NRC·CNRC

NRC Publications Archive
Archives des publications du CNRC

Identification of Vascular Breast Tumor Markers by Laser Capture Microdissection and Label-Free LC-MS.

Hill, Jennifer J.; Tremblay, Tammy-Lynn; Pen, Ally; Li, Jie; Robotham, Anna C.; Lenferink, Anne E. G.; Wang, Edwin; O'Connor-McCourt, Maureen; Kelly, John F.

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.
For the publisher's version, please access the DOI link below./ Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:
https://doi.org/10.1021/pr101267k
Journal of Proteome Research, 10, 5, pp. 2479-2493, 2011-05-06

NRC Publications Record / Notice d'Archives des publications de CNRC:
https://nrc-publications.canada.ca/eng/view/object/?id=230c95eb-0af9-4328-816e-4cfb4e812fa9
https://publications-cnrc.canada.ca/fra/voir/objet/?id=230c95eb-0af9-4328-816e-4cfb4e812fa9

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at https://nrc-publications.canada.ca/eng/copyright
READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site https://publications-cnrc.canada.ca/fra/droits
LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

National Research Council Canada | Conseil national de recherches Canada

Canada

Manuscript Title: Identification of Phosphoproteins in Arabidopsis thaliana Leaves Using Polyethylene Glycol Fractionation, Immobilized Metal-ion Affinity Chromatography, Two-Dimensional Gel Electrophoresis and Mass Spectrometry

Manuscript No: MCP/2011/012401

Manuscript Type: Research Article

Date Submitted by the Author: 15 Jul 2011

Complete List of Authors: Uma K. Aryal, Andrew R.S. Ross, and Joan E. Krochko

Keywords: 2-D Gel Electrophoresis; Peptides*; Phosphoproteins*; Post-translational modifications* ; Tandem Mass Spectrometry; Arabodopsis thaliana ; Phosphorylation; Polyethylene glycol; Protein IMAC ; Rubisco

Identification of Phosphoproteins in *Arabidopsis thaliana* Leaves Using Polyethylene Glycol Fractionation, Immobilized Metal-ion Affinity Chromatography, Two-Dimensional Gel Electrophoresis and Mass Spectrometry

Uma K. Aryal[#†], Joan E. Krochko and Andrew R.S. Ross[‡]

National Research Council, Plant Biotechnology Institute, 110 Gymnasium Place, Saskatoon, Saskatchewan, Canada S7N 0W9.

†Corresponding author: E-mail: Uma.Aryal@pnnl.gov

[#] Present address: Pacific Northwest National Laboratory, P.O. Box 999, 902 Battelle Boulevard, Richland, WA 99352.

[‡] Present address: Fisheries and Oceans Canada, Institute of Ocean Sciences, P.O. Box 6000, 9860 West Saanich Road, Sidney, British Columbia, Canada V8L 4B2.

Running title: Phosphoproteomic analysis of *Arabidopsis thaliana* leaf

Keywords: protein IMAC, phosphorylation, polyethylene glycol, two-dimensional gel electrophoresis, Rubisco, photosystem

Abbreviations: 1-D PAGE, one-dimensional polyacrylamide gel electrophoresis; 2-D PAGE, two-dimensional polyacrylamide gel electrophoresis; ADP, adenosine diphosphate; AtLFNR, *Arabidopsis thaliana* leaf-type ferredoxin-$NADP^+$-oxidoreductase; ATP, adenosine triphosphate; BPB, bromophenol blue; BR, brassinosteroid; CID, collision induced dissociation; DIGE, difference gel electrophoresis; DMSO, dimethylsulfoxide; DTT, dithiothreitol; ESI, electrospray ionization; FBA, fructose bisphosphate aldolase; IAA, iodoacetamide; IEF, isoelectric focusing; IMAC, immobilized metal-ion affinity chromatography; IPG, immobilized pH gradient strip; LC-MS/MS, liquid chromatography tandem mass spectrometry; LSU, large subunit (Rubisco); MD, multi-dimensional; MS, mass spectrometry; PEG, polyethylene glycol; PGK, phosphoglycerate kinase; PMSF, phenylmethylsulfonyl fluoride; PSI, photosystem I; PSII, photosystem II; PVPP, polyvinylpolypyrrolidone; Q-TOF, quadrupole time-of-flight; Rubisco, ribulose bisphosphate carboxylase/oxygenase; Ser, serine; SSU, small subunit (Rubisco); TBP, tributylphosphine; TEMED, N,N,N',N'-tetra-methyl-ethylenediamine; TFA, trifluoroacetic acid; Thr, threonine; Tyr, tyrosine.

Summary

Reversible protein phosphorylation is a key regulatory mechanism in cells. Identification and characterization of phosphoproteins requires specialized enrichment methods due to the relatively low abundance of these proteins, and in plants this is further complicated by the high abundance of Rubisco protein in green tissues. We present a novel method for plant phosphoproteome analysis that firstly depletes Rubisco by polyethylene glycol fractionation and then utilizes immobilized metal-ion affinity chromatography to enrich for phosphoproteins. Subsequent protein separation by one- and two-dimensional gel electrophoresis is further improved by extracting the PEG-fractionated protein samples with SDS/phenol and methanol/chloroform to remove interfering compounds. Using this approach we identified 132 phosphorylated proteins in a partial *Arabidopsis* leaf extract. These proteins are involved in a range of biological processes, including CO_2 fixation, protein assembly and folding, stress response, redox regulation and cellular metabolism. Both the large and small subunits of Rubisco were phosphorylated at multiple sites, while the depletion of Rubisco enhanced detection of less abundant phosphoproteins, including those associated with state transitions between photosystems I and II. The discoveries of a phosphorylated form of AtGRP7, a self-regulating RNA-binding protein that affects floral transition, as well as several previously uncharacterized ribosomal proteins confirm the utility of this approach for phosphoproteome analysis and its potential to increase our understanding of signal transduction during growth and development in plants.

INTRODUCTION

Recent improvements in protein analysis techniques, including mass spectrometry (MS), have brought proteomics to the forefront of methods for biological research (1). Characterization of posttranslational modifications of proteins is important for interpreting their biological functions and functional states (2). Phosphorylation is one of the most widespread of protein modifications, affecting approximately one third of proteins in the proteome and playing an important role in many cellular processes (3). Although the study of protein phosphorylation holds particular promise for dissecting signaling pathways, the identification of phosphorylated proteins and phosphorylation sites can be challenging because cellular levels of phosphorylated proteins are often too low for detection using conventional methods (4, 5). The size, charge and hydrophobicity of tryptic phosphopeptides also make it difficult to identify and determine the stoichiometry of phosphorylation on individual proteins using MS (4).

Two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) has the ability to resolve complex protein mixtures on the basis of two independent variables (isoelectric point and molecular mass) and has long been recognized as a key technology in proteome research (6). It is particularly well-suited to detecting changes in protein phosphorylation since each phosphorylation event produces an incremental reduction in pI, resulting in the characteristic horizontal row of protein spots on 2-D gels (5). The combination of 2-D PAGE and MS provides a powerful platform for visualizing relative abundance and for identifying proteins and their post-translational modifications in complex samples.

Good sample preparation is the key to achieving reproducible, high quality gels (7) and recent improvements in sample preparation, as well as protein visualization and MS techniques (8), have led to a significant increase in the number of plant proteomic studies employing 2-D PAGE (9). The plant proteome is highly dynamic, and more complex than the genome due to alternative splicing and posttranslational modifications (7). Plants also produce a large number of non-proteinaceous compounds that can interfere with 2-D PAGE separations (10). These interfering metabolites are particularly abundant in leaves and other green tissues, accumulate in the vacuole as soluble compounds (11, 12), and may be difficult to remove. Effective sample preparation procedures are therefore critical in obtaining protein extracts of sufficient purity for plant proteome analyses.

Although 2-D PAGE has been widely used to identify disease biomarkers and signaling molecules (13, 14) it has not always been effective in detecting low abundance proteins (5, 14). Even difference gel electrophoresis (DIGE), which is more sensitive than conventional 2-D PAGE, did not detect any of the known brassinosteroid (BR) signaling components in *Arabidopsis thaliana* total protein extracts (14). 2-D PAGE analysis of plant protein extracts is further complicated by the high abundance of ribulose bisphosphate carboxylase/oxygenase (Rubisco), which comprises about 30-60% of total soluble plant protein in leaves (15, 16). On polyacrylamide gels Rubisco co-migrates with less abundant proteins, thus inhibiting or preventing their detection (17). Its high relative abundance also limits the total amounts of other proteins that can be loaded for 2-D PAGE (16). Rubisco depletion columns recently have become available commercially (15); however, these products are expensive and, in our experience, the

buffer system supplied with the product is not very efficient in solubilizing hydrophobic proteins. Sub-cellular fractionation can also be effective in depleting Rubisco, although this approach involves additional steps that increase the salt and buffer content of the sample and require stringent sample clean-up prior to analysis. Polyethylene glycol (PEG) is a non-toxic, water-soluble synthetic polymer that has been used as a fractional precipitating agent to purify proteins from a variety of sources (18, 19). PEG fractionation has also been used to deplete Rubisco in plant extracts (20, 21) but is known to cause ion signal suppression and interference during MS analysis (22, 23). Residual PEG can also interfere with 2-D PAGE separations and may darken the background of stained gels.

To address these problems we have systematically evaluated a number of different purification methods in combination with 2-D PAGE and liquid chromatography-tandem mass spectrometry (LC-MS/MS) to analyze the *Arabidopsis thaliana* leaf proteome. Step-wise PEG fractionation significantly reduced the amount of Rubisco in the sample, while subsequent extraction of the PEG-fractionated samples with SDS/phenol (10) and methanol/chloroform (24) consistently enhanced the quality of the resulting 2-D gels. In combination with LC-MS/MS analysis, this protocol was used to identify proteins extracted from *Arabidopsis thaliana* leaves. In addition, for the first time, we have combined PEG fractionation and IMAC enrichment methods to achieve comprehensive mapping of *in vivo* protein phosphorylation sites in *Arabidopsis thaliana* proteins. During this study we found that both the large and small sub-units of Rubisco are phosphorylated in *Arabidopsis* and that depletion of Rubisco is necessary for efficient detection and identification of less abundant phosphoproteins in leaf tissue.

EXPERIMENTAL PROCEDURES

Chemicals – Acrylamide/bisacrylamide solution, IPG strips (pH 3-10, non-linear, 17 cm), carrier ampholytes, Precision Plus Protein™ standard markers, N,N,N',N'-tetra-methyl-ethylenediamine (TEMED), tributyl phosphine (TBP), dithiothreitol (DTT), and iodoacetamide (IAA) were purchased from Bio-Rad (Hercules, CA). Urea was obtained from Merck KGaA (Darmstadt, Germany), tris(hydroxymethyl) amino methane (tris base) from Roche Diagnostics (Indianapolis, IN), PPS Silent Surfactant from Protein Discovery Inc. (Knoxville, TN), and modified porcine trypsin (sequencing grade) from Promega (Madison, WI). Water was purified using the Milli-Q system (Millipore, Bedford, MA). All other chemicals were obtained from Sigma-Aldrich (St. Louis, MO) unless otherwise stated, and were of analytical research grade.

