Mapping phosphorylation sites in proteins by mass spectrometry

Wenying Shou[1], Rati Verma[1,2], Roland S. Annan[3], Michael J. Huddleston[3], Susan L. Chen[3], Steve A. Carr[3]**, and Raymond J. Deshaies[1,2]*

[1]Division of Biology 156-29 and [2]Howard Hughes Medical Institute
California Institute of Technology, Pasadena, CA 91125

[3]Proteomics and Biological Mass Spectrometry Laboratory
GlaxoSmithKline, 709 Swedeland Rd.,
King of Prussia, PA 19406

Running title: mass spectrometric analysis of phosphoproteins

* to whom correspondence should be addressed. Fax: (626) 449-0756; Phone: (626) 395-3162; E-mail: deshaies@its.caltech.edu

** Current address: Millennium Pharmaceuticals, 640 Memorial Drive, Cambridge, MA 02139

1. Introduction

Many intracellular pathways are regulated by protein phosphorylation. To understand how protein phosphorylation controls a pathway, two experiments are required (Fig. 1). The common starting point is to mutate the gene that encodes the relevant protein kinase (a *trans*-mutant) (Fig. 1B). There are four limitations to this experiment. First, the large number of protein kinases expressed in eukaryotic cells can make it difficult to identify the responsible enzyme. The budding yeast genome, for example, codes for 120 different protein kinases.[1] Second, homologous protein kinases can act redundantly to regulate a process. For example, the cyclin-dependent protein kinases Pho85 and Srb10 both phosphorylate Gcn4 and target it for ubiquitin-dependent proteolysis.[2] Third, most protein kinases have many substrates, and therefore, a protein kinase mutant may display many phenotypes that obscure the pathway of interest. Fourth, from the characterization of protein kinase mutants alone, it can be difficult to decipher the exact mechanism of regulation and untangle primary and secondary effects.

The second experiment required to establish that phosphorylation of a given protein plays a causal role in a process under study is the evaluation of the effect of non-phosphorylatable mutations in the candidate substrate (*cis*-mutations) (Fig. 1C). If one can establish that a *cis*-mutant in substrate Y (Fig. 1C) and a *trans*-mutant in protein kinase X

(Fig. 1B) both prevent the execution of process Z, then it is very likely that phosphorylation of Y by X activates Z.

A confounding problem for "closing the circle" by analysis of *cis* mutants is that it is often very difficult to map phosphorylation sites in substrate proteins. Many phosphoproteins are expressed in low abundance and can be recovered in only picomole quantities. In addition, most phosphoproteins are phosphorylated on more than one site and phosphorylation of any given site is often sub stoichiometric.

The serious constraints of low phosphopeptide yield and stoichiometry make it essential to have analytical methods that can preferentially detect and analyze phosphopeptides. By far the most common approach is to isolate phosphorylated substrate protein from ^{32}P-labeled cells, and separate tryptic digests of the labeled protein by thin-layer chromatography to reveal phosphopeptides.[3-6] One can then use a combination of secondary digestion with other proteases, deductive logic, and in some cases direct protein sequence to assign sites of phosphorylation.[7] In practice, this is not a trivial exercise. It is especially unappealing to contemplate the isolation of low abundance substrate from large cultures of ^{32}P-labeled cells.

Mass spectrometry (MS) on the other hand is ideally suited to the direct identification of protein phosphorylation sites. Phosphopeptides present in mixtures can be sequenced at the femtomole level without the

need for extensive purification. Over the past few years, a variety of mass spectrometry-based approaches have been employed to map phosphorylation sites in proteins.[8-13] A great advantage of these methods is that they do not require prior labeling of the target protein with ^{32}P. Thus, it is possible to isolate rare proteins from large-scale cultures for phosphopeptide mapping studies. Regardless of the method used to map phosphorylation sites, it is imperative that the native phosphorylation state of the target protein be preserved during isolation. Here, we describe general strategies for isolating phosphoproteins from budding yeast cells by drawing reference to two specific examples: the S-Cdk inhibitor Sic1, and the mitotic exit inhibitor Net1. We then provide an overview of the mass spectrometric methods used to map specific phosphorylation sites in these proteins.

2. General strategies for protein isolation

a. Strain Design

Complete phosphorylation site mapping by mass spectrometry typically requires approximately 5 - 200 pmol of purified protein (depending on the complexity and stoichiometry of phosphorylation) preserved in its native phosphorylation state. To obtain reasonable amounts of material (Your Favorite Phosphoprotein; Yfp) that is maximally phosphorylated , the

following molecular genetic manipulations can be performed with the strain from which Yfp is to be isolated.

I. *YFP* can be expressed from an inducible promoter (e.g. *GAL1,10*). A short 1-3 hour pulse of galactose may be enough to yield sufficient amounts of Yfp. Constitutive overproduction of Yfp on the other hand could overwhelm its cognate protein kinase,[14] or could circumvent other cellular regulatory systems, resulting in nonphysiological phosphorylation of the substrate. If a known kinase is being tested, then it could also be overproduced.

II. Determine physiological states under which Yfp is maximally phosphorylated. It may be beneficial to use mutants that stabilize the phosphorylation state of the protein. For example, if the counteracting phosphatase is known, Yfp can be isolated from a cell deficient in phosphatase activity.

III. Isolate Yfp from strains that can be conditionally inactivated for the kinase, and compare the phosphorylation pattern with that of Yfp isolated from wild type cells.

b. Substrate and Experimental Design

The purification of Yfp is facilitated by the introduction of one or more affinity tags at the N or C-terminus of the protein. Several commonly used epitopes which have proven to be effective include the His6, HA,

Myc, FLAG, polyoma, and 'ZZ' (IgG binding domain of protein A) tags. These epitopes can be combined in various permutations (e.g., His6-HA) to enable consecutive affinity purification steps.[15-16] Regardless of the tagging strategy, it is our bias that it is critical to inactivate phosphatases to prevent dephosphorylation of the target protein during purification. In both cases that we describe here, the extract was denatured prior to affinity purification of the target protein.

c. Growing and Harvesting cells

Growth. Cells are grown to exponential phase (OD_{600} = 0.5) before further manipulations such as temperature shift to 37^{o}C to impose a *ts* block, or induction of substrate or protein kinase expression with galactose.

Harvesting. After cells have reached the desired stage, they are chilled and harvested as quickly as possible to preserve the phosphorylation status of Yfp. We routinely fill 1-liter centrifuge bottles half-way with ice, place them on ice, and pour culture directly into the bottles with vigorous shaking. The cultures are centrifuged at 4^{o}C for 5- 10 min at 4000g to pellet cells. Pelleted cells are rapidly resuspended either with ice-cold water, or 25 mM Tris, pH 7.5 containing a phosphatase and protease inhibitor cocktail described below, transferred to 50-ml conical screw-cap tubes, and sedimented in a clinical centrifuge at 5000 rpm for 5 min at

4°C. It is important to work as quickly as possible during the cell harvest to minimize dephosphorylation.

d. Phosphatase and Protease Inhibitor Cocktails

Phosphatase and protease inhibitor cocktails are typically used as a prophylactic measure during lysis and purification to help preserve Yfp in its native state. Protease inhibitor cocktails consist of combinations of: 5 mM EDTA (stock = 0.5 M, pH 8), 2 mM EGTA (stock = 0.2 M, pH 8), 0.2 mM 4-(2-aminoethyl)-benzene sulfonyl fluoride hydrochloride (AEBSF: frozen stock = 100 mM in water), 25 μg/ml aprotinin (frozen stock = 10 mg/ml), 1 mM benzamidine (frozen stock = 1 M), 1 mM phenylmethylsulfonyl fluoride (PMSF, stock = 100 mM in 100% isopropanol), 5 – 10 μg/ml pepstatin, leupeptin, chymostatin (frozen stock = 5 – 10 mg/ml in DMSO). Phosphatase inhibitor cocktails consist of: 10 mM NaF (stock = 0.5 M), 60 mM β-glycerolphosphate (stock = 1M, pH 7.5, store at 4°C), 10 mM sodium pyrophosphate (stock = 100 mM), 2 mM sodium orthovanadate (stock = 200 mM; for method of preparation, see Ref. 17), and 3 μM microcystein-LR (frozen stock = 300 mM in DMSO).

