Journal of Biomolecular Techniques
13:119–130 © 2002 ABRF

Identification of Protein Phosphorylation Sites by a Combination of Mass Spectrometry and Solid Phase Edman Sequencing

David G. Campbell and Nicholas A. Morrice

MRC Protein Phosphorylation Unit, University of Dundee, Dundee, Scotland

The analysis of protein phosphorylation sites is one of the major challenges in the post-genomic age. To understand the role of reversible phosphorylation in cell signaling, the precise location of phosphorylation sites must be determined in a phosphoprotein as well as the effect that these post-translational modifications have on the function of the protein. The use of solid phase Edman degradation of ^{32}P-labeled phosphoproteins and peptides was described over 10 years ago as a method for the identification of phosphorylation sites. Since that time a number of laboratories have used this technique as the standard method for phosphorylation site analysis. In this report, we will describe how we routinely use this technology to perform hundreds of successful phosphorylation site analyses per annum. By combining mass spectrometry to identify the phosphopeptide and solid phase Edman degradation to localize the site of phosphorylation, subpmole quantities of phosphopeptides can be successfully characterized.

KEY WORDS: phosphorylation, mass spectrometry, Edman, Sequelon.

ADDRESS CORRESPONDENCE AND REPRINT REQUESTS TO: Nicholas A. Morrice, MRC Protein Phosphorylation Unit, MSI/WTB Complex, School of Life Sciences, University of Dundee, Dundee DD1 5EH, Scotland (phone: 44-1382348299, fax: 44-1382223778, email: n.a.morrice@dundee.ac.uk).

Protein phosphorylation is likely to be the most abundant post-translational modification in a mammalian proteome. Various estimates suggest that, on average, 30% of proteins are phosphorylated and on one of three major residues (serine, threonine, or tyrosine). Most phosphorylations are reversible, with dephosphorylation being catalyzed by protein phosphatases. A major focus of cell signaling laboratories throughout the world is the control of signal transduction cascades by reversible phosphorylation and the definition of the sites of phosphorylation/dephosphorylation that control enzyme/protein function in these cascades.[1]

The identification of protein phosphorylation sites has become one of the major challenges to proteomic analysis. Mass spectrometry has been shown to be a reliable and routine tool to identify proteins in a high throughput manner.[2] However, the identification of phosphorylation sites by mass spectrometry is not a trivial matter and to this day is not routine.[3] The detection of phosphopeptides by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) in a complex mixture, such as a tryptic mass fingerprint, is a rare occurrence. This is thought to be caused by suppression of the ionization of the mainly negatively charged phosphopeptide in the presence of a large excess of nonphosphorylated peptides. Phosphopeptides in "model proteins" can be enriched by metal chelate chromatography (Fe^{3+}/Ga^{3+}),[4] but this is notoriously unreliable for "real life" proteins. A recent report on the enrichment of methylated phosphopeptides by Fe^{3+}-IMAC may improve the reliability of this technique.[5]

Methods have been developed that rely on the base lability of phosphoserine and phosphothreonine to form the thioacyl derivative of the dehydroamino acid[6] and subsequent derivatives.[7,8] This thioalkylation is usually only efficient for phosphoserine, with the degree of efficiency being dependent upon the residues adjacent to the phosphoserine as well as the nature of the thioalkylating agent.

Electrospray-mass spectrometry (ES-MS) is the preferred method for detecting and sequencing phosphopeptides. Phosphopeptides can be detected by parent ion scanning (−79) or source-induced dissociation (−63/79) in the negative ion mode,[9,10], or by

neutral loss scanning from the loss of H_3PO_4 from phosphoserine/threonine-containing peptides in the positive ion mode. Detection of the immonium ion of phosphotyrosine at m/z 216.043 permits a selective detection method for phosphotyrosine-containing peptides.[11] Neutral loss scans, from MS/MS spectra, are often inspected post acquisition to look for the neutral loss of 98, 49, or 32.33 Da from singly, doubly, or triply charged ions, respectively. Either detection method usually requires that the sample be rerun once the identity of the phosphopeptide is deduced. This is because the collision energies required to fragment a phosphopeptide are often different than those used to fragment a nonphosphorylated peptide of the same m/z value in an ES-IT/ES-Q-TOF mass spectrometer. Multidimensional phosphopeptide mapping has been applied successfully to a number of complex phosphorylation projects by Annan et al.[10] and Loughrey Chen et al.[12] Though this approach is time consuming and requires picomole quantities of phosphoprotein, it is nevertheless robust enough to generate phosphorylation site information from nonradiolabeled phosphoproteins, which is a major advantage. One recent study found that MALDI-Q-TOF was also useful in determining protein phosphorylation sites, but this report, like so many before, only referred to known sites rather than the identification of any novel phosphorylation sites.[13]

The simplest method for determining the sites of phosphorylation in a peptide is to perform solid phase Edman degradation on a ^{32}P-labeled phosphopeptide linked covalently to an aminophenyl derivatised membrane. This was first demonstrated by Wettenhall and co-workers with the peptide covalently linked to an aminophenyl glass fiber disc, the anilinothiazolinone (ATZ) amino acids being extracted from the reaction chamber with 90% MeOH/10% water and diverted to a fraction collector.[14] The fractions were then counted in a scintillation counter and the released ^{32}P plotted against the primary sequence as determined by gas phase sequencing of another fraction of the sample to indicate the sites of phosphorylation. Our group adapted and modified this method and has been using it for many years, now using the commercially available Sequelon-AA membrane first described by Coull and co-workers.[15] Instead of Edman sequencing the peptide to identify the phosphopeptide, we now use either MALDI-TOF or ES-Q-TOF mass spectrometry routinely.[16,17] By using a combination of these techniques, subpicomole quantities of phosphopeptides can be identified and the sites identified unambiguously irrespective of the size of the phosphopeptide, the number of sites phosphorylated, or the amino acid(s) phosphorylated.

EXPERIMENTAL

Materials

Trypsin (modified, sequencing grade) and other proteases were obtained from Roche Diagnostics GmbH (Mannheim, Germany). Novex gels and reagents were from Invitrogen (Paisley, UK); Sequenal grade trifluoroacetic acid was from Pierce (Tattenhall, U.K.); and Sequelon-AA reagent kit (GEN920033), methanol (400470), and all other protein sequencing reagents were from Applied Biosystems (Warrington, U.K.). Other reagents were of laboratory grade or better from VWR (Lutterworth UK) or Sigma-Aldrich (Poole, Dorset, U.K.).

