Phosphorylation Analysis by Mass Spectrometry

MYTHS, FACTS, AND THE CONSEQUENCES FOR QUALITATIVE AND QUANTITATIVE MEASUREMENTS*

Hanno Steen‡§¶, Judith A. Jebanathirajah§‖, John Rush‡‡, Nicolas Morrice§§, and Marc W. Kirschner§**

The mass spectrometric analysis of protein phosphorylation is still far from being routine, and the outcomes thereof are often unsatisfying. Apart from the inherent problem of substoichiometric phosphorylation, three arguments as to why phosphorylation analysis is so problematic are often quoted, including (*a*) increased hydrophilicity of the phosphopeptide with a concomitant loss during the loading onto reversed-phase columns, (*b*) selective suppression of the ionization of phosphopeptides in the presence of unmodified peptides, and (*c*) lower ionization/detection efficiencies of phosphopeptides as compared with their unmodified cognates. Here we present the results of a study investigating the validity of these three arguments when using electrospray ionization mass spectrometry. We utilized a set of synthetic peptide/phosphopeptide pairs that were quantitated by amino acid analysis. Under the applied conditions none of the experiments performed supports the notions of (*a*) generally increased risks of losing phosphopeptides during the loading onto the reversed-phase column because of decreased retention and (*b*) the selective ionization suppression of phosphopeptides. The issue of ionization/detection efficiencies of phosphopeptides *versus* their unphosphorylated cognates proved to be less straightforward when using electrospray ionization because no evidence for decreased ionization/detection efficiencies for phosphopeptides could be found. *Molecular & Cellular Proteomics 5:172–181, 2006.*

Protein phosphorylation is one of the most prevalent intracellular protein modifications that is of pivotal importance in numerous cellular processes including cell differentiation, proliferation, and migration. It is estimated that ~30% of all proteins in a cell are phosphorylated at any given time. However, this number is in stark contrast to the actual number of phosphorylation sites found so far. For instance, the Phospho.ELM database (phospho.elm.eu.org) currently lists 1703 experimentally verified phosphorylation sites for 556 different proteins derived from eukaryotes, the human protein reference database (www.hprd.org) lists 3652 reported phosphorylation sites on 1240 human proteins, and PhosphoSite (www.phosphosite.org) lists 6084 non-redundant phosphorylation sites on 2430 human and mouse proteins. However, a recent study by Beausoleil *et al.* (1) gave some idea about the expected number of phosphorylation in the cell when the authors performed a large scale phosphoproteomics study on HeLa cell nuclei. Although their method was biased against basic phosphopeptides where His, Lys-Pro, and/or Arg-Pro residues are in close proximity to the phosphoamino acid residue, more than 2000 phosphorylation sites on 967 nuclear proteins were found. Considering the fact that protein phosphorylation analysis is of major interest in numerous laboratories around the world it is surprising that more information about protein phosphorylation sites has not been gathered since the discovery of protein phosphorylation.

This raises the question as to why it is still such a challenge to perform unbiased protein phosphorylation analysis. One inherent reason is the generally low phosphorylation stoichiometry of most of the proteins such that phosphopeptides are grossly underrepresented in the generated complex peptide mixtures (see below). In addition, numerous reasons can be listed for not being able to identify and localize a phosphorylation site in a given protein, which affects all stages of phosphorylation analysis, *i.e.* sample preparation, analysis, and data interpretation. Examples for problems during the sample preparation include *e.g.* omission of phosphatase inhibitor, overestimation of the degree of phosphorylation because of the lack of information about the phosphorylation stoichiometry when using phosphospecific antibodies, and/or underestimation of the heterogeneity of the phosphorylation due to nonspecific ^{32}P-labeling.

However, even if it was ensured that a particular protein sample is phosphorylated by using other methods such as radiolabeling it is often the case that the mass spectrometric analysis does not provide any information about the sites of phosphorylation. Three main reasons are commonly used to explain as to why the phosphopeptides were missed during

From the ‡Department of Pathology, Harvard Medical School and Children's Hospital, Boston, Massachusetts 02115, §Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, ‖Neurobiology Program, Children's Hospital, Boston, Massachusetts 02115, **MDS Sciex, Concord, Ontario L4K 4V8, Canada, ‡‡Cell Signaling Technology, Beverly, Massachusetts 01915, and §§MRC Protein Phosphorylation Unit, University of Dundee, Dundee, United Kingdom DD1 5EH

Received, May 12, 2005, and in revised form, September 30, 2005
Published, MCP Papers in Press, October 4, 2005, DOI 10.1074/mcp.M500135-MCP200

the mass spectrometric analysis: (i) increased hydrophilicity and hence reduced retention of phosphopeptides on reversed-phase materials, (ii) selective suppression of their ionization/detection efficiencies in the presence of large amounts of unphosphorylated peptides, and (iii) lower detection efficiencies of phosphopeptides as compared with their unphosphorylated cognates. In this study we tried to scrutinize these different arguments often brought forward with the aim to draw the attention to the most prevalent problem(s) of mass spectrometric phosphorylation analysis and to improve the analysis where it is most fruitful/promising.

MATERIALS AND METHODS

All solvents used were of HPLC-grade quality from Burdick and Jackson (VWR International). All chemicals were purchased from Sigma unless otherwise noted. The synthetic peptides were HPLC-purified and quantitated in duplicate by amino acid analysis. Stock solutions were prepared using freshly calibrated pipettes. The phosphopeptide and its unphosphorylated counterpart were mixed in at least two different defined ratios (final concentrations 30–150 nM). To avoid carry-over problems during the LC/MS analyses, the solutions of two different peptide pairs were analyzed in an alternating fashion. The standard protein digest was prepared by digesting 50 μM protein solution overnight at 37 °C using sequencing grade trypsin (w/w 1:10, Roche Biochemicals).

LC/MS Analysis—All electrospray ionization experiments were performed using a QSTAR XL hybrid mass spectrometer (AB/MDS Sciex) hyphenated with microscale capillary reversed-phase HPLC (Famos autosampler (LC Packings), Agilent 1100 HPLC pump (Agilent)). The columns were packed in-house using Magic C18 (5 μm, 200 Å, Michrom BioResources) beads. The buffer compositions are as follows: buffer A: 2.5% acetonitrile, 0.2% formic acid; buffer B: 2.5% H_2O, 0.2% formic acid. For the quantitation experiments a 5-min gradient was used with mass spectra being acquired every 0.15 s. Data analysis and quantitation was done using the Analyst software package provided by Applied Biosystems/MDS Sciex.

