# Ficarro, S.B. et al. Phosphoproteome analysis by mass spectrometry and its application to Saccharomycescerevisiae. Nat. Biotechnol. 20, 301–305

**ARTICLE** *in* NATURE BIOTECHNOLOGY · APRIL 2002

Impact Factor: 41.51 · DOI: 10.1038/nbt0302-301 · Source: PubMed

# Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*

Scott B. Ficarro[1], Mark L. McCleland[2], P. Todd Stukenberg[2], Daniel J. Burke[2], Mark M. Ross[4], Jeffrey Shabanowitz[1], Donald F. Hunt[1,3], and Forest M. White[4]*

Protein kinases are coded by more than 2,000 genes and thus constitute the largest single enzyme family in the human genome. Most cellular processes are in fact regulated by the reversible phosphorylation of proteins on serine, threonine, and tyrosine residues[1]. At least 30% of all proteins are thought to contain covalently bound phosphate. Despite the importance and widespread occurrence of this modification, identification of sites of protein phosphorylation is still a challenge, even when performed on highly purified protein[2]. Reported here is methodology that should make it possible to characterize most, if not all, phosphoproteins from a whole-cell lysate in a single experiment. Proteins are digested with trypsin and the resulting peptides are then converted to methyl esters, enriched for phosphopeptides by immobilized metal-affinity chromatography (IMAC)[3,4], and analyzed by nanoflow HPLC/electrospray ionization mass spectrometry. More than 1,000 phosphopeptides were detected when the methodology was applied to the analysis of a whole-cell lysate from *Saccharomyces cerevisiae*. A total of 216 peptide sequences defining 383 sites of phosphorylation were determined. Of these, 60 were singly phosphorylated, 145 doubly phosphorylated, and 11 triply phosphorylated. Comparison with the literature revealed that 18 of these sites were previously identified, including the doubly phosphorylated motif pTXpY derived from the activation loop of two mitogen-activated protein (MAP) kinases. We note that the methodology can easily be extended to display and quantify differential expression of phosphoproteins in two different cell systems, and therefore demonstrates an approach for "phosphoprofiling" as a measure of cellular states.

To develop a method for analysis of phosphoproteins in a complex mixture, we first prepared a standard mixture of tryptic peptides containing a single phosphopeptide and then analyzed the mixture before and after converting the peptides to the corresponding methyl esters. This peptide derivatization rendered the IMAC selective for phosphopeptides, and eliminated confounding binding through carboxylate groups. Equimolar quantities of glyceraldehyde 3-phosphate dehydrogenase, BSA, carbonic anydrase, ubiquitin, and β-lactoglobulin were digested with trypsin (~125 predicted cleavage sites) and then combined with the phosphopeptide DRVpYIHPF (lowercase p precedes a phosphorylated residue throughout this work), to give a mixture that contained tryptic peptides, each at the 2 pmol/μl level, and phosphopeptide at the 10 fmol/μl level (200:1 ratio of nonphosphorylated to phosphorylated peptides). All experiments were done on aliquots containing 1 pmol each of nonphosphorylated peptides and 5 fmol of the phosphorylated peptide.

Shown in Figure 1 are the results obtained when the nonderivatized mixture was analyzed by a combination of IMAC[3,4] and nanoflow-HPLC on an LCQ ion-trap mass spectrometer. In this experiment, the instrument was set to cycle between two different scan functions every 2 s throughout the HPLC gradient. Electrospray ionization spectra were recorded in the first scan, and MS/MS spectra on the $(M+2H)^{++}$ ion of the phosphopeptide, DRVpYIHPF (*m/z* 564.5) were recorded in the second scan of the cycle. Figure 1A shows a selected-ion chromatogram (SIC) or plot of the ion current observed for *m/z* 564.5 as a function of scan number. Note that a signal at this *m/z* value is observed at numerous points in the chromatogram. Only ions at *m/z* 564.5 in scans 610–616 fragment to generate MS/MS spectra characteristic of the phosphopeptide, DRVpYIHPF (Fig. 1B). We conclude that DRVpYIHPF elutes from the HPLC column in scans 610–616.

Displayed in Figure 1C is an electrospray ionization mass spectrum recorded during this same time period. Note that the spectrum contains signals of high intensity (ion currents of $1–3 \times 10^9$) corresponding to the nonphosphorylated tryptic peptides in the mixture but no signal above the chemical noise level for the phosphopeptide (*m/z* 564.5). We conclude that tryptic peptides containing multiple carboxylic acid groups can bind efficiently to the IMAC column, elute during the HPLC gradient, and suppress the signal from trace-level phosphopeptides in the mixture.