Three methods can be used to monitor how well the phosphorylation state of a protein is preserved throughout purification. Functional assays are the most reliable – for example, phosphorylated but not unphosphorylated Sic1 can serve as a substrate for SCF^{Cdc4} (see

Section 3a). However, functional assays are also the most demanding, because they require that the proteins remain competent after purification. If a phosphorylated protein migrates differently from its unphosphorylated counterpart in an SDS-polyacrylamide gel, then gel mobility can be used to track the degree of phosphorylation (see Section 3b). If these two methods failed, then the last resort would be to label yeast cells with [^{32}P] phosphate, and to use the radioactivity of the protein (normalized against the level of the protein) as an indicator.[15]

3. Isolation of Sic1 and Net1 for mapping sites of phosphorylation *in vivo*.

a. Sic1

Sic1 is an S-Cdk inhibitor that has to be degraded for cells to enter S-phase.[15,18] Following its ubiquitination by the SCF^{Cdc4}/Cdc34 pathway, Sic1 is recognized by the 26S proteasome and degraded.[19] Although it was known that G1-Cdk is required for both the ubiquitination and degradation of Sic1 and that Sic1 is phosphorylated by G1-Cdk,[18,20,21] it was not known whether phosphorylation of Sic1 itself (as opposed to phosphorylation of some other protein) triggered its ubiquitination and degradation. To address this key issue, we sought to construct a "non-phosphorylatable" mutant of Sic1. Sites at which G1-Cdk

phosphorylated Sic1 *in vitro* were mapped by nano-electrospray tandem mass spectrometry (nano-ESMS/MS),[13] and a mutant (Sic1- 3P) lacking a subset of the identified sites was constructed.[15] Sic1- 3P was not ubiquitinated by SCF^{Cdc4}/Cdc34 *in vitro*, and was stable *in vivo*, suggesting that G1-Cdk enabled the G1/S transition by phosphorylating Sic1, thereby targeting it for ubiquitination and degradation via the SCF^{Cdc4}/Cdc34 pathway.

To confirm our hypothesis, we sought to demonstrate that the same residues on Sic1 that are phosphorylated by G1-Cdk *in vitro* are also phosphorylated *in vivo.* At the time that we set out to map the sites of *in vivo* phosphorylation on Sic1 using nano-ESMS/MS, there was little precedent for using this technique to map phosphorylation sites on proteins isolated from their native environment. Thus, to maximize our likelihood of achieving success, we took great care to ensure that sufficient amounts of Sic1 were isolated to enable a thorough analysis, that the isolated Sic1 was of high purity, and that the native phosphorylation state of Sic1 was preserved as much as possible. To achieve these goals, we implemented the following experimental design:

1. Sic1 was transiently over-expressed from a chromosomally-integrated GAL promoter-driven cassette to yield enough material for analysis, but not so much as to overwhelm the cell.

2. The expressed Sic1 was tagged at the C-terminus with a bipartite haemagluttinin-hexahistidine (HAHis6 epitope). This tandem epitope provided a key advantage, because hexahistidine binds Ni-NTA even in the presence of strong denaturants. Thus, we were able to prepare the cell lysate and conduct the first purification step in the presence of 6 M guanidinium hydrochloride, which is expected to inactivate all phosphatases and proteases in the extract. However, single-step purification on Ni-NTA rarely yields material of sufficient purity unless the protein is expressed at very high levels, and therefore the HA domain of the tandem epitope was critical because it enabled consecutive affinity purification steps. Following these two steps, Sic1 was sufficiently pure to be submitted directly to mass spectrometric analysis without further fractionation by SDS-PAGE (Fig. 2A).

3. $Sic1^{HAHis6}$ was expressed in a *cdc34-2* strain held at the nonpermissive temperature. Cdc34 is the E2 enzyme that mediates the ubiquitination and degradation of phosphorylated Sic1. Mutant $cdc34^{ts}$ cells accumulate high levels of G1-Cdk activity,[20,22] but are unable to degrade Sic1,[18] and thus accumulate the phosphorylated protein.

Detailed Procedure: RJD 1044 (*GAL-SIC1*^{HAHis6} *cdc34-2*) cells were grown in 6 liters of YP plus raffinose at 24^{o}C to an optical density (OD_{600}) of 0.5. $Sic1^{HAHis6}$ synthesis was induced by the addition of

galactose to 2%, and after three hours, the culture was shifted to 37 C for three hours. After harvesting (as described above), frozen cells were ground to powder in liquid nitrogen.[23] The cell powder (40 g) was thawed in five volumes of denaturing lysis buffer (DLB: 100 mM sodium phosphate, 10 mM Tris pH 8.0, and 6 M guanidine hydrochloride), and the resulting slurry was stirred for 30 min at 24°C, and then centrifuged at 26,000g for 15 min in a Sorvall SS34 rotor. The supernatant was mixed with 2 ml of Ni-NTA resin (Qiagen) for 40 min at 24 C, after which the beads were washed twice with 20 ml DLB, twice with 100 mM sodium phosphate, pH 5.9, 10 mM imidazole, 2 M urea, and twice with 25 mM Tris, pH 8.0, 500 mM NaCl, and 0.2% Triton X-100. Bound proteins were eluted with 6 ml of buffer containing 250 mM imidazole, 50 mM Tris, pH 8.0, and 250 mM NaCl. The eluate was supplemented with NaCl (500 mM final), Triton X-100 (0.2%), and the protease and phosphatase inhibitor cocktail described above, and incubated with 0.5 ml 12CA5 resin (anti-HA antibody covalently crosslinked to Protein A beads) for 45 min at 4 C. The beads were collected by centrifugation and washed twice with binding buffer and three times with 10 mM Tris, pH 6.8. $Sic1^{HAHis6}$ (Fig. 2A) was eluted with 1.5 ml 0.1% trifluroacetic acid and processed for electrospray mass spectrometry as described in Sections 4 and 5. To confirm that the functionally-relevant phosphorylation state of Sic1 was preserved throughout isolation, we

demonstrated that the purified material was competent to serve as a substrate for ubiquitination by the SCF^{Cdc4}/Cdc34 pathway (Fig. 2B).

b. Net1

Net1 is a subunit of the nucleolar RENT complex. The disassembly of the RENT complex in late anaphase of the cell cycle culminates in the release of the protein phosphatase Cdc14 from the nucleolus, which then triggers the exit from mitosis.[24-26] The disassembly mechanism is unknown, but we hypothesize that it involves phosphorylation of Net1, because Net1 accumulates in a highly phosphorylated form in a *cdc14*^{ts} mutant that is unable to exit mitosis.[24] Moreover, the release of Cdc14 from Net1 requires the action of three protein kinases (Cdc5, Cdc15, and Dbf2), suggesting that protein phosphorylation may play a direct role in the release mechanism.

To address whether phosphorylation of Net1 regulates disassembly of the RENT complex, we sought to determine the phosphorylation state of Net1 in a *cdc14*^{ts} mutant. We specifically selected the *cdc14*^{ts} mutant because the $Cdc14^{ts}$ protein is released from Net1 in late mitosis with normal kinetics in *cdc14*^{ts} cells,[27] but since the phosphatase is inactive, the cells remain arrested in late mitosis with dispersed Cdc14. As a control, we sought to determine sites of Net1 phosphorylation in *cdc14*^{ts} *cdc5*^{ts} cells, because preliminary evidence suggested that Cdc5 is required for the

disassembly of the RENT complex. Thus, a comparison of the sites at which Net1 is phosphorylated in these two mutant strains might highlight amino acids whose phosphorylation correlates with release of Cdc14 from Net1.

Detailed Procedure: To isolate Net1 in a manner that preserved its native phosphorylation state, we pursued two alternative approaches. We first tried non-denaturing lysis by grinding frozen cells in liquid nitrogen, and thawing the cell powder in lysis buffer comprised of 20 mM Hepes, pH 7.2, 0.5 M NaCl, 2 mM DTT, 0.5% Triton X-100, protease inhibitors, and phosphatase inhibitors. Unlike the SDS-boiling method,[24] in which cells were boiled in SDS sample buffer to achieve fast inactivation of all proteins prior to lysis (Fig. 3A, Lanes 1 and 2), the native lysis method failed to preserve the phosphorylation state of Net1 isolated from *cdc14* cells (Fig. 3A, compare Lane 4 with Lane 2).