RESULTS AND DISCUSSION

"*Phosphopeptides often are lost during the loading onto the reversed-phase columns because the addition of anionic/acidic phosphate groups increases hydrophilicity resulting in reduced retention*." This is one reason often used to explain why a phosphorylation analysis using LC/MS failed. To test this we used a set of peptide/phosphopeptide pairs that varied in length from 7 to 17 amino acids, resembling the size of peptides commonly observed in tryptic digests. Cysteine and methionine-containing peptides were not included in this study to avoid problems associated with partial oxidation. All peptide pairs were analyzed using water/acetonitrile/0.2% formic acid buffers as mobile phases and Magic C18 as stationary phase; Magic C18 is a reversed-phase material commonly used in proteomics applications. Interestingly, despite the common belief that phosphopeptides are more hydrophilic than their unphosphorylated cognate, all singly phosphorylated peptides tested eluted off the reversed-phase column after the unmodified complement irrespective of the number of basic amino acid residues (His, Lys, Arg; Table I). One example is presented in Fig. 1*A*; it shows the selected ion chromatograms (XICs)[1] of RNYSVGS and RNYpSVGS (Table I, Peptide species 2), the shortest peptide/phosphopeptide pair investigated. Although there is a considerable overlap in the elution profiles, the phosphopeptide clearly elutes after the unphosphorylated cognate. Of the three doubly phosphorylated peptides in this test set, two phosphopeptides eluted significantly later than the singly phosphorylated or unphosphorylated peptide from the reversed-phase column; one example is shown in Fig. 1*B*. The third doubly phosphorylated peptide with the sequence DQAVpTEpYVATR (Table I, Peptide species 27) was the only peptide whose modified form eluted before its unmodified complement (Fig. 1*C*). It was noted that this particular peptide was the only phosphopeptide in our sample set in which the number of phospho moieties exceeded the number of basic amino acid residues. This led us to hypothesize that phosphorylation does increase the hydrophilicity of peptides; however, if the peptide contains basic amino residues that are positively charged under standard LC(MS) conditions, the increase of hydrophilicity can be overcompensated by charge neutralization, *i.e.* reduction of the net charge, thereby reducing the overall hydrophilicity.

This means that the introduction of phosphorylation sites can decrease the overall hydrophilicity/increase the retention time under the employed LC conditions as long as the positive net charge decreases. Once the number of phosphorylation sites exceeds the number of basic residues the hydrophilicity increases again whereby the most hydrophobic species are generated when the number of phosphoamino acid residues equals the number of basic amino acid residues, *i.e.* a net charge of 0 is reached. The hypothesis of charge compensation gained support when the doubly phosphorylated peptide DQAVpTEpYVATR (Table I, Peptide species 27) was partially dephosphorylated and the mixture of doubly phosphorylated, singly phosphorylated, and unphosphorylated species was analyzed by LC/MS. The selected ion chromatograms (XIC) of the different species are shown in Fig. 1*C*. Whereas the doubly phosphorylated peptide (net charge -1) eluted before the unphosphorylated peptide (net charge $+1$), the singly phosphorylated peptide (net charge 0) eluted after the unphosphorylated peptide.

The described elution order can be observed in the absence of strong ion pairing reagents such as trifluoroacetic acid (TFA) as these agents mask the positive charges on peptides, whereas the negative charges on peptides caused by carboxylic groups are neutralized at acidic pH. This is contrasted by peptides with highly acidic functional groups such as phospho moieties for which one negative charge on the phosphate group remains even at pH 2 such that phosphopeptides elute before their unphosphorylated peptide counterparts when ion pairing reagent is present to neutralize the positive charges. Because strong ion pairing reagents such as TFA are often not required (and not desired) in LC/MS

[1] The abbreviation used is: XIC, selected ion chromatogram.

TABLE I
Test peptides

Charge states, differences in peak retention (RT) time (RT (phosphopeptide)-RT (peptide)) and experimental ionization (Ion.)/detection (detec.) efficiency ratios (peptide *versus* phosphopeptide) for different peptide pairs. One standard deviation is given in brackets. P, peptide; pP, singly phosphorylated peptide; ppP, doubly phosphorylated peptide.

Peptide species	Peptide sequence	Z	ΔRT	Ion./detec. efficiency ratio
			min	
1	LLLRLpSENSG	2+	+ 0.1	1.4 (0.13)
2	RNYpSVGS	2+	+ < 0.05	1.08 (0.074)
3	IVADQpTPTPTRF	2+	+ < 0.05	0.28 (0.021)
4	FDSLPSpSPSSATPH	2+	+ < 0.05	1.74 (0.099)
5	GAHFpSVSSLAE	2+	+ 0.3	1.7 (0.034)
6	IGRRQpSLIEDA	2+	+ < 0.05	0.50 (0.038)
7		3+	+ < 0.05	0.86 (0.078)
8	KTQApSQGTLQTR	2+	+ 0.15	0.082 (0.006)
9		3+	+ 0.15	0.64 (0.076)
10	LQRQPSSpSPGPTPR	2+	+ < 0.05	0.93 (0.067)
11		3+	+ < 0.05	3.02 (0.12)
12	LRLSSpSSGRLR	2+	+ 0.1	0.37 (0.031)
13		3+	+ 0.1	0.59 (0.043)
14	GLGTRTGpSNLDRDKL	2+	+ 0.15	0.65 (0.053)
15		3+	+ 0.15	0.84 (0.093)
16		4+	+ 0.15	650 (320)
17	KFELLPpTPPLSPSRRSG	3+	+ 0.1	1.6 (0.24)
18	KDLKRLFpSGTQISTI	2+	+ 0.25	0.99 (0.066)
19		3+	+ 0.25	1.7 (0.086)
20		4+	+ 0.25	850 (600)
21	SRARIGpSDPLAYEPK	2+	+ 0.1	0.79 (0.038)
22		3+	+ 0.1	0.88 (0.14)
23		4+	+ 0.1	100 (10)
24	RYPRPVpSVPPSPSLSR	3+	+ 0.25	0.52 (0.064)
25	KRRQIpSIRGIV	2+	+ 0.25	0.35 (0.031)
26		3+	+ 0.25	0.76 (0.028)
27	DQAVpTEpYVATR	2+	− 0.15	2.0 (0.20)
28	KFELLPpTPPLpSPSRRSG (P *vs.* ppP)	3+	+ 0.18	2.3 (0.27)
29	RYPRPVpSVPPpSPSLSR (P *vs.* ppP)	3+	+ 0.55	0.79 (0.088)
30	KFELLPpTPPLpSPSRRSG (pP *vs.* ppP)	3+	+ 0.08	1.5 (0.15)
31	RYPRPVpSVPPpSPSLSR (pP *vs.* ppP)	3+	+ 0.3	1.5 (0.17)

experiments the elution order of phosphopeptides and their unmodified cognates are easily modulated.

