu. Rev. Biophys. Biomol. Struct.* 22, 283–298.
- [9] Wu, C.H. (1997) Artificial neural networks for molecular sequence analysis. *Comput. Chem.* 21, 237–256.
- [10] Kreegipuu, A., Blom, N. and Brunak, S. (1999) PhosphoBase, a database of phosphorylation sites: release 2.0. *Nucleic Acids Res.* 27, 237–239.
- [11] Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S. and Brunak, S. (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* 4, 1633–1649.
- [12] Hjerrild, M., Stensballe, A., Rasmussen, T.E., Kofoed, C.B., Blom, N., Sicheritz-Ponten, T., Larsen, M.R., Brunak, S., Jensen, O.N. and Gammeltoft, S. (2004) Identification of phosphorylation sites in protein kinase A substrates using artificial neural networks and mass spectrometry. *J. Proteome Res.* 3, 426–433.
- [13] Kim, J.H., Lee, J., Oh, B., Kimm, K. and Koh, I. (2004) Prediction of phosphorylation sites using SVMs. *Bioinformatics* 20, 3179–3184.
- [14] Plewczynski, D., Tkacz, A., Godzik, A. and Rychlewski, L. (2005) A support vector machine approach to the identification of phosphorylation sites. *Cell Mol. Biol. Lett.* 10, 73–89.
- [15] Huang, H.D., Lee, T.Y., Tzeng, S.W. and Horng, J.T. (2005) KinasePhos: a web tool for identifying protein kinase-specific phosphorylation sites. *Nucleic Acids Res.* 33, W226–W229.
- [16] Brinkworth, R.I., Breinl, R.A. and Kobe, B. (2003) Structural basis and prediction of substrate specificity in protein serine/threonine kinases. *Proc. Natl. Acad. Sci. USA* 100, 74–79.
- [17] Schwartz, D. and Gygi, S.P. (2005) An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat. Biotechnol.* 23, 1391–1398.
- [18] Nielsen, H., Brunak, S. and von Heijne, G. (1999) Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Eng.* 12, 3–9.
- [19] Manning, B.D., Tee, A.R., Logsdon, M.N., Blenis, J. and Cantley, L.C. (2002) 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.
- [20] Hjerrild, M., Stensballe, A., Jensen, O.N., Gammeltoft, S. and Rasmussen, T.E. (2004) Protein kinase A phosphorylates serine 267 in the homeodomain of engrailed-2 leading to decreased DNA binding. *FEBS Lett.* 568, 55–59.
- [21] Diella, F., Cameron, S., Gemund, C., Linding, R., Via, A., Kuster, B., Sicheritz-Ponten, T., Blom, N. and Gibson, T.J. (2004) Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. *BMC Bioinform.* 5, 79.
- [22] Hornbeck, P.V., Chabra, I., Kornhauser, J.M., Skrzypek, E. and Zhang, B. (2004) PhosphoSite: a bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* 4, 1551–1561.
- [23] Peri, S., Navarro, J.D., Amanchy, R., Kristiansen, T.Z., Jonnalagadda, C.K., Surendranath, V., Niranjan, V., Muthusamy, B., Gandhi, T.K., Gronborg, M., Ibarrola, N., Deshpande, N., Shanker, K., Shivashankar, H.N., Rashmi, B.P., Ramya, M.A., Zhao, Z., Chandrika, K.N., Padma, N., Harsha, H.C., Yatish, A.J., Kavitha, M.P., Menezes, M., Choudhury, D.R., Suresh, S., Ghosh, N., Saravana, R., Chandran, S., Krishna, S., Joy, M., Anand, S.K., Madavan, V., Joseph, A., Wong, G.W., Schiemann, W.P., Constantinescu, S.N., Huang, L., Khosravi-Far, R., Steen, H., Tewari, M., Ghaffari, S., Blobel, G.C., Dang, C.V., Garcia, J.G., Pevsner, J., Jensen, O.N., Roepstorff, P., Deshpande, K.S., Chinnaiyan, A.M., Hamosh, A., Chakravarti, A. and Pandey, A. (2003) Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res.* 13, 2363–2371.