Because Net1 was rapidly dephosphorylated in native cell extract, we sought to purify it from denatured cell extract. However, the strategy described above for Sic1 was deemed unsuitable, because we considered it unlikely that the large Net1 protein would efficiently refold upon removal of denaturant. Instead, we decided to isolate Net1 from SDS-denatured cells. A beaker containing 180 ml of H_2O was brought to boiling. Cell pellets (cells were harvested as rapidly as possible; see prior section) from

a 4.5-liter culture that was shifted to 37°C for 3h (final OD_{600}=1) were resuspended in ~100 ml ice-cold H_2O, and added to the boiling water bath in 20-ml aliquots with each addition being initiated when the water bath reached boiling. This ensured that the temperature of the bath never dropped below 90°C. After all cells were processed, the bath was boiled for 3 more min, and allowed to cool. The boiled cells were then harvested by centrifugation for 10 min at 5000 rpm in a clinical centrifuge. Cell pellet aliquots were frozen in liquid N_2, and stored at –80°C.

One pellet (~5 ml, from ~1,100 OD_{600}) was thawed on ice, and resuspended in 1.5 volume of lysis buffer (100 mM Tris, pH 7.5, 200 mM NaCl, 10 mM DTT, 2% SDS, plus protease and phosphatase inhibitors), and boiled immediately for 3 min. Aliquots (320 µl) of cell suspension were distributed in 2 ml flat-bottom microtubes (USA Scientific, 1420-2700), acid-washed 0.5 mm glass beads (200 µl per tube of Sigma G8772) were added, and tubes were vortexed at 4°C in a multi-bead vortexer (6 pulses of 3 min each, with 3 min intermissions). When glass bead lysis was carried out in a single large centrifuge tube, we found the lysis to be very inefficient. After vortexing, all tubes were boiled for 2 min to ensure that SDS had completely dissolved all extractable proteins.

The following steps (until after the addition of IP buffer) were carried out at room temperature to prevent precipitation of SDS. Supernatants were pooled into 1.7 ml microcentrifuge tubes, centrifuged at

14,000 rpm for 10 min, transferred to a 50-ml tube, and sonicated with a microtip at room temperature (Branson Sonifier 450, 6 cycles of 5 pulses per cycle at #5 output control and 50% duty cycle). Sonication appeared to reduce high molecular weight contaminants, possibly by shearing DNA into small fragments. Sonicated extracts were supplemented with 3 volumes of IP buffer (50 mM Tris, pH 7.5, 1% Triton X-100, 0.5 M NaCl, and protease inhibitors) ("Extracts" in Fig. 3B and C). Triton forms micelles that adsorb SDS not bound to proteins, thereby eliminating free SDS and creating an environment conducive for antibody-antigen interaction. Pansorbin® cells (100 µl; Calbiochem 507858) were incubated with extracts for 20 min to pre-absorb proteins that bound nonspecifically to protein A, and samples were then centrifuged at 160,000 g (40,000 rpm in Beckman Ti60 rotor) for 20 min at 12°C. The supernatant (~30 ml) ("Pre-IP" in Fig. 3B and C) was transferred to two 15-ml tubes, with each tube containing 0.32 ml of 9E10 resin (anti-myc antibody cross-linked to protein A beads). The antibody beads were incubated with extracts for 1 hour at 4°C, sedimented, resuspended in ice-cold wash buffer (20 mM Hepes, pH 7.2, 0.5% Triton X-100, 0.5 M NaCl, 1 mM DTT), and distributed to seven 0.6ml tubes. An aliquot of post-IP extracts ("Post-IP" in Fig. 3C) was set aside. The beads were washed seven times with 0.5 ml wash buffer/tube, transferred to fresh tubes (to avoid contaminants adsorbed to the side of the tube), and washed three

more times ("Beads" in Fig. 3B). The beads were then washed three times with 2 mM Tris, pH 8.8, 1 mM DTT, and eluted in 2 mM Tris, pH 8.8, 0.5 mM DTT, 0.1% SDS at 100°C for 3 min. The supernatant was collected, and elution was repeated. The eluates were pooled ("Eluate" in Fig. 3C), lyophilized, resuspended in 30 µl of SDS sample buffer, fractionated on a 7.5% SDS-polyacrymide gel, stained by Coomassie Brilliant Blue (0.1% Coomassie R-250, 20% MeOH, 0.5% acetic acid in water), and destained in 30% MeOH in water (Fig. 3D). Samples that were set aside throughout the preparation were fractionated by SDS-PAGE and evaluated by immunoblotting with anti-myc antibody. As shown in Fig. 3C, ~70% of Net1-myc was recovered in the immunoprecipitation step (compare Lanes 4,5 with 6,7). In the first preparation (Fig. 3B), the phosphorylation state of Net1 was apparently preserved throughout isolation. In the second preparation, Net1 from *cdc5 cdc14* was upshifted with respect to that from wild type cells, suggesting that the basal phosphorylation of Net1 was also preserved in *cdc5 cdc14* (compare Fig. 3C, Lanes 3 and 4). However, ~50% of the Net1 isolated from *cdc14* cells collapsed into a lower molecular weight species after the elution step (Fig. 3C, Lane 11), presumably as a result of dephosporylation. Note that the relative phosphoshifts are difficult to see in the Coomassie-stained gel (Fig. 3D) due to overloading. The source of variability in our preparations is unknown, but smaller scale preparations that require less time may better

preserve phosphorylation states. This procedure generated ~ 3 μg (~20 pmol) of Net1-Myc9 (Fig. 3D). Mass spectrometric analysis of the Net1 samples excised from the SDS-polyacrylamide gel revealed up to 20 sites of phosphorylation, which will be reported elsewhere.

4. Multi-Dimensional Electrospray Mass Spectrometry Based Phosphopeptide Mapping Method

Before any effort is made to determine the site of phosphorylation on a protein, the modification is usually first localized to a peptide from an enzymatic digest. A major challenge in phosphopeptide mapping is to isolate or identify the phosphorylated peptides from the usually overwhelming amount of nonphosphorylated peptides present in the digest. Although the phosphopeptide need not be purified prior to sequencing by MS/MS, it is necessary to identify which peptide in the mixture to sequence.

Earlier we reported on a method utilizing nanoelectrospray MS and a technique called precursor ion scanning, which takes advantage of the fact that in the mass spectrometer, under experimentally controllable conditions, phosphopeptides undergo facile loss of phosphate from serine, threonine and tyrosine residues to produce a PO_3^- ion with m/z 79.[28] During a precursor scan the mass spectrometer is set to detect only those

peptides which fragment to yield the diagnostic m/z 79 ion, thus allowing the selective detection of phosphopeptides. This high degree of selectivity allows the precursor scan to identify phosphopeptides which are present as very minor components of the sample. Once the phosphopeptides have been identified, the MS can be switched to the positive ion mode and the peptides sequenced by tandem MS.

Although this approach can successfully identify and sequence phosphopeptides from unfractionated mixtures, there are limitations. First, as the phosphorylation state of the protein becomes more complex, the least abundant phosphopeptides may go undetected. Second, large phosphoproteins can yield an overwhelming number of unmodified peptides that make it difficult to sequence the phosphopeptides even after they have been identified by the precursor ion scan. This difficulty stems from the unavoidable problems of wide dynamic range of peptide abundance, ion suppression effects, charge state overlap, and the possible need to sequence many phosphopeptides in a single sample. Recent studies of highly phosphorylated proteins illustrate the advantages of employing a separation step to fractionate the sample or enrich the phosphopeptide pool prior to site specific analysis.

In 1993, we reported a Liquid Chromatography (LC)-MS based method that used phosphopeptide-specific marker ions to selectively detect phosphopeptides in complex mixtures.[29] As components of a

proteolytic digest elute from an HPLC column into the MS, they are subjected to collision-induced dissociation conditions in the ion source. Peptides that contain phosphorylated residues fragment to produce highly diagnostic $[PO_2]^-$ and $[PO_3]^-$ marker ions at m/z 63 and m/z 79, respectively. Monitoring for these two marker ions permits selective collection of LC fractions which contain phosphopeptides (as well as co-eluting non-phosphorylated peptides). The marker ion elution 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.

