

Analysis of phosphorylation sites on proteins from *Saccharomyces cerevisiae* by electron transfer dissociation (ETD) mass spectrometry

An Chi^{*†}, Curtis Huttenhower[‡], Lewis Y. Geer[§], Joshua J. Coon^{*†¶}, John E. P. Syka^{||**}, Dina L. Bai^{*}, Jeffrey Shabanowitz^{*}, Daniel J. Burke^{††}, Olga G. Troyanskaya[‡], and Donald F. Hunt^{**§§}

^{*}Department of Chemistry and ^{||}Engineering Physics Program, University of Virginia, Charlottesville, VA 22904; [‡]Department of Computer Science and Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08540; [§]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894; ^{**}Thermo Fisher Scientific, San Jose, CA 95134; and Departments of ^{††}Biochemistry and Molecular Genetics and ^{†††}Pathology, Health Sciences Center, University of Virginia, Charlottesville, VA 22908

Edited by Fred W. McLafferty, Cornell University, Ithaca, NY, and approved December 6, 2006 (received for review August 15, 2006)

We present a strategy for the analysis of the yeast phosphoproteome that uses endo-Lys C as the proteolytic enzyme, immobilized metal affinity chromatography for phosphopeptide enrichment, a 90-min nanoflow-HPLC/electrospray-ionization MS/MS experiment for phosphopeptide fractionation and detection, gas phase ion/ion chemistry, electron transfer dissociation for peptide fragmentation, and the Open Mass Spectrometry Search Algorithm for phosphoprotein identification and assignment of phosphorylation sites. From a 30- μ g (\approx 600 pmol) sample of total yeast protein, we identify 1,252 phosphorylation sites on 629 proteins. Identified phosphoproteins have expression levels that range from <50 to 1,200,000 copies per cell and are encoded by genes involved in a wide variety of cellular processes. We identify a consensus site that likely represents a motif for one or more uncharacterized kinases and show that yeast kinases, themselves, contain a disproportionately large number of phosphorylation sites. Detection of a pHis containing peptide from the yeast protein, Cdc10, suggests an unexpected role for histidine phosphorylation in septin biology. From diverse functional genomics data, we show that phosphoproteins have a higher number of interactions than an average protein and interact with each other more than with a random protein. They are also likely to be conserved across large evolutionary distances.

yeast phosphoproteome | network analysis

In an earlier study of the yeast phosphoproteome (1), we digested proteins from a whole cell lysate with trypsin, used immobilized metal-affinity chromatography (IMAC) to enrich the sample for phosphopeptides, and analyzed the resulting mixture by nano-flow HPLC interfaced to electrospray ionization tandem mass spectrometry (MS/MS). Low-energy collision-activated dissociation (CAD) was used to fragment the peptide backbone and to produce ions of types b and y (Fig. 1) required for successful sequence analysis and identification of phosphorylation sites. We detected $>1,000$ phosphopeptides but defined only 383 sites of phosphorylation, largely because the CAD process often promoted elimination of phosphoric acid from Ser and Thr residues without breaking the amide bonds along the peptide backbone. The resulting MS/MS spectra were essentially devoid of sequence information.

To circumvent this problem, we have modified the LTQ mass spectrometer for ion/ion chemistry and now fragment both peptides (2) and intact proteins (3) by electron transfer dissociation (ETD). In this process, fluoranthene radical-anions are generated in a chemical ionization source and used as reagents to transfer an electron to a multiply charged peptide generated by electrospray ionization. This reaction is highly exothermic, reduces the peptide charge by one, and triggers fragmentation of the peptide backbone to produce a homologous series of complementary fragment ions of type c and z[•] (Fig. 1). Subtraction of m/z values for fragments within a given ion series that differ by a single amino acid affords the mass and thus the identity of the extra residue in the larger of

