Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome

Yoshiya Oda[1,2], Takeshi Nagasu[2], and Brian T. Chait[1]*

The current progression from genomics to proteomics is fueled by the realization that many properties of proteins (e.g., interactions, post-translational modifications) cannot be predicted from DNA sequence[1]. Although it has become feasible to rapidly identify proteins from crude cell extracts using mass spectrometry after two-dimensional electrophoretic separation, it can be difficult to elucidate low-abundance proteins of interest in the presence of a large excess of relatively abundant proteins[2,3]. Therefore, for effective proteome analysis it becomes critical to enrich the sample to be analyzed in subfractions of interest. For example, the analysis of protein kinase substrates can be greatly enhanced by enriching the sample of phosphorylated proteins. Although enrichment of phosphotyrosine-containing proteins has been achieved through the use of high-affinity anti-phosphotyrosine antibodies[4], the enrichment of phosphoserine/threonine-containing proteins has not been routinely possible. Here, we describe a method for enriching phosphoserine/threonine-containing proteins from crude cell extracts, and for subsequently identifying the phosphoproteins and sites of phosphorylation. The method, which involves chemical replacement of the phosphate moieties by affinity tags, should be of widespread utility for defining signaling pathways and control mechanisms that involve phosphorylation or dephosphorylation of serine/threonine residues.

Our approach to phosphoprotein enrichment and mapping is based on site-specific modification of phosphoseryl/phosphothreonyl residues. Using the property that the phosphate moiety on such residues is labile at high pH[5–14], we chemically replace these phosphates by biotinylated moieties (Fig. 1A). These biotin groups can then be used as affinity handles for immobilized avidin enrichment of the formerly phosphorylated proteins from complex mixtures of proteins. Individual protein components in the enriched fraction can be separated by gel electrophoresis, digested with trypsin, and characterized by mass spectrometry. Alternatively, the entire protein mixture can be digested with trypsin, whereupon the biotinylated tryptic peptide fraction can be enriched with immobilized avidin and characterized by mass spectrometry.

Under strongly alkaline conditions the phosphate moiety on phosphoseryl residues undergoes β-elimination to form the reactive dehydroalanyl residue[5–10] (Fig. 1B). This α,β-unsaturated residue is a Michael acceptor, which can readily react with a nucleophile (here ethanedithiol, or EDT) that in turn can be linked to a biotin affinity tag or other immobilizing agent. (An analogous reaction occurs for phosphothreonyl residues, albeit with a slower rate constant[9].) In our initial experiments, we found that the high concentration of hydroxyl ions required for efficient β-elimination (see Experimental Protocol) led to a significant protein degradation product[15]. To prevent this unwanted side reaction, we determined that the concentration of nucleophile should approach that of hydroxyl ion when the reaction is carried out in a single pot. A second potential problem in the present strategy arises from the reactivity of cysteine residues in the proteins. We used performic acid oxidation of the proteins as a means of protecting the cysteine residues[10]. (An additional consequence of this oxidation is quantitative conversion of methionine residues to methionine sulfones and oxidation of the tryptophan residues to yield two main forms containing, respectively, two and three oxygen atoms.) The reaction scheme that we ultimately settled on (performic acid oxidation followed by the chemistry indicated in Fig. 1B) allowed us to minimize unwanted side reactions while converting phosphoseryl and phosphothreonyl residues to biotinylated residues with high specificity and efficiency. (We note an additional complication that could arise by conversion of O-glycosylated residues, if present, to dehydroalanyl residues[9].)