Total Leaf Protein Extraction– Robust cauline leaves were harvested from 2 month old *Arabidopsis thaliana* (L.) Heynh (Col-0) plants cultivated in a growth chamber at 22ºC with a 16 h-light/8 h-dark cycle. Protein was extracted according to Kim et al. (20), with modifications. Briefly, 2 g of leaves were ground to a powder in liquid nitrogen with 0.5% (w/w) polyvinylpolypyrrolidone (PVPP) and homogenized in 10 ml of ice-cold Mg/TX-100 extraction buffer containing 2% v/v Triton X-100, 0.5 M Tris-HCl (pH 8.3), 20 mM $MgCl_2$, and 2% v/v β-mercaptoethanol (where Triton X-100 replaces NP-40 in the original buffer formulation; see reference (20)). Phenylmethylsulfonyl fluoride (PMSF), freshly prepared in dimethylsulfoxide (DMSO) at 500 mM, was added to the buffer to a final concentration of 1 mM. The buffer also contained 0.2% (v/v) of a protease inhibitor cocktail developed for plant cell and tissue extracts (P-9599, Sigma-Aldrich), along with

25 mM sodium fluoride, 1 mM sodium molybdate, 1mM sodium orthovanadate, and 1 mM sodium β-glycerophosphate as phosphatase inhibitors to maintain integrity at phosphorylation sites. The slurry was stirred for 1 h on ice and filtered through four layers of cheese cloth before centrifugation at 3200 g for 20 min at 4 ºC (Eppendorf 5810R Centrifuge, Hamburg, Germany). The cell-free supernatant (1st Supernatant) was transferred to a clean tube, combined with 5 volumes of ice-cold methanol containing 100 mM ammonium acetate, and proteins were precipitated overnight at -20 ºC. After centrifugation at 3200 g for 20 min at 4 ºC the resulting pellet was thoroughly washed twice with ice-cold 100% methanol.

One third of this non-fractionated sample (NF') was transferred to a new tube, washed 3 more times with ice-cold 80% methanol, and the pellet dried in a vacuum centrifuge (Model DNA 120; Thermo Savant, Colin Drive, NY). The pellet was re-dissolved in 0.5 mL of lysis buffer (7 M urea, 2 M thiourea, 4% CHAPS) and aliquots used for protein quantification (Bradford method; BioRad) and 2-D PAGE. In order to determine how additional sample cleanup would affect protein resolution on 2-D PAGE, the remaining two-thirds of the unpurified (NF') protein sample was extracted with SDS/phenol and methanol/chloroform (referred to as Phenol extraction in Fig. 1 and described in detail below) to obtain a purified (NF) sample.

PEG Fractionation of Proteins– Step-wise PEG fractionation of the protein sample was introduced at the stage of the first supernatant, following initial centrifugation of the crude protein extract (1st Supernatant, Fig. 1). Serial PEG fractionation was carried out using 5, 10 and 15% (w/v) PEG (av. mol. wt. 3350). PEG was added to the first

supernatant to a final concentration of 5% (w/v) and incubated on ice for 30 min with intermittent vortexing. The resultant protein precipitate (F1' pellet) was collected by centrifugation at 3200 g for 20 min. The F1' supernatant was mixed with PEG to 10% and treated as before to obtain the 10% PEG-precipitated protein fraction (F2' pellet). Likewise the F2' supernatant was mixed with PEG to 15% (w/v) to obtain the 15% PEG-precipitated fraction (F3' pellet). All three protein pellets were rinsed briefly twice, without breaking the pellet apart, with 2 ml ice-cold Milli-Q water to remove any residual Mg/TX-100 buffer before further purification using the SDS/phenol and methanol/chloroform protocol (Phenol extraction; see Fig. 1 and below).

The supernatant from the 15% PEG precipitation (F3' supernatant) was collected, mixed with 5 volumes of ice-cold methanol containing 100 mM ammonium acetate, and proteins precipitated overnight at -20ºC. The resulting precipitate (F4' pellet) was collected by centrifugation at 3200 g for 20 min and rinsed twice with ice-cold 100% methanol. All four protein pellets (F1' through F4') were split in the same way as the non-fractionated (NF') sample (see above) for a comparative evaluation of the effects of additional extractions with SDS/phenol and methanol/chloroform on 2-D PAGE protein separation.

SDS/Phenol and Methanol/Chloroform Extraction – Each protein pellet was re-suspended in 0.8 ml of SDS buffer containing 30% sucrose, 2% SDS, 0.1 M Tris-HCl (pH 8.0), and 2% β-mercaptoethanol. The buffer also contained several phosphatase inhibitors: 25 mM sodium fluoride, 1 mM sodium molybdate, 1mM sodium orthovanadate, and 1 mM sodium β-glycerophosphate. After complete suspension of the pellet in SDS buffer, 0.8 ml of Tris-buffered phenol (pH 7.9±0.2) was added and the

194 mixture kept on ice for 10 min with intermittent vortexing. The aqueous and phenol
195 phases were separated by centrifugation at 3200 g for 5 min and the upper phenol
196 phase carefully pipetted into a clean tube without disturbing the white SDS complex
197 formed at the solvent interface. The lower aqueous phase was re-extracted with an
198 additional 0.6 ml phenol. The pooled phenol aliquots were mixed with 5 volumes of
199 methanol containing 100 mM ammonium acetate and proteins precipitated at -20ºC
200 overnight. After centrifugation at 18500 g for 15 min the resulting pellet was washed
201 once with 100% methanol before further purification by treatment with
202 methanol/chloroform (24). Briefly, 600 μl of methanol (filtered) and 200 μl of chloroform
203 (filtered) were added to each tube in a flow hood and mixed thoroughly by vortexing.
204 Then 800 μl of water was added, the mixture vortexed and then centrifuged for 5 min at
205 18500 g. After carefully removing the upper phase with a gel-loading tip, to avoid
206 disturbing the protein disc formed at the solvent interface, another 500 μl of methanol
207 was added and the proteins collected by centrifugation at 18500 g This final pellet was
208 dried and re-dissolved in lysis buffer (7 M urea, 2 M thiourea, 4% CHAPS). Protein
209 concentrations were determined for all purified samples (NF and F1 through F4) and
210 volumes adjusted with lysis buffer to obtain aliquots of 2 mg/ml for 2D-PAGE and 6
211 mg/ml for IMAC**.**
212 *Immobilized Metal-Ion Affinity Chromatography (IMAC)*– IMAC was performed on each
213 purified protein extract (sample NF and F1 through F4) following SDS/phenol and
214 methanol/chloroform extraction. PHOS-Select^{TM} iron affinity gel beads (Sigma) were
215 carefully stirred until completely and uniformly suspended in the stabilizing buffer
216 supplied. Five hundred µl of the resulting slurry was transferred to a Sigma Prep spin

column (SC 1000; Sigma) and washed several times with 0.1% TFA in 30% acetonitrile to ensure complete removal of the stabilizing buffer (which contains glycerol). The beads were equilibrated three times with 500 µl of wash buffer (6 M urea, 0.25% CHAPS, 50 mM sodium acetate, pH 4.0) and centrifuged at 500 g for 1 min at each washing step.

For IMAC, protein samples at 6 mg/ml in lysis buffer were diluted to 1 mg/ml with wash buffer (6 M urea, 0.25% CHAPS, 50 mM sodium acetate, pH 4.0) (5). Five ml of diluted sample was loaded onto each spin column (i.e. ~ 5 mg protein per 500 µl of bead slurry) and incubated for 1 h at room temperature with gentle agitation in a shaker (Labquake, Barnstead International, Iowa). The residual sample flow-through from the column was removed by centrifugation at 500 g for 1 min. The column was washed five times with 500 µl of the wash buffer and centrifuged for 1 min at 500 g each time to remove unbound proteins. The washings were discarded and the bound phosphoproteins eluted with three 200 µl aliquots of elution buffer (6 M urea, 0.25% CHAPS, 50 mM tris-acetate pH 7.5, 0.1 M EDTA, 0.1 M EGTA). The column was incubated at room temperature for 10 min with gentle shaking for each aliquot of elution buffer and phosphoproteins collected by centrifugation at 1000 rpm for 1 min. The pooled eluates (containing phosphoproteins) were mixed with 5 volumes of 100% methanol containing 100 mM ammonium acetate and proteins precipitated overnight at -20°C. The resulting pellets were collected by centrifugation at 14,000 rpm for 15 min, then washed three times with 80% ice-cold methanol, vacuum-dried, and re-suspended in lysis buffer prior to gel electrophoresis.

240 *Gel Electrophoresis* – For 1-D PAGE, 10 µl (20 µg protein) from each protein sample in
241 lysis buffer was mixed with 10 µl of sample buffer (0.2 M Tris-HCl pH 6.8, 2% SDS, 10%
242 glycerol and 0.02% bromophenol blue) and electrophoresed on a 1.0 mm, 12.5%
243 Criterion Tris/HCl gel in a Criterion Cell (Bio-Rad) at a constant voltage of 150 V. The
244 separated proteins were then visualized using Bio-Safe Coomassie Blue stain (Bio-
245 Rad). For 2-D PAGE, 200 µl (400 µg protein) of each sample in lysis buffer was mixed
246 with 200 µl of the rehydration buffer (7 M urea, 2 M thiourea, 4% CHAPS, 20 mM DTT
247 and 0.5% IPG buffer, pH 3-10) and applied to a Ready Strip^{TM} IPG strip (pH 3-10, non-
248 linear, 17 cm; Bio-Rad). The strips were actively rehydrated for 12 h at 50V and 20ºC in
249 a Protein IEF Cell (Bio-Rad). Isoelectric focusing (IEF) was then performed at 20ºC at a
250 maximum 50 µA/strip using the gradient voltage program to reach a total of 80,000 V-h.
251 Following IEF the IPG strips were gently rinsed with 1× SDS electrophoresis buffer
252 (0.025 M Tris, 0.192 M glycine and 0.1% SDS). Reduction and alkylation of proteins
253 was performed by treating the strips sequentially in equilibration buffer (6 M urea, 2%
254 SDS, 375 mM Tris-HCl pH 8.8, 20% glycerol) containing 2% DTT and 2.5%
255 iodoacetamide, respectively, for 20 min each. The strips were transferred to 12% SDS
256 polyacrylamide gels mounted in an Ettan Dalt six-gel system (Amersham Biosciences,
257 San Francisco, CA) and SDS-PAGE performed first at 2 W/gel for 1 h, and then at 17
258 W/gel until the BPB dye front reached the lower edge of the gel. The separated proteins
259 were visualized by silver staining using a Hoefer Processor Plus unit (GE Healthcare)
260 and a modified version of the Vorum protocol (25) as described by Wang et al. (26). Gel
261 images were recorded using an ImageScanner (GE Healthcare) and Phoretix 2D

262 software (v2004) was used to measure the total number of protein spots visualized in
263 each 2-D gel image.
264
265 *In-Gel Protein Digestion*– Proteins were excised from the 2-D gels, arranged in 96-well
266 microtitre plates and automatically digested with trypsin using a MassPREP protein
267 digestion station and protocol recommended by the manufacturer (digestion 5.0;
268 Micromass, Manchester, U.K.). Fifteen µl of a 0.2% (w/v) PPS Silent Surfactant solution
269 prepared in 50 mM ammonium bicarbonate (pH 7.8) was added manually to each
270 sample during the 15-min dehydration period immediately following the gel de-staining
271 step and prior to the addition of DTT. Tryptic peptides were extracted automatically 3
272 times using 1% formic acid in 4% acetonitrile, and then twice manually using 1% TFA in
273 75% acetonitrile. The extracted peptides were vacuum-dried, re-suspended in 20 µl of
274 aqueous 1% TFA, and the microtitre plates sealed using an ALPS 300^{TM} heat sealer
275 (ABgene, UK) in preparation for liquid chromatography and MS analysis.
276
277 *Mass Spectrometry and Protein Identification*– Six µl of the tryptic peptides was
278 analyzed by LC/ESI-MS/MS using a nanoAQUITY UPLC system (Waters, Milford, MA,
279 USA) interfaced to a quadrupole time-of-flight (Q-TOF) Ultima Global hybrid tandem
280 mass spectrometer fitted with a Z-spray nanoelectrospray ion source (Waters,
281 Mississauga, ON, Canada). The peptides were separated using a Waters BEH130 C18
282 analytical column (75 µm, 1.75 mm × 100 mm) at a flow rate of 400 nl/min. Mobile
283 phase solvent A was 0.2% formic acid in water and solvent B was 0.2% formic acid in
284 100% acetonitrile. Separations were performed using the following 55-min solvent

program: 99:1 (%A:%B) for 1 min, changing to 90:10 at 16 min, 55:45 at 45 min, and 20:80 at 46 min, at which point the flow rate was increased from 400 nl/min to 800 nl/min and the solvent concentrations held until 52 min before reverting to 99:1 (%A:%B). A 5 min seal wash with 10% acetonitrile in water was carried out after the completion of each run.