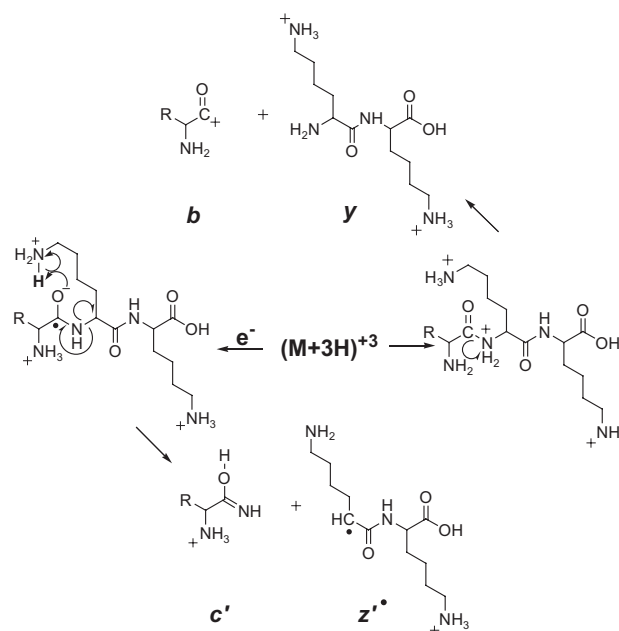


Fig. 1. Peptide fragmentation. Scheme for generating type b and y ions by CAD and type c and z[•] ions by ETD.

the two fragments. The complete amino acid sequence of a peptide is deduced by extending this process to all homologous pairs of fragments within a particular ion series. ETD, unlike CAD, works exceptionally well on large multiply charged peptides and is indifferent to most posttranslational modifications, including phosphorylation.

Author contributions: A.C. and C.H. contributed equally to this work; L.Y.G., J.S., D.J.B., O.G.T., and D.F.H. designed research; A.C. and C.H. performed research; A.C., C.H., L.Y.G., J.J.C., J.E.P.S., D.L.B., and J.S. contributed new reagents/analytic tools; A.C., C.H., L.Y.G., D.L.B., J.S., D.J.B., O.G.T., and D.F.H. analyzed data; and A.C., D.J.B., O.G.T., and D.F.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS direct submission.

Abbreviations: IMAC, immobilized metal-affinity chromatography; MS/MS, tandem MS; CAD, collision-activated dissociation; ETD, electron transfer dissociation.

[†]Present address: Merck Research Laboratories, Boston, MA 02115.

[¶]Present address: Department of Chemistry, University of Wisconsin, Madison, WI 53706.

^{§§}To whom correspondence should be addressed. E-mail: dfh@virginia.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0607084104/DC1.

© 2007 by The National Academy of Sciences of the USA

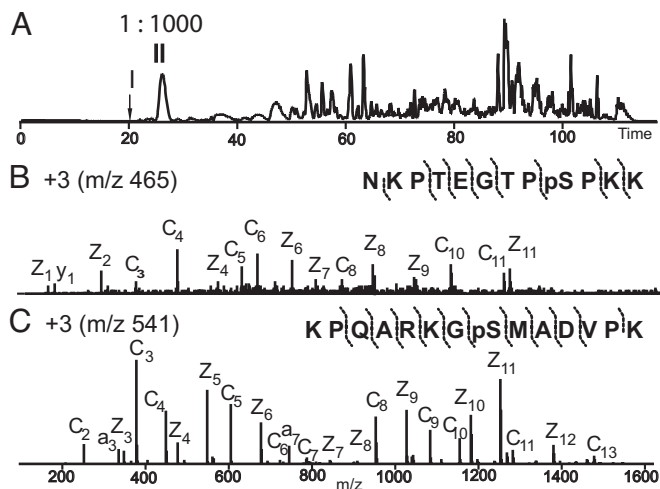


Fig. 2. Phosphopeptide chromatogram and mass spectra. (A) Base peak, ion chromatogram for a 90 min separation of yeast phosphopeptides enriched from a 30- μ g aliquot of yeast total cell protein. (B and C) ETD mass spectra recorded on $(M + 3H)^{+3}$ ions from peptides eluting under peaks I and II, respectively, in A. Observed c and z ions are indicated on the peptide sequence by $\bar{\quad}$ and $\underline{\quad}$, respectively.