To test our procedure under controlled conditions, we carried out the reactions on phosphorylated β-casein and added the biotinylated protein product to a large excess (60-fold by weight) of a mixture of unphosphorylated standard proteins. As demonstrated in Figure 2A, B, we can cleanly separate and enrich the phosphorylated protein from the excess of unphosphorylated proteins. In particular, it can be seen that although phosphorylated β-casein migrates to approximately the same gel position as the carbonic anhydrase standard (indicated by the arrow in Fig. 2A, lane 1), material eluted from the monovalent avidin beads (lanes 7 and 8) consists entirely of β-casein, as determined by matrix-assisted laser desorption/ionization mass spectrometric (MALDI-MS) tryptic peptide mapping of these gel bands[16] (Fig. 2B). It is noteworthy that the large excess of carbonic anhydrase in the original protein mixture precluded identification by peptide mapping of β-casein in the band indicated by

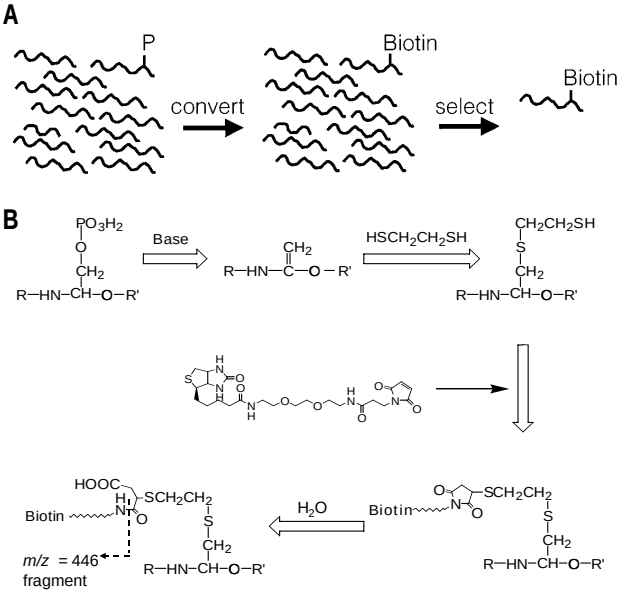

Figure 1. (A) Scheme for isolating/enriching phosphoproteins and/or phosphopeptides. Wavy lines represent proteins or peptides. Wavy line with "P" designates phosphoprotein or phosphopeptide. (B) Scheme for chemical conversion of a phosphoserine residue to a biotinylated residue. Indicated by the dashed line is the position of the facile cleavage (see text) that produces the ion observed at *m/z* = 446 in the MS/MS spectra.

[1]*The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA.* [2]*Eisai Co., Ltd., Tokodai 5-1-3, Tsukuba, Ibaraki 300-2635, Japan.*
**Corresponding author (chait@rockvax.rockefeller.edu).*