The column effluent was directed to the nanoES source of a Q-TOF mass spectrometer operating in the positive ion mode with a source temperature of 80ºC, and which had been calibrated using the product ion spectrum of Glu-fibrinopeptide B acquired over the mass-to-charge (*m/z)* range 50-1900. TOF MS spectra were acquired over the *m/z* range 400-1900 at the rate of one scan/s, and the 3 most abundant multiply-charged (2^+, 3^+, or 4^+) precursor peptide ions automatically selected for collision-induced dissociation (CID). Product-ion spectra were acquired over the *m/z* range 50-1900 in the MS/MS mode. A real-time exclusion window was used to prevent precursor ions with the same *m/z* from being selected for CID and TOF MS/MS within 2 min of their initial acquisition. Data were also acquired using pre-programmed exclusion lists for keratin and trypsin.

The MS/MS raw data were processed using MassLynx 4.1 (Waters, Milford, MA) and searched against the National Center for Biotechnology Information non-redundant (NCBInr) protein sequence database for *Arabidopsis thaliana* (thale cress) using an in-house Mascot server (Version 2.2, Matrix Sciences, UK). The datasets were searched using carbamidomethylation of cysteine as the fixed modification and oxidation of methionine as the variable modification, with mass tolerances of 0.2 Da for MS and 0.5 Da for MS/MS data. The database searches also included phosphorylation of serine (S),

threonine (T) and tyrosine (Y) with oxidation of methionine as variable modifications. Only tryptic peptides with up to one missed cleavage site were allowed. The following acceptance criteria were used for protein identification: (i) Mascot score greater than 95% confidence threshold, (ii) protein with at least four top-ranking unique peptides with significant ion scores ($P<0.05$); and (iii) protein sequence coverage by the matched peptides was >15%. If the same set of peptides matched multiple members of a protein family, or a protein appeared under a different name and accession number in the database, the entry with highest score and/or most descriptive name was reported. When protein isoforms were observed, the data were inspected manually. If several isoforms shared the same set of identified peptides, only the protein hit with the most matching peptides was accepted as a correct result. The presence of protein isoforms was confirmed and reported based on the identification of at least two unique peptides.

Phosphorylation (79.9663 Da) and sulfation (79.9568 Da) are often difficult to distinguish. Since the error tolerance of the current MS method (200 mDa) is greater than the mass difference between phosphorylation and sulfation (9.5 mDa), protein sulfation is often reported as phosphorylation in Mascot database searches. Therefore, all proteins identified as phosphorylated from the initial search were subjected to a second error-tolerant search, reporting masses to 0.1 mDa, and thus allowing sulfation and phosphorylation to be distinguished. Raw MS/MS spectra matched to phosphorylated peptides in the Mascot search were manually validated using MassLynx 4.1. The spectra were processed to give singly charged, monoisotopic, centroided peaks and compared with the *in silico* fragmentation masses for the matched peptide to confirm the neutral loss of phosphoric acid for serine and threonine phosphorylation, or

the mass increment of 80 Da associated with phosphorylated tyrosine. In order to confirm phosphorylation the peptide MS/MS spectrum had to be of good quality, with fragment ion intensities clearly above baseline (including the neutral-loss peaks associated with de-phosphorylation during CID) and contain three sequential y- or b-type ions.

RESULTS

PEG Fractionation and Purification

Plant proteomics using 2-D PAGE is challenging due to the high abundance of Rubisco, which limits visualization of many low-abundance proteins on 2-D gels (20). One way of overcoming this problem is to deplete Rubisco from the protein extract prior to electrophoresis; however, the additional steps involved often require the use of reagents (e.g. salts, detergents) that are incompatible with downstream protein separations and MS analysis. In this study we used a PEG-fractionation method to deplete Rubisco, followed by a multi-step purification procedure to remove PEG and other contaminants prior to 2-D PAGE and LC-MS/MS. A schematic representation of this fractionation procedure is shown in Figure 1.

The molecular weight distributions of proteins in PEG-fractionated and non-fractionated samples were investigated using 1-D PAGE (Fig. 2). The large and small subunits (LSU and SSU) of Rubisco (Fig. 2; boxed areas) predominated in the non-fractionated (NF) and in the 5% (F1), 10% (F2) and 15% (F3) PEG-fractionated protein samples, but were markedly reduced in the precipitated supernatant of the final PEG fraction (F4). In contrast, many other proteins (indicated by arrows) in the Rubisco-depleted fraction (F4) were enhanced relative to the non-fractionated (NF) control sample.

The effectiveness of contaminant removal using SDS/phenol and methanol/chloroform was assessed for non-fractionated (NF' and NF) and PEG-fractionated (F2' and F2) protein samples. Without the additional cleanup the NF' and F2' protein samples were poorly focused (particularly in the case of acidic proteins) and

showed vertical and horizontal streaking consistent with the presence of salt impurities (Supplementary Fig. S1, A and C). Extensive washing of the protein pellets with methanol or acetone prior to 2-D PAGE did not improve results (data not shown). However, extraction with SDS/phenol and methanol/chloroform resulted in NF and F2 gels with superior resolution and a significant reduction in vertical and horizontal streaking (Supplementary Fig. S1, B and D). Samples purified using the SDS/phenol and methanol/chloroform extraction protocol (i.e. NF and F1 through F4) were therefore used in all subsequent experiments.

2-D PAGE Separation and LC-MS/MS Identification

Purified NF, F2 and F4 protein extracts were resolved using 2-D PAGE (Fig. 3; F2 not shown). At least three biological replicates were prepared independently for each fraction, and the resulting gel images analyzed using Phoretix 2D (v2004) software to identify reproducible spot patterns. Reproducibility was confirmed by LC-MS/MS identification of the same protein in the same position on replicate gels. 2-D PAGE confirmed significant reductions in amounts of Rubisco LSU and SSU in the F4 fraction compared to the NF sample, as observed previously using 1-D PAGE (Fig. 2), thus revealing additional proteins in the F4 gel that had been masked by the subunits of Rubisco in the NF gel (Fig. 3, A and B). An average of 770, 580 and 990 protein spots were observed in the NF, F2 and F4 fractions, respectively, using Phoretix 2D. In one replicate experiment a total of 702 proteins were identified with high confidence in the three fractions with 481, 329, and 595 proteins identified in the NF, F2 and F4 fractions, respectively (Supplementary Fig. S2A). The pattern of distribution of identified proteins

400 across the NF, F2 and F4 fractions suggested that PEG fractionation significantly
401 enhances leaf proteome coverage. On average, Rubisco or its variants were identified
402 in 14% of the spots from the NF gel and 26% from the F2 gel, but only in about 2% of
403 the protein spots from the F4 gel (Supplementary Fig. S2B).
404
405 *Effect of PEG Fractionation on the Identification of Gel-separated Phosphoproteins*
406 The utility of PEG fractionation for plant phosphoproteome analysis was
407 evaluated by comparing the number of phosphorylated proteins identified in NF, F2 and
408 F4 protein samples following 2-D PAGE. In one experiment 36 phosphoproteins were
409 identified in the NF fraction whereas 37 and 77 phosphoproteins were identified in the
410 Rubisco-depleted F2 and F4 fractions, respectively (Fig. 4A). Of the 77 phosphoproteins
411 identified from the F4 sample, 43 were unique to that fraction compared with 7 and 5
412 unique phosphoproteins for the F2 and NF samples, respectively (Fig. 4A). The
413 phosphoproteins represented 12.9 % of the total proteins identified in F4, but only 7.5 %
414 of the proteins in the NF sample. Most of the phosphoproteins identified in NF and F2
415 were relatively common and abundant, whereas many of those identified in F4 were of
416 medium to low abundance, including several proteins that are known to be involved in
417 signal transduction such as paruvilin 1 (spot 229), legume lectin family protein (spot
418 230), rotamase cyp 5 (spot 222) and plasma membrane polypeptide family protein (spot
419 167). A total of 95 distinct phosphoproteins were identified in the three fractions
420 combined, with the majority (77) represented in the F4 sample (Fig. 4A). These data
421 demonstrate the effectiveness of PEG-fractionation for increasing both the number of

422 phosphoproteins identified from plant extracts, as well as the proportion of
423 phosphoproteins and number of low abundance phosphoproteins identified.
424 Protein phosphorylation changes the acidity of the protein and thus moves a
425 protein's final position horizontally across a 2-D gel (IEF/SDS-PAGE). A common
426 feature of 2-D gels is that some protein spots appear to be grouped in these horizontal
427 rows, suggesting that they may represent differentially phosphorylated isoforms of the
428 same protein. We identified such isoforms for several proteins (e.g. PGK, FBA and
429 AtLFNR1) using 2-D PAGE and LC-MS/MS and the relative acidities (positions) of these
430 were consistent with the degree of phosphorylation (Fig. 4B). LC-MS/MS analysis also
431 revealed phosphorylation at multiple sites on both the LSU and SSU of Rubisco
432 (Supplementary Table S1: protein spots 114 and 227). Representative spectra
433 confirming sites of phosphorylation in the LSU and SSU are shown in Fig. 5. Rubisco
434 LSU phosphorylation sites included Ser208, Ser321 (Fig 5A), and Ser341 (Fig. 5B),
435 Ser452, Thr466, and Tyr239, whereas SSU was phosphorylated at Ser71, Thr69,
436 Tyr113 (Fig. 5C) and Tyr133 (Fig**.** 5D). These results support our initial hypothesis that
437 depletion of Rubisco from protein extracts is a prerequisite for successful plant
438 phosphoproteome analysis.
439
440 *IMAC Enrichment of Phosphoproteins following PEG Fractionation*
441 To evaluate protein IMAC for the enrichment and subsequent identification of
442 phosphoproteins following PEG-fractionation, protein pellets from the non-fractionated
443 sample (NF) and three of the PEG-fractionated samples (F1, F2 and F4) were re-
444 suspended in lysis buffer (6 M urea, 2 M thiourea, 2% CHAPS) and subjected to IMAC

purification (see Experimental section). Proteins in the IMAC eluants were first separated by 1-D PAGE to examine the molecular weight distributions of the proteins recovered from each sample (Fig. 6). The most abundant proteins in the IMAC-enriched NF sample were the Rubisco LSU and SSU (Fig. 6). As expected from PEG-fractionation data and phosphorylation evidence, the LSU of Rubisco was more concentrated in the F1 and F2 samples than in the F4 sample after IMAC (Fig. 6). Additionally, there were different and distinct protein bands in the F4 sample on the 1-D gel as compared to the NF lane (Fig. 6). These data show for the F1 and F2 samples that the LSU and SSU of Rubisco were retained and specifically eluted from the IMAC column, and thus confirm the phosphorylation status of both subunits. The increased representation of additional non-Rubisco proteins in the IMAC eluant of the F4 sample where Rubisco had been depleted through PEG-fractionation (Fig. 6)) suggests that large quantities of Rubisco could have inhibited the binding of other less-abundant phosphoproteins through competition for available binding sites during IMAC (27).