Here we present a strategy for the analysis of the yeast phosphoproteome that uses endo-Lys C as the proteolytic enzyme, IMAC for phosphopeptide enrichment, ETD for peptide fragmentation, and the Open Mass Spectrometry Search Algorithm for phosphopeptide identification and assignment of phosphorylation sites. With this approach, we identify 1,252 phosphorylation sites on 629 proteins in a single experiment with 30 μ g (\approx 600 pmols) of protein from a yeast whole cell lysate. We find that the identified phosphoproteins are encoded by a sample of genes that is representative of a wide variety of cellular processes. Expression levels for the identified phosphoproteins range from <50 to $>1,200,000$ copies per cell. We note that most protein-kinase recognition motifs predicted by SCANSITE are significantly enriched in our phosphorylation data and present evidence for a motif recognized by one or more unidentified yeast kinases. We analyze the identified phosphoproteins in the context of interaction networks and find that they have a significantly higher number of interactions than expected and interact with other phosphoproteins more than expected. We note that the observed phosphoproteins, but not individual phosphosites, are likely to be conserved across very large evolutionary distances.

Results and Discussion

Fig. 2A shows the base peak chromatogram for a 90-min analysis of yeast phosphopeptides obtained from a 30 μ g (\approx 600 pmol) aliquot of total cell protein. Sample was digested with endo-Lys-C, enriched for phosphopeptides by IMAC, and analyzed by a combination of nanoflow-HPLC/electrospray ionization tandem MS, and ETD. From this experiment, we identify 1,252 phosphorylation sites on 629 proteins, $\approx 10\%$ of the proteome. [The complete data set of identified yeast phosphopeptides is available in [supporting information \(SI\) Table 2.](#)] A subset of this data, phosphopeptides identified from yeast protein kinases, is shown in Table 1. Note that we find phosphorylation sites on a disproportionately high number (33 of 126, 26%) of these proteins, even though the expression levels for the yeast kinases and the other 593 phosphoproteins in [SI Table 2](#) range from <50 to 1,200,000 copies per cell (4). It seems likely that many of the yeast kinases are, themselves, regulated by phosphorylation.

Fig. 2B and C shows the ETD spectra recorded on peptides eluting under peaks I and II, respectively, in Fig. 2A. Ion currents

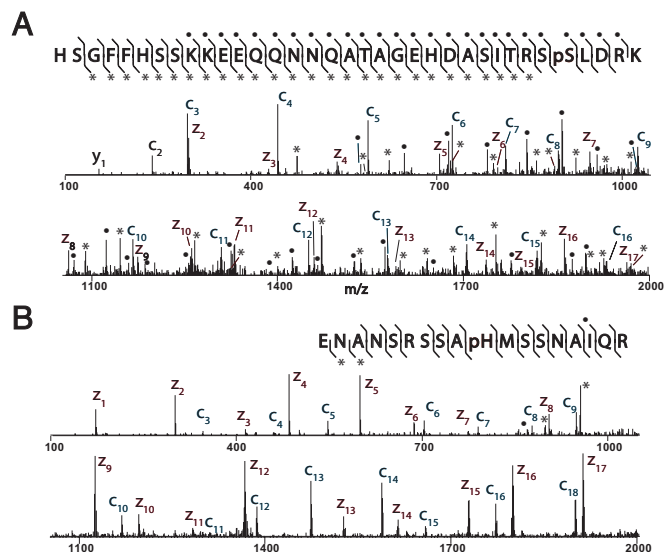


Fig. 3. Phosphopeptide mass spectra. ETD mass spectra recorded on $(M + 6H)^{+6}$ ions at m/z 683.3 for a 35 residue phosphopeptide of MW 4,093 (A), and $(M + 3H)^{+3}$ ions from a pHIS containing peptide at the C terminus of the septin protein, Cdc10 (B). Observed c and z ions are indicated on the peptide sequence by $\bar{\quad}$ and $\underline{\quad}$, respectively. Observed doubly charged c and z ions are indicated by an additional label, circle and asterisk, respectively.

carried by $(M + 3H)^{+3}$ ions for these two peptides differ by more than a factor of 1,000 (3E7 and 5E4), yet the resulting ETD spectra still contain a near full complement of c and z type ions that facilitate sequence assignment and location of the single site of phosphorylation in each peptide. Note that fragments of type c and z that contain the N and C termini of Pro, respectively, remain attached to the peptide backbone through other atoms in the proline ring and are thus not observed in ETD mass spectra (2).

