Predicting kinase-substrate interactions in the era of proteomics. Focus on “Identifying protein kinase target preferences using mass spectrometry”

Ewout J. Hoorn and Marcel E. Meima
Department of Internal Medicine - Nephrology, Erasmus Medical Center, Rotterdam, The Netherlands

WHEN THE “OMICS” BEGAN TO EMERGE in the life sciences, some investigators teasingly proposed to institute an “omics police” (2). This does not seem to have prevented a rapid growth of the “omics” (genomics, proteomics, metabolomics). Of these neologisms, proteomics refers to the concurrent identification, quantification, and analysis of a large number of proteins in a functional context (6). Mass spectrometry has been instrumental in the progress of proteomics because of its ability to reliably identify peptides based on their molecular weight (6). Because of the biological importance of phosphorylation, enrichment methods have been developed to specifically identify phosphopeptides by mass spectrometry (“phosphoproteomics”) (5). “Systems biology” has emerged as the term to describe the application of proteomics and other techniques to discover and model biological phenomena such as signaling networks (1). Now that the buzzwords have been introduced and briefly explained, you will appreciate the paper by Douglas, Gunaratne and colleagues in this issue of American Journal of Physiology-Cell Physiology (3). This group reports a mass spectrometry-based method to identify protein kinase target preferences. To do so, they developed a new bioinformatics tool that generates a visual language for kinase sequence preference motifs. In this “sequence logo” the phosphorylated residue is positioned at 0 and the predicted surrounding amino acids are displayed in their alphabet code with font size corresponding to probability (Fig. 1). The anticipated result from this in silico analysis is that some amino acids will be overrepresented on certain positions in the generated motifs so that this will point in the direction of a specific kinase or kinase family. The output comes with a statistical filter that corrects for multiple testing and reduces signal-to-noise ratio. The bioinformatics tool was coined “PhosphoLogo” to emphasize its utility in predicting phosphorylation events. The authors used in vitro experiments to test 16 protein kinases for their target sequence preferences. A tissue lysate containing rat kidney, liver, brain, and small intestine was first dephosphorylated and heat-denatured and subsequently incubated with ATP and a recombinant kinase of interest. The samples were then prepared for analysis by mass spectrometry using trypsin and immobilized ion affinity chromatography to enrich phosphopeptides. Of the 16 kinases, protein kinase A (PKA) was used as the model kinase. Indeed, PhosphoLogo generated a sequence logo that displayed the well-known R-R-X-S motif and was similar to sequence logos from existing sources such as PhosphoSitePlus and NetPhorest. Additional outputs of PhosphoLogo include a charged-grouped logo, size-grouped logo, and anti-logo. The charged-grouped logo shows positional preference for polar, nonpolar, acidic, or basic amino acids, while the size-grouped logo indicates the preferred size of the residue. Interestingly, the anti-logo shows disfavored residues, which should be especially helpful for investigators to reduce a list of candidate kinases. The authors translated the

![Information Content](http://www.ajpcell.org)

Fig. 1. We tested PhosphoLogo retrospectively by entering six peptide sequences that include known in vitro phosphorylation sites for the STE family kinases STK39/SPAK and OSR1 in two proteins in which our lab has an interest, namely, the sodium chloride cotransporter (NCC, Thr 46, Thr 55, and Thr 60) and the sodium potassium chloride cotransporter type 2 (NKCC2, Thr 95, Thr 100, Thr 105) (8, 9). The phosphorylated residue is positioned at 0, while positive positions represent amino acids on the COOH-terminal side of the phosphorylated residue. The strong preference for Thr as the phosphoacceptor and the Asn (N) at the –1 position shortlists OSR1, but not STK39/SPAK (Fig. 7 of Ref. 3) as candidate kinase. See text for details.
identified preferences for PKA substrate positions to the three-dimensional structure of PKA, which revealed a new preference for glycine at the −1 position. The authors speculate that a glycine at this position increases peptide chain flexibility to facilitate conformational selection during the phosphorylation process. The mass spectrometry analysis of PKA and the 15 other kinases generated several sequence logos and anti-logos that were summarized in a dendogram that was categorized by kinase family. The logos and anti-logos in the dendogram are likely to be of use as a reference for researchers to identify candidate kinases or substrates. The authors point out that the generated sequence logos in the dendogram already provided interesting insights. Namely, the serine-threonine kinase glycogen synthase kinase 3β (GSK3β) revealed a sequence logo that is in agreement with its position in the proline-directed group of kinases. Although the substrate specificity of GSK3β is commonly determined by a phosphoserine in the +4 position, the authors propose that it may switch to a proline in the +1 position depending on the regulatory mechanisms. This assumption was supported by the identified targets of GSK3β. These two examples, the glycine-preference of PKA and the proline-preference of GSK3β, illustrate how the hypothesis-free approach of PhosphoLogo can lead to new insights. Rightly so, however, the authors emphasize that motif specificity is just one determinant of kinase-substrate assignments, and that additional information is often necessary. For example, one needs to know whether the kinase is expressed in the tissue of interest and whether kinase and substrate are localized in the same cellular compartment. If so, the kinase and its substrate also need to be able to interact physically for which they may require docking and scaffolding proteins. This may explain why we did not identify SPAK as a candidate kinase for the sodium chloride cotransporter or the sodium potassium chloride cotransporter when we tested PhosphoLogo (Fig. 1).

SPAK is known to require docking to these substrates for efficient phosphorylation (8, 9). The authors therefore regard PhosphoLogo as “a complement to existing profiling approaches” such as combinatorial peptide array assays. The authors argue that the advantages of their mass spectrometry-based method is that it allows for the use of more representative substrates (a heterogeneous mixture of proteins instead of synthetic peptides) and the unbiased identification of novel candidates (3). This is true, but the method should also be compared with other emerging techniques such as kinome-wide siRNA screens (4). With kinome-wide siRNA screens, siRNA targeting kinases are transfected into cells after which they can be exposed to a stimulus of interest. Although the yield of such screens usually still includes multiple kinases, it provides a direct in vivo result rather than an in silico prediction and is also unbiased because all kinases are represented. Another question is how representative the substrate in the study by Douglas et al. was. The denaturing process may have uncovered residues that would normally not be phosphorylated (although the authors state that the logos for three kinases were “nearly identical” without heat denaturation). In addition, more proteins may have been phosphorylated during this in vitro method (24 h with ATP) than would normally occur in vivo. An indication that the yield of the in vitro experiments may not have been representative of in vivo phosphorylation is that none of the “natural” targets of the kinase Akt were identified, including GSK3β, the transcription factor forhead box O3 or the cyclin-dependent kinase inhibitors p21 and p27 (7). That being said, the aim of the study by Douglas et al. was probably more methodological than physiological. In that regard, the authors should be applauded for taking phosphoproteomics yet another step forward. The bioinformatics tool PhosphoLogo is likely to generate useful shortlists of candidate kinases and substrates out of endless mass spectrometry data. In addition, PhosphoLogo has been made freely available to the research community so that you can start your analysis today (http://helixweb.nih.gov/PhosphoLogo). Happy hunting!

DISCLOSURES
No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS
E.J.H. and M.E.M. interpreted results of experiments; E.J.H. and M.E.M. prepared figures; E.J.H. drafted manuscript; E.J.H. and M.E.M. edited and revised manuscript; E.J.H. and M.E.M. approved final version of manuscript.

REFERENCES