To prevent binding of nonphosphorylated peptides to the IMAC column, all peptides in the standard mixture were converted to the corresponding peptide methyl esters and then analyzed by the protocol outlined above. To detect the DRVpYIHPF phosphopeptide in which both carboxylic acid groups had been esterified, MS/MS spectra were recorded on the $(M+2H)^{++}$ ion at *m/z* 578.5. The SIC for *m/z* 578.5 (Fig. 1D) indicates that the phosphopeptide dimethyl ester elutes during scans 151–163. Indeed, MS/MS spectra (Fig. 1E) recorded in this time window all contain the predicted fragments expected for the dimethyl ester of DRVpYIHPF. Figure 1F shows an electrospray ionization mass spectrum recorded in the same area of the chromatogram (scan no. 154). Note that the parent ion, *m/z* 578.5 for the phosphopeptide dimethyl ester, is now observed with a signal:noise ratio of 3:1 and an ion current of $2 \times 10^7$. This signal level on the LCQ is typical for phosphopeptide samples at the 3–5 fmol level. Note also that signals above the chemical noise (ion current of $1 \times 10^7$) for nonphosphorylated tryptic peptides no longer appear in this electrospray ionization spectrum or in any other spectrum recorded throughout the entire chromatogram. We conclude that conversion of carboxylic acid groups to methyl esters reduces nonspecific binding by at least two orders of magnitude and allows detection of phosphopeptides in complex mixtures down to the level of at least 5 fmol with the LCQ instrument.

To further evaluate the above protocol for its ability to identify phosphorylation sites on proteins in a very complex mixture, we next analyzed a protein pellet (500 μg) obtained from a whole-cell lysate of *S. cerevisiae*. If one makes the assumption that 30% of expressed proteins contain at least one covalently bound phosphate[1], the total number of phosphoproteins in the sample could easily exceed 1,000. To evaluate the possibility of identifying all these phosphoproteins, the pellet was digested with trypsin and the resulting peptides were converted to peptide methyl esters. One fifth of the resulting mixture was then fractionated by IMAC and ana-

[1]*Department of Chemistry, University of Virginia, Charlottesville, VA 22901.* [2]*Departments of Biochemistry and Molecular Genetics and* [3]*Pathology, Health Sciences Center, University of Virginia, Charlottesville, VA 22908.* [4]*MDS Proteomics, Charlottesville, VA 22901.*
**Corresponding author (fwhite@mdsp.com).*