Endo-Lys-C cleaves proteins on the C-terminal side of Lys, but not Arg, and thus produces many fragments that are larger than those generated with trypsin. Of the phosphopeptides identified in the present study, 36% contain >20 residues. Fig. 3A shows the ETD mass spectrum recorded on the $(M + 6H)^{+6}$ ion at m/z 683.3 for a 35-residue phosphopeptide (MW 4,093) derived from the 76.5-kDa, Ser/Thr kinase, YPK1 (JKL126W). This peptide contains eight Ser/Thr residues. Phosphorylation of these residues increases their mass by 80 Da. In the spectrum shown in Fig. 3A, the fragment that contains the last four amino acids of the peptide, z₄, appears at the predicted m/z value of 543. The z₅ ion contains the same four amino acids plus Ser₃₁. Its m/z value is predicted to occur at m/z 630, but is observed at m/z 710. All larger singly and doubly charged z ions undergo shifts of 80 and 40 Da, respectively. We conclude that Ser₃₁ is the only site of phosphorylation in this peptide. Note that although the mass range of the LTQ instrument is currently only 2,000 Da, the presence of singly and doubly charged c and z ions in the ETD spectrum usually makes it possible to sequence peptides that are 40–50 residues in length.

Fig. 3B shows the ETD spectrum recorded on the $(M + 3H)^{+3}$ ion at m/z 690.5 for an 18-residue phosphopeptide (MW 2,067) derived from Cdc10 (YCR002C), one of five septin proteins that form a filamentous ring structure and play a conserved role in both cytokinesis and morphogenesis. This peptide contains five Ser residues. Fragments containing the first nine residues from the N terminus, c₉, and the last eight residues from the C terminus, z₈, both occur at the predicted m/z values, 948 and 904, respectively. In contrast, fragments c₁₀ and z₁₀, which contain His₃₁₄, are both observed at m/z values that are 80 Da higher than expected. We conclude that the Cdc10 peptide is phosphorylated on His rather than Ser.

