Anal. Chem. **2001,** *73,* 393−404

Accelerated Articles

A Multidimensional Electrospray MS-Based Approach to Phosphopeptide Mapping

Roland S. Annan,*,† Michael J. Huddleston,† Rati Verma,‡ Raymond J. Deshaies,‡ and Steven A. Carr*,†

Department of Physical and Structural Chemistry, SmithKline Beecham Pharmaceuticals, King of Prussia, Pennsylvania 19406, and Division of Biology, California Institute of Technology, Pasadena, California 91125

A new, multidimensional electrospray MS-based strategy for phosphopeptide mapping is described which eliminates the need to radiolabel protein with ^{32}P or ^{33}P. The approach utilizes two orthogonal MS scanning techniques, both of which are based on the production of phosphopeptide-specific marker ions at *m*/*z* 63 and/or 79 in the negative ion mode. These scan methods are combined with liquid chromatography−electrospray mass spectrometry and nanoelectrospray MS/MS to selectively detect and identify phosphopeptides in complex proteolytic digests. Low-abundance, low-stoichiometry phosphorylation sites can be selectively determined in the presence of an excess of nonphosphorylated peptides, even in cases where the signal from the phosphopeptide is indistinguishable from background in the conventional MS scan. The strategy, which has been developed and refined in our laboratory over the past few years, is particularly well suited to phosphoproteins that are phosphorylated to varying degrees of stoichiometry on multiple sites. Sensitivity and selectivity of the method are demonstrated here using model peptides and a commercially available phosphoprotein standard. In addition, the strategy is illustrated by the complete in vitro and in vivo phosphopeptide mapping of Sic1p, a regulator of the G1/S transition in budding yeast.

Among the thousands of proteins expressed in a typical mammalian cell, as many as one-third are now thought to be phosphorylated.[1] Reversible protein phosphorylation is probably the single most common intracellular signal transduction event. Knowledge of the particular residues phosphorylated on a protein can provide insight into a signaling pathway via an understanding of how the protein's activity is regulated and which enzymes are responsible for the regulation. However, isolating and sequencing phosphopeptides derived from protein digests to identify specific phosphorylated residues remains a labor-intensive, time-consuming challenge. Typical biochemical or molecular biology protocols for phosphopeptide mapping all involve first labeling the protein with radioactive [^{32}P]phosphate either in vitro using purified kinase or in vivo by metabolically labeling cells. Following enzymatic digestion of the labeled protein, phosphopeptides are isolated for further analysis. Perhaps the most common method for analyzing protein phosphorylation is two-dimensional (2D) phosphopeptide mapping.[2] Only a few hundred dpm are required for a ^{32}P-labeled phosphopeptide to be detected by autoradiography or phosphor imaging. Identification of a phosphorylated residue is usually accomplished by excising the ^{32}P-labeled spots from the 2D phosphopeptide map and performing a combination of phosphoamino acid analysis[2,3] and manual Edman sequencing, monitoring for the loss of radioactivity at each cycle.[4,5] The combination of these three techniques can identify both the position and the identity of the phosphorylated residue[3] within a given peptide sequence. However, the amount of peptide obtained from a 2D phosphopeptide map is generally insufficient to observe the

* Corresponding authors: (e-mail) Roland_S_Annan@sbphrd.com; Steven_A_Carr@sbphrd.com.

† SmithKline Beecham Pharmaceuticals.

‡ California Institute of Technology.

(1) Hubbard, M. J.; Cohen, P. *Trends Biochem. Sci.* **1993**, *18*, 172−177.
(2) Boyle, W. J.; Van Der Geer, P.; Hunter T. In *Protein Phosphorylation*; Hunter, T., Sefton, B. M., Eds.; Methods in Enzymology 201; Academic Press: San Diego, New York, 1991; pp 110−149.
(3) Van Der Geer, P.; Hunter, T. *Electrophoresis* **1994**, *15*, 544−554.
(4) Roach, P. J.; Wang, Y. In *Protein Phosphorylation*; Hunter, T., Sefton, B. M., Eds.; Methods in Enzymology 201; Academic Press: San Diego, New York, 1991; pp 169−185.
(5) Sullivan, S.; Wong, T. W. *Anal. Biochem.* **1991**, *197*, 65−68.

10.1021/ac001130t CCC: $20.00
Published on Web 12/29/2000

nonlabeled amino acids released by the Edman method; thus, the identification of the phosphorylation site is indirect, requiring prior knowledge or assumption as to the sequence of the phosphopeptide. Confirmation of the sequence is usually accomplished by preparing synthetic phosphopeptides[6] and comparing the migration in the 2D phosphopeptide map of the synthetic material with the naturally derived peptide. Alternatively, the suspected sites can be knocked out individually by site-directed mutagenesis and the native and mutant maps compared.[7]

In contrast, mass spectrometry is ideally suited for detecting the presence of posttranslational modifications or modified amino acids and locating the position of the specific modified residues.[8,9] When the protein sequence is known, the molecular weight of a peptide may be used to corroborate the presence of phosphate by the 80 Da difference between the observed mass of the peptide and that calculated on the basis of the sequence. Thus, phosphopeptides can be identified in ^{32}P-labeled HPLC fractions, even though the fractions contain other nonphosphorylated peptides. Any ambiguities can be resolved by sequencing the putative phosphopeptides by collision-induced dissociation (CID) tandem mass spectrometry.[10,11] This strategy has been successfully applied in a growing number of examples.[12−18] In a very useful variation of this approach, an aliquot of the sample is treated with a nonspecific phosphatase and then MALDI spectra are recorded of the treated and untreated sample to produce a phosphopeptide difference map.[19,20] Phosphopeptides are detected by the disappearance of a peak and the increase or sudden appearance of a peak 80 Da lower in mass.

Isolating phosphopeptides preferentially from among all of the other peptides present in a protein digest still relies, in most cases, on labeling the protein beforehand with radioactive [^{32}P]phosphate and Cherenkov counting of HPLC fractions from the protein digest. Efforts to isolate phosphopeptides that circumvent this bothersome process have centered mainly on two approaches: use of immobilized metal ion affinity techniques both off-line[21] and coupled on-line with an ES mass spectrometer[22,23] and selective detection in MS based on the unique fragmentation behavior of phosphopeptides.[24,31] Our laboratory previously reported an MS-based method to selectively detect phosphopeptides at the low-picomole level in complex mixtures, which relied on the fact that CID of phosphopeptides gave rise to phosphopeptide-specific marker ions in the negative ion mode.[24] By rapidly stepping voltages in the high-pressure region between the source and the quadrupole mass analyzer, we demonstrated that it was possible to detect phosphopeptide-specific marker ions and measure peptide molecular weights in the same scan. Alternatively, where the highest sensitivity for phosphopeptide detection is required, selected ion monitoring under continuous high-level CID conditions can be used. Plotting the phosphopeptide marker ions versus the total ion current (TIC) from an LC−ESMS experiment or the UV trace showed which regions of the chromatogram contained phosphopeptides. The marker ion profile is analogous to the output from an HPLC radioactivity detector or the autoradiogram from a 2D phosphopeptide map, but the MS-based method does not rely on $^{32/33}$P labeling.

Unfortunately, many phosphoproteins of interest, especially those involved in signaling, are present in cells at only low copy numbers. In addition, phosphorylation stoichiometry at any given site can frequently be quite low. Thus, the sensitivity with which any type of phosphopeptide mapping can be done remains an important issue. It soon became apparent to us that while the phosphate-specific marker ion LC−ESMS technique described above could efficiently isolate phosphopeptide-containing fractions, it suffered the same limitation as radioactivity detection and IMAC affinity chromatography. The isolated fractions never contained just a single peptide, and thus, it was still necessary to identify the molecular weight that corresponded to the phosphopeptide, which may be present at only a few percent. Furthermore, like radioactivity, the mass spectrometer's sensitivity to detect the phosphate-specific marker ion was better than the mass spectrometer's sensitivity to measure the molecular weights of the peptides giving rise to the marker ion signals.

We had encountered this problem previously using a very similar marker ion strategy which detected glycopeptides in complex mixtures.[25] To circumvent the problem, we developed a precursor ion scanning method on the triple quadrupole that selectively detected the molecular masses of glycopeptides that could fragment to produce carbohydrate-specific marker ions.[26] Unfortunately, the precursor scan mode used during standard low-pH on-line LC−MS analysis has insufficient sensitivity to detect phosphorylated peptides, and we have had only limited success to date using high-pH HPLC−ESMS.[27]