Combining these two orthogonal MS techniques – both of which selectively detect phosphopeptide-specific marker ions – with MS-based peptide sequencing yielded a multi-dimensional analytical method which is highly selective for phosphopeptides.[13] The overall strategy is outlined in Fig. 4. In the first step, the proteolytic digest of a protein is subjected to Reverse Phase (RP)-HPLC and collected into fractions. On-line monitoring of the phosphopeptide specific marker ion m/z 79 by ESMS in the negative ion mode identifies those fractions that contain phosphopeptides (those marked with *). This HPLC step greatly simplifies the complexity of phosphopeptide 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. Finally the phosphopeptides are sequenced by positive-ion nanoelectrospray tandem MS. This strategy, which we have developed and refined over the last few years, is particularly well suited to phosphoproteins that are phosphorylated with widely varying stoichiometry at many sites.

5. Analysis of Sic1 and Net1 Phosphorylation by Mass Spectrometry

a. Sic1

Sic1 was of sufficient purity that it could be analyzed directly from the 0.1% TFA elution from the anti-HA beads, without the need for further purification by SDS-PAGE. Approximately 70% of the purified Sic1 solution was lyophilized to remove the acid. The sample was then reconstituted in 20 µl of 8 M urea and vortexed. The urea concentration was reduced to 2 M by addition of 80 µl of 50 mM NH_4HC0_3 pH 8.5. Sic1 proved resistant to typical in-solution tryptic digestion conditions, finally succumbing to a 16-hr incubation at an enzyme:substrate ratio of 1:7. Sic1 contains no cysteines residues, thus reduction and alkylation were not performed. Half of the digest was loaded onto a 1 mm C18 trap cartridge, installed in place of a sample loop on the HPLC injector. After washing with 200 µl of 0.1% TFA-5% acetonitrile, the peptides were

backflushed off the trap cartridge onto a 0.5 mm x 15 cm C18 Reliasil HPLC column with an acetonitrile:water:TFA gradient run at 20 µl/min. The column eluate was split 4:1 after the UV detector with 4-5 µl/min going to the mass spectrometer, and the remainder going to a fraction collector taking half-minute fractions.

Electrospray mass spectra were acquired on a modified Perkin-Elmer Sciex API-III atmospheric pressure ionization triple quadrupole tandem mass spectrometer using the standard articulated, pneumatically-assisted nebulization electrospray probe. 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. The yield of the m/z 63 and 79 product ions was maximized by application of -350 V (m/z 63) and –300 V (m/z 79) to the orifice of the API III as has been described.[29,30] The MS is operated in a single ion monitoring mode for enhanced sensitivity.

The summed LC-ESMS ion trace for m/z 63 and 79 from the Sic1 sample showed more than 12 phosphopeptide-containing peaks which were collected in 30 fractions. To simplify the subsequent analysis we pooled adjacent fractions from clusters of phosphopeptide peaks and identified the phosphopeptides in each pool using a precursor scan for m/z 79. Precursor scans were acquired in the negative ion mode on the same

PE Sciex API III as was described previously, but using a nanoES source designed and built at the EMBL.[31] To enhance detection of the phosphopeptides, an aliquot of each fraction was made basic prior to precursor ion analysis as follows: 1/4 – 1/2 of a desired fraction was lyophilized and reconstituted in 50/50 methanol:water containing 5% concentrated ammonium hydroxide (30% by weight) and 1-2 µl of the sample was introduced into the nanoES needle.[28] After the molecular weights of the phosphopeptides were determined from the m/z 79 precursor ion scan, the mass spectrometer was switched into the positive ion mode. When sufficient signal was present, the phosphopeptides were sequenced from the same sample loading by acquiring a full scan CID product ion spectrum (MS/MS) on a chosen multiply-charged precursor. In some cases, an acidic aliquot of the sample in 50:50 MeOH:water with 5% formic acid was used for MS/MS sequencing of the phosphopeptides.

Precursor ion scan analysis and tandem MS of the pooled fractions revealed that Sic1 phosphorylation was localized on three tryptic peptides sequences: amino acids 2-8, 14-50, and 54-84. The 2-8 sequence appeared to be quantitatively mono-phosphorylated on Thr^{5}, while the latter two peptides were differentially phosphorylated (0-2 and 0-3 phosphates, respectively), with the lowest phosphorylation state being the most abundant. We sequenced the monophosphorylated form of 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. Tandem mass spectrometry of the 54-84 peptide failed 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.

Since the purification protocol described above yielded purified Sic1 in solution, we were able to use nanoESMS to determine the overall phosphorylation state of the intact protein.[13] To facilitate the nanoES, we passed the remaining protein (approximately 30% of the original material) over a 0.5mm C18 HPLC column and collected the Sic1 peak. The solvent was removed by lyophilization and the protein reconstituted in 4 µl of 50:50 MeOH:water with 5% formic acid and loaded into the nanoES needle. The distribution of phosphorylation on Sic1 corresponds to between one and six moles of phosphate per mole of Sic1, with quadruply- and quintuply-phosphorylated Sic1 being the predominant forms,[13] and suggests that Sic1 is phosphorylated by Cln/Cdc28 in a distributive manner *in vivo*.

The physiological significance of the three major *in vivo* phosphorylation sites determined by mass spectrometry was confirmed by analysis of a mutant in which Thr^{5}, Thr^{33} and Ser^{76} were substituted with Ala.[15] When this triple mutant was analyzed by nanoES, the mass spectrum showed that the most abundant phosphorylation state decreased from 4-5 of moles of phosphate (observed on wild type Sic1) to 2 moles of

phosphate.[13] This data suggests that these three sites are indeed major phosphoacceptor sites *in vivo.*

Whenever it is possible to generate sufficient amounts of purified protein (10-50 pmol), we advocate making a molecular weight measurement on the intact protein. 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. As was demonstrated here for Sic1, this information can be useful for interpreting the phosphopeptide mapping and mutagenesis data.

b. Net1

Proteins derived by SDS-PAGE require special preparation protocols to ready the sample for MS analysis. Elution of proteins from acrylamide gels is inefficient at best, and a hopeless proposition for a large protein like Net1. The most straightforward approach to preparing gel-fractionated proteins for MS analysis is direct digestion of the protein in the gel.[32] The Coomassie-stained Net1 band (Fig. 3D, Lane 2) was excised from the gel with a razor blade and the gel piece washed in a 1.5-ml eppendorf tube with several changes of 50:50 acetonitrile:50 mM NH_4HCO_3 until the Coomassie Blue was completely extracted. The wash solvent was discarded and replaced with 50 µl of fresh 50 mM NH_4HCO_3

pH 8.5 and 5 µl of 45 mM DTT. The sample was heated for 1 h at 38°C. After cooling to room temperature, 5 µl of 100 mM iodoacetamide was added and the sample was allowed to react for 1 h at RT in the dark. After 1 h, the gel piece was washed for 1 h with 50:50 acetonitrile:50 mM NH_4HCO_3 and then covered with acetonitrile and allowed to shrink until it turned white. The acetonitrile was removed and the gel piece dried in a speedvac. The gel piece was rehydrated with 15 µl of 50 mM NH_4HCO_3 containing 500 ng of modified trypsin. After the gel piece was completely reswollen, an additional 50 µl of 50 mM NH_4HCO_3 was added to cover the slice and the sample was incubated overnight at 38°C. After the overnight digestion, the NH_4HCO_3 solution was removed and transferred to a fresh 200 µl eppendorf tube. The gel piece was washed once with an additional 50 µl of fresh 50 mM NH_4HCO_3 and once with 50 µl of 5% formic acid. Both washes were combined with the original NH_4HCO_3 solution and 25-50% of the sample was injected onto a C18 trap cartridge which was installed in place of the sample loop on an HPLC injector as described above. Following a wash of the trap with 200 µl 0.1% TFA, the Net1 digest was analyzed using the multidimensional MS-based phosphopeptide mapping strategy described above.