Preparation of ^{32}P-Labeled Phosphopeptides

In vitro or in vivo ^{32}P-labeled proteins were denatured in 1× sample buffer plus 10 mM DTT, heated for 2 min at 95°C, and then treated with either 50 mM iodoacetamide or 50 mM 4-vinylpyridine. The proteins were separated on Novex NuPAGE 4–12% gels with a MOPS (3-morpholinopropane sulphonic acid) running buffer, stained with colloidal Coomassie, and the phosphoprotein localized by autoradiography. The ^{32}P-labeled proteins were excised and the gel pieces Cerenkov counted in a scintillation counter to assess the degree of phosphorylation. The gel pieces were cut into 2-mm^3 cubes and washed sequentially for 15 min with 0.5 mL of water:acetonitrile (1:1), 0.1 M NH_4HCO_3, 0.1 M NH_4HCO_3:acetonitrile (1:1), and finally acetonitrile. The gel pieces were dried under vacuum and then rehydrated in 0.1 mL 50 mM NH_4HCO_3, 0.05% Zwittergent 3–16 with 1 µg modified trypsin and incubated for 10–24 h at 30°C. An equal volume of acetonitrile was added to the digest which was then shaken for 15 min and the supernatant removed.

The gel pieces were extracted with 0.5 mL of 50% acetonitrile/0.1% TFA in water and the combined supernatants filtered through a 0.45-µm filter. The gel piece and supernatants were Cerenkov counted to assess the efficiency of the in-gel digestion. The supernatants were partially dried in a Savant Speedvac to remove the acetonitrile and injected onto a 250 × 4.6 mm Vydac C18 column equilibrated in 0.1% TFA at a flow rate of 0.8 mL/min coupled to an HPLC system with an on-line Berthold LD509 radioactivity detector. The column was developed with an acetonitrile/0.1% TFA gradient of 0–30% acetonitrile (0–90 min), 30–50% acetonitrile (90–110 min), and 50–100% acetonitrile (110–130 min). Fractions were collected every 0.5 min and radioactivity measured by Cerenkov counting.

Mass spectrometry

Isolated phosphopeptide fractions were dried and reconstituted in 50% acetonitrile/0.1% TFA in water to a final concentration greater than 100 fmol/µL. An aliquot was analyzed by MALDI-TOF-MS on a Perseptive Biosystems Elite STR mass spectrometer in both linear and reflector modes and in both positive and negative ion modes using saturated α-cyano-4-hydroxycinnamic acid (HCCA) in 50% acetonitrile/0.1% TFA in water as the matrix.

Q-TOF analyses were performed with either nanotips or by LC-MS on a PepMap C18 100 × 0.075-mm column connected to an LC packings Ultimate HPLC system interfaced to a Micromass (Manchester, U.K.) Q-TOF2 mass spectrometer. Protein sequences were digested with trypsin "in silico" using the MS-Digest program of Protein Prospector (UCSF, CA) allowing for modification of cysteine and methionine and for the phosphorylation of S, T, or Y.

Solid Phase Edman Sequencing

Aliquots of the ^{32}P-labeled phosphopeptide (50–5000 cpm) were spotted onto a disc of Sequelon-AA and allowed to dry. The peptide was coupled by spotting 4 µL 10 mg/mL N-(3-dimethylaminopropyl)-N′-ethylcarbodiimide hydrochloride (EDC) in coupling buffer onto the disc and allowing the disc to dry. An additional 4-µL aliquot was applied. The disc was then allowed to dry and washed in 1 mL of 50% acetonitrile/0.1% TFA in water. The disc was Cerenkov counted to assess the coupling yield and then placed in a vertical Blott cartridge of an Applied Biosystems 470/473/476/477/492 or 494 sequencer. For the 470 and 477A protein sequencers, the ATZ amino acid fractions were transferred to the internal fraction collector by using 90% methanol:10% water as solvent S3. For the Procise 492/494 sequencers, 90% methanol:10% water was used in the X3 position and a transfer line from a 476A (ABI part number 602930) was attached to port 39 of the Procise and connected to the arm of an external fraction collector. The fraction collector was interfaced electronically to the Procise and advanced by introducing a Relay 1 pulse function (function 253) into the sequencing program. For the 473/476 sequencers, solvent S2 was replaced with methanol as the sole reaction cartridge solvent. The ATZ amino acids were collected in an external fraction collector, advanced using the relay on/off functions.

In all cases the collection of the ATZ fraction was with 2 × 0.2 mL of the extraction solvent with a 20-sec wait between extractions. The cartridge was dried extensively with argon before proceeding with the next sequencing cycle. The collected fractions were Cerenkov counted and the ^{32}P released per cycle of Edman degradation was plotted versus the primary sequence of the peptide determined /inferred by mass spectrometry. Alternatively, the fractions were dried, reconstituted in 20 µL 20% acetonitrile, spotted onto a DEAE membrane (Pall Gelman part number 60040, Ann Arbor, MI), and phosphorimaged.

In the case of a phosphopeptide with an N-terminal glutamate residue, an aliquot of the peptide was spotted on a Biobrene precycled, TFA-etched glass fiber disc and one round of pulsed liquid Edman degradation performed. The disc was removed from the sequencer, cut into four sections, and extracted with 0.5 mL 50% acetonitrile/0.1% TFA in water. The supernatant was Cerenkov counted, dried down, and then coupled to a Sequelon-AA disc as described above.

RESULTS

Isolation of Phosphopeptides

It is rare that a phosphoprotein is labeled on a single residue. A simple profile was obtained from the in vitro phosphorylation of the protein kinase LKB1/STK11 by c-AMP dependent protein kinase, where a single phosphopeptide was obtained, comprising residues 429–434 of LKB1, with Ser 431 being the site of phosphorylation (Fig. 1A).[16] However, a far more complicated pattern of phosphorylation was obtained for the in vitro phosphorylation of a 200 kDa protein from *Xenopus laevis* extract (Fig. 1B) (N.A. Morrice, *unpublished data*). Although the latter profile is complicated, it is not that uncommon and in this experiment, most of the sites were not phosphorylated stoichiometrically (data not shown). The analysis of so many phosphopeptides from one protein is therefore challenging.

In general, it is usual to obtain between 3 and 10 phosphopeptides from the tryptic digest of proteins labeled in vitro or in vivo. This occurs due to (1) multiple sites of phosphorylation in the substrate for a particular protein kinase, (2) multiple protein kinases phosphorylating a substrate protein in vivo, (3) generation of bonds that are partially resistant to trypsin due to phosphorylation and (4) mixed oxidation states of methionine-containing peptides leading to multiple peaks. Phosphorylation generated trypsin resistant bonds can be summarised as Arg-Xaa-pSer, Lys-pSer, Arg-pSer, and Arg-pThr, but not Arg-Xaa-pThr and Lys-pThr, where pSer and pThr represent