(6) Carter, W. G.; Asamoah, K. A.; Sale, G. J. *Biochemistry* **1995**, *34*, 9488−9499.
(7) Najjar, S. M.; Philippe, N.; Suzuki, Y.; Ignacio, G. A.; Formisano, P.; Accili, D.; Tayor, S. I. *Biochemistry* **1995**, *34*, 9341−9349.
(8) Carr, S. A.; Biemann, K. In *Posttranslational Modification*, Part A; Wold, F., Moldave, K., Eds; Methods in Enzymology 106; Academic Press: Orlando, FL, 1984; pp 29−57.
(9) Annan, R. S.; Carr, S. A. *J. Protein Chem.* **1997**, *16*, 391−402.
(10) Gibson, B. W.; Cohen, P. In *Methods Enzymol.* **1990**, *193*, 480−501.
(11) Michel, H.; Hunt, D. F.; Shabanowitz, J.; Bennett, J. *J. Biol. Chem.* **1988**, *263*, 1123−1130.
(12) Erickson, A. K.; Payne, M.; Martino, P. A.; Rossomando, A. J.; Shabanowitz, J.; Weber, M. J.; Hunt, D. F.; Sturgill, T. W. *J. Biol. Chem.* **1990**, *265*, 19728−19735.
(13) Rossomando, A. J.; Wu, J.; Michel, H.; Shabanowitz, J.; Hunt, D. F.; Weber, M. J.; Sturgill, T. W. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5779−5783.
(14) Palczewski, K.; Buczylko, J.; Hooser, P. V.; Carr, S. A.; Huddleston, M. J.; Crabb, J. W. *J. Biol. Chem.* **1992**, *267*, 18991−18998.
(15) Hou, J.; McKeehan, K.; Kan, M.; Carr, S. A.; Huddleston, M. J.; Crabb, J. W.; McKeehan, W. L. *Protein Sci.* **1993**, *2*, 86−92.
(16) Resing, K. A.; Johnson, R. S.; Walsh, K. A. *Biochemistry* **1995**, *34*, 9477−9487.
(17) Ladner, R. D.; Carr, S. A.; Huddleston, M. J.; McNulty, D. E.; Caradonna, S. J. *J. Biol. Chem.* **1996**, *271*, 7752−7757.
(18) Palaty, C. K.; Kalmar, G.; Tai, G.; Oh, S.; Amankawa, L.; Affolter, M.; Aebersold, R.; Pelech, S. L. *J. Biol. Chem.* **1997**, *27*, 10514−21.
(19) Liao, P.; Leykam, J.; Andrews, P. C.; Gage, D. A.; Allison, J. *Anal. Biochem.* **1994**, *219*, 9−20.
(20) Zhang, X.; Herring, C. J.; Romano, P. R.; Szczepanowska, Brzeska, H.; Hinnebusch, A. G.; Qin, J. *Anal. Chem.* **1998**, *70*, 2050−2059.
(21) Posewitz, M. C.; Tempst, P. *Anal. Chem.* **1999**, *71*, 2883−2892.
(22) Nuwaysir, L. M.; Stults, J. T. *J. Am Soc Mass Spectrom.* **1993**, *4*, 662−669.
(23) Watts, J. D.; Affolter, M.; Krebs, D. L.; Wange, R. L.; Samelson, L. E.; Aebersold, R. *J. Biol. Chem.* **1994**, *269*, 29520−29529.
(24) Huddleston, M. J.; Annan, R. S.; Bean, M. F.; Carr, S. A. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 710.
(25) Huddleston, M. J.; Bean, M. F.; Carr, S. A. *Anal. Chem.* **1993**, *66*, 877−884.
(26) Carr, S. A.; Huddleston, M. J.; Bean, M. F. *Protein Sci.* **1993**, 2, 183−196.
(27) Stoney, K.; Huddleston, M. J.; Baldwin, S.; Carr, S. A. Abstracts of the Proceedings of the 44th Annual Meeting of the American Society for Mass Spectrometry, 1996; p 209.

The introduction of nanoelectrospray in 1994 by Wilm and Mann provided the ability to signal average very weak signals for long periods of time.[28,29] Our laboratory and Mann's quickly adopted the nanoelectrospray technique for phosphopeptide analysis using precursor ion scans on the entire unfractionated phosphoprotein digest.[30−32] By making the pH of the spray solution basic, we demonstrated selective detection and sequencing of phosphopeptides in complex mixtures at the low (<10)-femtomole range.[31] However, as we originally pointed out,[31] two limitations of this approach are that some phosphopeptides, particularly those with low site stoichiometry, may go undetected in the unfractionated mixture, and the digest must be desalted (partially or completely) prior to analysis.

Here we describe a new, multidimensional electrospray MS-based strategy for phosphopeptide mapping which, by combining elements of each of our previously described strategies, overcomes the individual limitations noted above. In the first dimension of the process, phosphopeptides present in the proteolytic digest of a protein are selectively detected and collected into fractions during on-line LC−ESMS, which monitors for the phosphopeptide-specific marker ion m/z 79 in the negative ion mode. The inclusion of the chromatographic step results in phosphopeptides being present in greatly simplified mixtures. In the second dimension of the analysis, the molecular weights of the phosphopeptides in each collected fraction are determined by precursor scans for m/z 79 in the negative ion mode. Phosphopeptides are sequenced in the third dimension of analysis by positive ion nanoelectrospray tandem MS of the molecular species identified. The strategy, which has been developed and refined in our laboratory over the past few years, is particularly well suited to phosphoproteins containing many sites of phosphorylation and variable stoichiometry at individual sites. Sensitivity and selectivity of the method are demonstrated here using model peptides and a commercially available phosphoprotein standard. In addition, the strategy is illustrated by the complete in vitro and in vivo phosphopeptide mapping of Sic1p, a regulator of the G1/S transition in budding yeast. Sic1p was found to contain between 6 and 12 phosphorylation sites with site stoichiometries ranging from 4.5 to 100%. Furthermore,the examples presented in this report clearly demonstrate why any phosphopeptide mapping strategy must incorporate peptide sequencing as the final step in characterizing phosphorylation sites.

MATERIALS AND METHODS

Protein Samples. α_S-Casein was purchased from Sigma. The protein was digested with trypsin (Boerhringer) in 50 mM ammonium bicarbonate buffer, pH 8.5, for 5 h at 38 °C at an enzyme-to-substrate ratio of 1:50 (w/w). MBP−Sic1p phosphorylated in vitro or Sic1p^{HAHis6} purified from *Saccharomyces cerevisiae* strain RJD 1044[33] was digested with modified trypsin (Promega) in 1.5 M urea and 50 mM ammonium bicarbonate buffer, pH 8.5, for 16 h at 38 °C at an enzyme-to-substrate ratio of 1:7 (w/w). Phosphorylated myoglobin tryptic digest was supplied as part of a study by the Association of Biomolecular Resource Facilities (ABRF). The sample consists of a tryptic digest of horse apomyoglobin spiked with two synthetic phosphorylated myoglobin peptides[34] LFTGHPEpTLEK (M_r 1350.6) and ApSEDLK (M_r 741.3) at a ratio of 1:1 with the nonphosphorylated sequences. The peptide KRPpSQRHGSKY (M_r 1422.7) was synthesized at the University of Michigan Peptide and Protein Facility.

LC−ESMS. Prior to analysis, all samples were acidified with 0.1% TFA and injected onto a 1-mm Peptide Trap Cartridge (Michrom BioResources Inc., Pleasanton, CA.) which had been installed in place of the sample loop on a Rheodyne model 8125 injector (Cotati, CA). After the trap was washed, the injector was rotated into the inject position and the sample back-flushed off the trap onto the analytical column using acetonitrile/water/TFA gradients. Separations were carried out on 0.5- or 1.0-mm-i.d. Reliasil HPLC columns using a Michrom UMA gradient HPLC system at flow rates of 20 and 50 μL/min, respectively. The column eluent was split after the UV detector with 4−5 μL/min going to the mass spectrometer and the remainder going to a fraction collector taking either 0.5- or 1-min fractions.

Electrospray mass spectra were acquired on a modified Perkin-Elmer Sciex API III atmospheric pressure ionization triple quadrupole tandem mass spectrometer (Thornhill, ON, Canada). This instrument is equipped with a high-pressure collision cell (PE-Sciex)[35] and either the standard articulated, pneumatically assisted nebulization electrospray probe or an articulated nanoelectrospray interface developed by Wilm and Mann[28] and built at the European Molecular Biology Laboratory in Heidelberg, Germany.

To selectively detect phosphopeptides in the first-dimension LC−ESMS analysis, the mass spectrometer, operating in the negative ion mode, was optimized to produce and detect m/z 63 and 79 product ions (PO_2^- and PO_3^-) produced by CID in quad 0, the high-pressure region located between the sampling cone and the quadrupole mass filter. These ions are highly specific for Ser, Thr, or Tyr phosphorylation.[24] The production of the m/z 63 and 79 product ions in the source was maximized by application of −350 (m/z 63) and −300 V (m/z 79) to the orifice of the API III as was previously described.[24] The MS is operated in a single ion monitoring mode for enhanced sensitivity. The signal from the UV detector is taken into the MS data system, so that the UV and MS chromatograms can be easily aligned. HPLC fractions that are shown by the first-dimension MS response to contain phosphorylated peptides are analyzed in the second dimension using nanoES and precursor ion scans.

Protein molecular weight measurements and peptide mapping were done in the positive ion mode under typical operating conditions[26] with the instrument scanning from m/z 400 to 2000 in 0.25 Da steps (MBP−Sic1p data) and 0.5 Da steps (Sic1p data). For intact Sic1p molecular weight data, the instrument was

(28) Wilm, M.; Mann, M. *Int. J. Mass Spectrom. Ion Processes* **1994**, *136*, 167−180.

(29) Wilm, M.; Mann, M. *Anal. Chem.* **1994**, *66*, 4390−4399.

(30) Huddleston, M. J.; Annan, R. S.; Carr, S. A. Femtomole Level Selective Detection of Phosphopeptides in Protein Digests By MS, Ninth Symposium of the Protein Society, Boston, MA, 1995.

(31) Carr, S. A.; Huddleston, M. J.; Annan, R. S. *Anal. Biochem.* **1996**, *239*, 180−192.

(32) Wilm, M.; Neubauer, G.; Mann, M. *Anal. Chem.* **1996**, *68*, 1−8.

(33) Verma, R.; Annan, R. S.; Huddleston, M. J.; Carr, S. A.; Reynard, G.; Deshaies, R. J. *Science* **1997**, *278*, 455−560.

(34) Swiderek, K.; Gharahdaghi, F.; Ericsson, L.; Hackett, M.; Fowler, B.; Hathaway, G.; Loo, R. O.; Johnson, R. A.; Stults, J. The 1997 ABRF Mass Spectrometry Committee Collaborative Study: Identification of Phosphopeptides in a Tryptic Digest of Apomyoglobin, ABRF'98: From Genomes to Function−Technical Challenges of the Post-Genome Era, San Diego, CA, 1998.

(35) Thomson, B. A.; Douglas, D. J.; Corr, J. J.; Hager, J. W.; Jollife, C. L. *Anal. Chem.* **1995**, *67*, 1696−1704.

scanned from m/z 900 to 2200 in 0.33 Da steps. The m/z windows and the dwell times per step were adjusted to give total scan times of ~5 s. Protein samples were redissolved in 50:50 MeOH/water with 5% formic acid and analyzed by positive ion nanoESMS.