Based on these results, the IMAC-enriched F4 fraction was selected for further analysis of phosphoproteins using 2-D PAGE (Fig. 7). A total of 233 proteins were unambiguously identified in the F4 gel by LC-MS/MS, of which 132 (57%) were phosphorylated. These phosphorylated proteins are listed, together with the corresponding matched phosphopeptides, in Supplementary Table S1. Functional categorization and relative numbers of phosphoproteins based on KEGG pathway information (28) and Gene Ontology (GO) analysis are shown in Fig. 7B. Thus, IMAC purification significantly increased the total number of phosphoproteins identified in the F4 sample by 2-D PAGE and LC-MS/MS analysis from 77 (12.9%) in the PEG-

fractionated F4 to 132 phosphoproteins (56.8%) in the PEG-fractionated IMAC-purified F4 (Fig 7A and C). Representative MS/MS phosphopeptides corresponding to several phosphoproteins are shown in Supplementary Figs.S3-S5. Phosphorylated forms of both the LSU and SSU of Rubisco also were identified in the IMAC-purified F4 fraction by 2-D PAGE and LC-MS/MS. Only 11 phosphoproteins that were identified without IMAC purification were missing from the pool of 132 phosphoproteins identified following IMAC enrichment (Supplementary Table S2).

Many of the phosphoproteins identified in the IMAC-enriched F4 fraction also included basic proteins including PYR4 (spot 119; pI 8.93), APX4 (spot 125; pI 8.59), PORB (spot 130; pI 9.23), L protein (spot 134; pI 9.30), EXGTA-1 (spot 135; pI 9.03), grPE family protein (spot 182; pI 9.44), RRF (spot 188; pI 9.39), PSBQ/PSBQ-1/PSBQA (spot 197; pI 9.72), RPL9 (spot 195; pI 9.48), RPL12 (spot 200; pI 9.05) and putative porin (spot 194; pI 8.88).The distribution of IMAC-enriched phosphoproteins also spanned the molecular weight range of the gel, suggesting there was minimal bias towards either acidic or basic proteins or high or low molecular weight proteins under these experimental conditions.

Several phosphoproteins identified in our study, including patellin 1 transporter (spot 104), lectin family protein (spot 230), DREPP family protein (spot 167), CA2 (spot 168), PGK (138), GRF3 (spot 173), immunophilin (spot 218), VMA10 (spot 124), 14-3-3 like protein (spot 206), V-type proton ATPase (spot 191), extracellular dermal glycoprotein (spot 178), prohibitin 4 (spot 148), ATPDIL2-1/MEE30/UNE5 (spot 152), have previously been reported in plasma membrane studies (5, 29-31). This is significant because many plasma membrane proteins are assumed to be

phosphorylated, yet these proteins have been difficult to identify using conventional extraction, gel separation and MS procedures due to their low abundance and hydrophobic nature. Similarly, no experimental data on the phosphorylation of AtLFNR1, PPL1, TGG2, prohibitin, ACC oxidase, TH1, and ribosomal proteins S1, S15, L1, L9 and L12 were found in The *Arabidopsis* Information Resource PhosPhAt 3.0 database (http://phosphat.mpimp-golm.mpg.de/phosphat.html) as of September 30, 2010, although all were identified as phosphoproteins in this study. Furthermore, additional sites of phosphorylation in previously known phosphoproteins were also identified using our methodology. Interestingly, the AtGRP7 (also known as CCR2; cold, circadian rhythm, and RNA binding 2) protein was phosphorylated at Ser132 (SGGGGGYSGGGGpSYGGGGGR). The AtGRP7 is a circadian-regulated RNA binding protein that auto-regulates its expression by influencing alternative splicing of its pre-mRNA and is part of a negative feedback loop (32). Taken together, these results suggest that our approach offers significant advantages for proteome and phosphoproteome analysis of green plant tissues.

DISCUSSION

The need for higher throughput and sensitivity in proteome analysis has resulted in a trend away from gel-based methods towards 'shotgun' or multi-dimensional (MD)LC-MS/MS-based techniques (33). However, these approaches are often complementary in terms of proteome coverage (5), and 2-D PAGE may have certain advantages (34), particularly when it comes to detecting post-translational modifications. For example, phosphorylation increases the acidity of a protein without significantly altering its molecular weight; hence, protein phosphorylation typically manifests itself on 2-D gels as a row of horizontal protein spots, of which the least acidic is usually the unphosphorylated form. This was clearly demonstrated for differentially phosphorylated isoforms of PGK, FBA and AtLFNR1 using our gel-based method (Fig. 4). Regardless of approach used, depletion of Rubisco using PEG fractionation should enhance any plant proteome or phosphoproteome investigation. Furthermore, the removal of reagents (including PEG) that suppress ionization during LC-MS/MS analysis is likely to be at least as important for gel-free as it is for gel-based approaches, both of which should benefit from the methodology described above.

A particular problem associated with gel-based proteomics, which can be addressed using PEG fractionation, is that of overlapping spots on 2-D gels (14, 35). Two possible situations can be envisaged; one in which the co-migrating proteins are of similar abundance, and one in which they differ significantly in this respect. In the latter case, one protein masks the other proteins and may well dominate in terms of spot intensity and/or MS analysis, inhibiting or even preventing detection and identification of

less abundant proteins in the same spot. It was found that about 14% of the protein spots from the non-fractionated sample (NF) on 2-D gels corresponded to Rubisco or its variants, but that PEG fractionation reduced this figure to 2% (F4 sample), allowing many low-abundance proteins to be identified (Fig. S2B). Unlike other sub-cellular fractionation techniques, PEG separates proteins by virtue of their solubility (36), acting as an inert solvent sponge that progressively reduces solvent availability with increasing concentration. This, in turn, increases the effective concentration of each protein until its specific solubility is exceeded and precipitation occurs in that fraction (37). Our results suggest that PEG fractionation has significant potential as a technique for routine plant (phospho)proteome analysis.

Earlier phosphoprotein enrichment studies using IMAC have suggested a bias towards high molecular weight phosphoproteins (5, 38) and this was attributed to the propensity for such proteins to be multiply phosphorylated (5). However, almost half of the phosphoproteins we identified using IMAC enrichment were singly phosphorylated (although some others were found to carry up to six phosphate groups) and included low as well as high molecular weight proteins. This suggests that IMAC is not obviously biased towards large or multiply phosphorylated proteins under these experimental conditions.

Rubisco and Associated Proteins

Rubisco is composed of eight large subunits encoded by the chloroplast *rbc*L gene, and eight small subunits encoded by the nuclear *rbc*S gene family (39). It is the key enzyme in photosynthetic carbon assimilation, and undergoes phosphorylation in

560 response to light (40); however, opinion is divided as to the sites of phosphorylation with
561 some groups reporting phosphorylation of the SSU only while others report
562 phosphorylation of the LSU in different species (41, 42). Our data revealed
563 phosphorylation of both Rubisco subunits in *Arabidopsis thaliana* with six
564 phosphorylated residues in the LSU (Ser208, Ser321, Ser341, Ser452, Thr466, and
565 Tyr239) and 4 in the SSU (Ser71, Thr69, Tyr113 and Tyr133) (Supplementary Table
566 S1) Except .Ser208, Tyr239 and Thr466 of LSU (Lohrig et al., ref), other phosphorylated
567 sites identified in this study for both LSU and SSU were not reported previously.
568 Phosphorylation of the LSU at Thr23, Thr147 and Thr330, and of the SSU at Tyr72,
569 Ser79 and Ser113 had been reported in *Arabidopsis thaliana* (43) but these sites were
570 not amongst those identified in our study, possibly because of the different
571 methodologies used. However, two of the tryptic phosphopeptides
572 (LSGGDHIHAGTVVGK and EHGNTPGYYDGR) previously reported as being
573 phosphorylated in the LSU were identified using our method, albeit with different sites of
574 phosphorylation. Another recent study (44) reported phosphorylation of the Rubisco
575 LSU at Ser208, Tyr239, Thr246, Thr330, Thr466 and Thr471, three of which (Ser208,
576 Tyr239 and Thr466) are confirmed by our analysis (Supplementary Table S1, spot 114).
577 Similarly, we found Rubisco activase (RCA) to be phosphorylated at Thr78 and Thr283
578 (Supplementary Table S1, spot 117), and phosphorylation at Thr78 has been reported
579 previously (43). RCA is a nuclear-encoded chloroplast protein that regulates the activity
580 of Rubisco, and the phosphorylated forms of RCA are also implicated in gibberellin (GA)
581 signaling (45). Detection and mapping of new phosphorylation sites on these proteins
582 demonstrates the utility of our approach for plant phosphoproteome analysis.

583 *Photosystem Proteins*– Reversible phosphorylation of PSII and PSI proteins is
584 important in maintaining the active electron transport chain under light stress (46). Using
585 our methodology we were able to identify phosphorylation of PSBO (spot 163), PSBQ
586 (spot 197) and other PSI and PSII proteins (spots 179, 180, 299, 221) (Supplementary
587 Table S1). We also confirmed phosphorylation of PSBP-1 (spot 187) at residues Thr149
588 and Ser153, and of PSBP-2 (spot 186) at Ser112, Ser147, Thr149 and Thr210. These
589 two nuclear-encoded extrinsic PSII proteins optimize water splitting reaction *in-vivo* (47).
590 Two other nuclear-encoded PsbP homologs, PPL1 and PPL2, were also identified in the
591 F4 fraction, and four phosphorylation sites (Ser120, Thr159, Thr221 and Thr229) were
592 confirmed in PPL1 using protein IMAC. PPL1 is required for efficient repair of photo-
593 damaged PSII complex (47). Similarly, NADH cytochrome B5 reductase, ferredoxin-
594 $NADP^+$ reductase, and AtLFNR1 NADH dehydrogenase, which work in series with other
595 proteins and mobile carrier (plastoquinone Qb, plastocyanin, ferredoxin) to convert
596 NADP to NADPH were also phosphorylated. This generates an electrochemical
597 gradient across the lumen powering conversion of ADP to ATP by ATP synthase (48).
598 Rieske-FeS, four isoforms of ATPase (CF1 α, β, δ subunits and VMA 10) and four 14-3-
599 3 family proteins (14-3-3 like protein AFT1, GF14 chi, GF14 psi and GRF3) were also
600 identified as phosphoproteins (Supplementary Table S1). Proteins of the 14-3-3 family
601 control many signal transduction events by controlling enzyme activities, sub-cellular
602 location, and protein-protein interactions (49), and phosphorylation of these proteins
603 might be important to activate their function.