The much smaller amount of protein typically available from a gel-derived sample necessitated some changes in the analysis. The sensitivity

of the method as described for Sic1 is limited by the HPLC column diameter. Because ESMS (at flow rates above ca. 50 nL/min) acts like a concentration sensitive detector, smaller internal diameter HPLC columns will provide higher sensitivity analyses. However the requirement to split the HPLC flow post-column to allow collection of fractions imposes a practical limitation on the HPLC column diameter which can be used. By switching the MS to a micro ionspray source flowing at 0.2- 0.5 µl/min, we were able to use a 180 µm i.d. HPLC column flowing at rates of 3-4 µl/min. By setting up an approximately 10:1 post column split, we were able to collect 3.5 µl/min fractions. Preliminary data obtained using this set up indicates that sensitivity in the first dimension analysis is 20-50 fold higher than the 0.5 mm i.d. column used for the Sic1 analysis (Zappacosta, unpublished results). The smaller HPLC column diameter also required us to use a smaller diameter C18 trap cartridge (300 µm i.d.).

The m/z 79 trace for Net1 from *cdc14*^{ts} *cdc5*^{ts} cells is shown in Fig. 5A. The Net1 digest was spiked with 500 fmol of a phosphopeptide standard (marked IS) prior to analysis. The peak heights of the Net1 phosphopeptides relative to the internal standard allows an estimate of the overall level of phosphorylation present in the sample. Ten fractions (collected by hand) labeled 3-12 in Fig. 5A were analyzed using precursor ion scans and tandem MS as described above. No changes were required in this part of the analysis to accommodate the lower sample amounts.

The negative ion MS and the precursor scan for m/z 79 for fraction 6 are shown in Fig. 5B and C respectively. The precursor ion scan shows five phosphopeptides present in this fraction, whereas none of these peptides are detectable in the negative ion scan. The sequences which have been assigned to these molecular weights were verified by tandem MS and in each case the specific phosphorylated residue was established. In all, 20 unique phosphorylation sites were determined from 22 different phosphopeptides.

Figure Legends:

FIG. 1. The significance of protein phosphorylation can be tested in two complementary experiments.

(A) Hypothesis: phosphorylation of substrate Y by protein kinase X leads to process Z, where kinase X also phosphorylates additional substrates such as a and b. (B, C) Either a loss-of-function mutation in kinase X (B) or a non-phosphorylatable mutation in substrate Y (C) should block process Z (blockage is indicated by double lines).

FIG. 2. Purification of phospho-Sic1 from yeast cells.

(A) Analysis of purity of the Sic1 preparation. Aliquots of eluates from Ni-NTA and 12CA5/Protein A resins were resolved by SDS-PAGE and visualized by Coomassie Blue staining. (B) Evaluation of the ability of purified phospho-Sic1 to serve as a substrate of the ubiquitin pathway. An aliquot of purified phospho-Sic1 was incubated with ATP, E1, and E2 (Cdc34), in the absence or presence of SCF^{Cdc4} ubiquitin ligase activity (supplied by yeast extract, see Ref. 21). Reaction products were visualized by immunoblotting with anti-Sic1.

FIG. 3. Purification of phospho-Net1 from yeast cells.

(A-C) Net1-Myc9 purifications from wild type (+), *cdc5 cdc14* (*5,14*), or *cdc14* (*14*) strains were analyzed by SDS-PAGE followed by Western blot

using 9E10 antibodies (against the Myc epitope). (A) The degree of phosphorylation of Net1 (as judged by its reduced mobility upon SDS-PAGE) was preserved less well in native lysis than in SDS-boil lysis method. (B, C) Two independent samples of phospho-Net1 prepared from yeast cells using denaturing lysis method (see text for details). Wild type extracts (+) in (C) were prepared by the SDS-boiling method. Net1 phosphorylation was the most extensive in *cdc14* and the least extensive in wild type cells (compare Lanes 3-5 in Panel C). (D) Eluate of phospho-Net1 from (C) was concentrated, fractionated on SDS-PAGE, and evaluated by Coomassie Blue staining.

FIG. 4. Schematic diagram of the multidimensional electrospray-mass spectrometry method for mapping phosphorylation sites in proteins.

FIG. 5. Multidimensional electrospray-MS phosphopeptide mapping of Net1.

A Coomassie blue stained band was excised from the gel (Fig. 3D, Lane 2) and digested *in situ* with trypsin. 500 fmol of phosphopeptide internal standard were added to the digest prior to injection onto the HPLC. (A) First dimension phosphopeptide specific LC-ESMS trace (m/z 79). The phosphopeptide internal standard peak is marked as “IS”. Twelve fractions were taken for further analysis and 100 fmol of internal standard

was added to each fraction. (B) NanoES (-) ion full scan MS spectrum of fraction 6 from the LC-ESMS run in (A). (C) NanoES (-) ion precursor scan for m/z 79 of fraction 6. Five phosphopeptides with listed molecular weights were detected.

References:

[1] H. Zhu, J.F. Klemic, S. Chang, P. Bertone, A. Casamayor, K.G. Klemic, D. Smith, M. Gerstein, M.A. Reed, and M. Snyder, *Nat Genet* **26**, 283 (2000).

[2] Y. Chi, M.J. Huddleston, X. Zhang, R.A. Young, R.S. Annan, S.A. Carr, and R.J. Deshaies, *Genes Dev* **15**, 1078 (2001).

[3] K.L. Gould and K.L. Nurse, *Nature* **342**, 39 (1989).

[4] W.J. Boyle, P. van der Geer, and T. Hunter, *Methods Enzymol* **201**, 110 (1991).

[5] T. Moll, G. Tebb, U. Surana, H. Robitsch, and K. Nasmyth, *Cell* **66**, 743 (1991).

[6] J.M. Sidorova, G.E. Mikesell, and L.L. Breeden, *Mol Biol Cell* **6**, 1641 (1995).

[7] W.J. Boyle, T. Smeal, L.H. Defize, P. Angel, J.R. Woodgett, M. Karin, and T. Hunter, *Cell* **64**, 573 (1991).

[8] P. Cohen, B.W. Gibson, and C.F. Holmes, *Methods Enzymol* **201**, 153 (1991).

[9] J.D. Watts, M. Affolter, D.L. Krebs, R.L. Wange, L.E. Samelson, and R. Aebersold, *J. Biol. Chem.* **269**, 29520 (1994).

[10] X. Zhang, C.J. Herring, P.R. Romano, J. Szczepanowska, H. Brzeska, A.G. Hinnebusch, and J. Qin, *Anal. Chem*. **70**, 2050 (1998).

[11] G. Neubauer and M. Mann, *Anal. Chem.* **71**, 235 (1999).

[12] M.C. Posewitz and P. Tempst, *Anal. Chem.* **71**, 2883 (1999).

[13] R.S. Annan, M.J. Huddleston, R. Verma, R.J. Deshaies, and S.A. Carr, *Anal Chem* **73**, 393 (2001).

[14] G. Alexandru, F. Uhlmann, K. Mechtler, M-A. Poupart, and K. Nasmyth, *Cell* **105**, 459 (2001).

[15] R. Verma, R.S. Annan, M.J. Huddleston, S.A. Carr, G. Reynard, and R.J. Deshaies, *Science* **278**, 455 (1997).

[16] G. Rigaut, A. Shevchenko, B. Rutz, M. Wilm, M. Mann, and B. Seraphin, *Nat Biotechnol* **17**, 1030 (1999).

[17] D.J. Brown and J.A. Gordon, *J Biol Chem* **259**, 9580 (1984).

[18] E. Schwob, T. Bohm, M.D. Mendenhall, and K. Nasmyth, *Cell* **79**, 233 (1994).

[19] R. Verma, S. Chen, R. Feldman, D. Schieltz, J. Yates, J. Dohmen, and R.J. Deshaies, *Mol Biol Cell* **11**, 3425 (2000).

[20] B.L. Schneider, Q.H. Yang, and A.B. Futcher, *Science* **272**, 560 (1996).

[21] R. Verma, R.M. Feldman, and R.J. Deshaies, *Mol Biol Cell* **8**, 1427 (1997).

[22] M. Tyers, *Proc Natl Acad Sci U S A* **93**, 7772 (1996).