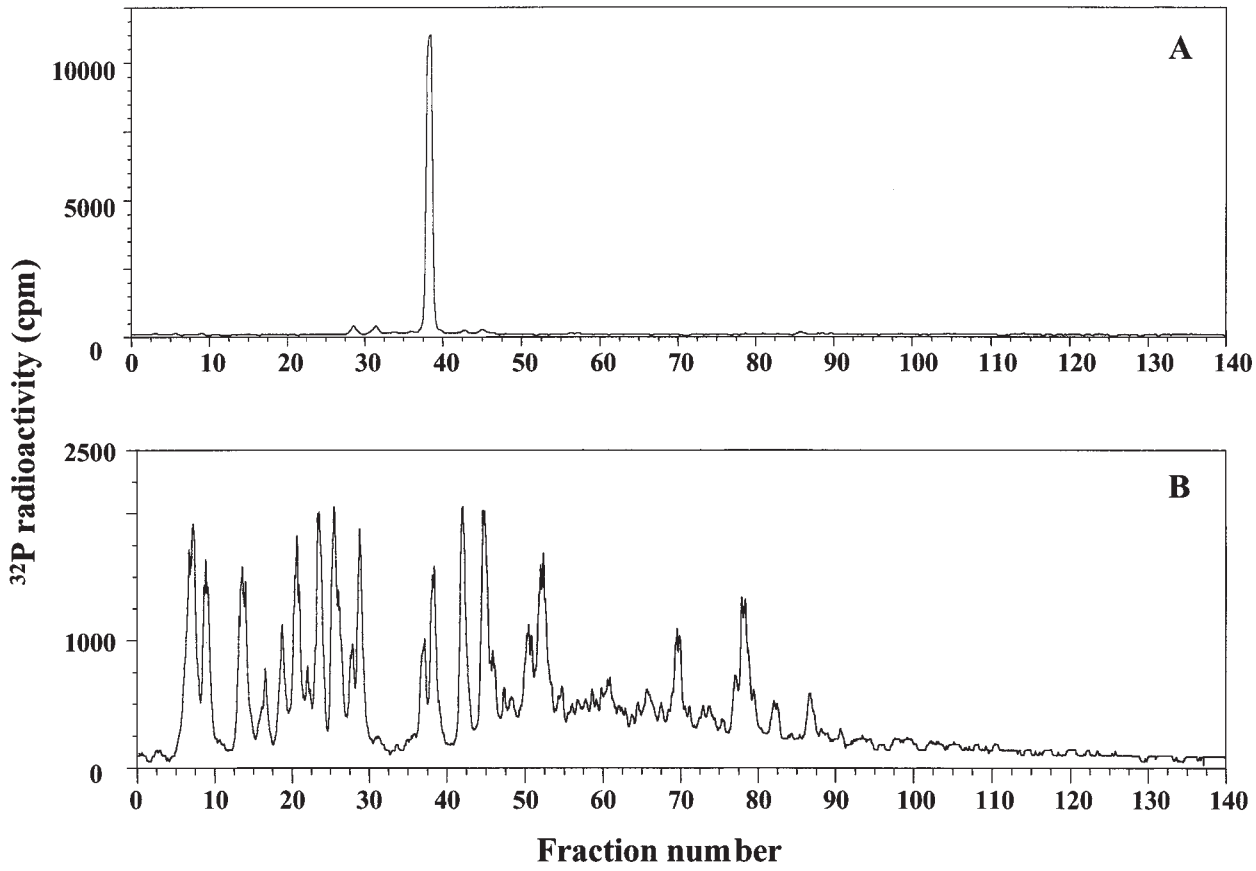

FIGURE 1

Reverse phase separation with on-line radioactive detection of ^{32}P-labeled tryptic digest of GST-LKB1 phosphorylated with cAMP-dependent protein kinase **(A)** and a 200-kDa Xenopus protein labeled with a Xenopus oocyte extract **(B).** Both tryptic digests were separated on a Vydac C18 column (4.6 x 250 mm) equilibrated in 0.1% TFA. The column was developed with an acetonitrile/0.1% TFA gradient of 0–30% acetonitrile (0–90 min), 30–50% acetonitrile (90–110 min); 0.5-min (0.4 mL) fractions were collected and Cerenkov counted in a scintillation counter.

phospho-serine and phospho-threonine respectively. There are, of course, exceptions to these rules, especially in proline rich sequences flanking the phosphorylation site, but this generalisation covers most situations. An example of a phosphorylation-induced missed cleavage was observed in the phosphorylation of the transcription factor FKHR by the protein kinase DYRK1a at Ser 329.[18]

Once the peptides have been isolated by HPLC and counted, they will need to be concentrated before mass spectral analysis. It is at this point that most peptides are lost. After removing the organic solvent under vacuum, phosphopeptides of 20 or more residues often adhere to the walls of a microfuge tube. It is therefore essential to monitor the ^{32}P radioactivity using a sensitive Mini monitor and/or scintillation counter when transferring samples from tube to tube. Tubes and tips should be monitored routinely before discarding. The amounts of phosphopeptides that are generated in our laboratory usually fall in the range of 0.1–10 pmol of total phosphate and so are normally reconstituted into a final volume of 10 µL to give a working concentration of between 0.01 and 1 pmol/µL.

Primary Structure Determination by Mass Spectrometry

Mass spectrometry is the most sensitive method to determine the primary sequence of the isolated phosphopeptides. We routinely use MALDI-TOF-MS analysis in linear and reflector modes and in positive ion and negative ion modes to assign the phosphopeptide in the primary sequence. Having performed an "in sil-

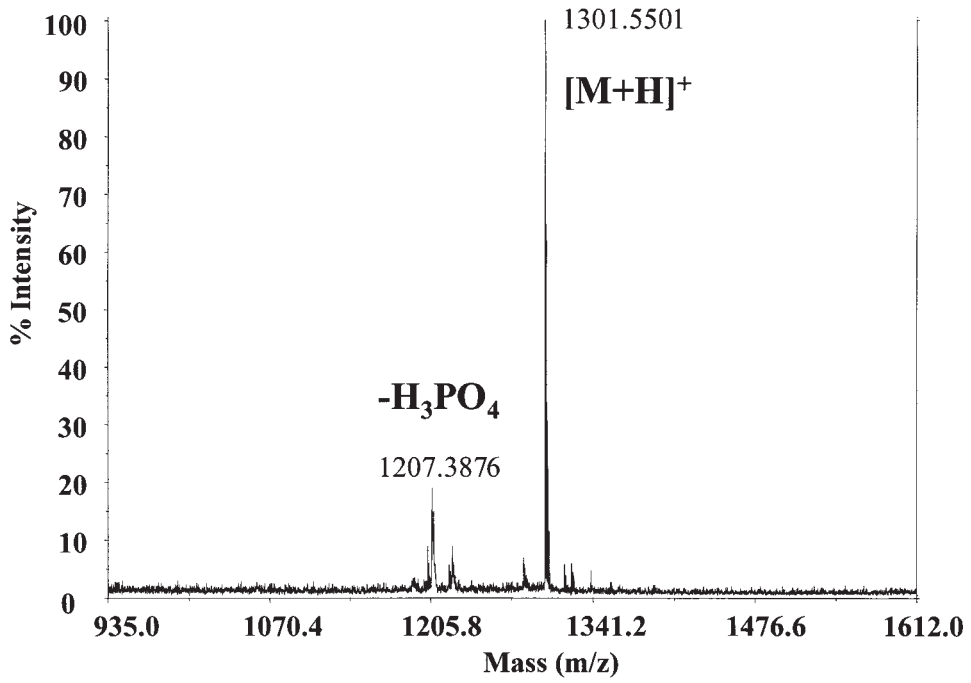

FIGURE 2

MALDI-TOF-MS of the phosphopeptide YPTAM**T**PTSPR with α-cyano-4-hydroxycinnamic acid as the matrix in the reflectron mode (**T** denotes phosphorylated residue).

ico" digest of the protein of interest using Protein Prospector /MS-Digest (http://prospector.ucsf.edu/ms-digest/) and allowing for the addition of phosphate, the theoretical masses of the phosphopeptides that could exist in the protein digest can be predicted. Phosphoamino acid analysis of both the phosphoprotein and the individual fractions from HPLC can narrow down the number of phosphopeptides that could exist in the phosphoprotein digest. An understanding of the substrate specificity of the protein kinase that was used in an in vitro experiment can reduce the number of possibilities even further[19] (see also http://www.cbs.dtu.dk/services/NetPhos).