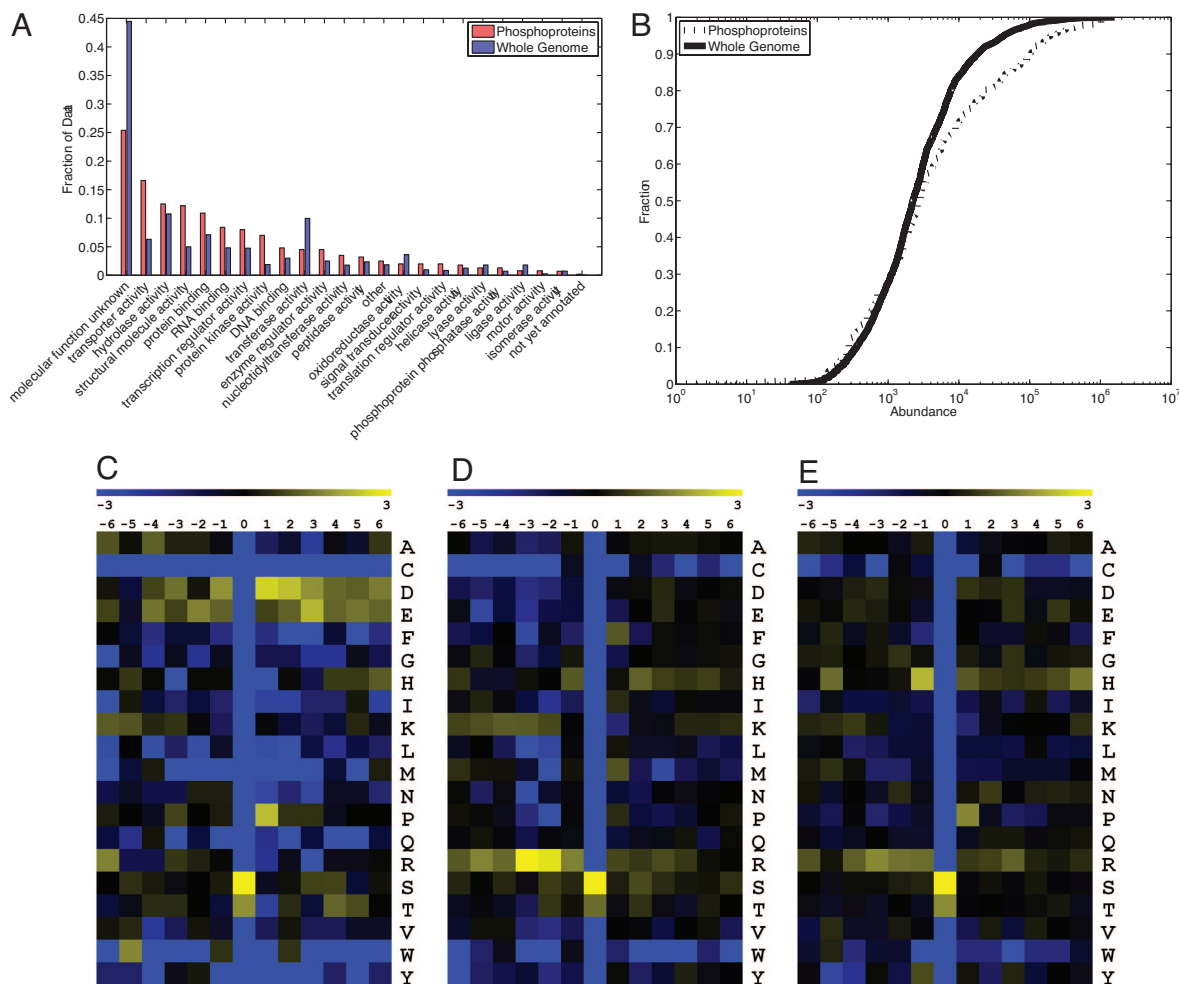


Fig. 4. Phosphoprotein and amino acid frequencies. (A) Comparison of GO Slim (33) term frequencies between the whole genome and the phosphoproteins. (B) Comparison of protein abundances between the entire genome and the phosphoproteins. The distribution of phosphoprotein abundances is comparable to that of the genomic background. (C–E) Log₂ ratios of per-site amino acid frequencies relative to the genomic background. (C) Sites identified as acidophilic by SCANSITE. (D) Sites identified as basophilic by SCANSITE. (E) Sites not recognized by SCANSITE.

SCANSITE was significantly enriched in our phosphorylation data (SI Fig. 6). A summary of the SCANSITE kinase target-groups found in our data, using medium stringency criteria, appears in SI Table 4. Basophilic sites make up the largest group of motifs, and acidophilic and proline-directed motifs are also well represented. These results are in agreement with data from other large datasets (11, 12). In addition to these known sites, our assay identifies 381 phosphorylation sites not found by SCANSITE. The relative occurrence of amino acids flanking the sites of phosphorylation appears as a heat map in Fig. 4 C–E. The acidophilic (Fig. 4C) and basophilic (Fig. 4D) maps are as predicted from SCANSITE. The heat map for the 381 sites not found by SCANSITE (Fig. 4E) shows a moderate enrichment of proline and histidine at positions that are +1 and –1, respectively, to the site of phosphorylation. We suggest that this likely represents a new motif for one or more protein kinases in yeast.

The majority of the phosphorylation sites in our data are on serine (82.3%); the remainder are on threonine (17.5%) and tyrosine (0.027%). This finding confirms the extremely low extent of tyrosine phosphorylation in yeast (13). There are no true protein tyrosine kinases in yeast, but there are seven dual-specificity kinases predicted on the basis of sequence similarity, including the MAPKK proteins (Ste7, Mkk1, Mkk2, and Pbs2) and three kinases that regulate the cell cycle (Rad53, Mps1, and Swe1) (14). None of the tyrosine-phosphorylated proteins are obviously related by func-

tional annotation. However, one of the proteins, Cnm67, is a structural component of the spindle pole body and is an excellent candidate for a substrate of Mps1, a protein kinase that regulates spindle pole body duplication (15, 16).