604 POR-encoded genes *PorA* and *PorB* catalyze the photoreduction of
605 protochlorophyllide (Pchlide) to chlorophyllide (50). PORB (spot 130) identified as a

606 phosphoprotein in our study, sustains light-dependent chlorophyll biosynthesis in green
607 plants in the absence of photo-transferrable pchlide-F635 (50). Mg protoporphyrin IX
608 methyltransferase (CHLM) plays an important role in chlorophyll biosynthesis and,
609 subsequently, the formation and stability of PSI, PSII and cytochrome b6f complexes
610 (51). Pbp1 and Jacalin-related lectin 30 are involved in the transport of Cu^{2+}, a metal ion
611 essential for maintaining stability and regulation of photosynthesis and respiratory
612 electron transport chains in chloroplasts (51). These results suggest that reversible
613 phosphorylation of these photosynthetic thylakoid membrane proteins is important in
614 maintaining photosystem stability and function.
615
616 *Kinases, Phosphatases and Hormonal Signaling Proteins*
617 Five protein kinases associated with photosynthetic membranes have so far
618 been implicated in the phosphorylation of thylakoid proteins; three TAKs (74), STN7 and
619 STN8 (46). We identified a chloroplast-specific serine/threonine kinase (gi|9294138)
620 and a pathogenesis related gene 5 (PR5) protein (gi|15222089), which is also related to
621 a protein serine/threonine kinase (52). However, phosphorylation of these proteins
622 could not be confirmed. In contrast, we observed changes in phosphorylation status of
623 PGK (gi|15219412; spot 138), ADK1 (gi|15232763; spot 159) and nucleoside
624 diphosphate kinase (gi|16396; spot 225), which are potentially localized in the
625 chloroplast. It is not known whether these kinases are involved in the phosphorylation of
626 thylakoid membrane proteins. PIN1, an integral membrane proteins directly involved in
627 auxin efflux from cells (53), was also identified as a phosphoprotein, and is known to be

involved in the regulation of several signaling pathways involving protein phosphorylation/dephosphorylation (53).

Three phosphatases, namely PP2C (gi|18399423; spot 153), phosphoglycolate phosphatase (gi|15239406; spot 162), and an acid phosphatase class B family protein (gi|15240067) were identified, and phosphorylation of PP2C and phosphoglycolate phosphatase were confirmed. PP2C is localized in the cytoplasm, and is known to be involved in ABA signaling in *Arabidopsis* (54). Additionally, phosphorylation of PATL1 (Supplementary Fig. S3A) and TGG2, proteins involved in membrane trafficking (55) and plant defense against bacteria, pathogens and herbivores (56), were also observed. TGG1, a plant myrosinase, is relatively abundant in *A. thaliana* guard cells (57) and was easily detected by gel-based proteomic analysis. TGG2 is much less abundant and was undetected in previous studies (57); however, we consistently detected this protein in our study, and we were able to capture the phosphorylated form using protein IMAC. A total of nine tryptic TGG2 peptides were subsequently identified, one of which (DFLpSQGVRPSALK) was found to be phosphorylated. DREPP, which has been identified in previous studies as a plasma membrane protein (5), was found to be phosphorylated at four locations (Thr32, Thr58, Thr152 and Ser127; Supplementary Fig. S3), in keeping with its potential role in hormone-mediated signal transduction.

Stress Response and Ribosomal Proteins

A total of 20 proteins implicated in stress responses and redox homeostasis were identified as phosphoproteins (Supplementary Table S1). These included several GSTs, chloroplast chaperonine HSPs, peroxidases, oxidases, reductases, PDI and

SODs. Peroxidases and SODs are important H_2O_2 scavenging enzymes, and their phosphorylation in *Arabidopsis* leaf extracts is consistent with their role in protecting plants from damage due to the production of reactive oxygen species (ROS) during photosynthesis and/or light stress. Using our methodology, we are able to identify and confirm phosphorylation status of ribosomal proteins S1, S15, L1, L9, L12 and RRF (Supplementary Table S1).While, phosphorylated isoforms of RPS6 (S6A and S6B) and the acidic P-proteins P0, P1, P2 and P3 have been reported previously, following isolation of the ribosomal proteome and TiO_2 enrichment (58), these additional ribosomal proteins had not been identified as phosphorylated previously. Ribosomal proteins are integral to the ribosome and are involved in the translation of transcripts encoded in cellular genomes and may exert a level of developmental control over cell/tissue fate and function (59).

CONCLUSION

PEG fractionation combined with protein IMAC, 2-D PAGE and LC-MS/MS provides an effective platform for comprehensive phosphoproteomic analysis in plants. Purification of PEG-fractionated samples using SDS/phenol and methanol/chloroform significantly enhances the quality of the resulting 2-D gels. With this methodology we identified numerous proteins with diverse cellular locations and functions, including signal transduction, cell-cell communication and membrane trafficking, in extracts of *Arabidopsis* leaves. A total of 132 phosphorylated proteins were identified from the Rubisco-depleted protein fraction and many of these were proteins of low relative abundance or previously uncharacterized as phosphoproteins. To our knowledge, this is the first time that PEG fractionation has been combined with protein IMAC to study plant leaf phosphoproteins. Conventional gel-based approaches have generally proven unsuccessful in identifying *in vivo* phosphorylation sites for low-abundance membrane proteins; however, we were able to identify and map sites of phosphorylation onto a number of key photosystem complex proteins associated with the thylakoid membrane. We believe our methodology holds significant promise for future research involving phosphorylation and signal transduction in plants.

ACKNOWLEDGEMENTS

We thank Doug Olson and Steve Ambrose of the Mass Spectrometry and Protein Research Group at the National Research Council-Plant Biotechnology Institute (NRC-PBI) for technical support. UKA gratefully acknowledges the generous support from NRC for a Visiting Fellowship at NRC-PBI. Funding for protein mass spectrometry equipment was provided by the Saskatchewan Provincial Government and NRC. This article is contribution number 50188 from the National Research Council of Canada

706

707 REFERENCES

708 1. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics.
709 *Nature* 422, 198-207.
710 2. Mann, M., and Jensen, O. N. (2003) Proteomic analysis of post-translational
711 modifications. *Nat Biotechnol* 21, 255-261.
712 3. Kosako, H., Yamaguchi, N., Aranami, C., Ushiyama, M., Kose, S., Imamoto, N.,
713 Taniguchi, H., Nishida, E., and Hattori, S. (2009) Phosphoproteomics reveals new ERK
714 MAP kinase targets and links ERK to nucleoporin-mediated nuclear transport. *Nat*
715 *Struct Mol Biol* 16, 1026-1035.
716 4. Jensen, O. N. (2004) Modification-specific proteomics: characterization of post-
717 translational modifications by mass spectrometry. *Curr Opin Chem Biol* 8, 33-41.
718 5. Tang, W., Deng, Z., Oses-Prieto, J. A., Suzuki, N., Zhu, S., Zhang, X.,
719 Burlingame, A. L., and Wang, Z. Y. (2008) Proteomics studies of brassinosteroid signal
720 transduction using prefractionation and two-dimensional DIGE. *Mol Cell Proteomics* 7,
721 728-738.
722 6. Gorg, A., Weiss, W., and Dunn, M. J. (2004) Current two-dimensional
723 electrophoresis technology for proteomics. *Proteomics* 4, 3665-3685.
724 7. Rose, J. K., Bashir, S., Giovannoni, J. J., Jahn, M. M., and Saravanan, R. S.
725 (2004) Tackling the plant proteome: practical approaches, hurdles and experimental
726 tools. *Plant J* 39, 715-733.
727 8. Rabilloud, T. (2002) Two-dimensional gel electrophoresis in proteomics: old, old
728 fashioned, but it still climbs up the mountains. *Proteomics* 2, 3-10.
729 9. Canovas, F. M., Dumas-Gaudot, E., Recorbet, G., Jorrin, J., Mock, H. P., and
730 Rossignol, M. (2004) Plant proteome analysis. *Proteomics* 4, 285-298.
731 10. Wang, W., Scali, M., Vignani, R., Spadafora, A., Sensi, E., Mazzuca, S., and
732 Cresti, M. (2003) Protein extraction for two-dimensional electrophoresis from olive leaf,
733 a plant tissue containing high levels of interfering compounds. *Electrophoresis* 24, 2369-
734 2375.
735 11. Granier, F. (1988) Extraction of plant proteins for two-dimensional
736 electrophoresis. *Electrophoresis* 9, 712-718.
737 12. Pirovani, C. P., Carvalho, H. A., Machado, R. C., Gomes, D. S., Alvim, F. C.,
738 Pomella, A. W., Gramacho, K. P., Cascardo, J. C., Pereira, G. A., and Micheli, F. (2008)
739 Protein extraction for proteome analysis from cacao leaves and meristems, organs
740 infected by Moniliophthora perniciosa, the causal agent of the witches' broom disease.
741 *Electrophoresis* 29, 2391-2401.
742 13. Kolkman, A., Dirksen, E. H., Slijper, M., and Heck, A. J. (2005) Double standards
743 in quantitative proteomics: direct comparative assessment of difference in gel
744 electrophoresis and metabolic stable isotope labeling. *Mol Cell Proteomics* 4, 255-266.
745 14. Deng, Z., Zhang, X., Tang, W., Oses-Prieto, J. A., Suzuki, N., Gendron, J. M.,
746 Chen, H., Guan, S., Chalkley, R. J., Peterman, T. K., Burlingame, A. L., and Wang, Z.
747 Y. (2007) A proteomics study of brassinosteroid response in Arabidopsis. *Mol Cell*
748 *Proteomics* 6, 2058-2071.