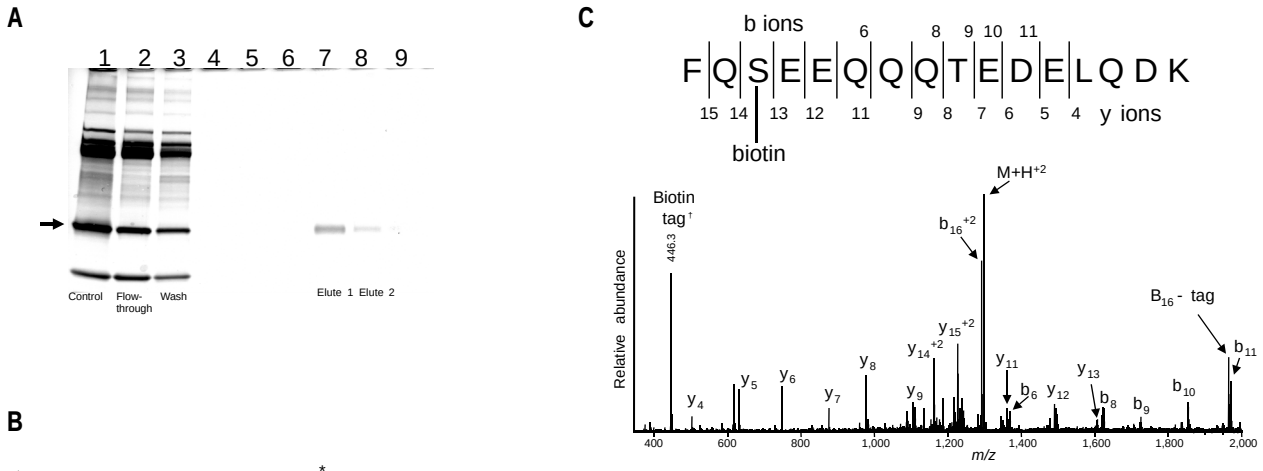

Figure 2. Enrichment and isolation of a phosphoprotein from a complex mixture of proteins. (A) Separation and enrichment of a biotinylated phosphoprotein (lanes 7 and 8) from a mixture containing a large excess of unphosphorylated proteins (lane 1) using immobilized monovalent avidin. Lane 1, phosphoprotein + mixture of unphosphorylated proteins; lane 2, flowthrough; lanes 3 and 4, buffer wash; lanes 5 and 6, 5 mM biotin wash; lanes 7 and 8, elution with 5 mM biotin in 5% CHAPS; lane 9, elution with 1% SDS at pH 1. (B) MALDI-MS analysis of tryptic digests of the ~29 kDa bands (indicated by the arrow in Fig. 2A) from lanes 1, 2, 3, and 7. The asterisk * indicates ion fragments from carbonic anhydrase. The bottom panel shows the mass spectrum of the tryptic peptides from the band in lane 7 after capture by immobilized NeutrAvidin and subsequent elution of a biotinylated phosphopeptide. Mass measurement indicates that the two peaks arise from the biotinylated phosphopeptide β-casein 33–48. (The two peaks, separated by 18 Da, correspond to the closed and open forms of the biotinylated products indicated in Figure 1.) (C) ESI-MS/MS analysis of the open form of the phosphopeptide β-casein 33–48, that is, the higher of the two peaks shown in the bottom panel of (B). The nomenclature for the peptide ion fragments is given by Biemann[19]. "tag†" designates the whole biotin tag, whereas "tag" designates the biotinylated portion of the tag up to and including the maleimide nitrogen. Note that control experiments showed no nonspecific binding of β-casein to the monovalent avidin resin.

the arrow (Fig. 2A, lane 1), although the intense β-casein peptide 184–202 (average molecular mass, $M_{average}$ = 2186.6 Da) can be readily discerned in the tryptic digest (Fig. 2B, top panel). This peptide is entirely absent from the flowthrough and wash, indicating that the biotin-labeled protein is efficiently retained on the avidin beads. Indeed, the intensity of this peak relative to the intensities of the various carbonic anhydrase peaks can be used to estimate losses in the flowthrough and washing steps. By comparing these peak intensities in Figure 2B (control, flowthrough, and wash), we conservatively estimate losses of <10% of the casein, indicating recoveries of >90%. Tryptic peptides eluted from the band in lane 7 were further incubated with NeutrAvidin beads to enrich the biotinylated (formerly phosphorylated) peptides. MALDI-MS analysis of the peptides eluted from the avidin beads (Fig. 2B, bottom panel) indicates the presence of biotinylated β-casein 33–48 (measured $M_{average}$ 2584.0 and 2602.2 Da, theoretical $M_{average}$ 2583.8 and 2601.8 Da, where the higher mass component of the pair of peaks arises from partial opening of the maleimide ring through hydrolysis (Fig. 1B). This pair of peaks separated by 18 Da also provides a signature for the biotinylated peptides. The identification of biotinylated β-casein 33–48 was confirmed[17] in a separate experiment by electrospray ionization–liquid chromatography–tandem mass spectrometry (ESI-LC-MS/MS) of the peptide having $M_{monoisotopic}$ 2600.4 (theoretical $M_{monoisotopic}$ 2600.1) (Fig. 2C). Note that the abundant ion fragment at mass-to-charge ratio (m/z) = 446.3 (arising from elimination of the portion of the biotin tag up to and including the maleimide nitrogen; Fig. 1B) provides an additional signature for the biotinylated peptide. The MS/MS analysis provides unambiguous localization of the biotinylated residue (and hence the site of phosphorylation).