More systematic work is necessary to evaluate the effects of different stationary phases and buffer compositions on the elution order of peptides and their phosphorylated cognates. This notion is underscored by a recently published example of a doubly phosphorylated peptide and its partially phosphorylated/unphosphorylated cognates; in this example, which employed only formic acid without TFA in the LC buffers, the peptides eluted in the following order: monophosphopeptide A < diphosphopeptide < unmodified peptide < monophosphopeptide B (2). Such an elution order is hard to rationalize irrespective of the presence or absence of any ion pairing reagents. Nevertheless, the data presented clearly show that phosphorylation does not necessarily lead to an overall increase in hydrophilicity. This is of particular interest for tryptic peptides, which generate peptides with at least one basic amino acid residue (exception: C-terminal peptide) such that singly phosphorylated peptides should elute after the unphosphorylated cognate; this minimizes the risk of losing phosphopeptides during the loading of the reversed-phase column. Multiple phosphorylations, however, can indeed lead to decreased retention if the number of phosphorylation sites is not matched by the number of basic amino acid residues. However, because phosphorylation affects the proteolysis kinetics of cleavage sites proximal to the site of phosphorylation (3), it can be expected that multiply phosphorylated peptides will show an increased frequency of more than one basic amino acid residues, thus partially compensating this problem. Nevertheless, alternative proteolysis strategies should be considered (*e.g.* Lys-C) which (*a*) increase the average number of basic residues within the proteolytic peptides and/or (*b*) increase the length of the proteolytic peptides, which generally also increases the retention time, thus reducing the potential loss of phosphopeptides during the loading of the reversed-phase column. That said, the loss of phosphopeptides because of increased hydrophilicity can be a serious problem if nonspecific proteases are used generating

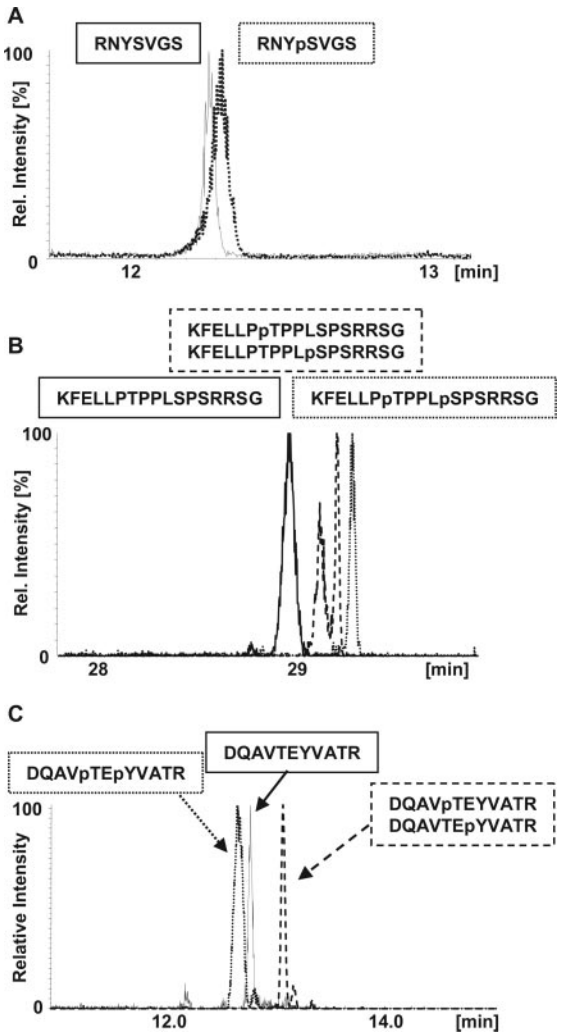

FIG. 1. **Testing the elution order analyzing numerous peptides and their different phosphorylated cognates by LC/MS.** *A*, normalized XICs of the peptide RNYSVGS (*unbroken line*) and its phosphorylated cognate RNYpSVGS (*dotted line*). *B*, normalized XICs of the peptide KFELLPTPPLSPSRRSG (*unbroken line*), its singly phosphorylated cognates KFELLPTPPLpSPSRRSG/KFELLPpTPPLSPSRRSG (*broken line*), and its doubly phosphorylated complement KFELLPpTPPLpSPSRRSG (*dotted line*). *C*, normalized XICs of the peptide DQAVTEYVATR (*unbroken line*), its singly phosphorylated cognates DQAVpTEYVATR/DQAVTEpYVATR (*broken line*), and its doubly phosphorylated complement DQAVpTEpYVATR (*dotted line*). *Rel.*, relative.

small peptides without any basic amino acid residues. This notion was confirmed when a highly acidic and hydrophilic peptide/phosphopeptide pair with the sequence EDADS(pY)ENMD was analyzed by LC/MS; whereas the unmodified form of the peptide was retained on the C18 column, the phosphorylated cognate was detected in the flow through under standard LC/MS conditions.

"Phosphorylated species are selectively suppressed in the presence of unmodified peptides" is another often used general statement that is made without detailed examination of this phenomenon. To test the validity of this generalization for LC/MS, constant amounts of several synthetic phosphopeptides and their unphosphorylated cognates were mixed with increasing amounts of a tryptic BSA digest, ranging from equal amount to 100-fold excess. The ratios of the signal intensities of peptide *versus* phosphopeptide were calculated for each pair and are shown in Fig. 2*A*. It is evident from the almost horizontal curves that there is no significant change in the relative ionization/detection efficiencies with increasing "background" of unmodified peptides. To ensure that this observation is valid not only for this particular set of peptides/phosphopeptides and tryptic BSA digest, another set of six different phosphopeptide/peptide pairs was spiked into seven individual tryptic digest of numerous commercially available proteins: (ubiquitin (8.5 kDa), avidin (17 kDa), α-casein (25 kDa), β-casein (25 kDa), carbonic anhydrase (29 kDa), alcohol dehydrogenase (40 kDa), ovalbumin (43 kDa)) and into a mixture thereof. This mixture simulated the digest of a $\sim$200 kDa protein. The normalized signal intensity ratios for the different peptide/phosphopeptide pairs present in those samples are shown in Fig. 2*B*. The horizontal trend of all curves again shows that the signal intensity ratios of peptide *versus* phosphorylated cognate did not significantly change, irrespective of the complexity of the digest, *i.e.* irrespective of the number of unmodified peptides co-eluting with the species of interest. This is exemplified in Fig. 3, *A–D*, which display mass spectra at the peak elution of one of the synthetic test peptides (marked with an *arrowhead*) spiked into various tryptic digests; differences in the sample complexity at the time of elution are obvious. The findings in either of the two sets of experiments do not support the notion that selective suppression of phosphopeptides occurs when analyzed by electrospray ionization mass spectrometry in the presence of unmodified peptides.