NanoES-MS and MS/MS. The configuration and operation of the nanoES source on the PE Sciex API III has been described previously.[29−31] To measure the molecular weights of the phosphopeptides in the fractions targeted by the first-dimension LC−ESMS analysis, the second dimension of the phosphopeptide mapping experiment utilizes negative ion nanoES to perform precursor ion scans for the m/z 79 marker ion. To enhance detection of the phosphopeptides, an aliquot of each fraction is made basic prior to precursor ion analysis. One-quarter to one-half of a targeted HPLC fraction is taken to dryness and reconstituted in 50:50 methanol/water containing 5% concentrated ammonium hydroxide (30 wt %) and 1−2 μL of the sample is introduced into the nanoES needle. First a full-scan negative ion spectrum is recorded with the resolution set so that it is possible to differentiate singly and doubly from higher charge state ions. Then, a negative ion precursor scan for m/z 79 is recorded. After the molecular weights of the phosphopeptides have been determined from the m/z 79 precursor ion scan, the mass spectrometer is switched into the positive ion mode, and if sufficient signal is present, the phosphopeptides are sequenced from the same sample loading by acquiring a full-scan CID product ion spectrum (MS/MS) on a chosen multiply charged precursor. Typically, a doubly or triply charged precursor ion is selected by quad 1 and undergoes low-energy collisions with Ar gas in the collision cell (quad 2). The product ion spectrum is then recorded by scanning quad 3. Alternatively, an acidic aliquot of the sample in 50:50 MeOH/water with 5% formic acid is used for MS/MS sequencing of the phosphopeptides. Peptide molecular weights derived from precursor ions scans are reported as average mass values. Otherwise, peptide and product ion masses are reported as monoisotopic values. Peptide fragment ion nomenclature is that of Biemann,[36] except that b_n^{2+} and y_n^{2+} refer to $(b_n + H)^{2+}$ and $(y_n + H)^{2+}$ and that $b_n^{\blacktriangle}$ or $y_n^{\blacktriangle}$ refer to $(b_n - H_3PO_4)^+$ and $(y_n - H_3PO_4)^+$.

RESULTS AND DISCUSSION

Multidimensional MS-Based Phosphopeptide Mapping Strategy. The multidimensional MS-based phosphopeptide mapping strategy presented in this paper relies mainly on two orthogonal MS scanning techniques, both of which selectively detect phosphopeptide-specific marker ions. In the first dimension of the process, phosphopeptides present in a proteolytic digest of the protein are selectively detected and collected during on-line LC−ESMS in the selected ion monitoring mode. Under collision-induced dissociation conditions in the high-pressure region located between the ion sampling cone and the quadrupole mass filter, peptides containing phosphorylated residues produce the highly diagnostic marker ions $[PO_2]^-$ and $[PO_3]^-$ at m/z 63 and 79, respectively.[24] Monitoring for these two ions permits selective collection of LC fractions that contain phosphopeptides (as well as coeluting nonphosphorylated peptides). Thus, the first dimension of our technique is the functional equivalent of conventional 2D phosphopeptide mapping with the critical difference being that no radiolabel is required.

In the second dimension of the analysis, the molecular weights of the phosphopeptides in each fraction are determined. Each phosphopeptide-containing fraction identified in the first-dimension experiment is analyzed by negative ion nanoelectrospray with precursor ion scanning for m/z 79.[30−32] From the m/z value of the precursor, the molecular weights of the various phosphopeptides can be determined. Because this method detects only precursors that can fragment to yield the specific marker ion, unmodified peptides in these fractions are not detected. This high degree of selectivity allows the precursor scan to identify phosphopeptides that are present as very minor components of the sample. Whenever possible, as a final dimension of analysis, the phosphopeptides are sequenced by positive ion nanoelectrospray tandem MS.

To illustrate the overall strategy, we used a mixture consisting of three synthetic phosphopeptides LFTGHPEpTLEK (M_r 1350.6), ApSEDLK (M_r 741.3), and KRPpSQRHGSKY (M_r 1422.7) spiked in equimolar amounts into an equivalent amount of a tryptic digest of horse apomyoglobin, which has no native sites of phosphorylation. Five picomoles of this test mixture was injected onto a 0.5-mm-i.d. HPLC column. The result of the first-dimension analysis is shown in Figure 1A, which compares the UV and the summed single ion current traces for m/z 79 and 63. Two major peaks can be observed with a ≥10:1 signal-to-noise ratio indicating that ~1−5 pmol of phosphopeptide injected is a conservative lower working limit for the first-dimension analysis when using a 0.5-mm-i.d. HPLC column with a flow rate of 20 μL/min. It is important to note that sensitivity in this first-dimension step is limited not by the MS but by the need to have a flow rate sufficient to permit fraction collection (see discussion of sensitivity, below).

Second-dimension analysis using nanoelectrospray precursor ion scanning for m/z 79 of the two fractions identified above as containing phosphopeptides indicated that peak 1 contained a phosphopeptide of M_r (av) = 1424.0 and that peak 2 contained a phosphopeptide of M_r (av) = 1351.1 (Figure 1B and C, respectively). These masses compare well with the calculated average mass for two of the spiked phosphopeptides KRPpSQRHGSKY (M_r (av) = 1423.7) and LFTGHPEpTLEK (M_r (av) = 1351.4). We did not recover the 741 Da phosphoserine peptide because it is unretained and elutes in the column void. Small, highly charged hydrophilic peptides are often problematic for reversed-phase HPLC. A recent report on the use of graphite carbon columns for capturing low molecular weight, hydrophilic peptides that elute in the RP C18 flow-through[37] offers some hope for the analysis of small phosphopeptides. We are currently investigating the practical utility of these columns.

The same sample loading in the spray needles used to produce the precursor ion spectra were next used to record a positive ion MS/MS spectra for each of the two phosphopeptides (Figure 1D and E). The spectrum obtained from the phosphopeptide eluting in peak 1 (Figure 1D) contains very few structurally useful fragment ions. The peptide in this fraction is not a tryptic peptide and contains four basic residues (K and R) scattered throughtout the sequence. Peptides of this nature tend to yield spectra that are not structurally informative, because the internal energy imparted by the CID process is distributed into too many different

(36) Biemann, K. In *Mass Spectrometry*; McCloskey, J. A., Ed.; Methods in Enzymology 193; Academic Press: San Diego, New York, 1990; pp 886−887.