We took advantage of diverse functional genomic and proteomic data to analyze protein kinase targets in the context of interaction networks. We began with affinity precipitation and yeast two-hybrid data consisting of 13,325 interaction pairs and 4,697 proteins (17–20). We found that phosphoproteins have a significantly higher number of physical interactions than would be expected by chance and, in particular, interact with other phosphoproteins more than with a random protein (Fig. 5B and SI Table 5). One explanation for this is that signaling cascades are often organized on molecular scaffolds that promote the physical association of proteins participating in the signal-transduction pathway (21). Among genetic interaction data (4,775 interaction pairs, 1,469 proteins), we found that genes encoding phosphoproteins also exhibit a strong tendency toward a high degree of interaction and exhibit enriched interactions with other phosphoproteins in genetic interaction networks. This effect has been previously observed for essential proteins and has been implicated for conserved proteins (22). Interestingly, phosphoproteins are not statistically likely to be essential, but they are highly enriched for strongly conserved genes (Fig. 5C). We propose that nonessential phosphoproteins play a central role in

this material was diluted 10-fold in 100 mM ammonium bicarbonate (pH 8.9), reduced with DTT, carboxyamidomethylated with iodoacetamide, and then digested with endo-Lys-C (Roche, Indianapolis, IN) (enzyme/substrate = 1/20) overnight at 37°C in a sand bath. To remove small molecules, the resulting solution of peptides was loaded onto a 360 μm o.d. \times 200 μm i.d. fused-silica column packed with irregular reverse-phase C₈ beads (10–30 μm , 120 Å ODS-AQ; Waters, Milford, MA). After the beads were rinsed with 20 column volumes of 0.1% acetic acid, peptides were eluted with solvent consisting of acetonitrile (70%) (Mallinckrodt, Paris, KY) and 0.1% acetic acid (30%) (Sigma–Aldrich, St. Louis, MO). Solvent was removed on a SpeedVac (Savant Instruments, Farmingdale, NY) and residues were dissolved in 100 μl of freshly prepared methanolic HCl. This reagent was prepared by adding 40 μl of anhydrous thionyl chloride (Sigma–Aldrich) dropwise to 1 ml of anhydrous methanol (Alltech, Deerfield, IL). Esterification was allowed to proceed for 1 h at room temperature, solvent was removed on a SpeedVac, and the resulting peptide methyl esters were then redissolved in 60 μl of a solution containing equal parts of methanol, acetonitrile and 0.01% acetic acid.

Chromatography. IMAC columns were constructed as described (1), with modification. Briefly, fused-silica columns (360 μm o.d. \times 200 μm i.d.) (Polymicro Technologies, Phoenix, AZ) were packed with 8 cm of Poros MC (PerSeptive Biosystems, Framingham, MA) and then washed with (i) 50 mM EDTA (pH 9) for 5 min to remove metal ions, (ii) NANOPure (Barnstead, Dubuque, IA) water for 5 min to bring the pH back to neutral, (iii) 100 mM FeCl₃ (Sigma–Aldrich) at a flow rate of 2 $\mu\text{l}/\text{min}$ for 10 min to activate the column, and (iv) 0.01% acetic acid, 5–10 column volumes, to remove excess FeCl₃. Peptides derived from the digestion of 30 μg (\approx 600 pmol) of yeast protein were loaded at a flow rate of 1 $\mu\text{l}/\text{min}$ and the column was then washed with (i) 1:1:1, methanol: acetonitrile: 0.01% acetic acid (60 μl) to remove peptides bound nonspecifically and (ii) 0.01% acetic acid (20 μl) to remove organic solvent. A fused-silica precolumn (360 μm o.d. \times 100 μm i.d.) packed with irregular C₈ beads (Waters) was then connected to the IMAC column using a Teflon sleeve (0.012 in o.d. \times 0.060 in i.d.; Zeus, Orangeburg, SC). Phosphopeptides were eluted from the IMAC column to the precolumn with 50 mM ascorbic acid (25 μl) (Sigma–Aldrich), and the precolumn was washed with 0.1% acetic acid (10 min) to remove ascorbic acid. A Teflon sleeve was used to connect to an analytical HPLC column (360 μm o.d. \times 75 μm i.d.) packed with 5 cm of regular C₈ beads (5 μm , Waters) and equipped