As control experiment, the absolute signal intensities of the peptides were investigated, and it was found that they were hardly affected by the presence of a large excess of unphosphorylated peptides over the concentration range tested. This is shown in Fig. 2*C* where the total ion currents (TIC) and the XIC for one exemplary peptide/phosphopeptide (*m/z* 683.35 *versus m/z* 710.0) pair in the presence of low amounts of BSA digest (*unbroken traces*) and in the presence of 100-fold excess of BSA digest (*dotted traces*) are displayed. The *dotted traces* are slightly shifted in the time dimension for better clarity. For both species, no obvious loss in signal intensity was observed, *i.e.* none of the experiments described above supports the general notion of selective ionization suppression of phosphopeptides in the presence of unmodified peptides when analyzing the sample by LC/MS. It should be noted that only samples of fairly low complexity and limited dynamic range were analyzed, and the level of saturated ionization was not reached. To test the issue of selective ionization suppression of phosphopeptides under saturated ionization conditions, *i.e.* the total number of charges is limited and

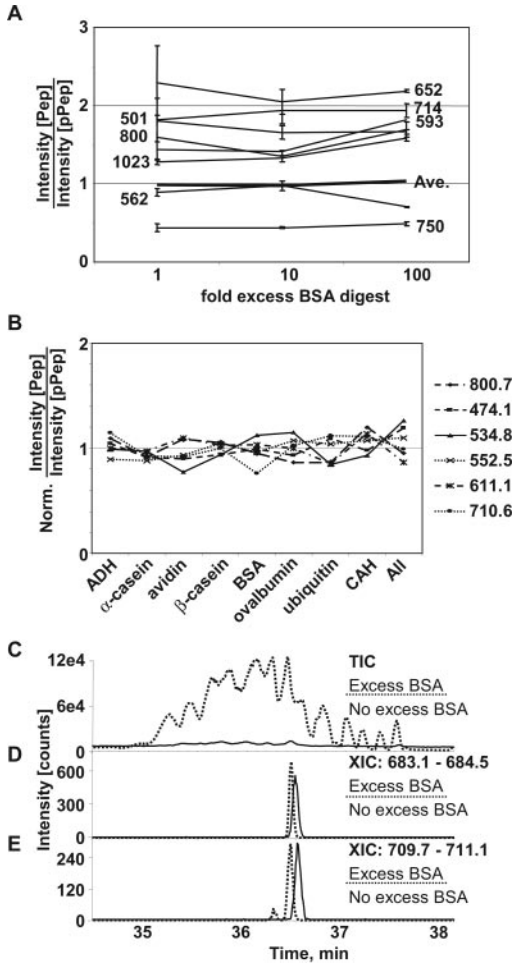

FIG. 2. **Testing the selective suppression of phosphopeptides in the presence of unphosphorylated peptides by analyzing numerous peptides and their phosphorylated cognates spiked into different protein digest by LC/MS.** *A*, eight different peptide/phosphopeptide pairs spiked into 1-fold, 10-fold, and 100-fold excess of BSA digest. The peptide ion signal ratios (peptide *versus* phosphopeptide) for each pair and the average are plotted as a function of excess BSA digest. The *error bars* represent one standard deviation after three measurements. *B*, six different peptide/phosphopeptide pairs were spiked into eight different protein digests and one mixture thereof. The normalized peptide ion signal ratios (peptide *versus* phosphopeptide) are plotted as a function of the digest. *CAH*, carbonic anhydrase; *ADH*, alcohol dehydrogenase; *Norm.*, normalized. *C*, total ion chromatogram of numerous peptide/phosphopeptide pairs and equal amount of BSA digest (*unbroken line*) and 100-fold excess of BSA digest (*broken line*). *TIC*, total ion current. *D*, XICs of the triply charged peptide in the presence of equal amounts of BSA digest (*unbroken line*) and 100-fold excess of BSA digest (*broken line*). The *latter trace* is slightly shifted for better visibility. *E*, XICs of the triply charged phosphorylated form of the peptide, whose XICs are shown in *D*, in the presence of equal amounts of BSA digest (*unbroken line*) and 100-fold excess of BSA digest (*broken line*). The *latter trace* is slightly shifted for better visibility.

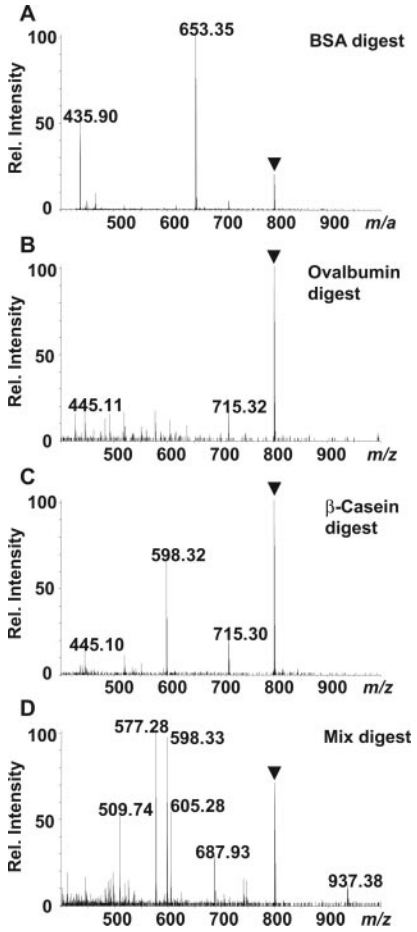

FIG. 3. **Mass spectra at the peak elution of the synthetic test peptides spiked into various digests.** A synthetic peptide (*m/z* 800.9, marked with an *arrowhead*) was spiked at equimolar levels into a tryptic BSA digest (*A*), ovalbumin digest (*B*), β-casein digest (*C*), and digest of an equimolar mixture of ubiquitin, avidin, α-casein, β-casein, carbonic anhydrase, alcohol dehydrogenase, and ovalbumin (*D*). *Rel.*, relative.

the analytes are competing for the available charges/protons, several peptide/phosphopeptide pairs were spiked into a 1000-fold excess of tryptic BSA digest (data not shown). As expected the total signal intensity of the peptides and the phosphopeptides were significantly reduced. However, no selective suppression of the phosphorylated species as compared with the unphosphorylated cognate was observed, *i.e.* the results of this experiment do not support the notion of selected suppression of phosphorylated species even under saturated ionization conditions. It should be noted, however, that the phosphopeptides are affected by the unspecific suppression observed for all unmodified and/or modified peptides of low/substoichiometric abundance under the conditions of saturated ionization. Therefore, analyzing phosphopeptides in very complex peptide mixtures such as unfractionated whole cell lysates will not be very successful because there is such a large excess (in number and relative amount) of unmodified peptides that it is unlikely that a phosphopeptide will be among the most abundant (by amount and/or signal intensity) species, *i.e.* those species that are detected under saturated

ionization conditions. This means that the most successful strategies for the analysis of phosphopeptides from complex protein/peptide mixtures will employ selective enrichment of the phosphopeptides.