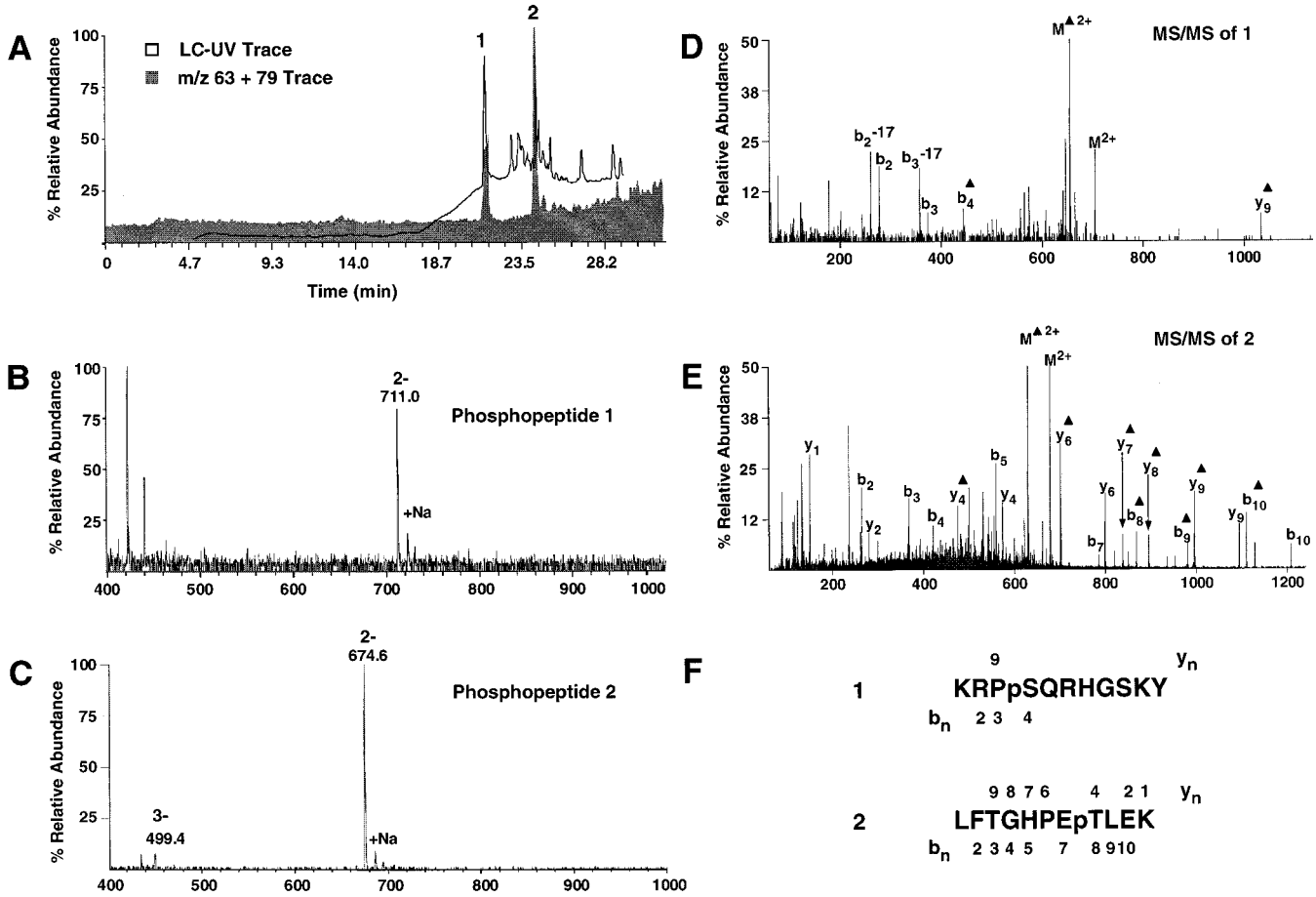

Figure 1. Multidimensional electrospray MS approach to phosphopeptide mapping. To demonstrate the strategy, a mixture composed of 5 pmol of horse apomyoglobin tryptic digest containing 5 pmol each of three synthetic phosphopeptides was injected onto a 0.5-mm-i.d. RP-HPLC column. The column eluent was split 5:1 to a fraction collector and the ESMS, respectively. (A) First-dimension LC−ESMS SIM trace (summed) for *m*/*z* 79 and 63. The MS is operated in the negative ion mode and is set to detect only *m*/*z* 63 and 79. Fractions containing peaks marked 1 and 2 were subjected to further analysis. (B) NanoES (−) ion precursor scan for *m*/*z* 79 of the HPLC fraction containing peak 1. A single phosphopeptide with an apparent average mass of 1424.0 was detected. The charge state of the *m*/*z* 711.0 ion was deduced from the presence of a sodium adduct 11 Da away. The unmarked ions at *m*/*z* 420−440 are not peptides. (C) NanoES (−) ion precursor scan for *m*/*z* 79 of the HPLC fraction containing peak 2. A single phosphopeptide with an apparent average mass of 1351.2 was detected. (D) NanoES (+) ion CID product ion spectrum of *m*/*z* 712.4, the doubly charged ion for the phosphopeptide found in peak 1. (E) NanoES (+) ion CID product ion spectrum of *m*/*z* 676.3, the doubly charged ion for the phosphopeptide found in peak 2. (F) Amino acid sequence coverage provided by b_n and y_n fragment ions for the phosphopeptides found in peaks 1 and 2. A third phosphopeptide with a mass of 741 was not recovered. Peptide fragment ion nomenclature is that of Biemann,[36] except that $b_n^{\blacktriangle}$ or $y_n^{\blacktriangle}$ refer to $(b_n - H_3PO_4)^+$ and $(y_n - H_3PO_4)^+$ and M^{2+} and $M^{\blacktriangle 2+}$ refer to $[M + 2H]^{2+}$ and $[(M + 2H) - (H_3PO_4)]^{2+}$, respectively.

fragmentation pathways. As can be seen in the spectrum shown in Figure 1D, only those few ions that are common to the various pathways are produced in good yield. However, despite the low information content of this spectrum, it is sufficient to confirm the sequence identification derived from the precursor scan data and to identify the specific residue phosphorylated (Figure 1F). The tryptic-like peptide eluting in peak 2 produced a very high quality, analytically useful product ion spectrum (Figure 1E), from which the sequence could easily be confirmed and the phosphorylated residue easily determined (Figure 1F).

The sensitivity of the method described here is currently limited by the first-dimension LC−ESMS analysis. Because ESMS (at flow rates above ~50 nL/min) is a concentration-sensitive detector, smaller internal diameter HPLC columns will provide higher sensitivity analyses. However, since the described method requires that the LC eluent be split postcolumn to allow fraction collection, the HPLC flow rate at which this can efficiently be done imposes a practical limitation on the diameter HPLC column that can be used for the LC separation. The conventional ion spray source on the Sciex API III operates most efficiently at flow rates between 2 and 5 μL/min. Whereas these flow rates would readily accommodate 180−320-μm-i.d. HPLC columns, the ability to collect fractions would be lost as the entire flow would be required for MS analysis. Thus the choice of a 500-μm-i.d. column operating at flow rate of 20 μL/min, with 4 μL/min going to the MS and 16 μL/min going for fraction collection, is a practical compromise. We have recently tested a micro ion spray source flowing at 200−500 nL/min and a 180-μm-i.d. HPLC column flowing at rates of 3−4 μL/min. Preliminary data obtained using this device indicated that sensitivity in the first-dimension analysis can be 10−50-fold higher for phosphopeptide detection than on a 0.5-mm-i.d. column

(37) Chin, E. T.; Papac D. E. *Anal. Biochem.* **1999**, *273*, 179−85.

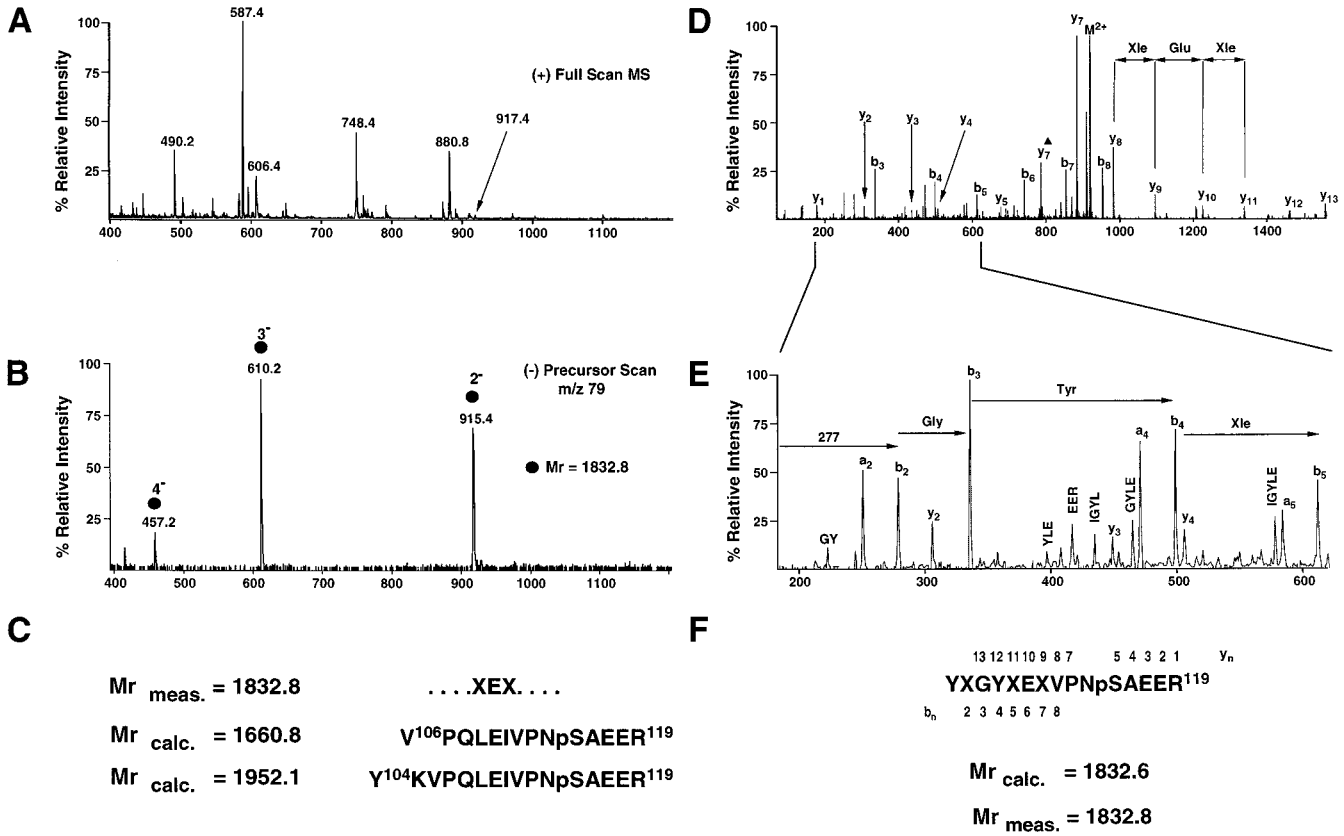

Figure 2. Detection and characterization of a novel phosphorylated sequence from αs1-casein. A heterogeneous protein mixture of α-casein was digested with trypsin and analyzed in the first dimension by LC−ESMS with SIM monitoring for *m*/*z* 79 and 63. (A) NanoES (+) ion full-scan MS spectrum from fraction 18 of the LC−ESMS run. An arrow points to the M^{2+} ion for the phosphopeptide present in this fraction. (B) NanoES (−) ion precursor scan for *m*/*z* 79 of fraction 18. A single phosphopeptide with an apparent average mass of 1833.6 was detected. (C) A partial sequence (XEX, where X = Leu or Ile) determined from the product ion spectrum of the phosphopeptide in fraction 18 matches a known phosphorylated sequence from αs1-casein. The complete and partial tryptic peptides encompassing the matched residues are shown. However, the mass of neither αs1-casein peptide matches the mass of the unknown. (D) NanoES (+) ion CID product ion spectrum of *m*/*z* 917.2, the doubly charged ion for the phosphopeptide found in fraction 18. E) Expansion of the low-mass region of the spectrum shown above. Internal fragment ions produced by cleavage of two amide bonds are represented by the single-letter amino acid code for the residues included in the internal fragment. (F) Final sequence determined from the product ion spectrum. The amino acid sequence coverage provided by b_n and y_n fragment ions is shown.

while retaining the ability to collect fractions for further characterization.[38]

Selectivity of the Method for Phosphopeptide Detection. Substoichiometric in vivo phosphorylation is the rule rather than the exception for many proteins involved in cellular signaling pathways. Thus, for a phosphopeptide mapping strategy to be truly useful, it must be able to selectively find small amounts of phosphopeptides in the presence of large amounts of nonphosphorylated peptides. The 2D mapping strategy described here has an inherently high degree of selectivity, since both the first- and second-dimension MS scanning techniques rely on the production of the PO_2^- or PO_3^- fragment ion, a property peculiar to phosphopeptides.