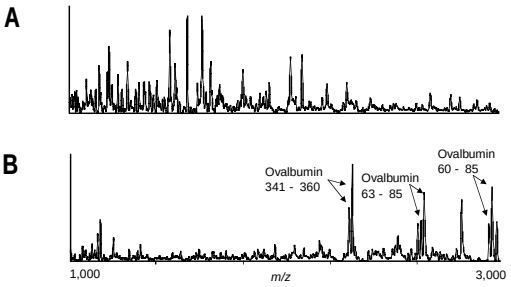

Figure 3. MALDI mass spectra of a tryptic digest of a whole-cell yeast (*Saccharomyces cerevisiae*) extract to which 2% by weight of ovalbumin was added. Before digestion with trypsin, the phosphoproteins were converted to biotinylated proteins using the procedure outlined in Figure 1B. (A) Before NeutrAvidin extraction. (B) After NeutrAvidin extraction. Biotinylated (formerly phosphorylated) ovalbumin peptides are clearly observed as pairs of peaks 18 Da apart. The residue numbers are indicated for the observed ovalbumin tryptic peptides.

A second approach that we tested for identifying and characterizing phosphoproteins involved direct isolation of the biotinylated peptides from a trypsin digest of an unfractionated protein mixture. In this case, we carried out our optimized chemistry (Fig. 1B) on a whole yeast cell extract, which we doped (for reasons of control) with a known concentration (2%) of exogenous phosphorylated protein (i.e., chicken ovalbumin). Figure 3 shows MALDI-MS spectra of the complex peptide mixture obtained before and after avidin purification. Three intense pairs of peaks dominate the spectrum obtained after avidin purification. The molecular masses of these

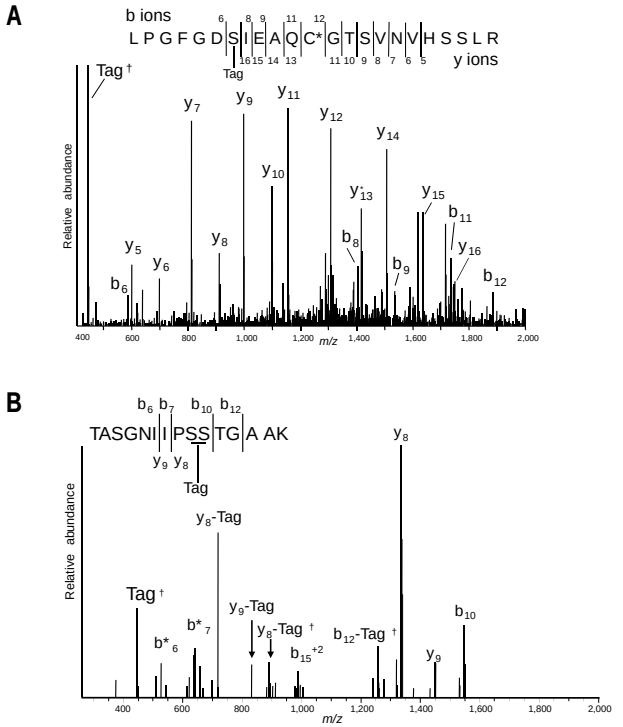

Figure 4. ESI-MS/MS spectra of biotinylated proteins allow identification of the formerly phosphorylated proteins and identification of the sites of biotinylation (i.e., phosphorylation). The asterisk * designates loss of NH_3 or H_2O from the ion fragments. "Tag" designates the whole biotin tag, whereas "Tag †" designates the biotinylated portion of the tag up to and including the maleimide nitrogen. C* designates cysteic acid. (A) Chicken ovalbumin, residues 63–85. (B) Yeast glyceraldehyde 3-phosphate dehydrogenase 2 (Tdh2p; YJR009c) residues 199–213.

peaks correspond to phosphorylated ovalbumin 341–360, 63–85, and 60–85. Their identities were confirmed in a separate experiment by ESI-LC-MS/MS (Fig. 4A), which also enabled definition of the sites of biotinylation.