"Phosphopeptides have lower ionization/detection efficiencies than their unphosphorylated counterpart" is one of the most commonly used arguments for not observing phosphopeptides in a particular sample, *i.e.* protein digests. This notion is based on the idea that the phosphate group of phosphorylated species is negatively charged, thereby affecting the ionization and detection efficiencies under acidic conditions used for mass spectrometric experiments in positive ion mode. To address this question, whether phosphopeptides indeed have lower detection/ionization efficiencies than their unphosphorylated cognates, the synthetic peptides and phosphopeptides used for the experiments described above were extensively purified by HPLC to homogeneity prior to amino acid analysis in replicate for quantitation. Stock solutions of all peptides were prepared using freshly calibrated pipettes to minimize the chances of pipetting errors, which would give rise to systematic deviations. Each peptide/phosphopeptide pair was mixed in at least two different defined ratios before performing repeated electrospray ionization mass spectrometric analyses. Microscale capillary HPLC was used to introduce the samples into the mass spectrometer. Magic C18 was used for the LC experiments because this reversed-phase material is a commonly used stationary phase in the field of proteomics. The buffers used were the same as described above. To determine the relative ionization/detection efficiencies for the peptide/phosphopeptide pairs, selected ion chromatograms were generated for each observed charge state of every peptide/phosphopeptide pair. Subsequently, the areas under the curves were calculated using the quantitation feature in Analyst QS. These integral values were then corrected for the defined concentration ratios (ranging from 4:1 to 1:4). The corrected ratios of those integrals reflect the ionization/detection efficiency ratio for the different peptide/phosphopeptide pairs. Throughout the article, the ionization/detection efficiency ratio for each peptide *versus* phosphopeptide is calculated such that a number >1 indicates that the unphosphorylated peptide has a better ionization/detection efficiency, whereas a ratio <1 indicates a better ionization/detection efficiencies for the phosphorylated form.

A quadrupole TOF mass spectrometer was used for the experiments. This type of mass spectrometer has a limited dynamic range. However, potential saturation-related problems are compensated for by the isotopic resolution provided by the instrument, which allowed choosing for the calculation of the ionization/detection efficiency ratios the signals of isotopes unaffected by saturation. The results of this set of experiments are listed in Table I, column 4. The peptides are grouped accounting for the number of basic amino acid residues and number of phosphorylation sites. The outcome of the experiments can be summarized as follows:

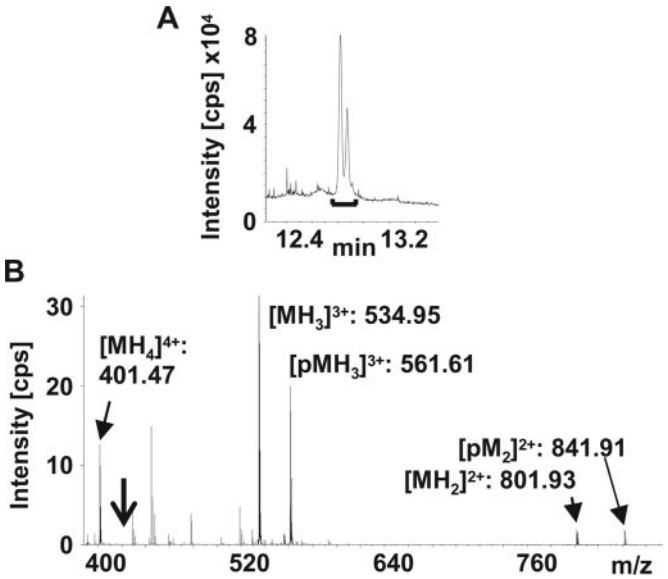

FIG. 4. **LC/MS analysis of a 2:1 mixture of GLGTRTGSNLDRDKL (M) and GLGTRTGpSNLDRDKL (pM).** TIC (*A*) and the corresponding mass spectrum (*B*) summed from 12.74 to 12.94 min.

More phosphopeptides species show better ionization/detection efficiencies than their unphosphorylated cognates as compared with the reversed situation. This trend is more pronounced when the peptide contains more than one basic amino acid residue. Even the phosphopeptides with only one basic amino acid residue which resemble as such His-free tryptic peptides show relative ionization/detection efficiencies (peptide *versus* phosphopeptide) ranging from 0.3 to 1.7 (Table I, Peptide species 1–5).

About 70% of the peptide pairs show relative ionization/detection efficiencies in the range of 0.5–2. However, it is not obvious (apart from the exceptions mentioned below) which peptide sequences give rise to ionization/detection efficiency differences >2 upon phosphorylation. This reflects the lack of general understanding of what determines the absolute ionization/detection efficiencies of peptides.

The ionization/detection efficiency ratio unphosphorylated peptide *versus* phosphopeptide increases with increasing the charge state. The extent of increase, however, varies significantly. The observed increases varied from 11% (SRARIG-(pS)DPLAYEPK: $2+ \rightarrow 3+$ (Table I, Peptide species 21 and 22)) to about 800% (KTQA(pS)QGTLQTR: $2+ \rightarrow 3+$ (Table I, Peptide species 8 and 9)). One example is presented in Fig. 4, which shows charge states of $2+$ to $4+$ of the peptide GLGTRTG(pS)NLDRDKL (Table I, Peptide species 14–16; please note that the peptide/phosphopeptide were combined in a 2:1 mixture). A clear decrease of the phosphopeptide ion signal relative to the unphosphorylated form is apparent with increasing charge state.

Although the number of doubly phosphorylated peptides investigated is too small to draw any general conclusions about their ionization/detection efficiencies as compared with

their unphosphorylated cognates, it is still interesting to note that all three test peptides show relative ionization/detection efficiencies similar to those of the singly phosphorylated species. Even in the one case where the number of phosphorylation sites is larger than the number of basic amino acid residues (*e.g.* DQAV(pT)E(pY)VATR (Table I, Peptide species 27)), the ionization/detection efficiency ratio of peptide *versus* phosphopeptides is only 2.0.