The importance of this capability to selectively detect minor amounts of phosphopeptides in complex mixtures became apparent during our earlier studies of the α-casein proteins.[31] Precursor ion scanning detected in an HPLC fraction a phosphopeptide of very low abundance which could not be assigned, based on the molecular weight of the peptide, to any of the known α-casein sequences. The first-dimension LC−ESMS analysis of the α-casein tryptic digest showed a very small peak in the *m*/*z* 79 trace eluting under a very large UV peak (data not shown). The coincidence of these peaks in the two different traces indicated the likelihood of a low-abundance phosphopeptide that had eluted in the presence of large amounts of nonphosphorylated peptides. Analysis of this phosphopeptide-containing HPLC peak by nanoelectrospray is shown in Figure 2. The *m*/*z* 79 precursor ion scan of this fraction gave an excellent spectrum, showing a multiply charged ion series for a single phosphopeptide with a determined average mass of 1832.8 (Figure 2B). The full-scan, positive ion nanoESMS spectrum of this fraction indicates that this phosphopeptide is a very minor component in the fraction (see the arrow in Figure 2A). Only a single charge state, the M^{2+} ion at *m*/*z* 917.4 is present in the positive ion spectrum. On the basis of the intensity of this ion, the mass 1832.8 phosphopeptide is present in this fraction at less than 2% of the most abundant nonphosphorylated peptide.

Using nanoESMS/MS, we were able to record an analytically useful product ion spectrum (Figure 2D) of this peptide. The

(38) Zappacosta, F.; Huddleston, M. J.; Annan, R. S.; Carr, S. A. Proceedings of the 47th Conference on Mass Spectrometry and Allied Topics, 1999; pp 469−470.

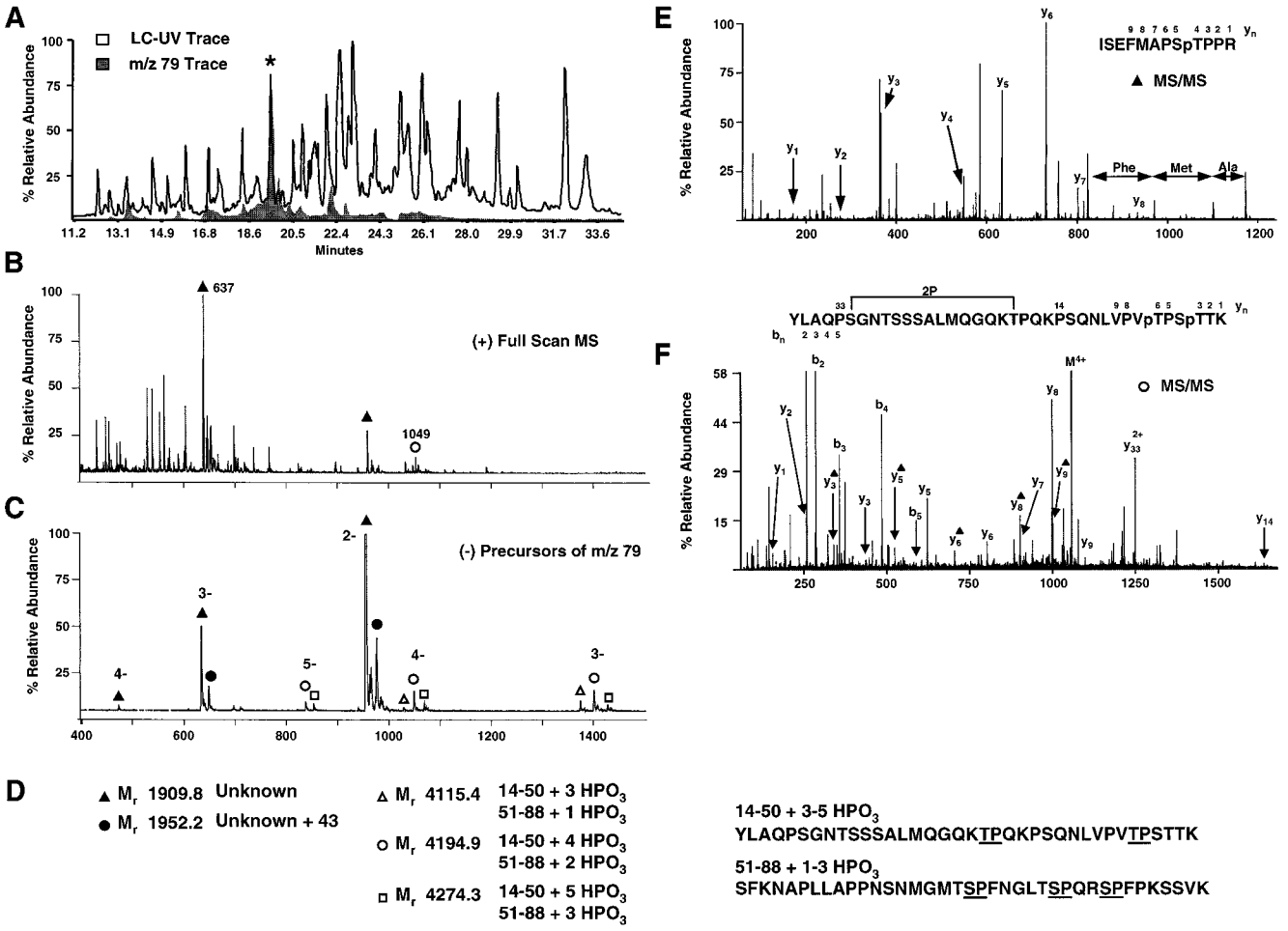

Figure 3. Multidimensional electrospray MS mapping of G_1 Cdk-dependent in vitro phosphorylation of Sic1p. Purified phosphorylated MBP−Sic1p was digested with trypsin and injected onto a 0.5-mm-i.d. RP-HPLC column. (A) First-dimension LC−ESMS SIM trace for *m*/*z* 79 and 63. The phosphopeptide-specific MS trace (gray) is compared with the LC−UV trace (white). Ten fractions were taken for further analysis. (B) NanoES (+) ion full-scan MS spectrum from fraction 19 of the LC−ESMS run. The major peak in fraction 19 is marked with an asteriskin (A). Peaks marked with symbols are phosphopeptide ions identified from the precursor ion scan. (C) NanoES (−) ion precursor scan for *m*/*z* 79 of fraction 19. Five phosphopeptides were detected. (D) Molecular weights and possible sequence assignments for the phosphopeptides detected in fraction 19. Cdk consensus sequences are underlined. (E) NanoES (+) ion CID product ion spectrum of *m*/*z* 637.6, the triply charged ion for the 1909.8 Da phosphopeptide found in fraction 19. The spectrum clearly indicates that the peptide contains the N-terminus of Sic1p and assigns the phosphorylation site to Thr^5 of Sic1p (numbering begins at the Met residue). (F) NanoES (+) ion CID product ion spectrum of *m*/*z* 1049.6, the quadruple charged ion for the 4194.9 Da phosphopeptide found in fraction 19. From this spectrum the sequence 14−50 + 4 HPO_3 could be assigned and phosphorylation at T^{45} and T^{48} was established.

series of ions marked y_8−y_{11} in the spectrum clearly indicate the partial sequence XEX (X = Leu or Ile), which matches residues 109−111 from the αs1-casein tryptic peptide, VPQLEIVPNSAEER (residues 106−119, calculated M_r (av) = 1660.8) a peptide that was identified in an earlier eluting fraction. Interpretation of the lower *m*/*z* region of the spectrum allowed us to extend the determined sequence toward the C-terminus using y_n ions, yielding the partial sequence XEXVPNpSAEER. However, the residual N-terminal mass of the unidentified peptide fits neither the remaining predicted αs1-amino acids nor any combination of N-terminal extensions, tryptic or otherwise, from the αs1-casein sequence (see Figure 2C).

Further analysis of the low-*m*/*z* region (Figure 2E) led to the identification of a series of b-ions (and some a-ions) that overlapped with the sequence already defined and extended toward the N-terminus of the peptide to yield (E,F or Y,X)GYLEIVPNpSAEER, where the choice and order of the first two residues was not yet defined. The incorporation of the Gly and Tyr residues was also supported by low-abundance y_{12} and y_{13} ions at the high-mass end of the spectrum. Presuming *m*/*z* 277 to be a b_2 ion, we searched for dipeptides with a residue mass of 276 Da and found that only two combinations of amino acids, EF and XY (where X is either Leu or Ile), fit this mass. The absence of the usually abundant *m*/*z* 120 immonium ion for Phe suggested to us that X,Y was the correct combination. A series of internal fragment ions that included XGYL and XGYLE confirmed the combination and fixed the order of the residues as YX (Figure 2E). A search of the NCBI nonredundant protein database failed to find any known protein containing the sequence YXGYLEIVPNSAEER. These data suggest that another previously undiscovered variant of αs1-casein exists (Figure 2F).

The identification of the novel αs1-casein sequence highlights the exquisite selectivity of the method described here. In addition to easily detecting phosphopeptides that are in low abundance relative to the nonphosphorylated components of the sample, it also is able to detect phosphopeptides that are in low abundance

Table 1. In Vitro and in Vivo Phosphorylation of Sic1p Determined by MS-Based 2D Phosphopeptide Mapping

HPLC fraction	M_r found[a]	found sequence[b] (mol % HPO_3)	M_r calc[a]	phosphorylation site[c]
		In Vitro		
12−13	1053.2	186−193 + 1HPO_3	1052.9	S191
14	1510.0	186−197 + 1HPO_3	1509.5	**S191**
16−17	1349.6	166−177 + 1HPO_3	1349.4	**T173**
16−17	2129.0	14−32 + 2HPO_3	2128.1	**T22-S25**
17	2048.8	14−32 + 1HPO_3	2068.1	**T22-S25**
16−17	2163.2	33−50 + 3HPO_3	2163.2	**T33**, **T45**, **T48**
18−20	4274.4	14−50 + 5HPO_3	4273.3	T22-S25, T33, **T45**, **T48**
18−20	4194.6	14−50 + 4HPO_3	4193.3	T22-S25, T33, **T45**, **T48**
18−19	1926.0	1−8 + 1HPO_3[d]	?	**T5**
19	4113.0	14−50 + 3HPO_3	4113.3	T33, **T45**, **T48**
22	1898.6	169−185 + 1HPO_3	1897.1	T173
24	3509.5	54−84 + 3HPO_3	3509.7	**S69,** S76, S80
		In Vivo		
24−25	1047.4	2−8 + 1HPO_3 (100)		**T5**
54	3951.5	14−50 + 1HPO_3 (28)		**T33**
53	4031.5	14−50 + 2HPO_3 (14)		nd[e]
70−71	3348.0	54−84 + 1HPO_3 (23)		**T75/S76**[f]
70−71	3426.3	54−84 + 2HPO_3 (4.5)		nd
70−71	3507.0	54−84 + 3HPO_3 (4.5)		nd
54	4264.5	11−50 + 1HPO_3		nd
53	4346.0	11−50 + 2HPO_3		nd
60−61	4630.5	11−53 + 1HPO_3		nd
60−61	4710.0	11−53 + 2HPO_3		nd
60−61	4314.0	14−53 + 1HPO_3		nd
60−61	4393.5	14−53 + 2HPO_3		nd