**Table 1. Phosphorylated peptide sequences from *S. cerevisiae***

| Protein | Phosphopeptide sequence | Protein | Phosphopeptide sequence |
|---|---|---|---|
| ABF1 | SNpSIDYAK | REG1, YDR028C | SGSpTNpSLYDLAQPSLSSATPQQK |
| ABP1 | KEPVKpTPpSPAPAAK | RPA190 | DKEpSDpSDpSEDEDVDMNEQINK |
| ABP1 | SFpTPSKpSPAPVSK | RPL12A | IGPLGLpSPK |
| ACC1 | AVpSVSDLSYVANSQSSPLR | RPL24A | VAATpSR |
| ACE2 | KLTpSPK | RPL25 | TpSATFR |
| AKL1 | DKDpSNSpSITISTSTPSEMR | RPL3 | AApSIR |
| AKL1 | STSpSYSSGGR | RPL7A, RPL7B | ILpTPEpSQLKK |
| BFR2, YDR299W | NGEpSDLpSDYGNSNTEETK | RPN8 | VpSDDSEpSESGDKEATAPLIQR |
| BLM3 | pSApTPTLQDQK | RPP1A and 2B | (E)EEAKEEpSDDDMGFGLFD |
| BNI5 | pSGGSpTPLDSQTK | RPS31 | VYTpTPKK |
| BNI5 | YDpSPVSpSPITSASELGSIAK | **RPS6A** | AS**pS**LKA |
| BOI2 | ALpSPIPSPpTR | **RPS6A** | RA**pSpS**LKA |
| BRE5 | ESGNNASpTPSpSSPEPVANPPK | SEC1 | SQDNSPKpSGTSpSPK |
| BUD4 | VNpSELEEpSPAAVHQER | SEC3 | TIpSGSpSAHHSR |
| CCC1 | HNDLpSSSpSSDIIYGR | SEC31 | APpSSVpSMVS*P*LHK |
| CDC11 | LNGpSSSpSSTTTR | SEC31 | APpSSVSMVpS*P*LHK |
| CDC39 | RQpTPLQSNA | SEC31 | VPpSLVATSEpSPR |
| CDC47 | FVDDGTMDpTDQEDSLVpSTPK | SEC31 | VPSLVATSEpSPR |
| CHD1 | NSVNGDGTAANpSDpSDDDSTSR | SEC4 | EGNIpSINpSGSGNS |
| CHO1, YER026C | DENDGYApSDEVGGTLpSR | SEC4 | TVpSASpSGNGK |
| CHS2 | TQFYRDpSAHNpSPVAPNR | SGV1 | YpTSVVVTR |
| CLA4 | GPMHPNNpSQRpSLQQQQQQQQQQK | SHP1 | KGpSTpSPEPTK |
| CRN1 | QEAPKpSPpSPLK | SIN3 | VpTTPMGTTTV*N*IpSPSGR |
| CRN1 | TKpSPEQEKSApTPPSSITAAK | SLA2 | TPpTPTPPVVAEPAIpSPRPVSQR |
| CTK1 | ADYpTNR | **SLT2** | GYSENPVENSQFL**pT**E**pY**VATR |
| CTK3 | DSITSpSSTpTTPPSSQQK | SMI1 | SQQGLSHVTSTGpSSSpSMER |
| CYK3 | LpSSSMPNpSPKKPVDSLTK | SMY2 | SQFQKpSPK |
| DBP10 | LQNSNNEADpSDpSDDENDR | SOK2, YIL055C | SIpSPR |
| DIP5 | MFTSTpSPR | SOL1 | STQMpSGTpSLNGNGNTESK |
| DIP5 | NSpSSLDpSDHDAYYSK | SOL1 | VNpSVRpSNASSR |
| ECM25 | SRpSPSPQR | SOL2 | SpTApSAAEGK |
| EDE1 | GVATpTPK | SPC98 | pSMVpSSPNR |
| EDE1 | RANSNEDDGEpSVpSSIQEpSPK | SRA1 | SRpSSVMFK |
| EDE1 | TTPLpSANSpTGVSSLTR | SSD1 | SSpTINNDSDSLSpSPTK |
| ENO1, ENO2 | pSVYDSR | **STE2** | EGEVEPVDMY**pT**PD**pT**AADEEAR |
| ERG6 | VARKPENAETPpSQTpSQEATQ | STE2 | YQLPpTPpTpSSKNTR |
| FEN1 | NVPpTPpSPpSPKPQHR | STF2 | RGpSNLQSHEQK |
| FPR4 | LEEDEpSEpSEQEADVPKR | **SUI2** | ELDNR**pS**D**pS**EDDEDE**pS**DDE |
| GCS1 | SApTPANSpSNGANFQK | SUR7 | pSHERPDDVpSV |
| GLY1 | SESpTEVDVDGNAIR | SUR7 | SHERPDDVpSV |
| GNP1 | KSpSYIpTVDGIK | SWI4 | STpSETSpSPK |
| GPA2 | NGSTPDTQTApSAGpSDNVGK | SYG1 | RRpSpSVFENISR |
| GPD1 | SSpSSVpSLK | TAT1 | QDEVpSGQpTAEPR |
| GPD2 | SDpSAVpSIVHLK | TIF5 | AAKPFITWLETAEpSDDDEEDDE |
| GPD2 | SDpSAVSIVHLK | TPO1, YLL028W | TTTMNpSAAEpSEVNITR |
| GSY1 | IARPL**pS**VPGpSPK | TPS2 | SApSYTGAKV |
| GSY2 | VARPL**pS**VPGpSPR | TPS3 | TSpSSMpSVGNNK |
| GTS1 | SHpSFYK | TRS120 | ATASTpTASSpTPR |
| HOG1 | IQDPQM**pT**G**pY**VSTR | TSL1 | SApTRpSPSAFNR |
| HSP26 | KIEVSpSQEpSWGN | TYS1 | GYPVApTPQK |
| HSP26 | KIEVSSQEpSWGN | UBP1 | AAQQDpSpSEDENIGGEYYTK |
| HSP26 | QLANpTPAK | UGP1 | THpSTYAFESNTNSVAASQMR |
| HSP30 | EAVPEpSPR | VPS13 | TApTPQpSLQGSNK |
| HTA1 | ATKApSQEL | VRP1 | NPTKpS*PP*SPSTMDTGTSNpSPSK |
| HXK1/HXK2 | KGpSMADVPK | VRP1 | S*PP*SPSTMDTGTSNpSPSK |
| HXT2 | LEpTDEpSPIQTK | **YAK1** | RA**pS**LNSK |
| HXT2 | VESGpSQQTpSIHpSTPIVQK | **YAK1** | RK**pS**pSLVVPPAR |
| HXT2 | VESGSQQTpSIHpSTPIVQK | YBR235W | pSQNNLQK |
| IRA2 | NSDNVNpSLNSpSPK | YBT1 | SSILpSRANpSSANLAAK |
| IST2 | TTESSSpSSpSAAK | YBT1 | ETSNEASpSpTNSENVNK |
| IST2 | VPpTVGpSYGVAGATLPETIPTSK | YCR023C | SpSLSpSLSNQR |
| KSP1 | MEGGDNESpSSTpSPDER | YDL113C | NGVGpSPKKpSPK |
| MCM3 | RSTASpSVNATPpSSARR | YDL166C | MWLEQHPDGVTNEYQGPRpSDDEDDEDpSE |
| MCM3 | VRQPApSNSGpSPIK | YDL189W | RApSVEGpSPSSR |
| MLF3 | TTELPATpSPYVpSPQQSAR | YDL222C | AHSTYSDHDMYAQYESPpSVDpTGAQMEK |
| MOB1 | MpSPVLTpTPKR | YDL223C | GSDYDYNNSTHpSAEHpTPR |
| MON2, YNL297C | NIpSTSpSVTTSPVESTK | YDL223C | KVSVGpSMGpSGK |
| MRH1 | APVApSPRPAApTPNLSK | YDR090C | YSRLpSV |
| MSC3 | STAG*N*DpSRANpSITVK | YDR262W | NVpSRTpSPNTpTNTK |
| MSL5 | NRpSPpS*P*VYDAQGK | YDR262W | NVpSRTpSPNTTNTK |