It was observed that for the three highly basic peptides, the most highly charged species of the unmodified cognate had ionization/detection efficiencies 2 to 3 orders of magnitude better than the phosphorylated complement (peptide species 16, 20, and 23). One example is given in Fig. 4. The triply charged phosphopeptide GLGTRTGpSNLDRDKL (Table I, Peptide species 16) is more efficiently ionized and detected than the unmodified cognate GLGTRTGSNLDRDKL. However, the quadruply charged unmodified species shows a vastly better ionization/detection efficiency as compared with its phosphorylated form, which is almost completely absent. A significant fraction of the molecules of these "exceptional" species show this very high charge state because of the overall proton affinity of the peptides. As soon as the proton affinity is reduced by *e.g.* phosphorylation, harboring four protons becomes energetically unfavorable, and the ion signal intensity of this charge state becomes negligible. The rather large standard deviations in these exceptional cases result from the limited dynamic range of the instrument used for the study.

The findings in this part of the study clearly show a general statement, stating that phosphopeptides show lower ionization/detection efficiencies than their unphosphorylated cognates, does not apply to electrospray ionization mass spectrometry as the majority of the phosphopeptides tested actually show better ionization/detection efficiencies than their unphosphorylated cognates. Similar ionization/detection efficiencies were also found for tryptic-like peptides, *i.e.* those with only one basic amino acid residue. This finding is supported by other studies from our own and other groups, which determined similar ionization/detection efficiencies for truly tryptic peptide/phosphopeptide pairs, *i.e.* peptides that were generated by tryptic digestion of proteins (4–6); all these studies report similar ionization/detection efficiencies for tryptic peptides and their unphosphorylated cognates.

This raises the questions (or highlights the lack of knowledge about) what determines the ionization/detection efficiencies of peptides. Some preliminary studies have been published investigating the effect of different physicochemical properties of peptides on the ionization/detection efficiencies during electrospray ionization. These studies highlighted a positive correlation between the hydrophobicities and the ionization/detection efficiencies of peptides as well as between pI values and ionization/detection efficiencies (*e.g.* Refs. 7–9). However, when following their line of argument that the phosphorylated form of a particular peptide should

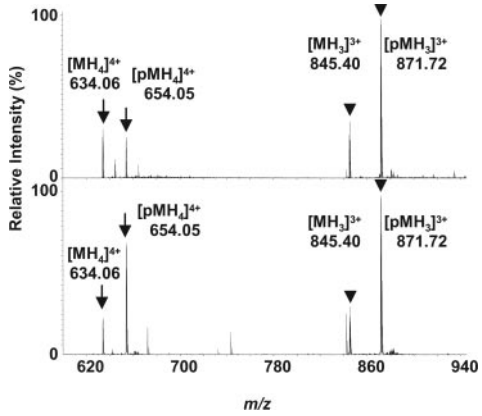

FIG. 5. **LC/MS analysis of a tryptic digest of the yeast transcription factor Pho4 analyzed under two different LC/MS conditions.** The data shown are the two spectra summed over the elution time of the peptide TSSSAEGVVVASESPVIAPHGSTHAR and its phosphorylated cognate TSSSAEGVVVASEpSPVIAPHGSTHAR without (*top panel*) and with heptafluorobutyric acid as ion pairing reagent (*bottom panel*; for details, see "Results and Discussion").

have lower ionization/detection efficiencies than their unmodified cognate because both the hydrophobicity and pI are thought to decrease with the covalent attachment of a phosphate group. Although the pI of a peptide is certainly reduced upon phosphorylation, the effect of phosphorylation on the hydrophobicity is not as clear. As described above, some kind of charge compensation/internal salt bridges/zwitterion formation might actually decrease the hydrophilicity of basic amino acid-containing peptides in diluted formic acid/acetonitrile/water solutions upon phosphorylation. However, a more detailed understanding of the determinants of the ionization/detection efficiencies of peptides still has to be established. Such a study must include the composition of the peptides as well as instrumental parameters and the composition of the spray solution. The importance of the latter was underscored by a phosphorylation analysis on the yeast transcription factor Pho4 performed in our laboratory. The same sample was analyzed twice. The following parameters were varied in the two LC/MS experiments: (i) a different electrospray emitter was used, and (ii) the LC buffer system was changed from water/acetonitrile/0.4% acetic acid/0.005% heptafluorobutyric acid to water/acetonitrile/0.2% formic acid. When observing the signal intensities of the ions corresponding to the phosphopeptide TSSSAEGVVVASE(pS)PVIAPHGSTHAR and its unmodified cognate it was noted that the triply charged species show almost identical intensity ratios for both sets of experimental conditions, whereas the signal intensity ratio of the quadruply charged peptide/phosphopeptide is significantly different, changing from 1.2 to 0.33 (Fig. 5). It should be noted that measuring relative ionization/detection efficiencies of peptides and their phosphorylated cognates is not only of educational importance in the context of this study but can be used for stable-isotope-free quantitation of protein phosphorylation stoichiometries as we have

recently shown (4). Once such an ionization/detection efficiency ratio has been determined the degree of phosphorylation can easily be calculated based on the ion signal intensity ratio corrected by the relative ionization/detection efficiencies. Because there is a lack of information regarding the factors influencing ionization/detection efficiencies, the ratios have to be determined empirically. However, it is foreseeable that this property can be theoretically estimated in the future once a better understanding of ionization/detection efficiencies has been provided. This would greatly simplify the task of quantitation of protein phosphorylation stoichiometries.

Conclusions and Perspectives—The data presented in this study shed some light onto some of the often used reasons explaining why mass spectrometric phosphorylation analysis is so difficult. The reasons investigated included the issue of increased hydrophilicity because of phosphorylation, the concern of selective suppression of phosphopeptides in the presence of unphosphorylated peptides, and the anecdotal lower ionization/detection efficiencies of phosphopeptides as compared with their unphosphorylated cognates. A set of synthetic peptide/phosphopeptide pairs was purified to homogeneity and quantitated by amino acid analysis. These peptide/phosphopeptide pairs were then used for numerous mass spectrometric experiments, which allowed one to address these issues separately. Several conclusions could be drawn from the results of these experiments.