[a] All values are reported as average mass. [b] Sequence numbering begins with Met^1 of Sic1p and does not include the MBP fusion protein sequence; mole percent indicates the fraction of a given peptide found in a particular phosphorylation state. [c] Numbers reported in boldface type were verified by tandem MS. [d] N-terminus of the peptide contains the fusion junction and does not match the predicted sequence. [e] nd, specific phosphorylation sites on these peptides were not determined. [f] MS/MS could not distinguish between T75 and S76; see text for details.

compared to the total phosphopeptide pool. This is in direct contrast to alternative phosphopeptide mapping methods that attempt to identify phosphopeptides in the total peptide pool or those that measure the entire phosphopeptide pool in a single experiment.

Analysis of Sic1p Phosphorylation. Degradation of the S-phase cyclin-dependent kinase (Cdk) inhibitor protein Sic1p has been shown to be essential for the transition from G_1 to S phase in the cell cycle of *S. cerevisiae*.[33] The multi-protein SCF complex ubiquitinates Sic1p that has been phosphorylated by G1-Cdk,[39] thereby targeting Sic1p for destruction by the 26S proteasome. To prove that Sic1p is a direct substrate for G1 Cdk, we mapped the in vitro phosphorylation sites on a maltose-binding protein−Sic1p fusion construct (MBP−Sic1p) after treatment with immunoaffinity-purified kinase complex.[33] To determine which of the in vitro phosphoacceptor sites are utilized in vivo, we mapped the phosphorylation sites on Sic1p purified from *S. cerevisiae*. To accomplish both of these tasks, we employed the novel phosphopeptide mapping strategy described above.

In Vitro Phosphorylation of Sic1p. MBP−Sic1p phosphorylated in vitro with immunoaffinity-purified G1-Cdk was digested with trypsin and analyzed for phosphopeptide content using the first-dimension phosphate-specific LC−ESMS experiment. Figure 3A shows the negative ion, m/z 79 ion chromatogram from the MS superimposed on the UV trace from the HPLC. Ten HPLC fractions eluting between 12 and 23 min are indicated by the m/z 79 trace to contain phosphopeptides. Aliquots of each of the 10 phosphopeptide-containing fractions were then analyzed in the second dimension by nanoESMS with precursor ion scan for m/z 79 to obtain the molecular weights of the phosphopeptides.

The determined phosphopeptide molecular weights from the precursor ion scans were then compared with the molecular weights of predicted Sic1p tryptic peptides (plus phosphate) that contained Cdk consensus sites. The Sic1p sequence contains nine Ser/Pro or Thr/Pro candidate G1 Cdk phosphoacceptor sites. An initial evaluation of the data resulted in 10 peptides which matched Sic1p tryptic fragments that contained consensus sites and 8 peptides for which no reasonable match against either the Sic1p or the maltose-binding protein sequence could be found.

Five of the eight unmatched phosphopeptides were found in HPLC fraction 19 (marked with an asterisk in Figure 3A). Figure 3C shows the second-dimension precursor ion scan from this fraction. The full-scan positive ion spectrum is shown for comparison in Figure 3B. A deconvolution of the ion series indicated that there were five phosphopeptides present in this fraction. Two of the peptides, M_r (av) = 1909.8 (labeled ▲) and M_r (av) = 1952.2 (labeled ●) differ in mass by 43 Da. We were not able to assign either of these peptide masses to a probable Sic1p sequence. However, we observed pairs of peptides in several other fractions that differed by 43 Da. Since the fusion protein was trypsinized in 2 M urea, we hypothesized that the peptides have the same sequence and that the +43 Da difference was the result of carbamylation at the N-terminus. Tandem MS of the triply charged ions from the M_r 1909.8 (Figure 3E) and 1952.2 peptides confirmed they were the same sequence and contained the N-terminus of the Sic1p protein beginning with Met^1 (see the three residue

(39) Verma, R.; Feldman, R.; Deshaies, R. J. *Mol. Biol. Cell.* **1997**, *8*, 1427−1436.

partial sequence derived from a b_n ion series labeled in Figure 3E) These peptides also contained several amino acids that constitute the fusion junction between MBP and Sic1p. However, it was clear from the MS/MS spectrum that the actual peptide sequence in the fusion junction region did not match the sequence predicted from the DNA sequence. Having established the reason the peptide mass did not match any predicted values, we did not attempt to elaborate the exact nature of the difference. The tandem mass spectra also confirmed that the +43 Da was located on the N-terminus of the peptide, consistent with carbamylation. The spectra also confirmed that both peptides were phosphorylated on the first consensus site, Thr^5 (See Figure 3E).

The three other peptides identified in this fraction differed in mass from each other by increments of 80 Da, suggesting differential phosphorylation of the same peptide. However, the determined average molecular masses of 4115.4, 4194.9, and 4274.3 fit two possible Sic1p tryptic peptide sequences: residues 14−50 + 3−5 HPO_3 or residues 51−88 + 1−3 HPO_3 (Figure 3D, sequence numbering does not include the MBP fusion protein and begins with the first Sic1p amino acid). Unfortunately, secondary criteria such as kinase specificity and enzyme cleavage specificity were not helpful in distinguishing between these two possibilities. For example, the partial sequence 14−50 has only one missed cleavage site and contains two Cdk consensus sites for phosphorylation. However the mass of the peptide requires that it contain up to 5 mol of phosphate. The second sequence, 51−88, contains three missed tryptic cleavages, but has three consensus sites for Cdk phosphorylation, matching the suggested number of moles of phosphate on the peptide. Partial sequencing of each peptide by tandem MS was essential for unambiguous assignment. Tandem MS of the quadruply charged ion for each of the three peptides clearly indicated that the correct sequence for these peptides was 14−50, containing 3−5 mol of phosphate. The consensus sites S69, S76, and S80, which are present in the alternative sequence (51−88), were recovered in fraction 24 in a peptide spanning residues 54−88 that contained 3 mol of phosphate (see Table 1).

Complete interpretation of the MS/MS spectra from the M^{4+} ion from the triply, quadruply, and quintuply phosphorylated 14−50 peptide revealed that all three peptides were phosphorylated on the consensus site Thr^{45} and on a nonconsensus site, Thr^{48} (see Figure 3F). The locations of the other phosphates could not be defined by these data but were determined by tandem MS of the two primary tryptic fragments contained within this sequence (data not shown). These peptides were found eluting earlier in fractions 16 and 17. The second Cdk consensus site within the 14−50 sequence occurs just after a tryptic cleavage site (K^{32}). Although we would have expected this cleavage position to have been greatly hindered by the presence of phosphate, both the 14−32 and the 33−50 peptides were recovered in good yield, and we were able to sequence both peptides by MS/MS. The 33−50 peptide was only recovered as a triply phosphorylated peptide. The product ion spectrum clearly indicated phosphorylation at the consensus site Thr^{33} in addition to the previously determined sites Thr^{45} and Thr^{48} (data not shown). The 14−32 sequence was found in singly and doubly phosphorylated forms, and the product ion spectrum showed that the 2 mol of phosphate were distributed over Thr^{22}, Ser^{23},Ser^{24}, and Ser^{25}, all nonconsensus sites (data not shown). Utilization of nonconsensus phosphorylation sites during an in vitro kinase reaction is not uncommon and is the result of the high enzyme concentrations and long reaction times that are frequently employed to drive the kinase reaction to stoichiometric labeling. These data demonstrate why it is best to assign phosphorylation sites on the basis of sequence information and not just molecular weight information.

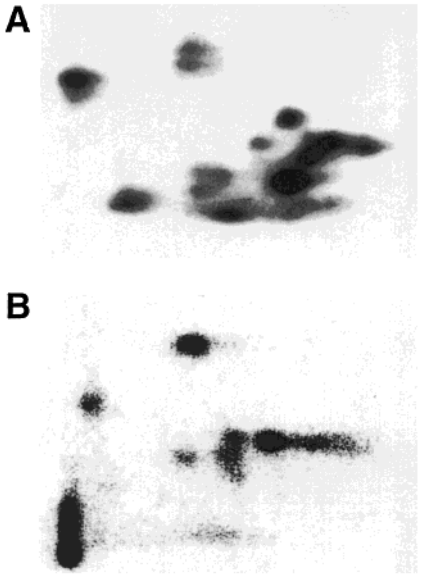

Figure 4. Two-dimensional phosphopeptide mapping of Mbp-Sic1p phosphorylated: (A) in vitro; (B) in vivo. Mbp−Sic1p was either phosphorylated in vitro by G1-Cdk or isolated from cdc34ts cells as described in Verma et al.[33] Phosphorylated Sic1 was resolved by SDS−PAGE and transferred to nitrocellulose. After identification of the phosphorylated Sic1, the band was excised from the membrane and digested with trypsin. All subsequent procedures involving phosphopeptide mapping by 2D separation on thin-layer cellulose plates were as described in ref 2.