749 15. Cellar, N. A., Kuppannan, K., Langhorst, M. L., Ni, W., Xu, P., and Young, S. A.
750 (2008) Cross species applicability of abundant protein depletion columns for ribulose-
751 1,5-bisphosphate carboxylase/oxygenase. *J Chromatogr B Analyt Technol Biomed Life*
752 *Sci* 861, 29-39.
753 16. Xi, J., Wang, X., Li, S., Zhou, X., Yue, L., Fan, J., and Hao, D. (2006)
754 Polyethylene glycol fractionation improved detection of low-abundant proteins by two-
755 dimensional electrophoresis analysis of plant proteome. *Phytochemistry* 67, 2341-2348.
756 17. Corthals, G. L., Wasinger, V. C., Hochstrasser, D. F., and Sanchez, J. C. (2000)
757 The dynamic range of protein expression: a challenge for proteomic research.
758 *Electrophoresis* 21, 1104-1115.
759 18. Polson, A., Potgieter, G. M., Largier, J. F., Mears, G. E., and Joubert, F. J. (1964)
760 The Fractionation of Protein Mixtures by Linear Polymers of High Molecular Weight.
761 *Biochim Biophys Acta* 82, 463-475.
762 19. Yamamoto, K. R., Alberts, B. M., Benzinger, R., Lawhorne, L., and Treiber, G.
763 (1970) Rapid bacteriophage sedimentation in the presence of polyethylene glycol and
764 its application to large-scale virus purification. *Virology* 40, 734-744.
765 20. Kim, S. T., Cho, K. S., Jang, Y. S., and Kang, K. Y. (2001) Two-dimensional
766 electrophoretic analysis of rice proteins by polyethylene glycol fractionation for protein
767 arrays. *Electrophoresis* 22, 2103-2109.
768 21. Ahsan, N., Lee, D. G., Lee, S. H., Kang, K. Y., Bahk, J. D., Choi, M. S., Lee, I. J.,
769 Renaut, J., and Lee, B. H. (2007) A comparative proteomic analysis of tomato leaves in
770 response to waterlogging stress. *Physiol Plant* 131, 555-570.
771 22. Hesse, A. M., Marcelo, P., Rossier, J., and Vinh, J. (2008) Simple and universal
772 tool to remove on-line impurities in mono- or two-dimensional liquid chromatography-
773 mass spectrometry analysis. *J Chromatogr A* 1189, 175-182.
774 23. Shen, J., and Buko, A. (2002) Rapid identification of proteins in polyethylene
775 glycol-containing samples using capillary electrophoresis electrospray mass
776 spectrometry. *Anal Biochem* 311, 80-83.
777 24. Wessel, D., and Flugge, U. I. (1984) A method for the quantitative recovery of
778 protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* 138,
779 141-143.
780 25. Mortz, E., Krogh, T. N., Vorum, H., and Gorg, A. (2001) Improved silver staining
781 protocols for high sensitivity protein identification using matrix-assisted laser
782 desorption/ionization-time of flight analysis. *Proteomics* 1, 1359-1363.
783 26. Wan, L., Ross, A. R., Yang, J., Hegedus, D. D., and Kermode, A. R. (2007)
784 Phosphorylation of the 12 S globulin cruciferin in wild-type and abi1-1 mutant
785 Arabidopsis thaliana (thale cress) seeds. *Biochem J* 404, 247-256.
786 27. Ahn, N. G., and Resing, K. A. (2001) Toward the phosphoproteome. *Nat*
787 *Biotechnol* 19, 317-318.
788 28. Kanehisa, M., and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and
789 genomes. *Nucleic Acids Res* 28, 27-30.
790 29. Alexandersson, E., Saalbach, G., Larsson, C., and Kjellbom, P. (2004)
791 Arabidopsis plasma membrane proteomics identifies components of transport, signal
792 transduction and membrane trafficking. *Plant Cell Physiol* 45, 1543-1556.

30. Nuhse, T. S., Stensballe, A., Jensen, O. N., and Peck, S. C. (2004) Phosphoproteomics of the Arabidopsis plasma membrane and a new phosphorylation site database. *Plant Cell* 16, 2394-2405.
31. Nuhse, T. S., Stensballe, A., Jensen, O. N., and Peck, S. C. (2003) Large-scale analysis of in vivo phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry. *Mol Cell Proteomics* 2, 1234-1243.
32. Staiger, D., Zecca, L., Wieczorek Kirk, D. A., Apel, K., and Eckstein, L. (2003) The circadian clock regulated RNA-binding protein AtGRP7 autoregulates its expression by influencing alternative splicing of its own pre-mRNA. *Plant J* 33, 361-371.
33. Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., and Yates, J. R., 3rd (1999) Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 17, 676-682.
34. Dowell, J. A., Frost, D. C., Zhang, J., and Li, L. (2008) Comparison of two-dimensional fractionation techniques for shotgun proteomics. *Anal Chem* 80, 6715-6723.
35. Hunsucker, S. W., and Duncan, M. W. (2006) Is protein overlap in two-dimensional gels a serious practical problem? *Proteomics* 6, 1374-1375.
36. Fahie-Wilson, M., and Halsall, D. (2008) Polyethylene glycol precipitation: proceed with care. *Ann Clin Biochem* 45, 233-235.
37. Atha, D. H., and Ingham, K. C. (1981) Mechanism of precipitation of proteins by polyethylene glycols. Analysis in terms of excluded volume. *J Biol Chem* 256, 12108-12117.
38. Machida, M., Kosako, H., Shirakabe, K., Kobayashi, M., Ushiyama, M., Inagawa, J., Hirano, J., Nakano, T., Bando, Y., Nishida, E., and Hattori, S. (2007) Purification of phosphoproteins by immobilized metal affinity chromatography and its application to phosphoproteome analysis. *Febs J* 274, 1576-1587.
39. Spreitzer, R. J., and Salvucci, M. E. (2002) Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annu Rev Plant Biol* 53, 449-475.
40. Budde, R. J., and Randall, D. D. (1990) Light as a signal influencing the phosphorylation status of plant proteins. *Plant Physiol* 94, 1501-1504.
41. Lucero, H. A., Lin, Z. F., and Racker, E. (1982) Protein kinases from spinach chloroplasts. II. Protein substrate specificity and kinetic properties. *J Biol Chem* 257, 12157-12160.
42. Guitton, C., and Mache, R. (1987) Phosphorylation in vitro of the large subunit of the ribulose-1,5-bisphosphate carboxylase and of the glyceraldehyde-3-phosphate dehydrogenase. *Eur J Biochem* 166, 249-254.
43. Reiland, S., Messerli, G., Baerenfaller, K., Gerrits, B., Endler, A., Grossmann, J., Gruissem, W., and Baginsky, S. (2009) Large-scale Arabidopsis phosphoproteome profiling reveals novel chloroplast kinase substrates and phosphorylation networks. *Plant Physiol* 150, 889-903.
44. Lohrig, K., Muller, B., Davydova, J., Leister, D., and Wolters, D. A. (2009) Phosphorylation site mapping of soluble proteins: bioinformatical filtering reveals potential plastidic phosphoproteins in Arabidopsis thaliana. *Planta* 229, 1123-1134.

836 45. Sharma, A., and Komatsu, S. (2002) Involvement of a Ca(2+)-dependent protein
837 kinase component downstream to the gibberellin-binding phosphoprotein, RuBisCO
838 activase, in rice. *Biochem Biophys Res Commun* 290, 690-695.
839 46. Bellafiore, S., Barneche, F., Peltier, G., and Rochaix, J. D. (2005) State
840 transitions and light adaptation require chloroplast thylakoid protein kinase STN7.
841 *Nature* 433, 892-895.
842 47. Ishihara, S., Takabayashi, A., Ido, K., Endo, T., Ifuku, K., and Sato, F. (2007)
843 Distinct functions for the two PsbP-like proteins PPL1 and PPL2 in the chloroplast
844 thylakoid lumen of Arabidopsis. *Plant Physiol* 145, 668-679.
845 48. Ishikawa, Y., Schroder, W. P., and Funk, C. (2005) Functional analysis of the
846 PsbP-like protein (sll1418) in Synechocystis sp. PCC 6803. *Photosynth Res* 84, 257-
847 262.
848 49. Bridges, D., and Moorhead, G. B. (2005) 14-3-3 proteins: a number of functions
849 for a numbered protein. *Sci STKE* 2005, re10.
850 50. Runge, S., Sperling, U., Frick, G., Apel, K., and Armstrong, G. A. (1996) Distinct
851 roles for light-dependent NADPH:protochlorophyllide oxidoreductases (POR) A and B
852 during greening in higher plants. *Plant J* 9, 513-523.
853 51. Pontier, D., Albrieux, C., Joyard, J., Lagrange, T., and Block, M. A. (2007) Knock-
854 out of the magnesium protoporphyrin IX methyltransferase gene in Arabidopsis. Effects
855 on chloroplast development and on chloroplast-to-nucleus signaling. *J Biol Chem* 282,
856 2297-2304.
857 52. Wang, X., Zafian, P., Choudhary, M., and Lawton, M. (1996) The PR5K receptor
858 protein kinase from Arabidopsis thaliana is structurally related to a family of plant
859 defense proteins. *Proc Natl Acad Sci U S A* 93, 2598-2602.
860 53. Mravec, J., Skupa, P., Bailly, A., Hoyerova, K., Krecek, P., Bielach, A., Petrasek,
861 J., Zhang, J., Gaykova, V., Stierhof, Y. D., Dobrev, P. I., Schwarzerova, K., Rolcik, J.,
862 Seifertova, D., Luschnig, C., Benkova, E., Zazimalova, E., Geisler, M., and Friml, J.
863 (2009) Subcellular homeostasis of phytohormone auxin is mediated by the ER-localized
864 PIN5 transporter. *Nature* 459, 1136-1140.
865 54. Ma, Y., Szostkiewicz, I., Korte, A., Moes, D., Yang, Y., Christmann, A., and Grill,
866 E. (2009) Regulators of PP2C phosphatase activity function as abscisic acid sensors.
867 *Science* 324, 1064-1068.
868 55. Peterman, T. K., Ohol, Y. M., McReynolds, L. J., and Luna, E. J. (2004) Patellin1,
869 a novel Sec14-like protein, localizes to the cell plate and binds phosphoinositides. *Plant*
870 *Physiol* 136, 3080-3094; discussion 3001-3082.
871 56. Barth, C., and Jander, G. (2006) Arabidopsis myrosinases TGG1 and TGG2
872 have redundant function in glucosinolate breakdown and insect defense. *Plant J* 46,
873 549-562.
874 57. Zhao, Z., Zhang, W., Stanley, B. A., and Assmann, S. M. (2008) Functional
875 proteomics of Arabidopsis thaliana guard cells uncovers new stomatal signaling
876 pathways. *Plant Cell* 20, 3210-3226.
877 58. Carroll, A. J., Heazlewood, J. L., Ito, J., and Millar, A. H. (2008) Analysis of the
878 Arabidopsis cytosolic ribosome proteome provides detailed insights into its components
879 and their post-translational modification. *Mol Cell Proteomics* 7, 347-369.

59. Whittle, C. A., and Krochko, J. E. (2009) Transcript profiling provides evidence of functional divergence and expression networks among ribosomal protein gene paralogs in Brassica napus. *Plant Cell* 21, 2203-2219.

899 **FIGURE CAPTIONS**

900 **Figure 1. Extraction and purification of *Arabidopsis* leaf proteins**. Proteins in the
901 supernatant from the crude extract were either precipitated directly to obtain a non-
902 fractionated (NF') sample or treated sequentially with increasing concentrations of
903 polyethylene glycol (PEG) to obtain PEG-fractionated samples (F1' through F4'). In
904 addition, each of these samples was further extracted with SDS/phenol and
905 methanol/chloroform (abbreviated to Phenol extraction* in the Figure and text) to obtain
906 purified PEG-fractionated (F1 through F4) and non-fractionated (NF) samples for IMAC
907 purification and/or 2-D PAGE.

908 **Figure 2. One-dimensional gel electrophoresis of PEG-fractionated and non-**
909 **fractionated protein samples.** Lane markings correspond to the purified samples (F1
910 through F4 and NF) described in Fig. 1. Molecular weight markers (M) are shown on the
911 left of the gel. Boxed areas enclose the large (LSU) and small subunits (SSU) of
912 Rubisco. Arrows indicate protein bands that are enriched in the Rubisco-depleted F4
913 fraction relative to the other samples shown.