A characteristic spectrum is obtained for a tryptic phosphopeptide analyzed in the reflector mode using HCCA as the matrix (Fig. 2). It is common to observe a daughter ion corresponding to the loss of H_3PO_4 from the parent peptide ion because of a post-source decay fragmentation event. This ion is not 98 Da less than the parent because it is not focused to the reflector detector. By changing the potential (mirror ratio) on the reflector, the ion can be focused and then observed as an ion 98 Da lighter than the parent ion.[20]

When a phosphopeptide cannot be assigned in the reflector mode, linear mode analysis should be used (though more sensitive, it offers lower resolution). This is especially useful in detecting large phosphopeptides, and the peptide identity is inferred from the parent mass plus 80 Da for one phosphate group. By changing to the negative ion mode, some selectivity towards phosphopeptides can be obtained (Fig. 3). However, negative ion MALDI is much less sensitive than positive ion mode for peptide analysis. Short and/or hydrophilic phosphopeptides (normally those that elute between 0% and 5% acetonitrile from a C18 HPLC column) are difficult to detect in MALDI-TOF-MS. Changing the matrix to 2,5-dihydroxybenzoic acid (DHB) can help in the detection and assignment of these peptides. However, the parent ion for peptides of less than 6 residues would normally be lost under the excess matrix /buffer derived ions at $m/z < 650$. Therefore, these peptides are best analyzed by ESI-MS, especially in conjunction with parent or neutral loss scanning. It should be noted, however, that these peptides rarely, if ever, bind to reverse phase microcolumns (ZipTips, microguards, etc.) if formic acid or acetic acid are used as the ion-pairing agent. This is especially important if LC-MS or microcolumn cleanup experiments are to be performed on these peptides. Collection of the flowthrough of a precolumn equilibrated in these buffers is recommended and the ^{32}P content of the flow through should be assessed.

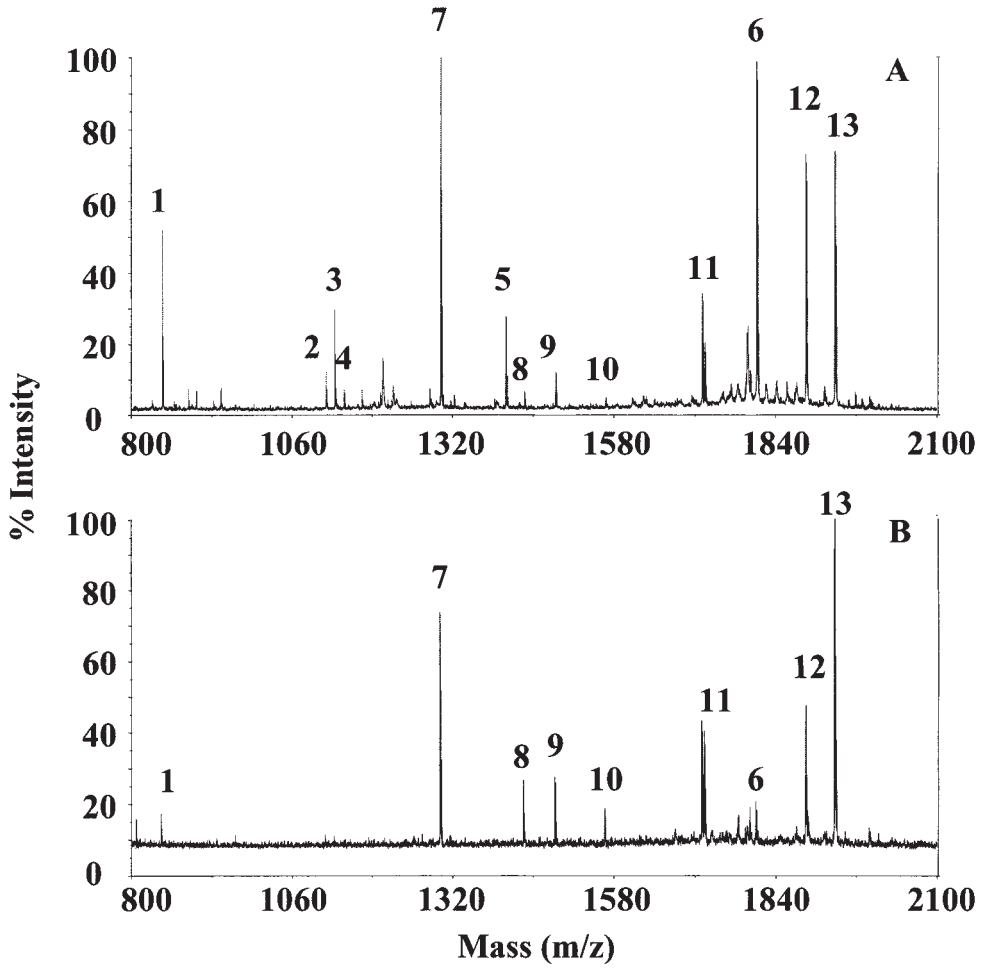

FIGURE 3

MALDI-TOF reflectron mode mass spectrum of a mixture of nonphosphorylated (1–6) and phosphorylated (7–13) peptides in the positive ion **(A)** and negative ion **(B)** modes. Both spectra were acquired from the same sample prepared with α-cyano-4-hydroxycinnamic acid as the matrix. The identities of the peptides were 1 = RPRTSSF, 2 = KKRNRTLTV, 3 = KKLNRTLSVA, 4 = KKKNRTLSVA, 5 = ALLPGASPKSPGLKA, 6 = KPPLRRSPSRTEKQE, 7 = YPTAM**T**PTSPR, 8 = KDGATMK**T**FCGTP, 9 = ALLPGA**S**PKSPGLKA, 10 = ALLPGA**S**PK**S**PGLKA, 11 = KGPGDTTS**T**FCGTPNY, 12 = KPPLRR**S**PSRTEKQE, and 13 = NGHITTTP**T**PTQFLCPK. Phosphorylated residues are denoted in underlined bold.