In this manner, we analyzed all of the phosphopeptides detected in the fractions from the first-dimension LC−ESMS analysis of Sic1p. The unique peptide sequences that were found to contain Cdk-dependent phosphoacceptor sites are summarized in Table 1. In all, 19 unique phosphopeptide molecular weights were identified in 9 fractions. We found that all of the 9 Cdk consensus sites were fully utilized and could be accounted for in 12 unique peptide sequences. In addition, nearly all of the unknown phosphopeptides could be accounted for as random enzymatic cleavages, carbamylation products, or methionine oxidation products. It is interesting to note that, apart from the Thr^{22}-Ser^{25} cluster and Thr,[48] we did not find any additional phosphorylation at nonconsensus sites. In addition, we did not detect any phosphate on the MBP domain of the fusion protein, despite there being S/T P Cdk consensus sites within the MBP sequence.

In Vivo Phosphorylation of Sic1p. We had hoped to be able to show by conventional 2D phosphopeptide mapping[2] that Sic1p is phosphorylated in vivo on the same sites as those utilized by the Cdk kinase complex in vitro, but the images were not sufficiently similar to be able to make any such conclusions (see Figure 4). No effort was made to identify the peptides that correspond to the spots on either of the gels. Rather, we purified $Sic1p^{HAHis6}$ to homogeneity from *S. cerevisiae* $cdc34^{ts}$ cells[33] by combining two

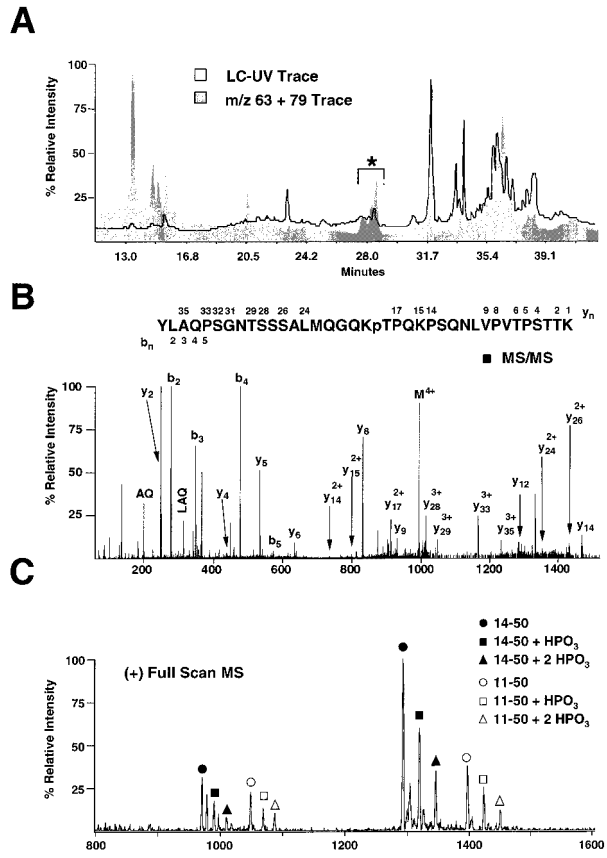

Figure 5. Multidimensional electrospray MS mapping of in vivo phosphorylated Sic1p from *S. cerevisiae*. The purified protein was digested with trypsin and injected onto a 0.5-mm-i.d. RP-HPLC column. (A) First-dimension LC−ESMS SIM trace for *m*/*z* 79 and 63. The phosphopeptide-specific MS trace (gray) is compared with the LC−UV trace (white). (B) NanoES (+) ion CID product ion spectrum of *m*/*z* 994.4, the quadruply charged ion for the oxidized form of the 3953.2 Da phosphopeptide found in the pooled fractions identified with an asterisk in (A). This peptide matched the predicted average M_r for 14−50 + HPO_3 (3953.3 Da). The product ion spectrum confirms this match and assigns the phosphorylation site to Thr[33]. (C) LC−ESMS (+) ion MS spectrum summed over a 1.2-min period during which the peptides containing Thr[33] eluted.

affinity purification steps (Ni-NTA and anti-HA immunoaffinity chromatography) in tandem and subjected the purified phospho-Sic1 to the same MS-based multidimensional phosphopeptide mapping strategy employed above. At their nonpermissive temperature, cdc34[ts] arrest cell division at the end of G1 phase with high levels of G1-Cdk activity. Due to the deficit of Cdc34 activity, these arrested cells are unable to attach ubiquitin to and degrade phosphorylated Sic1. We confirmed that the purified Sic1 retained biologically relevant phosphorylations, as ∼75% of the purified molecules were converted to ubiquitinated species in the presence of recombinant SCF.

As was to be expected, based on the comparison of the 2D maps shown in Figure 4, the summed ion trace for *m*/*z* 63 and 79 from the LC−ESMS analysis of the in vivo derived phosphorylation sites (Figure 5A) was quite different from that of the in vitro derived profile (Figure 3A).

Applying precursor ion scans and tandem MS to the various fractions, we determined that in vivo phosphorylation of Sic1p was localized on three tryptic peptide sequences (2−8, 14−50, and 54−84). The sequence 2−8 appeared to be quantitatively phosphorylated on Thr[5], whereas the latter two peptides were differentially phosphorylated (0−2 and 0−3 phosphates, respectively), with the lowest phosphorylation state being the most abundant in each case. We sequenced the monophosphorylated form of both the 14−50 peptide and the 54−84 peptide by MS/MS and determined that the major sites of phosphorylation are Thr[33] and Ser[76], respectively.

Figure 5B shows that, for the peptide 14−50, the tandem MS spectrum of the quadruply charged ion at *m*/*z* 994 yielded fragment ions covering most of the sequence, leaving only one large gap of six residues (MQGQKT[33]) in the center of the peptide. Since the fragment ion data show conclusively that none of the other serine or threonine residues in the sequence are modified, we concluded that Thr[33] is phosphorylated. Unfortunately, tandem mass spectrometry of the peptide 54−84 was unable to distinguish between Thr[75] and Ser[76], and assignment of the phosphorylation site in this case was based on the fact that Ser[76] is in a Cdk consensus sequence (data not shown).

Site-Specific Stoichiometry. The amount of phosphate incorporated at each site is an important guide for further biological experimentation aimed at defining function. Phosphoacceptor sites that are more heavily utilized are the most relevant ones to target for mutational analysis, a common way to assess the functional relevance of a phosphorylation site. The need for site-specific quantitation of phosphate incorporation is an important but frequently overlooked aspect of phosphorylation mapping, especially in proteins that contain a significant number of modification sites. We have previously shown that the ratio of MS response for phosphorylated to nonphosphorylated peptides (in the full-scan, positive ion ESMS data obtained at acidic pH) may be used to determine phosphorylation stoichiometry with a accuracy of ∼±30%.[31]

For Sic1p we reserved ∼20% of the tryptic digest from the in vivo sample and used the positive ion LC−ESMS approach described above to define the phosphorylation stoichiometry at the various sites. For example, Figure 5C shows the various peptides containing T[33] that eluted over a 1.2-min period in the positive ion LC−ESMS experiment. The ion intensities for all observed charge states of the nonphosphorylated peptide and each phosphorylated form of the peptide (i.e., mono-, di-, and triphosphorylated) were summed and the stoichiometry was calculated as the percentage of any given form of the peptide to the entire population of that peptide. In cases where the same phosphorylation site was observed on more than one peptide sequence due to incomplete tryptic cleavage, the stoichiometry was determined for each phosphopeptide and the results were averaged. The data are summarized in Table 1.

As a final check of the overall phosphorylation state of the protein, we determined the distribution of phosphorylation on intact Sic1p protein by nanoESMS. Figure 6A shows that the predominant molecular species detected by mass spectrometry correspond to various phosphoforms of Sic1p with 1−6 mol of phosphate attached. The essentiality of the three major in vivo phosphorylation sites determined by mass spectrometry was demonstrated by preparation of a mutant in which Thr[5], Thr[33], and Ser[76] were substituted with Ala. The most abundant phos-

(40) Nugroho, T. T.; Mendenhall, M. D. *Mol. Cell. Biol.* **1994**, *14*, 3320−3328

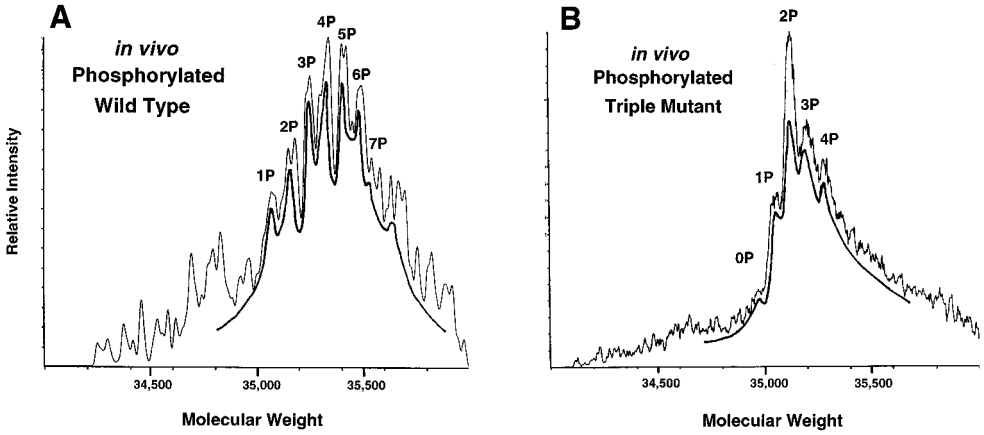

Figure 6. Distribution and relative stoichiometry of in vivo phosphate incorporation in Sic1p from *S. cerevisiae* measured by MS. Raw MS data were smoothed and transformed into the molecular weight domain. The dark line showing the individual phosphorylation states for each protein was hand fitted to the data for clarity. The molecular weight of Sic1p with zero phosphate does not match the molecular weight predicted by the cDNA sequence. It has been suggested,[40] on the basis of the behavior of the purified protein, that Sic1p may contain modifications other than phosphorylation. (A) NanoES (+) ion MS spectrum of wild-type Sic1p. (B) NanoES (+) ion MS spectrum of a Sic1p triple mutant. Residues Thr^{5}, Thr^{33}, and Ser^{76} were substituted with Ala.

phorylation state of the in vivo phosphorylated triple mutant decreased from 4 to 5 mol of phosphate (observed on wild-type Sic1p) to 2 mol of phosphate (Figure 6B). These data suggest that these three sites are in indeed major G1 Cdk phosphoacceptor sites. That they are both sufficient and necessary for targeting Sic1p destruction via ubiquitin-mediated proteolysis is demonstrated by the severe functional defect exhibited by the mutant. The triple mutant can no longer be ubiquitinated by the SCF pathway and thereby causes cells to arrest in G1 phase.[33]

CONCLUSIONS

We have presented a novel multidimensional electrospray mass spectrometry-based phosphopeptide mapping strategy. The approach developed does not require that the protein be labeled with radioactive ^{32}P. Instead, two orthogonal MS scanning techniques, both of which are based on the production of a phosphopeptide-specific marker ion in the negative ion mode at m/z 79, are combined with liquid chromatography−electrospray mass spectrometry and nanoelectrospray MS/MS to selectively detect and identify phosphopeptides in complex mixtures.