914 **Figure 3. Two-dimensional gel electrophoresis of PEG-fractionated and non-**
915 **fractionated protein samples.** The large (LSU) and small (SSU) subunits of Rubisco
916 are depleted in the PEG-fractionated (F4) sample (A) relative to the non-fractionated
917 (NF) sample (B). The boxed areas of the F4 (a, b) and NF (a', b') gels show additional
918 changes in relative concentrations of less abundant proteins between the two samples.
919 **Figure 4. Identification of phosphorylated and nonphosphorylated isoforms of**
920 ***Arabidopsis* leaf proteins**. (A) Venn diagram comparing the number of

phosphoproteins identified in NF, F2 and F4 fractions without IMAC purification. The complete list of 77 phosphoproteins and component phosphopeptides identified in fraction F4 are presented in Supplementary Table S1. (B) An expanded view of a portion of the F4 (Rubisco-depleted) 2-D gel from Fig. 3A (Box a) showing protein spots corresponding to phosphorylated and non-phosphorylated isoforms of PGK (i), FBA (ii) and ATFNR1 (iii). The black and blue arrows indicate the phosphorylated and non-phosphorylated forms of each protein, respectively, as summarized in the table below the Fig. 4B. PP, phosphorylated peptides.

Figure 5. Phosphorylation of Rubisco in *Arabidopsis* leaves. Product-ion tandem mass spectra for doubly-charged phosphopeptide ions of mass-to-charge (*m/z*) 510.26 (A) and 951.55 (B) indicate phosphorylation of the Rubisco large subunit (gi|7525041) at Ser321 and Ser341, respectively, whereas those for *m/z* 723.29 (C) and 814.91 (D) show phosphorylation of the small subunit (gi|16194) at Tyr113 and Tyr133. Product ions corresponding to sequential loss of intact amino acid residues from the C or N terminus are labeled as b- or y-type ions, respectively, whereas peaks arising from neutral loss of either HPO_3 (80 Da) or H_3PO_4 (98 Da) are labeled in bold type. The Rubisco LSU (spot 114) was also found to be phosphorylated at Ser208, Ser452 and Thr466, and the SSU (spot 227) at Thr69 (see also Supplementary Table S1).

Figure 6. One-dimensional polyacrylamide gel electrophoresis of IMAC-enriched PEG-fractionated (F1, F2, F4) and non-fractioned proteins (NF). Lane markings correspond to the purified samples (NF and F1 through F4) shown in Fig. 1.

Figure 7. Two-dimensional gel electrophoresis of proteins in the IMAC-enriched F4 fraction. Numbered arrows on the gel indicate phosphoproteins identified by LC-MS/MS (A). Phosphoproteins were functionally categorized and the number of phosphoproteins in each category shown in parentheses (B). Venn diagram comparing the number of phosphoproteins identified in F4 fraction with and without IMAC purification (C). Extractable proteins from the same biological sample was used for both the analysis. All phosphoproteins identified following IMAC are listed in Supplementary Table S1.

Supplementary Figures

Figure S1. Effect of combined phenol/SDS and methanol/chloroform extraction on quality of protein separations with 2-D PAGE. Protein separations by two-dimensional gel electrophoresis (2-D PAGE) are shown for (A) non-fractionated *Arabidopsis* leaf protein samples without the additional SDS/phenol and chloroform/methanol clean-up (NF') and (B) non-fractionated *Arabidopsis* leaf protein samples with the SDS/phenol and methanol/chloroform extraction (NF). Similarly, protein separations are shown for (C) PEG-fractionated proteins without phenol extraction (F2') and (D) PEG-fractionated proteins after phenol extraction (F2). Streaking was significantly reduced and resolution much improved in the phenol extracted samples (compare NF' and NF, and F2' and F2).

Figure S2. Identification of *Arabidopsis* leaf proteins. (A) Venn diagram summarizing the number of proteins identified in 2-D gels of PEG-fractionated (F2, F4) and non-fractionated (NF) samples. (B) Percentage of protein spots containing Rubisco or its variants in the 2-D gel maps of F2, F4 and NF samples. Spots containing multiple proteins were counted if Rubisco was one of the top 5 protein hit and identified with at least two unique peptides with $p < 0.05$.

Figure S3. CID spectra of representative phosphopeptides. Product-ion tandem mass spectra for doubly charged peptide ions of mass-to-charge (*m/z*) 807.17 (A), 728.51 (B), 707.44 (C) and 611.01 (D), respectively, indicating phosphorylation of patellin 1 transporter (gi|15218382) at Thr29, DREPP (gi|15235363) at Ser127, legume

lectin family protein (gi|15242724) at Ser38, and ankyrin repeat containing protein 2 (gi|4205079) at Ser289 (see also Supplementary Table S.1).

Figure S4. CID spectra of representative phosphopeptides. Product-ion tandem mass spectra for doubly charged peptide ions of mass-to-charge (*m/z*) 982.52 (A), 790.96 (B), 886.05 (C) and 689.77 (D), respectively, indicating phosphorylation of PIN1AT (gi|15227956) at Ser103, PPL1 (gi|15233245) at Thr158, ATPHB4 (gi|15232129) at Ser223, and PBP1 (gi|15228198) at Ser196 (see also Supplementary Table S.1).

Figure S5. Additional CID spectra of phosphopeptides identified in protein spots from the 2-DE gel of the IMAC purified PEG supernatant (F4) fraction.

1001 **Supplementary Table S1:** *Arabidopsis thaliana* leaf phosphoproteins identified in the PEG-
1002 fractionated F4 sample following phosphoprotein enrichment by IMAC. One hundred and thirty-
1003 two phosphoproteins were identified (see Figure 7).

1004

Spot No.	NCBI Sequence	Protein name	Peptides Matched[1]	CV. (%)[2]	Matched Phosphopeptides[3]	Ion Score	w/o IMAC[4]
Carbohydrate and energy metabolism (20)							
102	gi|7329685	Transketolase like protein	28	40	ANpSYSVHGAALGEK ALPpTYTPESPGDATR	65 43	
122	gi|7769871	Malate dehydrogenase	24	46	pTQDGGTEVVEAK	70	
128	gi|15227752	Peroxisomal NAD-malate dehydrogenase (PMDH1)	14	41	AIVNIIpSNPVNSTVPIAAEVFK	42	
129	gi|79327392	Peroxisomal NAD-malate dehydrogenase (PMDH2)	23	59	IQNGGpTEVVEAK	44	
138	gi|15219412	Phosphoglycerate kinase (PGK)	29	57	AHApSTEGVAK pYSLKPLVPR GVpSLLLPTDVVIADK	44 35 73	
139	gi|15221631	3-Isopropyl malate dehydrogenase like protein	9	21	NLANPpTALLLSGVMMLR	38	√
140	gi|11131561	Fructose bisphosphate aldolase, putative	22	47	pTVVSSIPNGPSALAVK TIApSPGHHIMAMDESNATCGK GLVPLVGpSNNESWCQGLDGLSSR YAAIpSQDSGLVPIVEPEILLDGEHD	56 61 78 48	
141	gi|15226185	Fructose bisphosphate aldolase	19	43	LGDGAAEpSLHVK VpSPEVIAEHTVR IGENEPpSEHSIHENAYGLAR TVPAAVPAIVFLpSGGQSEEEATR	35 58 54 100	
142	gi|18420348	Fructose bisphosphate aldolase, putative	26	44	RLDpSIGLENTEANR pTVVSIPNGPSALAVK GILAMDEpSNATCGK DRATPEQVAAYpTLK GLVPLVGpSNNESWCQGLDGLSSR	40 39 33 33 79	√ √ √

					YAAIpSQDSGLVPIVEPEILLDGEHD	61	
143	gi|15229781	Fructose bisphosphate aldolase	15	51	pTVPAAVPAIVFLSGGQSEEEATR	71	
144	gi|15229231	Glyceraldehyde 3-phosphate dehydrogenase C-subunit (GAPC)	15	38	pTLLFGEKPVTVFGIR	81	√
					FGIVEGLMTpTVHSITATQK	72	
					FGIVEGLMpTTVHSITATQK	38	
145	gi|15222848	Glyceraldehyde 3-phosphate dehydrogenase C-2 (GAPC-2)	36	51	pTLLFGEKPVTVFGIR	78	√
					FGIVEGLMTpTVHSITATQK	71	
					FGIVEGLMpTpTVHSITATQK	46	
146	gi|166702	Glyceraldehyde 3-phosphate dehydrogenase A subunit	24	53	GpTMTTTHSYTGDQR	62	
150	gi|336390	Glyceraldehyde 3-phosphate dehydrogenase B subunit	17	46	GpTMTTTHSYTGDQR	73	
156	gi|15218869	Isocitrate dehydrogenase, putative	17	38	LVPGWpTKPICIGR	45	
166	gi|15219721	Malate dehydrogenase	22	50	LSpSALSAASSACDHIR	85	√
					NVIIWGNHpSSSQYPDVNHAK	37	√
168	gi|42573371	Carbonic anhydrase 2 (CA2)	17	51	IpTAELQAASSSDSK	68	
					pYAGVGAAIEYAVLHLK	85	√
					VLAESEpSSAFEDQCGR	72	
					EAVNVpSLANLLTYPFVR	40	√
					VCPpSHVLDFHPGDAFVVR	56	
183	gi|15226479	Triosephosphate isomerase (TIM)	14	28	GGAFpTGEISVEQLK	59	√
					EAGKpTFDVCFAQLK	36	
					AAYALpSEGLGVIACIGEK	98	
202	gi|414550	Cytosolic triosephosphate isomerase	12	53	VIACVGEpTLEER	80	√
					AILNEpSSEFVGDK	31	
					VApSPAQAQEVHDELRK	27	
203	gi|1655482	Delta subunit of mitochondrial F1 ATPase	12	40	NFLpSLLAENGK	53	
					ELpTETLQEIIGAGK	86	
					IEpTDLSEMIEAMK	48	
Photosynthesis and electron transfer chain (20)							
114	gi|7525041	Rubisco large subunit (LSU)	26	42	WpSPELAAACEVWK	55	
					GGLDFTKDDENVNpSQPFMR	61	√
					LEGDREpSTLGFVDLLR[5]	45	√