[23] R. Verma, Y. Chi, and R.J. Deshaies, *Methods Enzymol* **283**, 366 (1997).

[24] W. Shou, J.H. Seol, A. Shevchenko, C. Baskerville, D. Moazed, S.Z.W. Chen, J. Jang, A. Shevchenko, H. Charbonneau, and R.J. Deshaies, *Cell* **97**, 233 (1999).

[25] A.F. Straight, W. Shou, G.J. Dowd, C.W. Turck, R.J. Deshaies, A.D. Johnson, and D. Moazed, *Cell* **97**, 245 (1999).

[26] R. Visintin, E.S. Hwang, and A. Amon, *Nature* **398**, 818 (1999).

[27] S.L. Jaspersen and D.O. Morgan, *Curr Biol* **10**, 615 (2000).

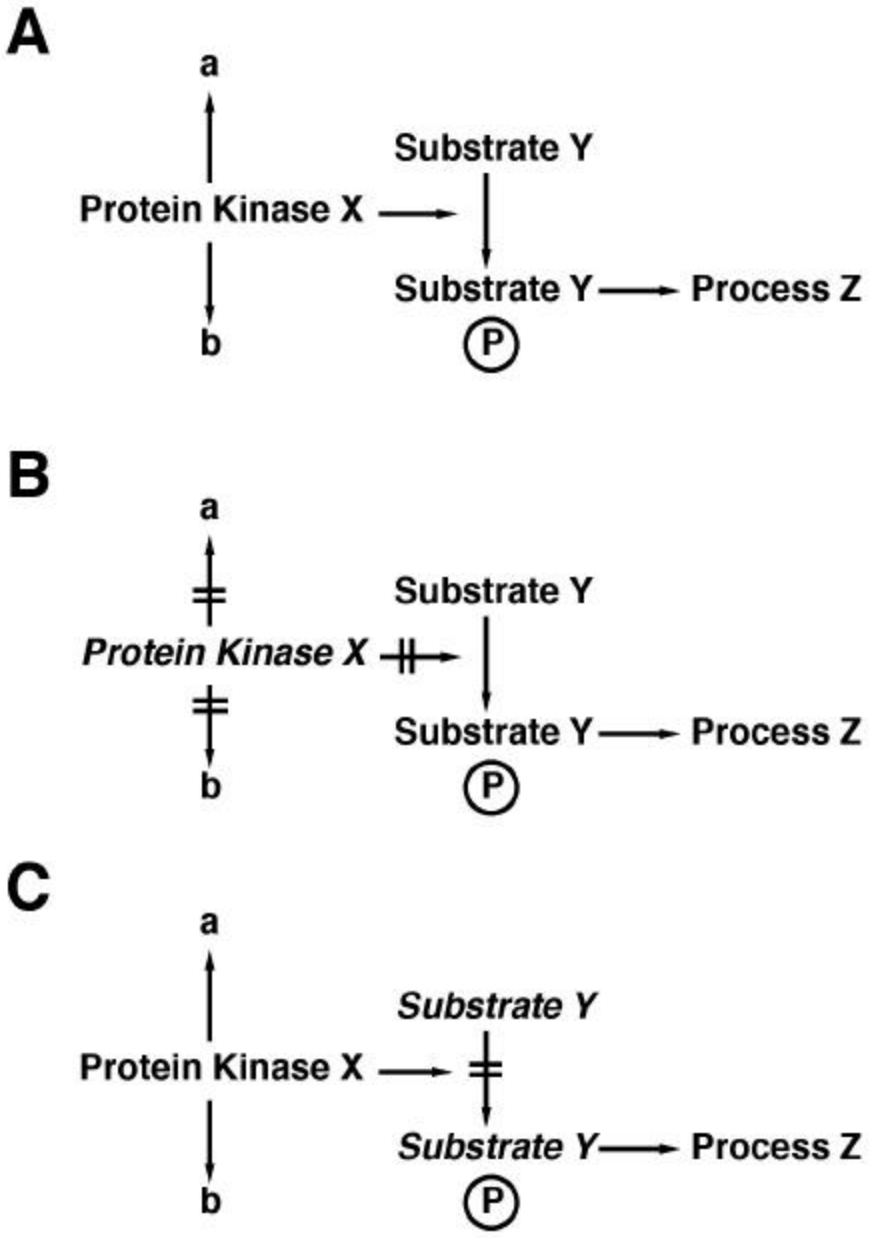
[28] S.A. Carr, M.J. Huddleston, and R.S. Annan, *Anal Biochem* **239**, 180 (1996).

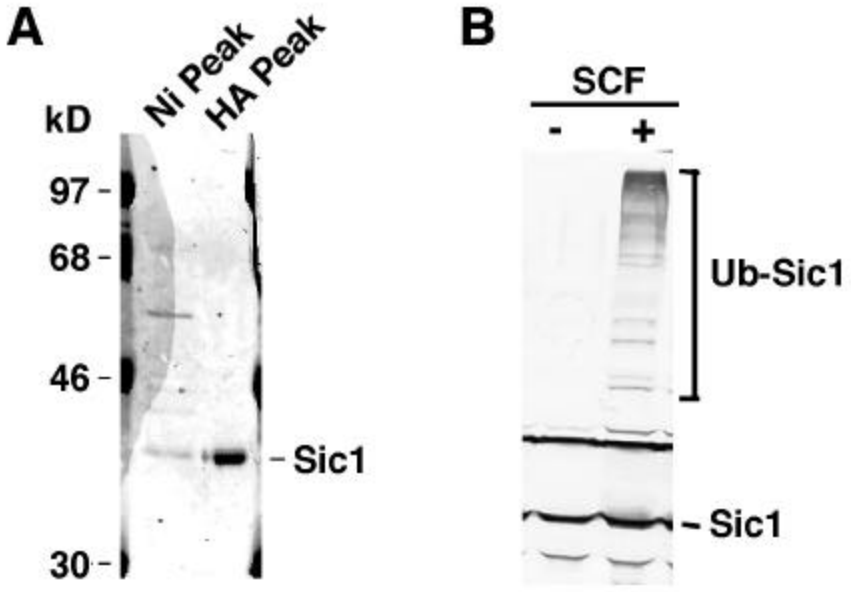
[29] M.J. Huddleston, R.S. Annan, M.F. Bean, and S.A. Carr, *J. Am. Soc. Mass Spectrom.* **4**: 710 (1993).