**Table 1. Phosphorylated peptide sequences from *S. cerevisiae***

| | | | |
|---|---|---|---|
| MSN2 | RPpSYR | YDR372C | ADpSGDTSpSIHSSANNTK |
| MYO3 and 5 | SMpSLLGYR | YDR372C | ADSGDTSpSIHSpSANNTK |
| NCB2 | LHHNpSVSDPVKSEDS*pS | YDR384C | TpSSApSpSPQDLEK |
| NCB2 | LHHNSVSDPVKpSEDS*pS | YDR466W | ApSSEPSpS*P*ISR |
| NEW1 | GTPKPVDpTDDEED | YEF3 | KKELGDAYVpSpSDEEF |
| NIP1 | SSNYDpSpSDEEpSDEEDGKK | YFR016C | SKTPEpSPK |
| NOT3 | IGpSALNpTPK | YFR017C | RRpSTNYMDALNSR |
| NOT3 | TPTTAAATTTSpSNANpSR | YFR017C | RSpSGPMDFQNTIHNMQYR |
| NPL3 | GGYDpSPR | YFR024C | LAPTNpSGGpSGGKLDDPSGASSYYASHR |
| NPR1 | QSpSIYSASR | YGR138C | LTKpTEpTVK |
| NRD1 | NRpSRpSPPAPFSQPSTGR | YGR138c | TApSALpSR |
| **NTH1** | RG**pS**EDDTYSSSQGNR | YHR052W | SNpSKKSpTPVpSTPSKEK |
| NTH1 | RLSpSLpSEFNDPFSNAEVYYGPPTDPR | YHR052W | SNpSKKSpTPVSTPSK |
| NUP145 | AYEPDLpSDADFEGIEApSPK | YHR097C | ANSpSTpTTLDAIKPNSK |
| NUP2 | EpTYDpSNEpSDDDVTPSTK | YHR132W-A | RMpSpSSSGGDSISR |
| NUP2 | ETYDpSNEpSDDDVTPSTK | YHR186C | AGpSIQpTQSR |
| OAF1 | SApSPINTNNASGDpSPDTKK | YHR186C | FAVANLpSTMpSLVNNPALQSR |
| S45866 | AHNVpSTSNNSPpSTDNDSISK | YML029W | SQpSPVpSFAPTQGR |
| S45866 | SYpTNTTKPK | YML072C | ASpSFAR |
| PAM1 | SDQGNNpSpSGNDSR | YML072C | SPpSNLNSTSVpTPR |
| PAN1 | DASApSSTSpTFDAR | YMR196W | IGGTHSGLpTPQSSISpSDKAR |
| PAN1 | SSpSPSYSQFK | YMR295C | SSIpSNTpSDHDGANR |
| PAT1 | RRpSSpYAFNNGNGATNLNK | YNL136W | HSSSTGNTpSNETpSPK |
| PBS2 | SASVGpSNQSEQDKGSSQpSPK | YNL156C | ILpSASpSIHENFPSR |
| **PDA1** | YGGH**pS**MSDPGTTYR | YNL156C | SVpSIDpSTK |
| PDR5 | TLTAQSMQNpSTQpSAPNK | YNL321W | SHpSVPDLNTATPpSSPKR |
| PEA2 | NTSpSPPIpSPNAAAIQEEDSSK | YOR042W | VVAETTYIDpTPDpTETKKK |
| PFK2 | VHpSYTDLAYR | YOR052C | SSpSNSpSVTSTGQSSR |
| PMA1, PMA2 | VSpTQHEK | YOR175C | KMSFpSGYpSPKPISK |
| POM34 | YAYMMNpSQpSPR | YOR220W | GGSSLpSPDKSSLEpSPTMLK |
| POM34 | YAYMMNSQpSPR | YOR273C | TMEpTDPpSTR |
| PRS5 | KTTpSTSpSTpSSQSSNSSK | YPL247C | SpSISFGSSQR |
| PRS5 | TTpSTSpSTSSQSSNSSK | YPR156C | TEpTVKpSLQDMGVSSK |
| PTR2 | ANDIEILEPMEpSLRpSTTKY | YPR156C | TSpTAIpSR |
| PTR2 | DSYVpSDDVANpSTER | YRO2 | KAQEEEEDVApTDpSE |
| **RAS2** | KM**pS**NAANGK | YSC84 | GYGDFDpSEDEDYDYGR |
| RAS2 | NApSIEpSKTGLAGNQATNGK | ZUO1 | NHTWpSEFER |
| REG1 | RpSDSGVHpSPITDNSSVASSTTSR | TyB protein | TDSSpSADpSDMTSTKKY |