By testing how hydrophilicity of peptides is altered upon phosphorylation, compelling evidence is provided that calculating nominal hydrophilicities by simply adding up hydrophilicity coefficients for each amino acid side chain is a suboptimal descriptor for the expected elution order of peptides and their phosphorylated cognates. Although phosphorylation clearly increases the nominal hydrophilicity of peptides with respect to their unphosphorylated cognate, this effect is overcompensated under certain LC conditions if the overall net charge of the peptide is decreased by the addition of phosphate groups. In other words, conditions can be used under which phosphopeptides are better retained on the reversed-phase column as long as the number of phosphorylation sites does not exceed the number of basic amino acid residues (Arg, His, and Lys) within the peptides. Under our chosen conditions the risk of losing singly phosphorylated tryptic peptides caused by reduced retention can be minimized. However, retention problems associated with increased hydrophilicity can still be problematic for multiply phosphorylated tryptic peptides without additional basic amino acid residues or phosphorylated peptides without basic residues.

The issue of selective ionization suppression of phosphopeptides in the presence of unphosphorylated peptides was tested a) by mixing increasing amounts of a standard protein digest with constant amounts of various peptide/phosphopeptide pairs and b) by mixing different protein digests with numerous peptide/phosphopeptide pairs. None of the data in our test series supported the notion of selective ionization suppression of phosphopeptides in the presence of unphosphorylated species; instead the signal intensity ratio peptide/phosphopeptide and the total ion signal even remained constant within the experimental error over the concentration and complexity range tested.

Similarly, the common belief that phosphopeptides show lower ionization/detection efficiencies than their unphosphorylated cognates could not be substantiated as a general fact. When testing the peptide/phosphopeptide pairs by using online LC electrospray ionization MS, the majority of the phosphopeptides tested actually showed better ionization/detection efficiencies than their unmodified cognates and even the tryptic-like peptides showed similar ionization/detection efficiencies corroborating earlier reports (4–6). This indicates that LC combined with electrospray ionization mass spectrometry is a good choice for routine phosphorylation analyses.

If several of the most commonly used arguments for not observing phosphorylation sites are questionable in light of this study, this raises the question as to why so many phosphopeptides are not observed. There are numerous reasons that can be grouped into biological and technical ones:

The human genome encodes not only 518 protein kinases, but also up to 65 protein phosphatases that can dephosphorylate proteins rapidly with high specificity (10–12) such that proteins phosphorylated in *e.g.* a signal transduction cascade in response to an extracellular stimuli are usually dephosphorylated rapidly back to the resting state once the stimuli has been removed (13). Also only a subset of the cellular complement of a particular protein may be involved in a particular signaling pathway, and hence only a small fraction of the total cellular pool of that protein may be phosphorylated at any one time. So in the preparation of phosphoproteins from cell lysates by immunoprecipitation or organelle isolation, the majority of the isolated protein is not likely to be stoichiometrically phosphorylated, and the subsequent proteolytic digest will then be dominated by unmodified peptides. These reasons underscore the notion that most of the phosphorylation sites are substoichiometrically modified.

This substoichiometric nature of protein modification highlights one of the limitations of automated, data-dependent MS/MS routines to select precursors for fragmentation based on ion signal intensity. In the situation where the stoichiometry of phosphorylation at any given site is $<10\%$, the ion intensity of the unphosphorylated peptide will always dominate the survey mass spectrum, and hence it is these precursors that will be automatically selected for MS/MS and not the lower abundance phosphopeptides. This applies especially to highly complex peptide mixtures in which the number of peptides species eluting off the column exceeds the number of peptides that can be sequenced in the same time period.

Thus, to improve the detection of phosphopeptides from proteolytic digests either phosphoprotein enrichment, phosphopeptide enrichment, or phosphopeptide selective mass

spectrometric detection methods must be employed. This selective enrichment of phosphoproteins can involve the use of metal chelate affinity of phosphoproteins prior to digestion (14) or phosphospecific antibody enrichment of phosphoproteins, such as those phosphorylated on tyrosine residues (15). Affinity purification of phosphopeptides from proteolytic digests can be performed by immunoaffinity methods (16), metal chelate affinity chromatography (17–20), ion exchange fractionation (1), or a combination thereof (21) prior to reverse phase LC-MS. The result of this enrichment of phosphoproteins and/or phosphopeptides is that the proteolytic digest analyzed by LC/MS is highly simplified, and phosphopeptide ions are more likely to be selected for MS/MS fragmentation because they now represent the more abundant peptide ions. Alternatively, mass spectrometric scanning features for the selective detection of phosphorylated species on a triple quadrupole or hybrid triple quadrupole linear ion trap mass spectrometer such as precursor ion scanning (22, 23) or neutral loss scanning can detect the generation of either PO_3^- ions or the neutral loss of H_3PO_4 from collision induced dissociation of phosphorylated precursor ions (24). Once these ions are detected they can be either automatically selected for MS/MS, or if necessary, manually assigned for MS/MS on a subsequent LC/MS analysis.

Furthermore, improved sequence coverage upon proteolysis of the proteins of interest is of pivotal importance to identify and locate the protein phosphorylation sites. For protein identification purposes any set of peptide suffices, whereas for more comprehensive protein characterization which includes phosphorylation analysis, significantly improved sequence coverage is necessary. This means that more effort should be directed toward (*a*) miniaturizing and minimizing of the sample handling to reduce losses (25) and (*b*) the analysis of smaller and/or larger proteolytic peptides, *i.e.* those outside the normally observed *m/z* range. Alternatively, in solution protein separation, handling and data interpretation tools should be improved such that FTICR-based top-down analyses of proteins become more routine. Top-down analyses by definition provide 100% sequence coverage. If following this approach the dynamic range of the mass spectrometer becomes very important to be able to analyze protein isoforms of lower abundance in the presence of high abundance proteins.

Another obstacle for the analysis of phosphopeptides and the exact localization of the phosphorylation site is the facile loss of phosphoric acid that is commonly observed for serine- and threonine-phosphorylated species. This facile loss leads to information-poor MS/MS spectra dominated by the neutral loss of phosphoric acid. Hence, some phosphopeptides might have been picked during data-dependent acquisition routines, but the following database search did not provide any meaningful hit or the unambiguous localization of the modification site was not possible. However, this limitation can be partially compensated for by using automated routines that trigger MS/MS/MS experiments if a significant neutral loss of 98/z (*z* being the charge state of the selected precursor ion) is observed (1); nevertheless, such an approach still requires that the modified peptides has been selected for fragmentation based on its intensity (see above).

In conclusion, the data presented above suggest that phosphopeptides are difficult to identify in phosphoprotein digests because of the substoichiometric nature of this modification and not because they have poor ionization efficiencies. Thus, most of the technical limitations discussed above would profit from mass spectrometers with improved detection limits and dynamic ranges and the development of more robust and reliable selective detection and enrichment procedures.