with a laser-pulled, electrospray, emitter tip (2–4 μm in diameter). An Agilent 1100 Series binary HPLC system (Palo Alto, CA) was used to generate a gradient 0–7% B in 5 min, 7–45% B in 70 min, 45–100% B in 15 min, 100–0% B in 5 min (A = 0.1 M acetic acid; B = 70% acetonitrile in 0.1 M acetic acid) and to elute peptides to the mass spectrometer at a flow rate of 60 nl/min.

Instrument Modification and Operation. All experiments were performed on a Finnigan LTQ mass spectrometer (Thermo Electron, San Jose, CA) equipped with a nano-flow HPLC micro-electrospray ionization source and modified to perform ETD (2). The instrument was operated in the data-dependent mode and cycled through acquisition of a full MS spectrum plus ETD/MS/MS spectra on the 10 most abundant ions in the initial full MS spectrum every 5 s (mass window, 3 Da; dynamic exclusion, 45-s; repeat counts, 1). Instrument control software (ITCL) was modified to facilitate the following sequence of events after selection and storage of the precursor peptide cation; (i) injection of fluoranthene radical anions (\approx 2 ms) (3); (ii) isolation of fluoranthene radical anions (m/z 202) (10 ms); (iii) electron transfer (ion/ion reaction time of 65 ms); (iv) removal of excess fluoranthene radical anions, and (v) mass analysis of positively charged fragment ions.

Data Analysis. MS/MS spectra were preprocessed with in-house software to determine the charge states of peptide precursor ions and to eliminate charged reduced, but nondissociated, ions as well as fragments formed from the parent ion by loss of small molecules rather than cleavage of the peptide backbone. The resulting spectra were then searched against the yeast protein database (www.yeast-genome.org) with Open Mass Spectrometry Search Algorithm software (version 1.0.5) (32) (<http://pubchem.ncbi.nlm.nih.gov/omssa>). Parameters for the search included the following: endo-Lys C specificity; static modifications of 14 Da on Asp, Glu, and the peptide C terminus (formation of methyl esters) and 57 Da on Cys (alkylation with iodoacetamide); differential modification of 80 Da on Ser, Thr, and Tyr (phosphorylation) and 16 Da on Met (oxidation). The window for the precursor ion mass was set as \pm 1.5 Da and the fragment–ion mass tolerance was set as \pm 0.5 Da. Peptide sequence assignments in SI Table 2 were validated by manual interpretation of the corresponding ETD mass spectra. Nonphosphorylated peptides were not detected under the above experimental conditions.

This work was supported by National Institutes of Health Grant GM37537 and National Science Foundation Grant MCB-0209793 (to D.F.H.).