lyzed by nanoflow HPLC on the LCQ ion-trap mass spectrometer. Spectra were acquired with the instrument operating in the data-dependent mode throughout the HPLC gradient. Every 12–15 s, the instrument cycled through acquisition of a full-scan mass spectrum and five MS/MS spectra recorded sequentially on the five most abundant ions present in the initial MS scan. More than 1,500 MS/MS spectra were recorded in this mode of operation during the chromatographic separation (data not shown).

Data acquired in the above experiment were analyzed both by an in-house computer algorithm, the Neutral Loss Tool, and also by SEQUEST (ref. 5). The Neutral Loss Tool searches MS/MS spectra for fragment ions formed by loss of phosphoric acid, 32.6, 49, or 98 Da from the $(M+3H)^{+++}$, $(M+2H)^{++}$, and $(M+H)^{+}$ ions, respectively. Phosphoserine and phosphothreonine, but not phosphotyrosine, lose phosphoric acid readily during the collision activation dissociation process in the ion-trap mass spectrometer. Thus, appearance of fragment ions 32.6, 49, or 98 Da below the triply, doubly, or singly charged precursor ions in peptide MS/MS spectra strongly indicates that the peptide contains at least one phosphoserine or phosphothreonine residue. In the above experiment, more than 1,000 different phosphoserine- or phosphothreonine-containing peptides were detected in the yeast whole-cell lysate with the Neutral Loss Tool.

To identify phosphopeptides in the above sample, MS/MS spectra were searched with the SEQUEST algorithm[5] against the yeast protein database (obtained from the *Saccharomyces* Genome Database (SGD) at http://genome-www.stanford.edu/Saccharomyces/). To date we have manually confirmed assignments made for the 216 sequences shown in Table 1. Of these entries, 60 (28%) are singly phosphorylated, 145 (67%) are doubly phosphorylated, and 11 (5%) are triply phosphorylated. The large percentage of multiply phosphorylated peptides is most likely due to increased avidity of the IMAC column for these peptides relative to singly phosphorylated peptides. Phosphorylation sites that were previously identified are shown in boldface in the table. Two of the sequences derive from MAP kinases encoded by HOG1 and SLT2 and contain the doubly phosphorylated motif pTXpY derived from the activation loop in their respective catalytic domains. This clearly indicates the potential of the phosphoprofiling approach as a measure of cellular activation states. In fact, the list contains 171 different proteins, including abundant species such as the heat-shock proteins (Hsp26p, Hsp30p) and those involved in carbohydrate metabolism (Hxk1p, Hxt2p) and protein synthesis (Rp112AP, Rpl24Ap). Rare proteins such as the cell cycle–regulatory molecules Mob1p and Ash1p, and cytoplasmic proteins such as Myo3p and Pea2p, also appear in the table. Of the 216 peptides in Table 1, 66 have sequences that correspond to a codon bias of <0.1 and are therefore likely to be poorly translated and present in low copy number[6].

Recording MS/MS spectra on the phosphopeptides isolated by the IMAC column and subsequently treated with alkaline phosphatase to remove covalently bound phosphate identified 85 additional phosphopeptides (data not shown). In this experiment, peptide methyl esters were eluted from the IMAC column directly to a second column packed with F7m Polyvinyl spheres containing immobilized alkaline phosphatase. Dephosphorylated peptides were then eluted to