Although the results shown in Figure 3 confirmed that the chemistry proceeded as expected and that a known exogenous phosphorylated protein can be readily detected in a complex mixture, our ultimate goal is to enrich, detect, and identify endogenous phosphorylated proteins that are extracted from whole cells, subcellular compartments, and protein complexes. We thus subjected the NeutrAvidin-purified peptide subfraction obtained from the whole-cell yeast protein extract to ESI-LC-MS/MS and used the abundant ion fragment at m/z = 446 (arising from loss of a portion of the biotin tag from the biotinylated peptides) as a definitive signature of the biotin-tagged peptides. The resulting MS/MS spectra can be used both to identify the protein from which the peptide originated and to determine the site(s) of biotinylation (and hence phosphorylation). An example of such an analysis is provided in Figure 4B, which shows the MS/MS spectrum of a doubly protonated tryptic peptide with m/z = 997.3. Input of this parent ion mass together with the masses of the intense fragment ions into the search engine "PepFrag"[17] identifies the peptide TASGNIIPSSTGAAK as arising from the abundant protein glyceraldehyde 3-phosphate dehydrogenase 2 (Tdh2p; YJR009c; NCBI nonredundant database). In addition to this unambiguous identification, the fragment ions enable localization of the phosphorylated residue to one of the two central serine residues. Absence of a fragment between these two residues precludes more specific identification of this previously unknown phosphorylation site. Because low-abundance phosphopeptides yield data of lower quality than that shown in Figure 4, we are currently refining the method to include incorporation of a cleavage site (e.g., a disulfide linkage) between the biotin moiety and the peptide. This refinement should not only allow almost quantitative recovery of the proteins/peptides of interest from the NeutrAvidin beads but also preclude the facile fragment losses of the biotin tag that are observed in the MS/MS spectra (Fig. 4), leading to simpler, more intense, and more easily interpretable fragmentation patterns.

In summary, we describe a method for enriching phosphoproteins and phosphopeptides (phosphorylated on serine/threonine residues) and for subsequently identifying the phosphoproteins as well as the sites of phosphorylation. Simple variations of the chemical procedure (involving, for example, the use of unlabeled and deuterium-labeled EDT) should prove effective for quantifying site-specific changes in phosphorylation.

Experimental protocol

Materials. Bovine β-casein, chicken egg ovalbumin, and the "standard" proteins, urease, transferrin, conalbumin, albumin, carbonic anhydrase, and β-lactoglobulin (Fig. 2A) were obtained from Sigma Chemical Co. (St. Louis, MO).

Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE). Proteins were separated by SDS–PAGE on Tris-glycine gradient gels (Novex, San Diego, CA) and visualized with zinc stain (Bio-Rad Laboratories, Hercules, CA). The protein bands were cut out, placed in Eppendorf tubes, and destained with Bio-Rad destaining buffer. In-gel trypsin digestion was performed as described[18].

Oxidation of proteins. Quantitative conversion of cysteine residues to cysteic acid residues was achieved by adding 0.5 ml of 30% hydrogen peroxide (Wako Pure Chemicals, Osaka, Japan) to 4.5 ml of 88% formic acid (Wako) containing 25 mg phenol. The solution was maintained at room temperature (RT) for at least 30 min, whereupon it was cooled to 4°C. At this point, 0.5 ml of the resulting performic acid solution was added to 5 mg of dry protein and left overnight at 4°C. The sample was diluted with 0.5 ml of ice-cold water and twice dialyzed against 1 L of water (4°C). The resulting sample was dried by vacuum evaporation in a SpeedVac (Savant, Holbrook, NY).