Phosphorylation Site Determination

Once the primary sequence of the phosphopeptide has been deduced by mass spectrometry, the sites of phosphorylation need to be defined. Tandem mass spectrometry can be used at this point, especially if the phosphopeptide is relatively short (less than 15 amino acids) and the collision conditions are tuned for each phosphopeptide.[5,9–13] However, the simplest and most reliable method of identifying the sites of phosphorylation in a ^{32}P-labeled phosphopeptide is to perform solid phase Edman degradation and Cerenkov count the ATZ-amino acid fraction. The ^{32}P released per cycle of Edman degradation can then be plotted against the primary sequence of the peptide determined by mass spectrometry. The peptide has to be immobilized because the ATZ amino acid fraction is collected with methanol. This solubilizes ^{32}P-Tyr and $H_3{}^{32}PO_4$ (the latter is released because of β-elimination from P-Ser/P-Thr during the formation of the ATZ amino acid); both are insoluble in ethyl acetate or butyl chloride, which are normally used to extract ATZ amino acids in a microsequencer.[14–18] If the peptide was not immobilized, it would also be extracted by the methanol.

Phosphopeptides are coupled to Sequelon-AA—an aminophenyl derivatized PVDF membrane—by forming an amide bond between it and either side

TABLE 1

Applied Biosystems 494A Protein Sequencer Cycle Listing for Solid Phase Edman Degradation of ^{32}P-Labeled Peptides Coupled to Sequelon-AA*

Step	Function	Function name	Time(s)	Elapsed time
1–41	—	Gas phase cycle	—	0–15 min
42	34	Del R3g waste	30	15:30
43	31	Del R3g	720	27:30
44–47	—	Block wash	30	28:55
48	131	Dry cart (top)	40	29:35
49	142	Set cart temperature	48	29:35
50	119	Transfer to FC (X3)	45	30:20
51	148	Cartridge wait	20	30:40
52	122	Transfer to FC (gas)	20	31:00
53	119	Transfer to FC (X3)	45	31:45
54	148	Cartridge wait	20	32:05
55	122	Transfer to FC (gas)	20	32:25
56	148	Cartridge wait	10	32:35
57	122	Transfer to FC (gas)	20	32:55
58	131	Dry cart (top)	120	34:55
59–61	—	Block wash	20	36:05
62	253	490A relay 1 pulse	1	36:06
63	61	Del S3 cart (top)	15	36:21
64	148	Cartridge wait	5	36:26
65	131	Dry cart (top)	120	38:26
66	259	End	0	38:26

*The cycle listing is for gas phase TFA cleavage (steps 42–43).

chain (Glu/Asp) or C-terminal carboxylates of the peptide.[15] This reaction is catalyzed with EDC and is performed at pH 4.3, which was found to be the optimum pH to neutralize most carboxylates and not protonate the aminophenyl membrane.[14,15] The coupling efficiency should be monitored routinely by measuring the ^{32}P radioactivity on the membrane after the coupling reaction, and before and after washing with 50% acetonitrile/0.1% TFA. Coupling efficiencies of 70–100% are common for tryptic peptides with a carboxy-terminal arginine residue. If the sample is applied to the membrane in strong acid or base (>1%), the pH during the coupling reaction will be compromised and poor coupling efficiencies will be observed. Peptides that have C-terminal lysines couple inefficiently (<50%) unless there are internal Glu/Asp residues, and peptides with C-terminal proline or C-terminal amides do not couple at all.

To perform solid phase sequencing on a microsequencer only minor alterations are required, as detailed earlier in Phosphorylation Site Determination. Highly abbreviated cycles can be performed because excessive valve block washing is not required. A coupling temperature of 55°C is used, with only two deliveries of PITC, followed by two 200-sec deliveries of N-methylpiperidine. An abbreviated solid phase cycle listing (modified gas phase cycle) that is used on our 494A is shown in Table 1. The gas phase cycle is used routinely for peptides that have multiple proline residues, to reduce the cleavage lag in these peptides. For all other peptides, a modified, pulsed liquid program (26 min) is used and the cartridge is only dried minimally after the delivery of liquid TFA (not shown).

Phosphopeptides sequenced from the in vivo labeling of the protein kinase PDK1 were determined to be singly phosphorylated by mass spectrometry and at serine residues by phosphoamino acid analysis (Fig. 4).[21] Solid phase Edman degradation established that PDK1 was phosphorylated at Ser 241 (A), Ser 393 and Ser 396 (B), Ser 25 (C), and Ser 410 (D). The release of ^{32}P radioactivity observed in cycles 1 and 7 (Fig. 4B and C, respectively) corresponds to the release of peptide that was only coupled through the side chain of aspartate and not through the carboxyl terminal residue. This also occurs with glutamate residues and is more evident if the peptide has a lysine residue at the carboxyl terminus. The tryptic

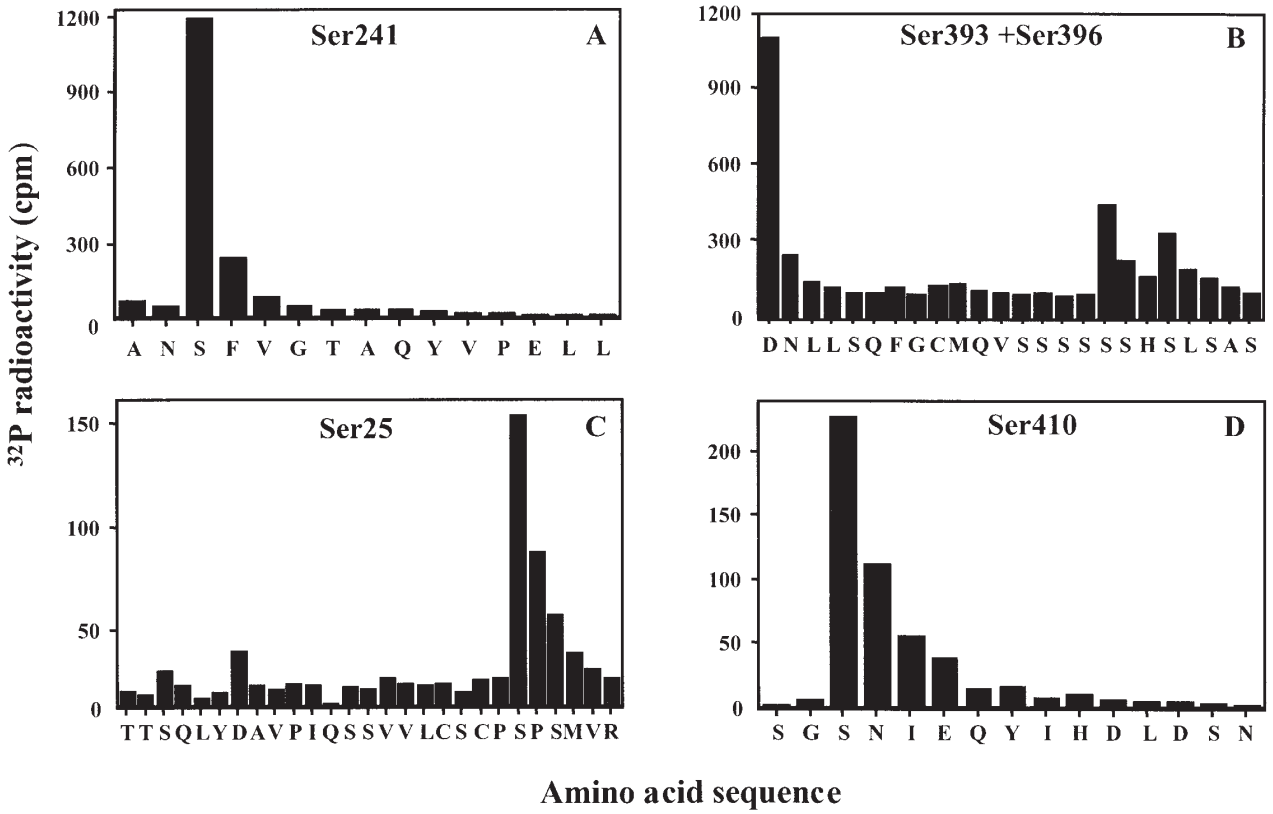

FIGURE 4

Solid phase sequence analysis of tryptic phosphopeptides derived from the in vivo labeling of the protein kinase PDK1. Phosphopeptides sequenced were **(A)** residues 239–257, **(B)** Asp-N subdigest of residues 358–407 corresponding to residues 377–400, **(C)** residues 4–30, and **(D)** residues 408–435.