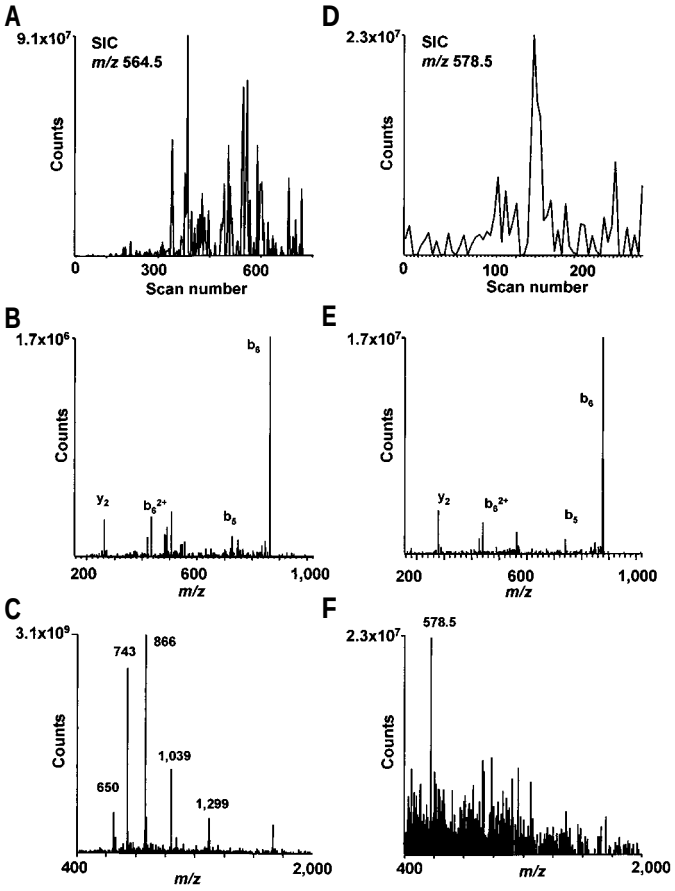


**Figure 1.** Phosphopeptide enrichment by IMAC–LC/MS/MS after blocking of carboxylate groups. (A) Selected-ion chromatogram (SIC) or plot of the ion current versus scan number for *m/z* 564.5 corresponding to the $(M+2H)^{++}$ of the phosphopeptide DRVpYIHPF, present at the level of 10 fmol/μl in a mixture containing tryptic peptides from five proteins, each at the level of 2 pmol/μl. (B) MS/MS spectrum characteristic of the sequence DRVpYIHPF, recorded on ions of *m/z* 564.5 in scans 610–616. (C) Electrospray ionization mass spectrum recorded during this same time interval. Abundant ions from tryptic peptides nonspecifically bound to the IMAC column obscure the signal at *m/z* 564.5 for DRVpYIHPF. (D) SIC for *m/z* 578.5 corresponding to the $(M+2H)^{++}$ ion for the dimethyl ester of DRVpYIPF. (E) MS/MS spectrum characteristic of the sequence DRVpYIPF, recorded on ions of *m/z* 578.5 in scans 151–163. (F) Electrospray ionization mass spectrum recorded in scan 154 showing the parent ion, *m/z* 578.5 for the phosphopeptide dimethyl ester, and the absence of signals for tryptic peptides nonspecifically bound to the IMAC column.

the standard nanoflow HPLC column and analyzed on the LCQ ion-trap mass spectrometer using the data-dependent scan protocol described above. This approach has the advantage that the resulting MS/MS spectra usually contain a larger number of abundant, sequence-dependent fragment ions than those recorded on the corresponding phosphorylated analogs. This, in turn, improves the likelihood that the SEQUEST algorithm will find a unique match in the protein database. The disadvantage of the protocol is that the resulting MS/MS spectra no longer contain information on the number and location of the phosphorylated residues within the peptide.

Recently, two groups have described protocols to identify phosphorylation sites on proteins from complex mixtures. The approach employed by Chait[7] involves digestion of cellular proteins with trypsin and replacement of phosphate groups of serine and threonine phosphopeptides by ethanedithiol. The resulting derivatives are combined with a biotin affinity tag, separated from nonphosphorylated tryptic peptides by affinity chromatography on NeutrAvidin beads, and then characterized by mass spectrometry. Disadvantages of this approach include the failure to detect phosphotyrosine-containing peptides and generation of diastereoisomers in the derivatization step. The approach suggested by the Aebersold group circumvents these problems but involves a six-step derivatization/purification protocol for tryptic peptides that requires more than 13 h to complete and affords only a 20% yield[8]. Mass spectrometry was employed to characterize the phosphopeptides isolated in this approach as well. Application of this technology to a total protein extract from yeast identified 24 phosphopeptides derived from 12 of the most abundant proteins in the yeast genome.

Finally, we note that the above methodology can be modified easily to allow quantitation and/or differential display of phosphoproteins expressed in two different samples. For this experiment, peptides are converted to methyl esters from one sample with methanol and from the other sample with deuterated methanol. The two samples are combined and fractionated by IMAC, and the resulting mixture of labeled and unlabeled phosphopeptides is then analyzed by nanoflow HPLC–electrospray ionization on a home-built Fourier transform mass spectrometer[9]. This instrument operates with a detection limit in the low-attomole level. Signals for peptides present in both samples appear as doublets separated by $n(3\ \mathrm{Da})/z$ (where $n$ = the number of carboxylic acid groups in the peptide and $z$ = the charge on the peptide). The ratio of the two signals in the doublet changes as a function of expression level of the particular phosphoprotein in each sample. Peptides of interest are then targeted for sequence analysis in a subsequent analysis performed on the ion-trap instrument as discussed above. Initial data from these experiments are promising, and results will be reported in a subsequent paper.

In conclusion, this methodology virtually eliminates nonspecific binding to the IMAC column, thereby allowing for the identification of hundreds of phosphorylation sites on proteins from whole-cell lysate in a single analysis. The sensitivity of the method was shown by the detection of 5 fmol of a phosphorylated peptide in a standard mixture, by the mapping of phosphorylation sites on proteins with very low codon bias, and by the identification of sites of tyrosine phosphorylation on proteins in whole-cell lysate.