1. Ficarro SB, McLeland ML, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM (2002) *Nat Biotechnol* 20, 301–305.
2. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF (2004) *Proc Natl Acad Sci USA* 101:9528–9533.
3. Coon JJ, Ueberheide B, Syka JEP, Dryhurst DD, Ausio J, Shabanowitz J, Hunt DF (2005) *Proc Natl Acad Sci USA* 102:9463–9468.
4. Ghaemmaghami S, Huh W-K, Bower K, Howson RW, Belle A, Dephoure N, O’Shea EK, Weissman JS (2003) *Nature* 425:737–741.
5. West AH, Stock AM (2001) *Trends Biochem Sci* 26:369–376.
6. Posas F, Wurler-Murphy SM, Maeda T, Witten EA, Thai TC, Saito H (1996) *Cell* 88:865–875.
7. Li S, Ault A, Malone CL, Raitt D, Dean S, Johnston LH, Deschenes RJ, Fassler JS (1998) *EMBO J* 17:6952–6962.
8. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. (2000) *Nat Genet* 25:25–29.
9. Qi M, Elion EA (2005) *J Cell Sci* 118:3569–3572.
10. Obenaus JC, Cantly LC, Jaffe MB (2003) *Nucleic Acids Res* 31:3635–3641.
11. Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villen J, Li J, Cohn MA, Cantly LC, Gygi SP (2004) *Proc Natl Acad Sci USA* 101:12130–12135.
12. Nuhse TS, Stensballe A, Jensen ON, Peck SC (2004) *Plant Cell* 16:2394–2405.
13. Modesti A, Bini L, Carraresi L, Magherini F, Liberatori S, Pallini V, Manao G, Pinna LA, Raugi G, Ramponi G (2001) *Electrophoresis* 22:576–585.
14. Hunter T, Plozman GD (1997) *Trends Biochem Sci* 22:18–22.
15. Lauze E, Lauze E, Stoelcker B, Luca FC, Weiss E, Schutz AR, Winey M (1995) *EMBO J* 14:1655–1663.
16. Schaerer F, Morgan G, Winey M, Philippsen P (2001) *Mol Biol Cell* 12:2519–2533.
17. Bader GD, Donaldson I, Wolting C, Francis Ouellette BF, Pawson T, Hogue GWV (2001) *Nucleic Acids Res* 29:242–245.
18. Breitkreutz BJ, Stark C, Tyers M (2003) *Genome Biol* 4:R23.
19. Gavin A-C, Bosche M, Krause R, Grandl P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon A-M, Cruciat C-M, et al. (2002) *Nature* 415:141–147.
20. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams S-L, Millar A, Taylor P, Bennett K, Boutiller K, et al. (2002) *Nature* 415:180–183.
21. Bhattacharyya RP, Reményi A, Good MC, Bashor CJ, Falick AM, Lim WA (2006) *Science* 311:822–826.
22. Jeong H, Mason SP, Barabási AL, Oltvai ZN (2001) *Nature* 411:41–42.
23. Balakrishnan R, Christie KR, Costanzo MC, Dolinski K, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hong EL, Nash R, et al. (2005) *Nucleic Acids Res* 33:D374–D377.
24. Moritz RL, Ji H, Schutz F, Connolly LM, Kapp EA, Speed TP, Simpson RJ (2004) *Anal Chem* 76:4811–4824.
25. Malmström J, Lee H, Nesvizhskii AI, Shteynberg D, Mohanty S, Brunner E, Ye M, Weber G, Eckerskorn C, Aebersold R (2006) *J Proteome Res* 5:2241–2249.
26. Salomon, A.R. Ficarro SB, Brill LM, Brinker A, Phung QT, Ericson C, Sauer K, Brock A, Horn DM, Schultz, P.G. et al. (2003) *Proc Natl Acad Sci USA* 100:443–448.
27. Ficarro SB, Salomon AR, Brill LM, Mason DE, Stettler-Gill M, Brockl A, Peters EC (2005) *Rapid Commun Mass Spectrom* 19:57–71.
28. Kim J-E, White FM (2006) *J Immunol* 176:2833–2843.
29. Gruhler A, Olsen JV, Mohammed S, Mortensen P, Faergeman NF, Mann M, Jensen ON (2005) *Mol Cell Proteomics* 4:310–327.
30. Moser K, White FM (2006) *J Proteome Res* 5:98–104.
31. Kim, J-E., Tannenbaum SR, White FM (2005) *J Proteome Res* 4:1339–1346.
32. Geer LY, Markey SP, Kowalak JA, Wagner L, Maynard DM, Yang X, Shi W, Bryant SH (2004) *J Proteome Res* 3:958–964.
33. Dwight SS, Balakrishnan R, Christie KR, Costanzo MC, Dolinski K, Engel SR, Feierbach B, Fisk DG, Hirschman J, Hong EL, et al. (2004) *Brief Bioinform* 5:9–22.
34. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M (2006) *Nucleic Acids Res* 34:D354–D357.
35. Meng F, Cargile BJ, Miller LM, Forbes AJ, Johnson JR, Kelleher NL (2001) *Nat Biotechnol* 19:952–957.