- [24] Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S. and Schneider, M. (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* 31, 365–370.
- [25] Salih, E. (2005) Phosphoproteomics by mass spectrometry and classical protein chemistry approaches. *Mass Spectrom. Rev.* 24, 828–846.
- [26] Jensen, H.H., Hjerrild, M., Guerra, B., Larsen, M.R., Hojrup, P. and Boldyreff, B. (2001) Phosphorylation of the Fas associated factor FAF1 by protein kinase CK2 and identification of serines 289 and 291 as the in vitro phosphorylation sites. *Int. J. Biochem. Cell Biol.* 33, 577–589.
- [27] Hjerrild, M., Milne, D., Dumaz, N., Hay, T., Issinger, O.G. and Meek, D. (2001) Phosphorylation of murine double minute clone 2 (MDM2) protein at serine-267 by protein kinase CK2 in vitro and in cultured cells. *Biochem. J.* 355, 347–356.
- [28] Aebersold, R., Watts, J.D., Morrison, H.D. and Bures, E.J. (1991) Determination of the site of tyrosine phosphorylation at the low picomole level by automated solid-phase sequence analysis. *Anal. Biochem.* 199, 51–60.
- [29] Campbell, D.G. and Morrice, N.A. (2002) Identification of protein phosphorylation sites by a combination of mass spectrometry and solid phase edman sequencing. *J. Biomol. Tech.* 13, 119–130.
- [30] Czernik, A.J., Girault, J.A., Nairn, A.C., Chen, J., Snyder, G., Keabian, J. and Greengard, P. (1991) Production of phosphorylation state-specific antibodies. *Meth. Enzymol.* 201, 264–283.
- [31] Gronborg, M., Kristiansen, T.Z., Stensballe, A., Andersen, J.S., Ohara, O., Mann, M., Jensen, O.N. and Pandey, A. (2002) A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate. *Mol. Cell Proteomics* 1, 517–527.
- [32] Harada, H., Becknell, B., Wilm, M., Mann, M., Huang, L.J., Taylor, S.S., Scott, J.D. and Korsmeyer, S.J. (1999) Phosphorylation and inactivation of BAD by mitochondria-anchored protein kinase A. *Mol. Cell* 3, 413–422.
- [33] Lizcano, J.M., Morrice, N. and Cohen, P. (2000) Regulation of BAD by cAMP-dependent protein kinase is mediated via phosphorylation of a novel site, Ser155. *Biochem. J.* 349, 547–557.
- [34] Davies, S.P., Reddy, H., Caivano, M. and Cohen, P. (2000) Specificity and mechanism of action of some commonly used protein kinase inhibitors. *Biochem. J.* 351, 95–105.
- [35] Morgan-Lappe, S., Woods, K.W., Li, Q., Anderson, M.G., Schurdak, M.E., Luo, Y., Giranda, V.L., Fesik, S.W. and Levenson, J.D. (2006) RNAi-based screening of the human kinome identifies Akt-cooperating kinases: a new approach to designing efficacious multitargeted kinase inhibitors. *Oncogene* 25, 1340–1348.
- [36] Jensen, O.N. (2004) Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol.* 8, 33–41.
- [37] Blagoev, B., Kratchmarova, I., Ong, S.E., Nielsen, M., Foster, L.J. and Mann, M. (2003) A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nat. Biotechnol.* 21, 315–318.
- [38] Salomon, A.R., Ficarro, S.B., Brill, L.M., Brinker, A., Phung, Q.T., Ericson, C., Sauer, K., Brock, A., Horn, D.M., Schultz, P.G. and Peters, E.C. (2003) Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry. *Proc. Natl. Acad. Sci. USA* 100, 443–448.
- [39] Kane, S., Sano, H., Liu, S.C., Asara, J.M., Lane, W.S., Garner, C.C. and Lienhard, G.E. (2002) A method to identify serine kinase substrates. Akt phosphorylates a novel adipocyte protein with a Rab GTPase-activating protein (GAP) domain. *J. Biol. Chem.* 277, 22115–22118.