Acknowledgments—We thank Michael Springer, Michael Brenner, Nagi Ayad (Department of Systems Biology, Harvard Medical School), Scott Gerber, Wilhelm Haas, Steve Gygi (Department of Cell Biology, Harvard Medical School), and Benedikt Kessler (Department of Pathology, Harvard Medical School) for fruitful discussions and instrument support. J. R. acknowledges Jeffrey Knott and Jason Reynolds (Cell Signaling Technology) for amino acid analysis. N. M. thanks the Medical Research Council for financial support and Dr. Graham Bloomberg, University of Bristol, UK, for peptide synthesis.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed: Children's Hospital Boston, Department of Pathology/Enders 1130, 300 Longwood Ave., Boston, MA 02115. Tel.: 617-919-2629; E-mail: hanno.steen@childrens.harvard.edu.

REFERENCES

1. Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villen, J., Li, J., Cohn, M. A., Cantley, L. C., and Gygi, S. P. (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 12130–12135
2. Ballif, B. A., Roux, P. P., Gerber, S. A., MacKeigan, J. P., Blenis, J., and Gygi, S. P. (2005) Quantitative phosphorylation profiling of the ERK/p90 ribosomal S6 kinase-signaling cassette and its targets, the tuberous sclerosis tumor suppressors. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 667–672
3. Schlosser, A., Pipkorn, R., Bossemeyer, D., and Lehmann, W. D. (2001) Analysis of protein phosphorylation by a combination of elastase digestion and neutral loss tandem mass spectrometry. *Anal. Chem.* **73,** 170–176
4. Steen, H., Jebanathirajah, J. A., Springer, M., and Kirschner, M. W. (2005) Stable isotope-free relative and absolute quantitation of protein phosphorylation stoichiometry by MS. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 3948–3953
5. Tsay, Y. G., Wang, Y. H., Chiu, C. M., Shen, B. J., and Lee, S. C. (2000) A strategy for identification and quantitation of phosphopeptides by liquid chromatography/tandem mass spectrometry. *Anal. Biochem.* **287,** 55–64
6. Hegeman, A. D., Harms, A. C., Sussman, M. R., Bunner, A. E., and Harper, J. F. (2004) An isotope labeling strategy for quantifying the degree of phosphorylation at multiple sites in proteins. *J. Am. Soc. Mass Spectrom.* **15,** 647–653
7. Cech, N. B., Krone, J. R., and Enke, C. G. (2001) Predicting electrospray response from chromatographic retention time. *Anal. Chem.* **73,** 208–213
8. Nielsen, M. L., Savitski, M. M., Kjeldsen, F., and Zubarev, R. A. (2004) Physicochemical properties determining the detection probability of tryptic peptides in Fourier transform mass spectrometry. A correlation

study. *Anal. Chem.* **76,** 5872–5877
9. Pan, P., Gunawardena, H. P., Xia, Y., and McLuckey, S. A. (2004) Nano-electrospray ionization of protein mixtures: solution pH and protein pI. *Anal. Chem.* **76,** 1165–1174
10. Cohen, P. (2002) The origins of protein phosphorylation. *Nat. Cell Biol.* **4,** E127–E130
11. 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
12. Berwick, D. C., and Tavare, J. M. (2004) Identifying protein kinase substrates: hunting for the organ-grinder's monkeys. *Trends Biochem. Sci.* **29,** 227–232
13. Cohen, P. T. W. (2003). Protein serine/threonine phosphatases and the PPP family, In Handbook of Cell Signaling, R. A. Bradshaw, and E. A. Dennis, eds., pp. 593–600, Academic Press, San Diego, CA
14. Collins, M. O., Yu, L., Coba, M. P., Husi, H., Campuzano, I., Blackstock, W. P., Choudhary, J. S., and Grant, S. G. (2005) Proteomic analysis of *in vivo* phosphorylated synaptic proteins. *J. Biol. Chem.* **280,** 5972–5982
15. Hinsby, A. M., Olsen, J. V., and Mann, M. (2004) Tyrosine phosphoproteomics of fibroblast growth factor signaling: a role for insulin receptor substrate-4. *J. Biol. Chem.* **279,** 46438–46447
16. Rush, J., Moritz, A., Lee, K. A., Guo, A., Goss, V. L., Spek, E. J., Zhang, H., Zha, X. M., Polakiewicz, R. D., and Comb, M. J. (2005) Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat. Biotechnol.* **23,** 94–101
17. Posewitz, M. C., and Tempst, P. (1999) Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal. Chem.* **71,** 2883–2892
18. Stensballe, A., Andersen, S., and Jensen, O. N. (2001) Characterization of phosphoproteins from electrophoretic gels by nano-scale Fe(III) affinity chromatography with off-line mass spectrometry analysis. *Proteomics* **1,** 207–222
19. Ficarro, S. B., McCleland, 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
20. Cole, A. R., Knebel, A., Morrice, N. A., Robertson, L. A., Irving, A. J., Connolly, C. N., and Sutherland, C. (2004) GSK-3 phosphorylation of the Alzheimer epitope within collapsin response mediator proteins regulates axon elongation in primary neurons. *J. Biol. Chem.* **279,** 50176–50180
21. Gruhler, A., Olsen, J. V., Mohammed, S., Mortensen, P., Faergeman, N. J., Mann, M., and Jensen, O. N. (2005) Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol. Cell. Proteomics* **4,** 310–327
22. Annan, R. S., Huddleston, M. J., Verma, R., Deshaies, R. J., and Carr, S. A. (2001) A multidimensional electrospray MS-based approach to phosphopeptide mapping. *Anal. Chem.* **73,** 393–404
23. Le Blanc, J. C., Hager, J. W., Ilisiu, A. M., Hunter, C., Zhong, F., and Chu, I. (2003) Unique scanning capabilities of a new hybrid linear ion trap mass spectrometer (Q TRAP) used for high sensitivity proteomics applications. *Proteomics* **3,** 859–869
24. Steen, H., Küster, B., and Mann, M. (2001) Quadrupole time-of-flight versus triple-quadrupole mass spectrometry for the determination of phosphopeptides by precursor ion scanning. *J. Mass Spectrom.* **36,** 782–790
25. Ficarro, S. B., Salomon, A. R., Brill, L. M., Mason, D. E., Stettler-Gill, M., Brock, A., and Peters, E. C. (2005) Automated immobilized metal affinity chromatography/nano-liquid chromatography/electrospray ionization mass spectrometry platform for profiling protein phosphorylation sites. *Rapid Commun. Mass Spectrom.* **19,** 57–71