peptide, corresponding to residues 358–407, was sequenced for 23 Edman cycles with no release of ^{32}P radioactivity (data not shown). A further aliquot was subdigested with endoproteinase Asp-N and resequenced (Fig. 4B). The results showed that the peptide (residues 377–400) was phosphorylated on either Ser 393 or Ser 396, but not both, as MALDI-MS analysis of both the original tryptic peptide and the endoproteinase Asp-N peptide were observed as singly phosphorylated species.[21] Therefore, it can be deduced that the original tryptic peptide fraction was a mixture of two singly phosphorylated peptides.

The tryptic phosphopeptide isolated from p70 S6 kinase, which had been phosphorylated by PDK1, was determined to have an N-terminal glutamate residue (residues 219–244).[22] Direct solid phase sequencing of this peptide resulted in a very limited release of ^{32}P radioactivity (Fig. 5A). This was the result of the formation of pyroglutamate after treatment of the peptide with the carbodiimide in the coupling reaction. To overcome this, an aliquot of the peptide was sequenced by conventional pulsed liquid phase Edman degradation to remove one amino acid. The peptide was extracted from the TFA etched glass fiber disc with 50% acetonitrile/0.1% trifluoroacetic acid and coupled to Sequelon-AA. This peptide gave a clear release of ^{32}P radioactivity corresponding to Thr 229 (Fig. 5B). The overall efficiency of this removal of the N-terminal glutamate is approximately 50% and has been used successfully many times in our laboratory (not shown).

Solid phase sequencing data for the tryptic peptide derived from p38α MAP kinase phosphorylated with the dual specificity protein kinase MKK6 showed that both Thr 180 and Tyr 182 were labeled, and again there was a minor release of ^{32}P radioactivity corresponding to the release of the peptide coupled through Glu 178 alone (Fig. 6). A 58-residue tryptic peptide derived from the in vivo labeling of the protein kinase LKB-1/STK 11 was determined to be doubly phosphorylated at threonine residues by a combination of mass spectrometry and by phosphoamino acid analysis.[23] The solid phase sequencing analysis of this peptide showed two clear releases of ^{32}P radioactivity at cycle

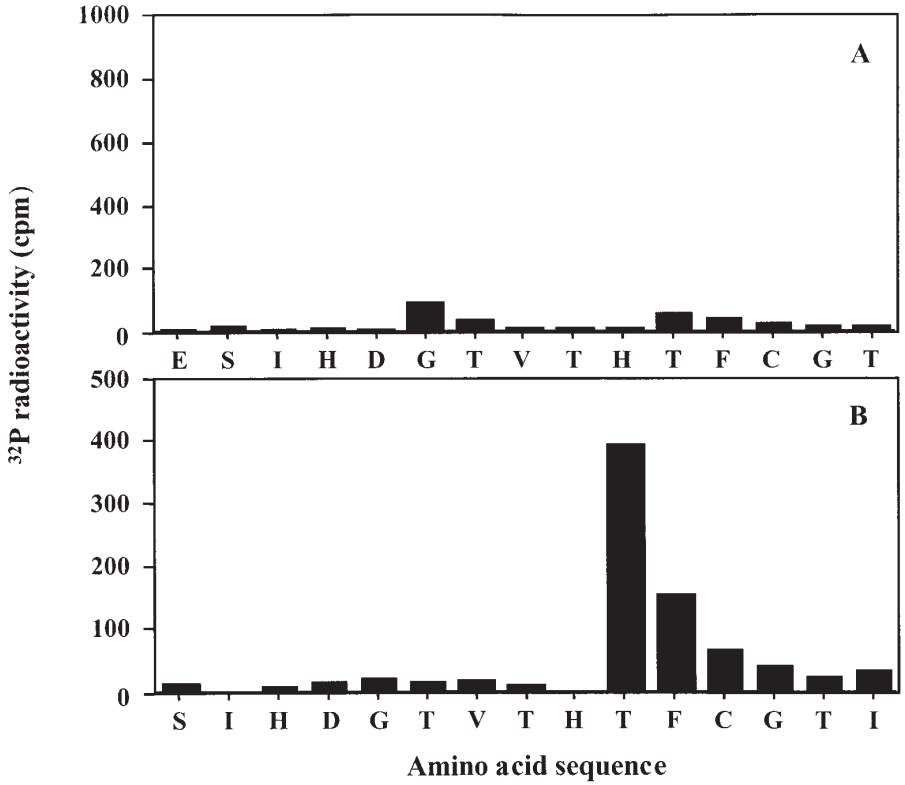

FIGURE 5

(A) The tryptic phosphopeptide ESIHDGTVTHTFCGTIEYMAPEILMR (p70 S6 kinase residues 219–244) was coupled to Sequelon-AA (4000 cpm) and solid phase sequence analysis performed. Only a very minor release of ^{32}P was detected with 3450 cpm remaining on the disc after 17 cycles of Edman degradation. **(B)** A second fraction of this peptide was processed to remove the N-terminal glutamate and subjected to solid phase sequence analysis (1880 cpm coupled).

3 and cycle 33 (Fig. 7). This corresponded to phosphorylation at Thr 336 and Thr 366. The sensitivity of the technique is not limited simply to the efficiency of a scintillation counter. If samples of less than 100 cpm are coupled to a Sequelon membrane, the collected ATZ fractions can be spotted onto a DEAE membrane and the ^{32}P radioactivity detected by phosphorimaging (Fig. 8). We have found this to be the membrane of choice as the ^{32}P-phosphate does not diffuse through the membrane, as has been found for PVDF and glass fiber membranes (not shown). By integrating the phosphorimage response, a bar graph can be obtained that mimics the data obtained using Cerenkov counting.