Biotinylation and enrichment of phosphoproteins. Oxidized proteins (0.05–1 mg) were dissolved in 240 µl of 5% 3-((cholamidopropyl)-dimethylammonio)-1-propanesulfonate (CHAPS; Pierce Chemical Co., Rockford, IL) solution. A 20 µl aliquot of 1,2-ethanedithiol (EDT; Tokyo Kasei Kogyo Co., Tokyo) was diluted into 50 µl ethanol. To this solution were added 140 µl of 4 M LiOH, 50 µl acetonitrile, and finally the oxidized protein–CHAPS solution (total volume 500 µl). The resulting reaction solution was incubated at 37°C for 1 h. After neutralizing with acetic acid, 1 ml of ice-cold acetone was added, the solution kept at –40°C for 1 h, centrifuged at 20,000 *g* for 3 min, the precipitant washed with ethanol, and redissolved in 500 µl of 8 M urea, 0.5% CHAPS, 1 mg Tris (2-carboxy ethylphosphine) hydrochloride (TCEP; Pierce) in 0.1 M sodium phosphate buffer (pH 6.8). After 1 h at RT, 5 mg of (+)-biotinyl, 3-maleimidopropionamidyl-3,6-dioxoctanediamine (Pierce) was added. This solution was twice dialyzed against 3 L of 50 mM NH_4HCO_3.

Enrichment of biotinylated proteins. The biotinylated proteins were enriched by incubation with 0.5 ml immobilized monomeric avidin gel (Pierce) for 1 h at RT. The gel solution was poured into an empty column (2 ml volume) and washed with 5 ml of 50 mM NH_4HCO_3. Biotinylated proteins were eluted with 1 ml of 5 mM biotin in 2% CHAPS solution.

Enrichment of biotinylated peptides. After tryptic digestion, the peptide solution was passed through an immobilized NeutrAvidin column (Pierce) with a bed volume of 50 µl. The column was first washed with 2 ml of 200 mM sodium phosphate buffer (pH 7) containing 2% CHAPS and 150 mM NaCl, and then with 2 ml of water. Finally, biotinylated peptides were eluted with 200 µl of acetonitrile:water:trifluoroacetic acid (TFA) (50:50:0.1, vol/vol/vol).

MALDI-MS. Each dried sample was redissolved in 5 µl of acetonitrile: 0.1% aqueous TFA (1:1, vol/vol). Aliquots (0.5 µl) of sample solution were loaded onto the MALDI-MS sample plate together with 0.5 µl of 2,5-dihydroxyben-

zoic acid (DHB; Aldrich, Milwaukee, WI) matrix solution (60 g/L) in acetonitrile:water 1:2 (vol/vol). The MALDI-MS measurements shown were obtained using a delayed-extraction linear time-of-flight mass spectrometer (Model Voyager Linear DE; PerSeptive Biosystems, Framington, MA).

LC-MS/MS. High-performance liquid chromatography (HPLC) was done with a 50 mm × 0.2 mm C_{18} capillary column (Magicms, Michrom BioResources, Auburn, CA), and the eluent directed to an ESI ion trap mass spectrometer (Finnigan Model LCQ, San Jose, CA) at a flow rate of 1 µl/min after flow splitting. MS/MS spectra were obtained in data-dependent mode in which the highest intensity peaks in each MS scan were chosen for collision-induced dissociation.

Acknowledgments

This work was supported by the National Center for Research Resources, National Institutes of Health (Grant RR00862 to B.T.C.), and the Eisai Co., Ltd. We thank Fred Cross, Angus Nairn, Derek McLachlin, Yasutaka Takase, and Takashi Owa for useful discussions, Wenzhu Zhang and David Fenyö for development of the protein identification software, and Kappei Tsukahara for assistance in yeast culture and cell handling.