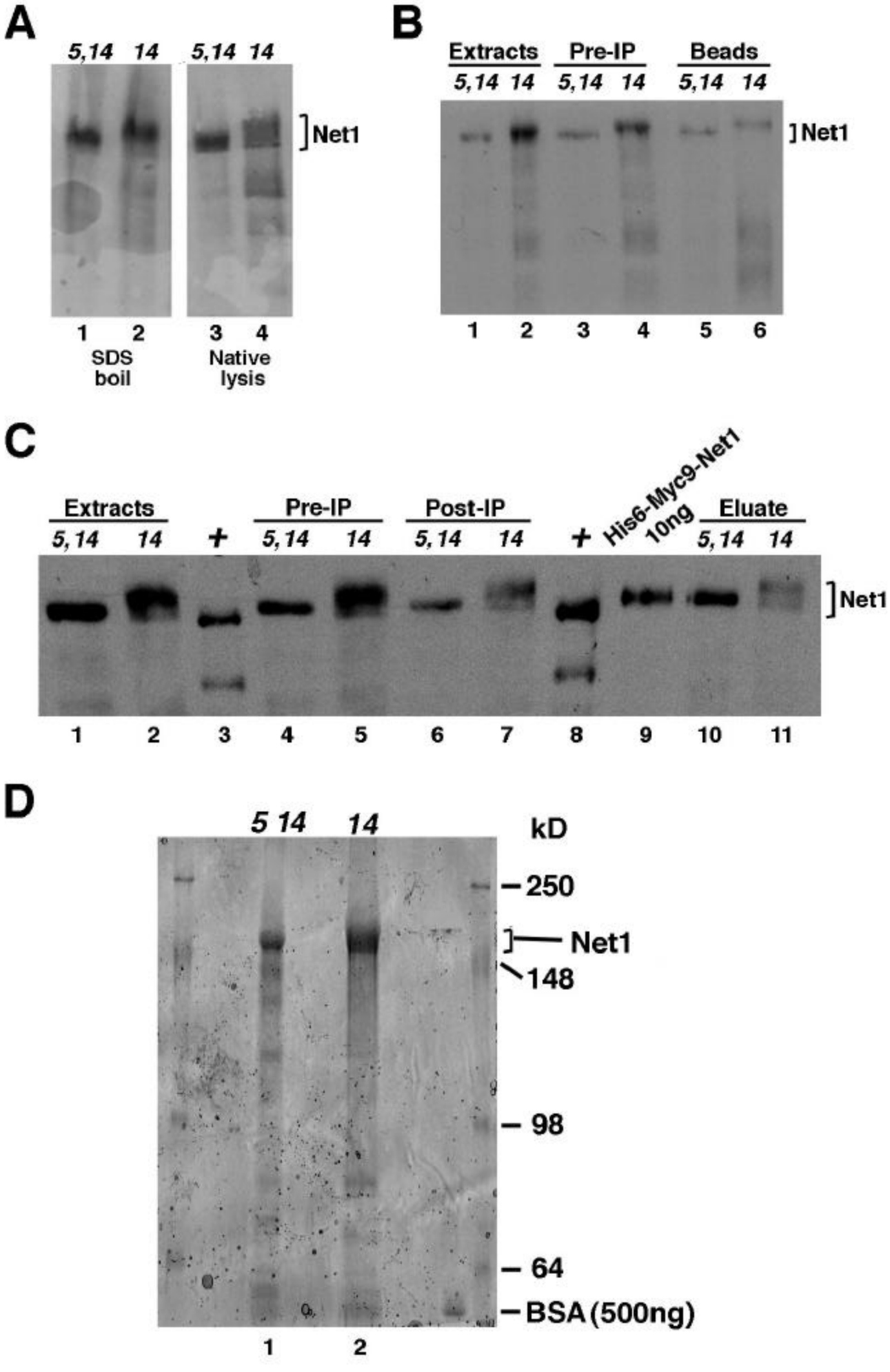
[30] S.A. Carr, R.S. Annan, and M.J. Huddleston, Methods Enzymol. in press (2002).

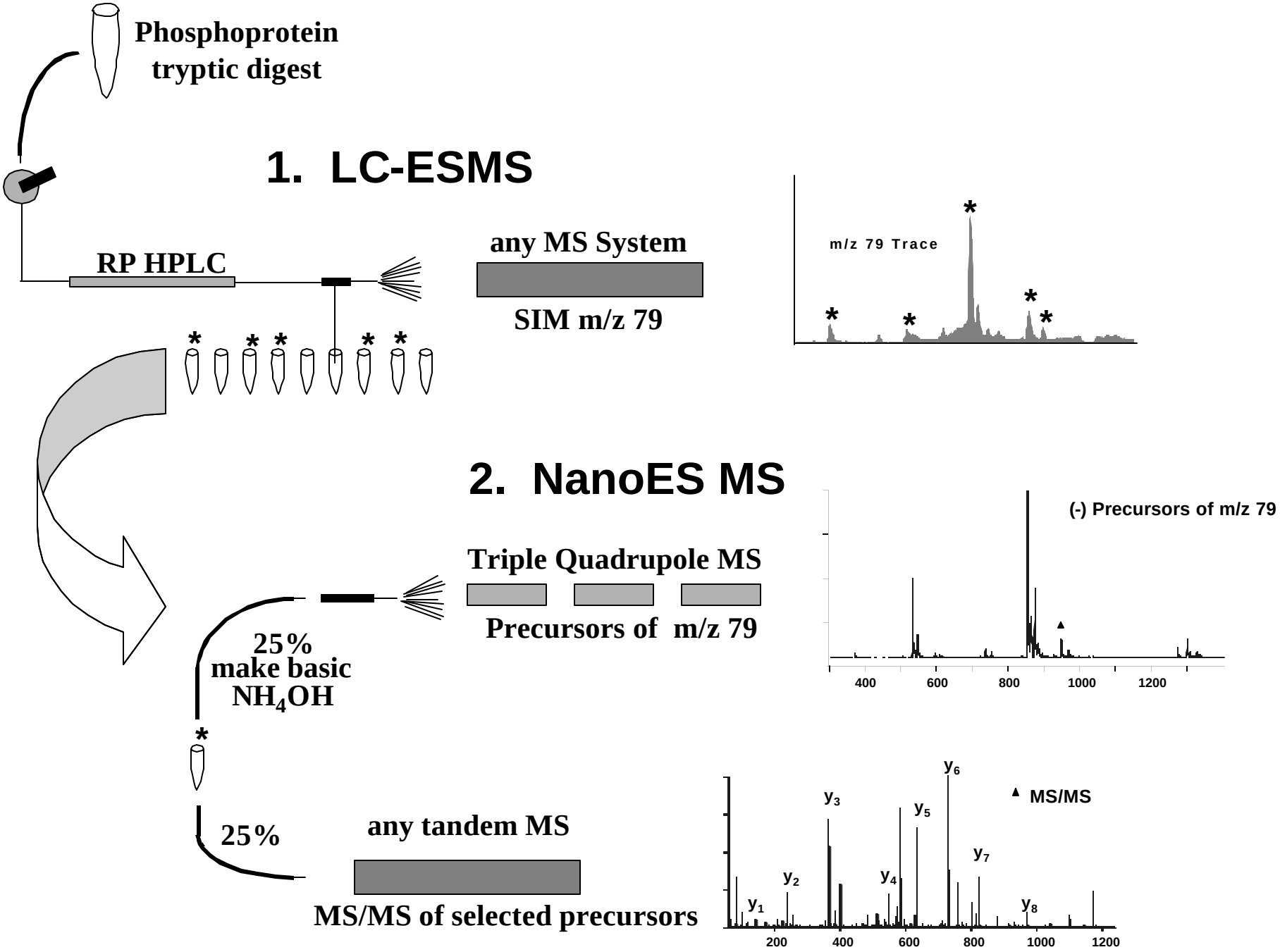
[31] M. Wilm and M. Mann, *Int. J. Mass Spectrom. and Ion Proc.***136**, 167-180 (1994).