A selected ion monitoring trace for the sum of m/z 63 and 79 produced by the mass spectrometer in the first-dimension step is the functional equivalent of a radioactivity trace, pointing the investigator to HPLC fractions containing phosphopeptides. This trace also serves as a fingerprint for the phosphorylation profile of a protein. Changes in the phosphorylation state of a protein would be reflected by a change in the phosphorylation profile. Only those components of the profile that change would need further analysis.

The second dimension of the phosphopeptide mapping strategy utilizes negative ion nanoES to perform precursor ion scans for the m/z 79 marker ion on fractions identified by the first-dimension LC−ESMS analysis. Precursor ion scanning is a highly selective and sensitive method for detecting phosphopeptides in mixtures. The precursor scan produces molecular ions only for those peptides that can produce the phosphopeptide-specific product ion of m/z 79. Therefore, the molecular weights of phosphorylated peptides can be determined in this dimension, even though other, more abundant nonphosphorylated peptides are present in the same HPLC fraction, and even in cases where the signal from the phosphopeptide is indistinguishable from background in the conventional molecular weight scan. This permits detection of very low abundance, very low stoichiometry phosphorylation sites that could otherwise go undetected. Some peptides yield only a single charge state for the precursor. In these cases, because the mass resolution of the precursor experiment is low, the m/z spacing between the peptide and ions formed by common modifications (e.g., oxidation) or by gas-phase adduction (e.g., with sodium) is used to define the charge states of the ions and therefore the molecular mass of the peptide.

In this study, we tested the validity of our method by mapping 10 in vitro and 3 in vivo phosphorylation sites in the yeast protein Sic1p. In the case of the in vivo phosphorylated protein, we were able to detect and identify peptides that contained less than 5% incorporation of phosphate and sequence peptides as large as 4032 Da to identify sites of phosphorylation. The highly specific nature of the technique was demonstrated by the identification of a very low abundance, novel variant of the milk phosphoprotein α-casein.

Both the Sic1p mapping and the identification of the novel α-casein variant point to the need to include some type of phosphopeptide sequencing in any mapping strategy. In addition to peptides that contain more than one potential phosphorylation site, there are invariably those peptides for which no reasonable assignment can be made based on the determined molecular weight or where the assignment is ambiguous. In our study, the peptide sequencing was accomplished using nanoESMS/MS. Usually the sequencing is performed with the same aliquot of sample that provided the precursor ion scan data. On-line LC−ESMS/MS, under optimal conditions, may be somewhat more sensitive than nanoES-MS/MS due to the higher effective concentration of peptide in high-resolution LC−ESMS versus nanoES. However, the long analysis times provided by nanoES allow for optimization of the collision energy, selection of the most appropriate charge state, and building up weak areas of the spectrum by accumulating ion counts. This is especially important for large

peptides and whenever a more complete sequence is required (for example, in cases where the protein sequence is not known and is not in a database). In fractions where there are more than one phosphopeptide that requires sequencing (such as in the Sic1p study reported here), the somewhat lower sensitivity of nanoES is greatly offset by the fact that several peptides can be sequenced without resorting to use of another aliquot of sample.

With our current configuration (0.5-mm-i.d. HPLC column), the practical sensitivity of the method is ~5 pmol of phosphopeptides. Five picomoles of a 50 kDa protein represents 100 ng of protein. Thus, this technique is sufficiently sensitive to detect biologically relevant phosphorylations on signaling molecules of moderate to low abundance. For example, a protein present at 1 part in 10 000 could be adequately analyzed if phosphopeptides were recovered at 10% efficiency from a standard immunoprecipitation conducted with 10 mg of initial protein extract. An important caveat is that many such proteins might be phosphorylated to only 10% stoichiometry, which would necessitate 10 times more protein. However, preliminary data generated in our laboratory indicate that a sensitivity increase of between 10- and 50-fold may be achieved for the first-dimension step using 180-μm-i.d. columns at flow rates of 3−4 μL/min.[38] Such an increase in sensitivity will enable mapping of phosphorylation sites on proteins present at moderate (1/10 000) to low (1/100 000) abundance. These studies will be reported elsewhere.

The goal of phosphorylation site mapping (or any posttranslational site mapping for that matter) is to provide as complete a map as possible of the phosphorylation sites in a protein. In the method reported here, complex mixtures are fractionated to maximize the potential to detect the majority of phosphopeptides present, and phosphopeptide-containing fractions are collected during this analysis. Since only those phosphopeptides in a given fraction are being consumed by subsequent nanoESMS/MS analyses, other phosphopeptides in other fractions are preserved. The most significant drawback of the method is that not all phosphopeptides will chromatograph on reversed-phase (or similar-behaving phases such as Poros) HPLC. Methods employing direct analysis of unfractionated peptide digests minimize losses that invariably occur in sample handling at low levels. However, most of the literature that describes the complete (including sequencing) analysis of unfractionated phosphoprotein digests (e.g., refs 20 and 41) describe the analysis of fairly simple samples, containing only a few, high-stoichiometry phosphorylation sites. As the complexity of the phosphorylation pattern increases, the benefits of techniques that analyze unfractionated mixtures diminish rapidly due to problems with dynamic range, suppression effects, charge state overlap, and the need to sequence many phosphopeptides in a single sample. Several recent studies of more highly phosphorylated proteins illustrate the advantages of employing a separation stage prior to analysis.[45−46] Clearly, phosphorylation sites may go undetected by any of these approaches, and no one phosphorylation mapping technique appears to be universally able to accomplish the goal of complete mapping by itself. Rather a combination of approaches is almost always required. Techniques that employ a separation component prior to MS have clear advantages as the complexity and extent of phosphorylation increase and should be used in these cases.

In this study, we have not illustrated the use of the multidimensional mapping approach for identifying phosphorylation sites from SDS−PAGE derived proteins. However, there are no inherent obstacles in doing so. Once a protein digest is in solution, it is irrelevant to the current method whether the protein was derived from a gel or from a solution digest. The first dimension of analysis described here involves a reversed-phase HPLC separation, which accomplishes desalting as well as affecting separation. Desalting is required for analysis of in-gel digests by nanoES and is highly desirable for analysis of such digests by MALDI. Neubauer and Mann have pointed out[41] that recovering phosphopeptides from in-gel digests does not present any special challenge. In fact, since phosphopeptides are generally more hydrophilic than their nonphosphorylated counterparts, recovery in aqueous media should be higher. Clearly other factors such as size, folding, and the presence hydrophobic patches are at play in peptide recovery yields, but phosphorylation by itself should not present a significant problem. Although the in-gel digestion process is fairly efficient (70−85%), our current sensitivity would require a modest Coomassie Blue-stained band on a gel (200−500 ng of a 50 kDa protein) to be successful. In fact, most phosphorylation sites mapping studies reported in the literature that used mass spectrometry as the analytical method for identifying unknown phosphorylation sites in gel-derived proteins started from Coomassie Blue-stained gels.[41−46]

Finally when it is possible to generate a sufficient amount of protein (50 pmol), we advocate making a molecular weight measurement on the intact protein and performing a positive ion LC−ESMS analysis on the original tryptic digest. The intact molecular weight provides the total number of moles of phosphate added to the protein and the relative distribution of protein molecules with different numbers of phosphate on them. The positive ion LC−ESMS analysis provides the approximate stoichiometry of the various phosphorylation states for a given peptide sequence. Since it is possible to separate peptides that have the same phosphorylation state but are phosphorylated on different residues, the positive ion LC−ESMS data can, in favorable cases, also provide the stoichiometry for different forms of the same phosphorylation state.[47]

Received for review September 20, 2000. Accepted December 7, 2000.

AC001130T

(41) Neubauer, G.; Mann, M. *Anal. Chem.* **1999**, *71*, 235−242.
(42) Betts, J. C.; Blackstock, W. P.; Ward, M. A.; Anderton, B. H. *J. Biol. Chem.* **1997**, *272*, 12922−12927.
(43) Cortez, D.; Wang, Y.; Qin, J.; Elledge, S. J. *Science* **1999**, *286*, 162−1166.
(44) Immler, D.; Gremm, D.; Kirsch, D.; Spengler, B.; Presek, P.; Meyer, H. E. *Electrophoresis* **1998**, *19*, 1015−1023.
(45) Watty, A.; Neubauer, G.; Dreger, M.; Zimmer, M.; Wilm, M.; Burden, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 4585−4590.
(46) Wu, X.; Ranganahan, V.; Weisman, D. S.; Heine, W. F.; Ciccone, D. N.; O'Neil, T. B.; Crick, K. E.; Pierce, K. A.; Lane, W. A.; Rathburn, G.; Livingston, D. M.; Weaver, D. T. *Nature* **2000**, *405*, 477−482.
(47) Zhang, X.; Nguyen, K.-H.; Jackson, J. R.; Annan, R. S.; Carr, S. A. *Proceeding of 48th Conference on Mass Spectrometry and Allied Topics*, 2000; pp 1014−1015.