Received 28 August 2000; accepted 2 February 2001

1. Lander, E.S. The new genomics: global views of biology. *Science* **274**, 536–539 (1996).
2. Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y. & Aebersold, R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci. USA* **97**, 9390–9395 (2000).
3. Corthals, G.L., Wasinger, V.C., Hochstrasser, D.F. & Sanchez, J.C. The dynamic range of protein expression: a challenge for proteomic research. *Electrophoresis* **21**, 1104–1115 (2000).
4. Pandey, A. *et al.* Analysis of receptor signaling pathways by mass spectrometry: identification of vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors. *Proc. Natl. Acad. Sci. USA* **97**, 179–184 (2000).
5. Taborsky, G. Phosphoproteins. *Adv. Protein. Chem.* **28**, 1–210 (1974).
6. Annan, W.D., Manson, W. & Nimmo, J.A. The identification of phosphoseryl residues during the determination amino acid sequence in phosphoproteins. *Anal. Biochem.* **121**, 62–68 (1982).
7. Meyer, H.E., Hoffmann-Posorske, E., Korte, H. & Heilmeyer, L.M., Jr. Sequence analysis of phosphoserine-containing peptides. Modification for picomolar sensitivity. *FEBS Lett.* **204**, 61–66 (1986).
8. Holmes, C.F. A new method for the selective isolation of phosphoserine-containing peptides. *FEBS Lett.* **215**, 21–24 (1987).
9. Mega, T., Nakamura, N. & Ikenaka, T. Modifications of substituted seryl and threonyl residues in phosphopeptides and a polysialoglycoprotein by beta-elimination and nucleophile additions. *J. Biochem. (Tokyo)* **107**, 68–72 (1990).
10. Meyer, H.E., Hoffmann-Posorske, E. & Heilmeyer, L.M., Jr. Determination and location of phosphoserine in proteins and peptides by conversion to *S*-ethylcysteine. *Methods Enzymol.* **201**, 169–185 (1991).
11. Meyer, H.E. *et al.* Strategies for nonradioactive methods in the localization of phosphorylated amino acids in proteins. *FASEB J.* **7**, 776–782 (1993).
12. Moorhead, G., MacKintosh, R.W., Morrice, N., Gallagher, T. & MacKintosh, C. Purification of type 1 protein (serine/threonine) phosphatases by microcystin–Sepharose affinity chromatography. *FEBS Lett.* **356**, 46–50 (1994).
13. Fadden, P. & Haystead, T.A. Quantitative and selective fluorophore labeling of phosphoserine on peptides and proteins: characterization at the attomole level by capillary electrophoresis and laser-induced fluorescence. *Anal. Biochem.* **225**, 81–88 (1995).
14. Jaffe, H., Veeranna & Pant, H.C. Characterization of serine and threonine phosphorylation sites in beta-elimination/ethanethiol addition-modified proteins by electrospray tandem mass spectrometry and database searching. *Biochemistry* **37**, 16211–16224 (1998).
15. Patchornik, A. & Sokolovsky, M. Nonenzymatic cleavages of peptide chains at the cysteine and serine residues through their conversion into dehydroalanine. I. Hydrolytic and oxidative cleavage of dehydroalanine residues. *J. Am. Chem. Soc.* **86**, 1206–1212 (1964).
16. Zhang, W. & Chait, B.T. ProFound: an expert system for protein identification using mass spectrometric peptide mapping information. *Anal. Chem.* **72**, 2482–2489 (2000).
17. Fenyo, D., Qin, J. & Chait, B.T. Protein identification using mass spectrometric information. *Electrophoresis* **19**, 998–1005 (1998).
18. Sechi, S. & Chait, B.T. Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. *Anal. Chem.* **70**, 5150–5158 (1998).
19. Biemann, K. Contributions of mass spectrometry to peptide and protein structure. *Biomed. Environ. Mass Spectrom.* **16**, 99–111 (1988).