DISCUSSION

The methods outlined in this report are routinely used in our laboratory to successfully determine hundreds of protein phosphorylation sites every year. The key to the whole technique is the use of ^{32}P-labeling of the phosphoproteins. The radioactive labeling of proteins is the most sensitive method for determining which proteins/peptides are phosphorylated. The efficiency of the "in-gel" proteolytic digestion of the phosphoprotein can be assessed by the release of ^{32}P radioactivity from the gel piece. All subsequent steps that have been outlined can be followed and quantified by detection of the ^{32}P radioactivity. The ultimate sensitivity of the method is not limited to the detection of the ^{32}P radioactivity but rather to the sensitivity of the mass spectrometry techniques used to define the primary sequence of the phosphopeptide. Routinely, at least 500 fmol of a phosphopeptide is required to perform a complete characterization, including defining the phosphorylation sites. There will be examples where less material is required, but the largest proportion of any isolated phosphopeptide fraction will be used in defining the primary sequence of the phosphopeptide by mass spectrometry.

Many laboratories either cannot or will not label proteins radioactively before analysis. This removes

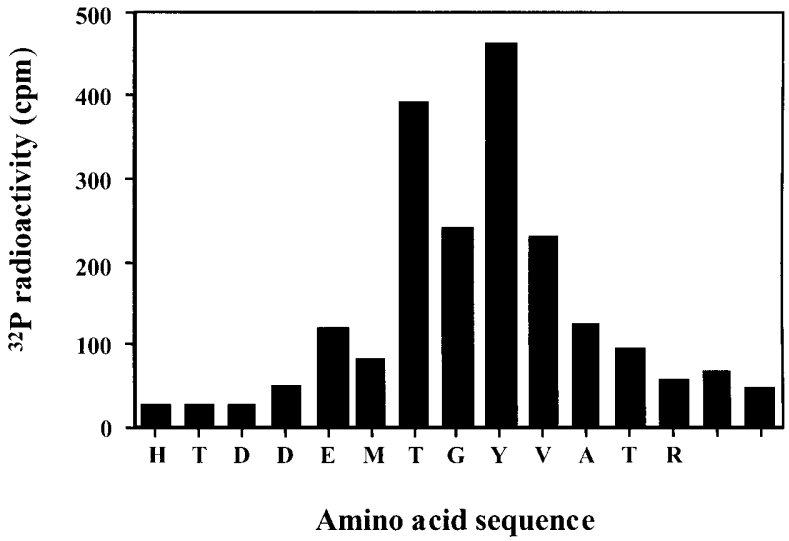

FIGURE 6

The tryptic diphosphopeptide HTDDEMTGY-VATR, derived from the tryptic digestion of p38αMAP kinase (which had been phosphorylated by the protein kinase MKK6 in vitro), was subjected to solid phase sequence analysis on an Applied Biosystems Procise 494A sequencer. The peptide was phosphorylated at Thr 180 and Tyr 182.

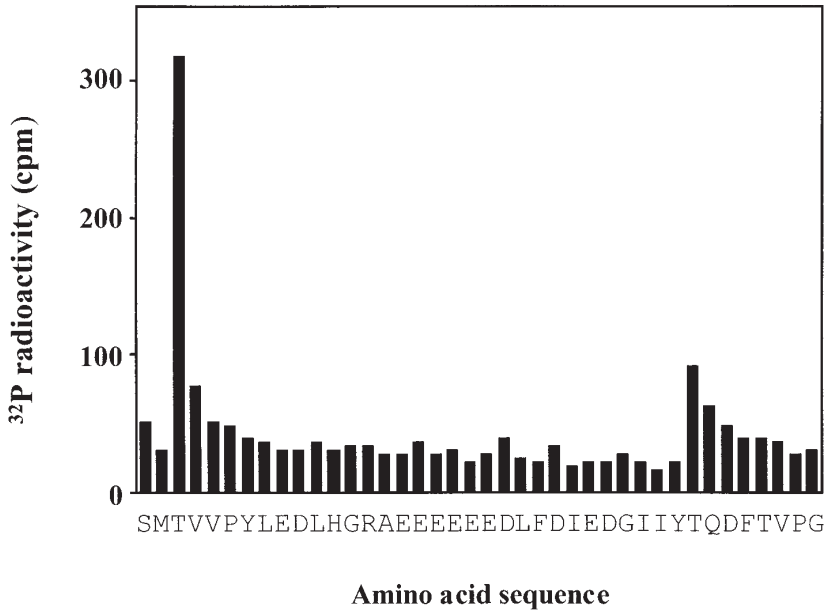

FIGURE 7

Solid phase sequencing of the tryptic peptide residues 333–391 from the protein kinase LKB1/STK11 that was labeled in vivo. It was determined to be doubly phosphorylated at threonine residues by MALDI-TOF-MS and phosphoamino acid analysis. The release of ^{32}P-radioactivity corresponds to Thr 336 and Thr 366.

the most specific and sensitive technique used to define or detect phosphoproteins and their phosphorylation sites. Defining the stoichiometry of phosphorylation of a substrate protein by a protein kinase in vitro and determining which sites are labeled is made much more difficult without the use of ^{32}P. Annan and co-workers have shown that it is possible to define phosphorylation sites from nonradiolabeled phosphoproteins.[10,12] The method they describe requires picomole quantities of the phosphopeptides, because multiple mass spectrometry experiments are performed on HPLC fractionated phosphoprotein digests. In the primary separation of the phosphopeptides, cone-induced dissociation of the peptides leads to the generation of PO_3^- and PO_2^- ions, whose ion currents are combined to produce a selected ion chromatogram. This chromatogram mimics that obtained when using an on-line radioactivity detector for ^{32}P-labeled peptides, but the method does consume sample, unlike detection by radioactivity.

The numerous mass spectrometry-based methods for defining protein phosphorylation sites from nonradiolabeled proteins are invariably based on defining known phosphorylation sites from model phosphoproteins.[3–4,7–9,13] The detection of phosphopeptides from proteolyic digests can be performed by parent

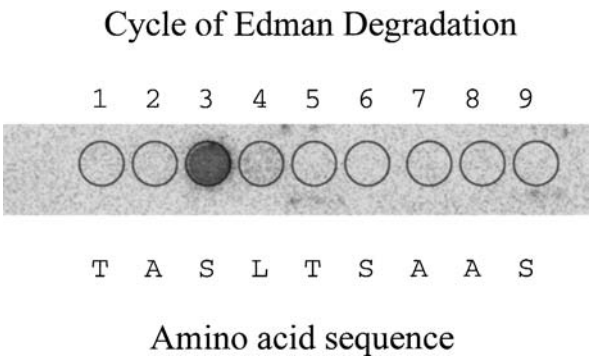

FIGURE 8

Solid phase sequencing of a tryptic peptide TASLAASIDGSR. Ten femtomoles of the peptide (45 cpm) was coupled to Sequelon-AA membrane and the collected ATZ–amino acid fractions were concentrated and spotted onto a DEAE membrane (Pall Gelman), and the ^{32}P radioactivity detected by phosphorimaging.

ion scanning in the negative ion mode;[9,10] neutral loss of H_3PO_4 in the positive ion mode;[5,13] or in the case of tyrosine phosphorylated proteins, immonium ion scanning.[11] All these techniques have their merits for defining phosphopeptides, as long as the phosphopeptide is ionized in the mass spectrometer. This is not always the case, as phosphopeptide ionization is often suppressed in both electrospray and MALDI-TOF-MS, when analyzing a complex mixture of peptides, such as a proteolytic digest. Enrichment of the phosphopeptides has been described, and this does lead to enhanced detection of the phosphopeptides.[4,5,7,13] However, these enrichment techniques are highly dependent upon the primary sequence of the phosphopeptide and without using a specific tag on the peptide, such as ^{32}P radioactivity, losses of phosphopeptides would go undetected.