## Experimental protocol

Preparation of standard mixture. Five nonphosphorylated proteins (glyceraldehyde 3-phosphate dehydrogenase, BSA, carbonic anhydrase, ubiquitin, and β-lactoglobulin, 100 nmol each, all from Sigma Chemical Co., St. Louis, MO) in 1.1 ml of 100 mM ammoniun bicarbonate (pH 8) were digested with trypsin (20 μg) (Promega, Madison, WI) for 24 h at 37°C. The reaction was quenched with 65 μl of glacial acetic acid, and the mixture was then diluted to a final volume of 50 ml with 0.1% acetic acid (2 μl of a 250 pmol/μl stock solution). To this solution were added 500 pmol of HPLC-purified phosphopeptide DRVpYIHPF (Novabiochem, San Diego, CA) in 0.1% acetic acid. An aliquot of the standard mixture (100 μl) was lyophilized and redissolved in 100 μl of 2 M methanolic HCl. This latter solution was prepared by dropwise addition of 160 μl of acetyl chloride with stirring to 1 ml of methanol. Esterification was allowed to proceed for 2 h at room temperature. Solvent was removed by lyophilization and the resulting sample was redissolved in 100 μl of solution containing equal volumes of methanol, water, and acetonitrile. Phosphate methyl esters are not observed under these conditions.

Protein extraction from S. cerevisiae. Yeast strain 2124 MATa ade2-1, ade6-1, leu2-3,112, ura3-52, his3Δ1, trp1-289, can1 cyh2 bar1:KAN (40 ml) was grown in YPD at 23°C to a density of $1 \times 10^7$ cells/ml. The cell pellet was resuspended in 1.5 ml of Trizol (Gibco-BRL, Carlsbad, CA) and cell lysis was performed by homogenization with glass beads in three consecutive sessions of 45 s each in a Fastprep FP120 shaker (Savant, Holbrook, NY). Total yeast protein, free of nucleic acids, was extracted from this yeast lysate using Trizol according to the manufacturer's directions (Gibco-BRL). The protein pellet was resuspended in 1% (v/v) SDS, and dialyzed against 1% SDS using a Slyde-A-Lyzer, 10,000 MW cutoff (Pierce, Rockford, IL) to remove small mol-

ecules and stored at –80°C. To follow the removal of nucleotide, 0.1 µl of a [$^{32}$P] CTP (Amersham-Pharmacia, Amersham, UK) was added to a 10 ml equivalent of lysed cells. Aliquots were removed after each step in purification, and the amount of nucleotide was quantitated by scintillation with Scintisafe EconoF (Fischer, Pittsburgh, PA). Yeast protein, 500 µg (~10 nmol), in 500 µl of 100 mM ammonium acetate (pH 8.9) was digested with trypsin (20 µg; Promega) overnight at 37°C. Solvent was removed by lyophilization and the residue was reconstituted in 400 µl of 2 M methanolic HCl and allowed to stand at room temperature for 2 h. Solvent was lyophilized and the resulting peptide methyl esters were dissolved in 120 µl of a solution containing equal parts of methanol, water, and acetonitrile. An aliquot corresponding to 20% of this material (2 nmol of yeast protein) was subjected to chromatography and mass spectrometry as described below.

Chromatography. Construction of IMAC columns has been described[10]. Briefly, a 360 µm o.d. × 100 µm i.d. fused-silica column (Polymicro Technologies, Phoenix, AZ) was packed with 8 cm POROS 20 MC (PerSeptive Biosystems, Framingham, MA). Columns were activated with 200 µl 100 mM $FeCl_3$ (Aldrich, Milwaukee, WI) and loaded with either 0.5 µl of the above standard mixture or sample corresponding to peptides derived from 100 µg (10 nmol) of protein extract from *S. cerevisiae*. To remove nonspecific binding peptides, the column was washed with a solution containing 100 mM NaCl (Aldrich,) in acetonitrile (Mallinckrodt, Paris, KY), water, and glacial acetic acid (Aldrich) (25:74:1, vol/vol/vol). For sample analysis by mass spectrometry, the affinity column was connected to a fused-silica precolumn (6 cm of 360 µm o.d. × 100 µm i.d.) packed with 5–20 µm C18 particles (YMC, Wilmington, NC). All column connections were made with 1 cm of 0.012-inch i.d. × 0.060-inch o.d. Teflon tubing (Zeus, Orangeburg, SC). Phosphopeptides were eluted to the precolumn with 10 µl 50 mM $Na_2HPO_4$ (pH 9.0; Aldrich), and the precolumn was then rinsed with several column volumes of 0.1% acetic acid to remove $Na_2HPO_4$. The precolumn was connected to the analytical HPLC column (360 µm o.d. × 100 µm i.d. fused silica) packed with 6–8 cm of 5-µm C18 particles (YMC). One end of this column contained an integrated laser-pulled electrospray ionization emitter tip (2–4 µm in diameter)[9]. Sample elution from the HPLC column to the mass spectrometer was carried out with a gradient consisting of 0.1% acetic acid and acetonitrile. For removal of phosphate from the tryptic peptides, the IMAC column was connected to a fritted 360 µm o.d. × 200 µm i.d. fused-silica capillary packed with F7m (Polyvinyl spheres), containing immobilized alkaline phosphatase (MoBiTech, Marco Island, FL). Phosphopeptides were eluted from the IMAC column through the phosphatase column onto a precolumn with 25 µl of 1 mM EDTA (pH 9.0), and the precolumn was then rinsed with several column volumes of 0.1% acetic acid to remove EDTA. The precolumn was connected to an analytical HPLC column. Sample elution from the HPLC column to the mass spectrometer was carried out with a gradient consisting of 0.1% acetic acid and acetonitrile.