- [40] Andersson, L. and Porath, J. (1986) Isolation of phosphoproteins by immobilized metal (Fe³⁺) affinity chromatography. *Anal. Biochem.* 154, 250–254.
- [41] Stensballe, A., Andersen, S. and Jensen, O.N. (2001) Characterization of phosphoproteins from electrophoretic gels by nanoscale Fe(III) affinity chromatography with off-line mass spectrometry analysis. *Proteomics* 1, 207–222.
- [42] Ficarro, S.B., McClelland, M.L., Stukenberg, P.T., Burke, D.J., Ross, M.M., Shabanowitz, J., Hunt, D.F. and White, F.M. (2002) Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat. Biotechnol.* 20, 301–305.

- [43] Pinkse, M.W., Uitto, P.M., Hilhorst, M.J., Ooms, B. and Heck, A.J. (2004) Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal. Chem.* 76, 3935–3943.
- [44] Larsen, M.R., Cordwell, S.J. and Roepstorff, P. (2002) Graphite powder as an alternative or supplement to reversed-phase material for desalting and concentration of peptide mixtures prior to matrix-assisted laser desorption/ionization-mass spectrometry. *Proteomics* 2, 1277–1287.
- [45] Jaffe, H., Veeranna, Shetty, K.T. and Pant, H.C. (1998) Characterization of the phosphorylation sites of human high molecular weight neurofilament protein by electrospray ionization tandem mass spectrometry and database searching. *Biochemistry* 37, 3931–3940.
- [46] Annan, R.S. and Carr, S.A. (1996) Phosphopeptide analysis by matrix-assisted laser desorption time-of-flight mass spectrometry. *Anal. Chem.* 68, 3413–3421.
- [47] Wilm, M., Neubauer, G. and Mann, M. (1996) Parent ion scans of unseparated peptide mixtures. *Anal. Chem.* 68, 527–533.
- [48] Steen, H., Kuster, B., Fernandez, M., Pandey, A. and Mann, M. (2001) Detection of tyrosine phosphorylated peptides by precursor ion scanning quadrupole TOF mass spectrometry in positive ion mode. *Anal. Chem.* 73, 1440–1448.
- [49] Larsen, M.R., Sorensen, G.L., Fey, S.J., Larsen, P.M. and Roepstorff, P. (2001) Phospho-proteomics: evaluation of the use of enzymatic de-phosphorylation and differential mass spectrometric peptide mass mapping for site specific phosphorylation assignment in proteins separated by gel electrophoresis. *Proteomics* 1, 223–238.
- [50] Goshe, M.B., Veenstra, T.D., Panisko, E.A., Conrads, T.P., Angell, N.H. and Smith, R.D. (2002) Phosphoprotein isotope-coded affinity tags: application to the enrichment and identification of low-abundance phosphoproteins. *Anal. Chem.* 74, 607–616.
- [51] Knight, Z.A., Schilling, B., Row, R.H., Kenski, D.M., Gibson, B.W. and Shokat, K.M. (2003) Phosphospecific proteolysis for mapping sites of protein phosphorylation. *Nat. Biotechnol.* 21, 1047–1054.
- [52] Zhou, H., Watts, J.D. and Aebersold, R. (2001) A systematic approach to the analysis of protein phosphorylation. *Nat. Biotechnol.* 19, 375–378.
- [53] Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A. and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell Proteomics* 1, 376–386.
- [54] Zhang, Y., Wolf-Yadlin, A., Ross, P.L., Pappin, D.J., Rush, J., Lauffenburger, D.A. and White, F.M. (2005) Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol. Cell Proteomics* 4, 1240–1250.
- [55] Stemmann, O., Zou, H., Gerber, S.A., Gygi, S.P. and Kirschner, M.W. (2001) Dual inhibition of sister chromatid separation at metaphase. *Cell* 107, 715–726.