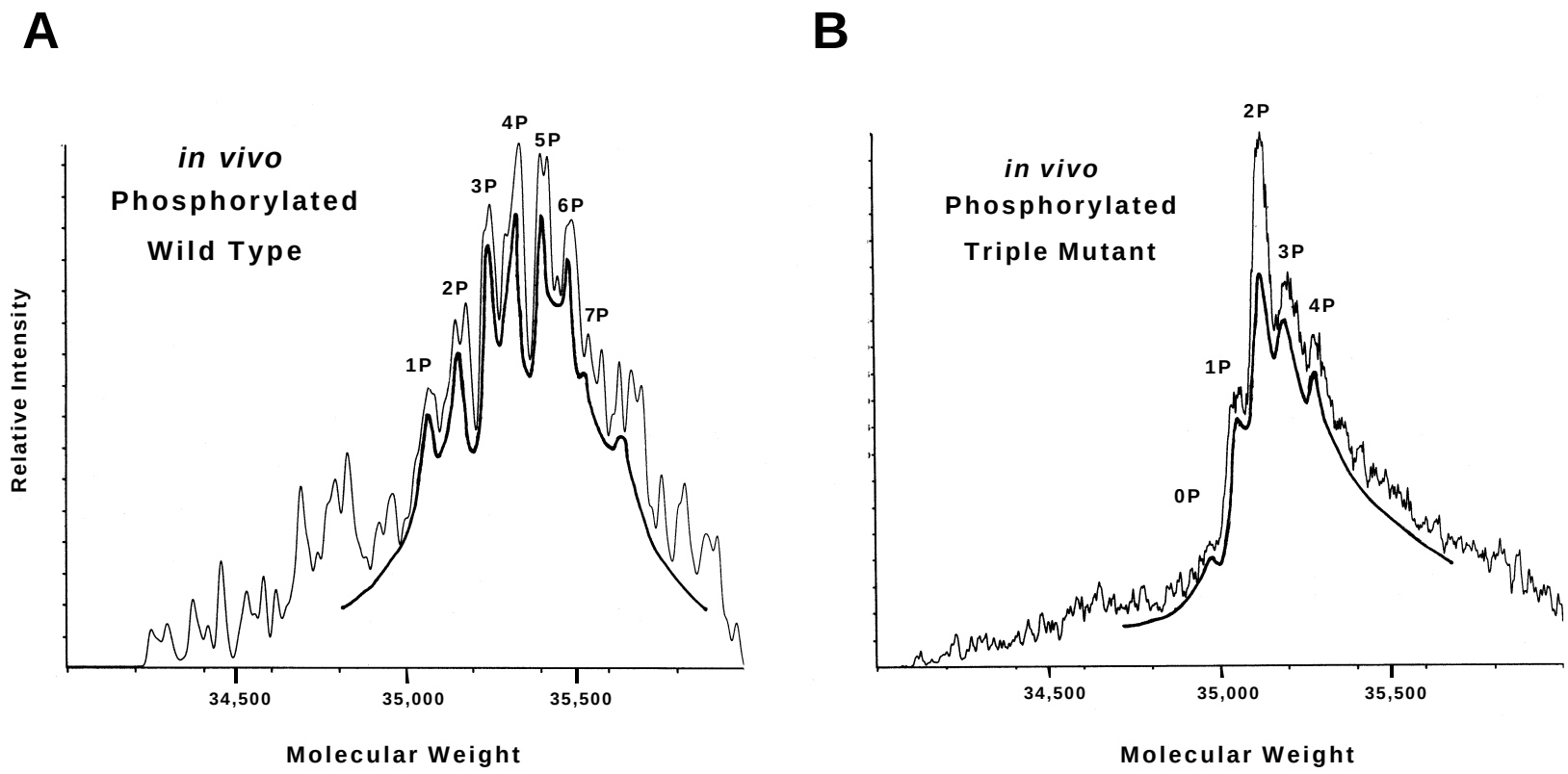
[32] J. Rosenfeld, J. Capdevielle, J. Guillemot, and P. Ferrara, *Anal. Biochem.* **203**, 173 (1992).

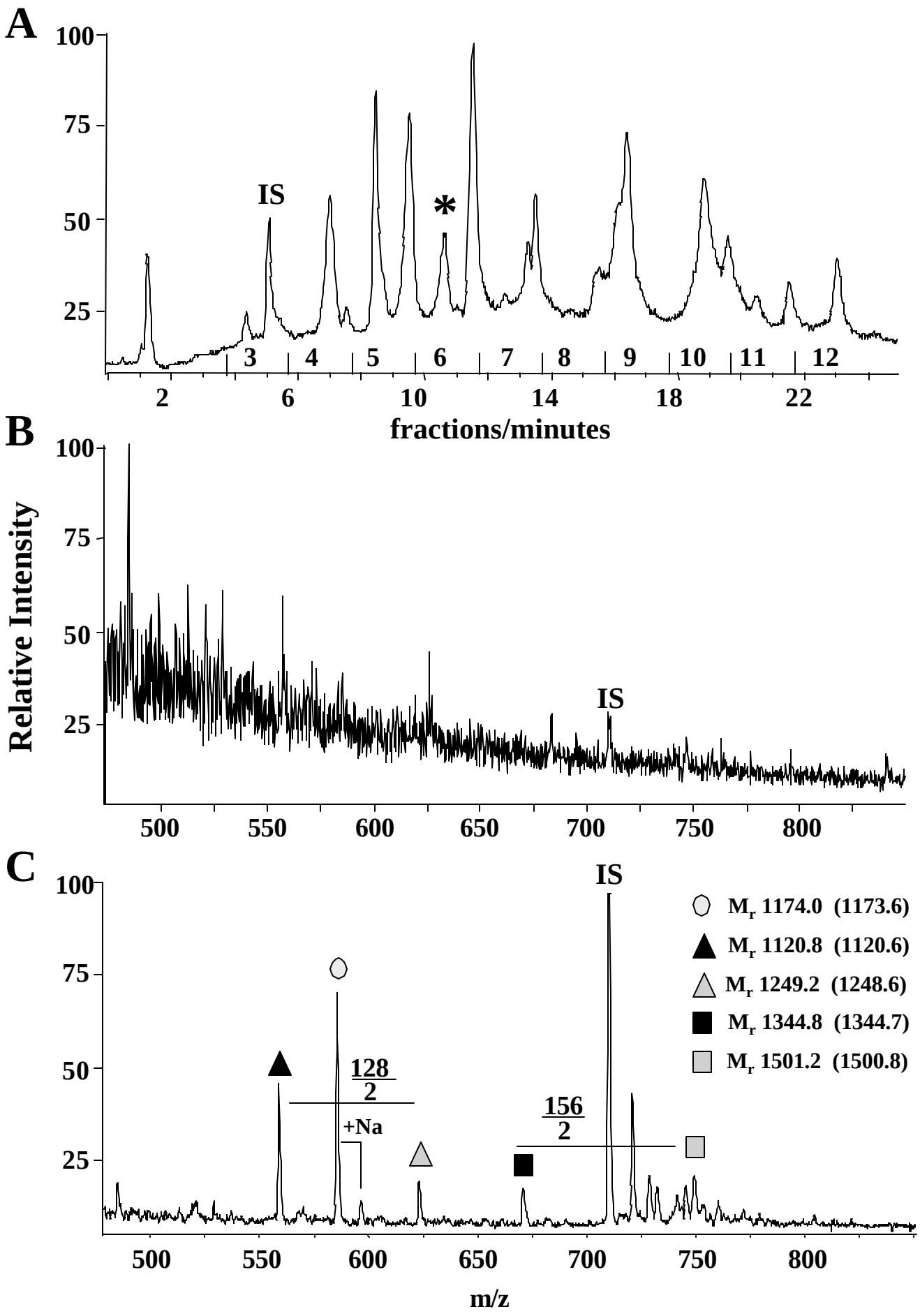

A
a
Substrate Y
Protein Kinase X
Substrate Y
Process Z
b
P
B
a
Substrate Y
Protein Kinase X
Substrate Y
Process Z
b
P
C
a
Substrate Y
Protein Kinase X
Substrate Y
Process Z
b
P


A
Ni Peak
HA Peak
kD
97
68
46
30
Sic1
B
SCF
-
+
Ub-Sic1
Sic1


A
5,14 14 5,14 14
]Net1
1 2 3 4
SDS boil
Native lysis
B
Extracts Pre-IP Beads
5,14 14 5,14 14 5,14 14
]Net1
1 2 3 4 5 6
C
Extracts Pre-IP Post-IP Eluate
5,14 14 + 5,14 14 5,14 14 + His6-Myc9-Net1 10ng 5,14 14
]Net1
1 2 3 4 5 6 7 8 9 10 11
D
5 14 14
kD
250
]— Net1
148
98
64
BSA (500ng)
1 2


Phosphoprotein
tryptic digest
1. LC-ESMS
RP HPLC
any MS System
SIM m/z 79
m/z 79 Trace
2. NanoES MS
Triple Quadrupole MS
Precursors of m/z 79
(-) Precursors of m/z 79
400
600
800
1000
1200
25%
make basic
NH4OH
25%
any tandem MS
MS/MS of selected precursors
MS/MS
y1
y2
y3
y4
y5
y6
y7
y8
200
400
600
800
1000
1200


A
in vivo
Phosphorylated
Wild Type
Relative Intensity
1P
2P
3P
4P
5P
6P
7P
34,500
35,000
35,500
Molecular Weight
B
in vivo
Phosphorylated
Triple Mutant
0P
1P
2P
3P
4P
34,500
35,000
35,500
Molecular Weight


A
100
75
50
25
IS
*
3
4
5
6
7
8
9
10
11
12
2
6
10
14
18
22
fractions/minutes
B
Relative Intensity
100
75
50
25
IS
500
550
600
650
700
750
800
C
IS
100
75
50
25
Mr 1174.0 (1173.6)
Mr 1120.8 (1120.6)
Mr 1249.2 (1248.6)
Mr 1344.8 (1344.7)
Mr 1501.2 (1500.8)
128/2
+Na
156/2
500
550
600
650
700
750
800
m/z