Mass spectrometry. All samples were analyzed by nanoflow HPLC–microelectrospray ionization on a Finnigan LCQ ion trap (San Jose, CA). A gradient consisting of 0–40% B in 60 min, 40–100% B in 5 min (A = 100 mM acetic acid in water, B = 70% acetonitrile, 100 mM acetic acid in water) flowing at ~10 nl/min was used to elute peptides from the reverse-phase column to the mass spectrometer through an integrated electrospray emitter tip[9]. Spectra were acquired with the instrument operating in the data-dependent mode throughout the HPLC gradient. Every 12–15 s, the instrument cycled through acquisition of a full-scan mass spectrum and five MS/MS spectra (3 Da window; precursor $m/z$ ± 1.5 Da, collision energy set to 40%, dynamic exclusion time of 1 min) recorded sequentially on the five most abundant ions present in the initial MS scan. To do targeted analysis of the phosphopeptide in the standard mixture, the ion-trap mass spectrometer was set to repeat a cycle consisting of a full MS scan followed by an MS/MS scan (collision energy set to 40%) on the $(M+2H)^{++}$ of DRVpYIHPF or its methyl ester ($m/z$ 564.5 and 578.5, respectively). The gradient employed for this experiment was 0–100% B in 30 min for the underivatized sample, 0–100% B in 17 min for derivatized sample (A = 100 mM acetic acid in water, B = 70% acetonitrile, 100 mM acetic acid in water).

Database analysis. All MS/MS spectra recorded on tryptic phosphopeptides derived from the yeast protein extract were searched against the *S. cerevisiae* protein database by using the SEQUEST algorithm[5]. Search parameters included a differential modification of +80 Da (presence or absence of phosphate) on serine, threonine, and tyrosine and a static modification of +14 Da (methyl groups) on aspartic acid, glutamic acid, and the C terminus of each peptide.

*Acknowledgments*

*This work was supported by US Public Health Service Grant, GM 37537 to D.F.H.*

*Competing interests statement*

*The authors declare that they have competing financial interests: see the* Nature Biotechnology *website (http://biotech.nature.com) for details.*

Received 28 November 2001; accepted 9 January 2002

1. Hubbard, M.J. & Cohen, P. On target with a new mechanism for the regulation of protein phosphorylation. *Trends Biochem. Sci.* **18**, 172–177 (1993).
2. Annan, R., Huddleston, M., Verma, R., Deshaies, R. & Carr, S. A multidimensional electrospray MS-based approach to phosphopeptide mapping. *Anal. Chem.* **73**, 393–404 (2001).
3. Andersson, L. & Porath, J. Isolation of phosphoproteins by immobilized metal ($Fe^{3+}$) affinity chromatography. *Anal. Biochem.* **154**, 250–254 (1986).
4. Michel, H., Hunt, D.F., Shabanowitz, J. & Bennett, J. Tandem mass spectrometry reveals that three photosystem II proteins of spinach chloroplasts contain *N*-acetyl-*O*-phosphothreonine at their $NH_2$ termini. *J. Biol. Chem.* **263**, 1123–1130 (1988).
5. Eng, J., McCormack, A.L. & Yates, J.R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
6. Bennetzen, J.L. & Hall, B.D. Codon selection in yeast. *J. Biol. Chem.* **257**, 3026–3031 (1982).
7. Oda, Y., Nagasu, T. & Chait, B. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat. Biotechnol.* **19**, 379–382 (2001).
8. Zhou, H., Watts, J. & Aebersold, R. A systematic approach to the analysis of protein phosphorylation. *Nat. Biotechnol.* **19**, 375–378 (2001).
9. Martin, S.E., Shabanowitz, J., Hunt, D.F. & Marto, J.A. Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **72**, 4266–4274 (2000).
10. Zarling, A.L. *et al.* Phosphorylated peptides are naturally processed and presented by major histocompatibility complex class I molecules *in vivo*. *J. Exp. Med.* **192**, 1755–1762 (2000).