Analyzing known phosphorylation sites is best carried out by immunoblotting with phosphospecific antibodies. In our laboratory, once phosphorylation sites have been defined from a ^{32}P-labeled protein, phosphospecific antibodies are then raised.[16–18,22,23] These reagents permit selective analysis of phosphorylation status of sites in a phosphoprotein, in cells and in vivo, in response to drugs, mitogens, or other extracellular stimuli. By defining the sites of phosphorylation from a ^{32}P-labeled protein, further studies can be performed on that phosphoprotein from non-radiolabeled cell lysates.

The method of defining unknown phosphorylation sites outlined in this report is the result of over ten years of method development. The technique has evolved from the original reports by Wettenhall et al.[14] and Coull et al.[15] by integrating mass spectrometry with solid phase Edman sequencing. It is not a high throughput method, but does result in a high percentage of phosphorylation sites identified from a single proteolytic digest. The method is adaptable to any protein sequencer, and is independent of the size of the phosphopeptide, the residue that is phosphorylated, or the number of phosphorylation sites.

ACKNOWLEDGMENTS

We thank Professor Sir Philip Cohen for critical reading of the manuscript and past and present members of the MRC Unit for the many phosphopeptides they have provided. This work was funded by the Medical Research Council (U.K.) and by the companies supporting the Division of Signal Transduction Therapy in Dundee (AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Novo-Nordisk and Pfizer).

REFERENCES

1. Cohen P. The origins of protein phosphorylation. *Nature Cell Biol* 2002;4:E127–E130.
2. Kuster B, Mortensen P, Andersen JS, Mann M. Mass spectrometry allows direct identification of proteins in large genomes. *Proteomics* 2001;1:641–650.
3. McLachlin DT, Chait BT. Analysis of phosphorylated proteins and peptides by mass spectrometry. *Curr Opin Chem Biol* 2001;5:591–602.
4. Stensballe A, Anderson S, Jensen ON. Characterisation of phosphoproteins from elecrophoretic gels by nanoscale Fe(III) affinity chromatography with off line mass spectrometry analysis. *Proteomics* 2001;1:207–222.
5. Ficarro SB, McCleland ML, Stukenberg PT, Shabanowitz J, Hunt DF, White FM. Phosphoproteome analysis by mass spectrometry and its application to *Saccbromyces cerevisiae*. *Nat Biotechnol* 2002;20:301–305.
6. Meyer HE, Hoffmann-Posorske E, Heilmeyer, LMG Jr. Determination and location of phosphoserine in proteins and peptides by conversion to S-ethylcysteine. *Methods Enzymol X* 1991;201:169–185.
7. Oda Y, Nagasu T, Chait BT. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat Biotechnol* 2001;19:379–382.
8. Goshe MB, Veenstra TD, Panisko EA, Conrads TP, Angell NH, Smith RD. Phosphoprotein isotope-coded affinity tags: Application to the enrichment and identification of low-abundance phosphoproteins. *Anal Chem* 2002;74:607–616.
9. Neubauer G, Mann M. Mapping of phosphorylation sites of gel-isolated proteins by nanoelectrospray tandem mass spectrometry: Potentials and limitations. *Anal Chem* 1999;71:235–242.
10. Annan RS, Huddleston MJ, Verma R, Deshaies RJ, Carr SA. A multidimensional electrospray MS-based approach

to phosphopeptide mapping. *Anal Chem* 2001;73: 393–404.
11. Steen H, Kuster B, Fernandez M, Pandey A, Mann M. Detection of tyrosine phosphorylated peptides by precursor ion scanning quadrupole TOF mass spectrometry in positive ion mode. *Anal Chem* 2001;73:1440–1448.
12. Loughrey Chen S, Huddleston MJ, Shao W, Deshaies RJ, Annan RS, Carr SA. Mass spectrometry based methods for phosphorylation site mapping of hyperphosphorylated proteins applied to Net1, a regulator of exit from mitosis in yeast. *Mol Cell Proteomics* 2002;1:186–196.
13. Bennet KL, Stensballe A, Podtelejnikov AV, Moniatte M, Jensen ON. Phosphopeptide detection and sequencing by matrix assisted laser desorption/ionisation quadrupole time of flight tandem mass spectrometry. *J Mass Spectrom* 2002;37:179–190.
14. Wettenhall REH, Aebersold RH, Hood LE. Solid-phase sequencing of 32P-labeled phosphopeptides at picomole and subpicomole levels. *Methods Enzymol* 1991;201:186–199.
15. Coull JM, Pappin DJC, Mark J, Aebersold R, Koster H. Functionalized membrane supports for covalent protein microsequence analysis. *Anal Biochem* 1991;194:110–120.
16. Sapkota GP, Kieloch A, Lizcano JM, et al. Phosphorylation of the protein kinase mutated in Peutz-Jeghers cancer syndrome, LKB1/STK11, at Ser431 by p90RSK and cAMP-dependent protein kinase, but not its farnesylation at Cys433, is essential for LKB1 to suppress cell growth. *J Biol Chem* 2001;276:19469–19482.
17. Knebel A, Morrice N, Cohen P. A novel method to identify protein kinase substrates; eEF2 kinase is phosphorylated and inhibited by SAPK4/p38δ. *EMBO J* 2001;20:4360–4343.
18. Woods YL, Rena G, Morrice N, et al. The kinase DYRK1A phosphorylates the transcription factor FKHR at Ser(329) *in vitro*, a novel *in vivo* phosphorylation site. *Biochem J* 2001;355:597–607.
19. Kemp BE, Pearson RB. Protein kinase recognition sequence motifs. *Trends Biochem Sci* 1990;15:342–346.
20. Annan RS, Carr SA. Phosphopeptide analysis by matrix assisted laser desorption time-of-flight mass spectrometry. *Anal Chem* 1996;68:3413–3421.
21. Casamayor A, Morrice NA, Alessi DR. Phosphorylation of Ser-241 is essential for the activity of 3–phosphoinositide-dependent protein kinase-1: Identification of five sites of phosphorylation *in vivo*. *Biochem J* 1999; 342:287–292.
22. Alessi DR, Kozlowski MT, Weng QP, Morrice N, Avruch J. 3 Phosphoinositide-dependent protein kinase 1 (PDK1) phosphorylates and activates the p70 S6 kinase in vivo and in vitro. *Curr Biol* 1998;8:69–81.
23. Sapkota GP, Boudeau J, Deak M, Kieloch A, Morrice N, Alessi DR. Identification and characterisation of four novel phosphorylation sites (Ser^{31}, Ser^{325}, Thr^{336}, Thr^{366}) on LKB1/STK11, the protein kinase mutated in Peutz-Jeghers cancer syndrome. *Biochem J* 2002;362:481–490.