					LpSGGDHIHAGTVVGK[5]	42	√
					EIpTFNFPTIDKLDGQE	28	
					GHpYLNATAGTCEEMIK	32	
117	gi|18405145	Rubisco activase (RCA)	39	55	GLAYDTSDDQQDITR	86	
					VPIICTGNDFSTLYAPLIR	71	
123	gi|15239282	ATLFNR1 (LEAF FNR1) NADH dehydrogenase)	32	51	EGQpSIGVIPEGIDK	56	√
					GVCpSNFLCDLKPGDEAK	80	
					LpYSIASSAIGDFGDSK	99	√
					DPNApTIIMLGTGTGIAPFR	69	√
126	gi|18420117	Putative NADH cytochrome b5 reductase	26	45	VpSHNTQLFR	31	√
					MpSQHFASLKPGDVLEVK	42	
					HIGMIAGGSGIpTPMLQVIDAIVK	81	√
147	gi|2289095	WD-40 repeat protein	15	49	YpTISEGGEGHR	38	
					pYWLCAATEHGIK	44	
157	gi|15450379	Rubisco activase (RCA)	17	40	VPIICpTGNDFSTLYAPLIR	62	
163	gi|15230324	Photosystem II subunit O-2 (PSBO-2/PSBO2)	21	54	QLEApSGKPESFSGK	56	
					GGSpTGYDNAVALPAGGR	90	
					GTGpTANQCPTIDGGSETFSFK	42	
					LpTYTLDEIEGPFEVGSDGSVK	68	
					SKPETGEVIGVFEpSLQPSDTDLGAK	37	
177	gi|15229349	Ribose 5-phosphate isomerase-related	10	39	pSLGIPLVGLDTHPR	75	√
					LLSpSGELYDIVGIPTSK	62	√
179	gi|430947	PSI type III chlorophyll a-b binding protein	14	27	GLAGpSGNPAYPGGPFFNPLGFGK	64	
180	gi|15235029	Light harvesting complex of photosystem II 5 (LHC B5)	6	22	pYQAFELIHAR	38	
186	gi|15222166	Oxygen evolving enhancer protein 2 (PSBP-2)	38	36	pTADGDEGGKHQLITATVNGGK	66	√
					WNPpSKEIEYPGQVLK	31	
					pSITDYGSPEEFLSQVNYLLGK	114	√
					SIpTDpYGSPEEFLSQVNYLLGK	87	
187	gi|186478207	Oxygen evolving enhancer protein 1 (PSBP-1)	10	34	SIpTDYGSPEEFLSQVNYLLGK	116	√
					SITDYGpSPEEFLSQVNYLLGK	62	√
197	gi|15234637	PSBQ/PSBQ-2/PSII-Q	16	45	FpYIQPLSPTEAAR	58	

					YYpSETVSSLNNVLAK	109	√
					SSPDAEKYYpSETVSSLNNVLAK	61	√
198	gi|15233587	PSBQ/PSBQ-1/PSBQA; calcium ion binding	13	38	LpYLNTIISSKPK	60	
					LFDpTIDNLDYAAK	79	
					YpYAEpTVSALNEVLAK	75	√
					VGPPPAPpSGGLPGTDNSDQAR	51	
199	gi|15235503	Photosystem I, subunit D-1 (PSAD-1)	18	42	EQIFEMPpTGGAAIMR	46	
					EAPVGFpTPPQLDPNTPSPIFAGSTGGLLR	81	√
					pTDpSpSAAAAAAPATK	42	√
201	gi|9843639	Rieske FeS protein	6	14	FLCPCHGpSQYNAQGR	24	
					GPAPLpSLALAHADIDEAGK	68	
219	gi|15233245	PSBP-like protein 1 (PPL1)	18	46	LHpTVVDSFK	54	√
					LHTVVDSFKIpTV	31	
					pTTLIDASEHDVDGK	70	√
					DVIEPLEpSVSVNLVPTSK	56	√
221	gi|18379115	Photosystem II Family protein	9	32	pTDPNVADAVAELR	72	√
					DIpYSALNAVSGHYVSFGPTAPIPAK	69	√
227	gi|16194	Rubisco small subunit (SSU)	18	47	**EHGNpTPGYYDGR**[5]	47	√
					LPLFGCpTDSAQVLK[5]	37	√
					FEpTLSYLPDLSDVELAK	76	√
					KFEpTLSYLPDLSDVELAK	63	√
228	gi|13926229	F1O19.10/F1O19.10	15	72	EHGNpTPGYYDGR	39	
					LPLFGCpTDSAQVLK	34	
					FEpTLSYLPDLSDVELAK	71	√
					KFEpTLSYLPDLSDVELAK	73	√
					KFEpTLpSYLPDLSDVELAK	67	√
Photorespiration (5)							
132	gi|5225026	AGT (Alanine glyoxylate aminotransferase	20	59	ALpSLPTGLGIVCASPK	71	
160	gi|15235745	Serine hydroxymethyl- transferase 1 (SHM1)	19	38	NpTVPGDVSAMVPGGIR	61	
162	gi|15239406	Phosphoglycolate phosphatase	16	38	LIEGVPEpTLDMLR	27	

					TLLVLSGVpTSISMLESPENK	77	
210	gi|15220620	Glycerate dehydrogenase (HPR)	13	28	EGMApTLAALNVLGR	40	√
224	gi|15226973	Glycine decarboxylase complex H (GDCH)	12	35	EYpTKFLEEEDAAH	38	
					VKPpSSPAELESLMGPK	64	
Chlorophyll biosynthesis (4)							
130	gi|15234129	PORB (Photochlorophyllide oxidoreductase B)	26	46	DSYTVMHLDLASLDSVR	72	
					TETPLDVLVCNAAVYFPTAK	59	
137	gi|15229018	Glutamate 1 semialdehyde 2 (GSA2)	26	49	AGpSGVATLGLPDSPGVPK	84	√
					ELpTNGILEAGKK	62	√
151	gi|15242822	Glutamate 1 semialdehyde 1 (GSA1)	15	33	AGSGVApTLGLPDSVGVPK	32	
164	gi|15234905	Magnesium-protoprophyrin IX methyltransferase (CHLM)	12	31	ADGMIAHLApSLAEK	35	
Protein folding and assembly (8)							
103	gi|15233779	Chaperonine heat shock protein (CPHSP 70-1)	25	32	NQADpSVVYQTEK	65	√
					IELSSLpTQTNMSLPFITATADGPK	36	√
105	gi|1522729	Chaperonin HSP 60B (CPN60B)	27	40	CCLEHAApSVAK	40	
					pSQYLDDIAILTGATVIR	107	√
106	gi|16221	Chaperonin HSP 60	25	45	GIpSMAVDAVVTNLK	80	√
107	gi|15226314	Chloroplast 60 kDa chaperonin alpha subunit (CPN60B)	27	38	HGLLpSVTSGANPVSLK	46	
108	gi|397482	HSP 70 cognate	42	56	pSTVHDVVLVGGSTR	67	√
					ApTAGDTHLGGEDFDNR	54	√
111	gi|20453106	At2g04030/F3C11.4	26	40	TFDMIQEISEpSENK	24	√
182	gi|18400095	Co-chaperone grPE family protein	6	20	VpSVIATVGKPFDPLLHEAISR	56	√
214	gi|15242045	Chaperonine 20 (CPN20)	13	55	YpTSIKPLGDR	39	√
Stress and redox homeostasis (16)							
112	gi|8052534	F3F9.11	11	12	pSPLLLQMNPIHK	47	
					pSLPDPEKVTEFVSELR	87	
124	gi|18408627	Haloacid dehalogenase-like hydrolase family protein	18	47	AVpSAIVSCLLGPER	51	
					KKPDPAIpYNLAAETLGVDPSK	46	
125	gi|15236678	Ascorbate peroxidase 4 (APX4)	16	45	GGPIpSYADIIQLAGQSAVK	92	√
127	gi|15229095	ATPCB/ATPERX34/PERX34/PRXCB (peroxidase 34)	13	29	NVGLDRPSDLVALpSGGHTFGK	66	

149	gi|8778996	Contains similarity to ferredoxin-NADP+ reductase	16	33	DPNApTVIMLATGTGIAPFR	58	
152	gi|15226610	ATPDIL2-1/MEE30/UNE5	11	39	DLDDFVpSFINEK	55	
181	gi|15229806	2-Cys peroxiredoxin	13	42	EVVIQHpSTINNLGIGR	54	
					pSGGCGDLNYPLISDVTK	78	
184	gi|15223049	Ascorbate peroxidase 1 (APX1)	12	30	ELLpSGEKEGLLQLVSDK	66	√
185	gi|11761812	Glutathione dependent dehydroascorbate reductase	13	40	AEpTEDVIAGWRPK	51	√
204	gi|20197312	Glutathione S-transferase 6 (GST6)	18	52	AIpTQYLAEEYSEK	59	
					GMFGMpTTDPAAVQELEGK	77	
					QEAHLALNPFGQIPALEDGDLpTLFESR	35	
205	gi|15228407	Manganese superoxide dismutase 1 (MSD1)	10	37	LVVDpTTANQDPLVTK	58	
					GpSLGSAIDAHFGSLEGLVK	103	√
					HHQAYVTNpYNNALEQLDQAVNK	79	√
213	gi|3273753	Copper/zinc SOD	7	34	GGHELpSLTTGNAGGR	34	√
215	gi|15231718	Peroxiredoxin type 2, putative	17	50	VLNLEEGGAFpTNSSAEDMLK	94	√
					LPDpSTLSYLDPSTGDVK	75	
					pTILFAVPGAFTPTCSSQK	73	
216	gi|1061036	Glutathione peroxidase	15	39	VDVNGPpSTAPIYEFLK	31	√
226	gi|15230982	Perodoxin Q, putative, antioxidant (ATPRXQ)	10	42	AGAEVIGIpSGDDSASHK	55	√
223	gi|15242674	Glutaredoxin, putative	10	54	LVPLLpTEAGAIGK	64	
					pTVPNVFIGGNHIGGCDA	59	
Membrane and transport (10)							
104	gi|15218382	PATELLIN 1 transporter (PATL1)	24	37	EVpTIPTPVAEKEEVAAPVSDEK	65	√
					VEEEKDAVPAAEEEKpSSEAAPVETK	38	√
115	gi|7525040	ATP synthase CF1 beta subunit	31	50	TNPpTTSNPEVSIR	61	√
					VGLpTALTMAEYFR	45	√
					IVGEEHYEpTAQQVK	38	√
116	gi|7525018	ATP synthase CF1 alpha subunit	42	45	LIEpSPAPGIISR	59	√
					EApYPGDVFYLHSR	59	√
					EQHpTLIIYDDLSK	69	√
190	gi|1755154	Germin-like protein	8	29	GDpSMVFPQGLLHFQLNSGK	93	√

191	gi|1143394	V-type proton ATPase	16	43	IDYpSMQLNASR pSNDPHGLHCSGGVVLASR	38 67	 √
192	gi|15242210	Porin, putative	10	36	VSTDSpSLLTTLTFDEPAPGLK	81	
193	gi|15232074	Porin, putative	11	46	FpSITTFSPAGVAITSTGTK	38	
194	gi|15240765	Porin, putative	14	51	LNNHGpTLGALLQHEVLPR	78	
218	gi|15240623	Immunophilin	24	47	GGpSFEGDDKEFFK	61	√
124	gi|15236211	Vacuolar membrane ATPase 10 (VMA10)	10	63	IpSKDVVEMLLK KLEEpTSGDSGANVK LEQEpTDTKIEQLK	32 54 26	
Protein synthesis and translation (13)							
133	gi|23397095	Putative chloroplast translation elongation factor EF-Tu precursor	28	46	TpTLTAALTMALASIGSSVAK	112	
134	gi|7572940	Ribosomal protein L1 protein	20	42	EpYDVNTAISLLK pSAHICSSMGPSIK pSAGADIVGSDDLIEQIK	62 43 82	
158	gi|30692346	Ribosomal protein S1 (RPS1)	14	28	SpSAYLSVEQACIHR	77	
170	gi|25299531	Probable elongation factor 1 beta	9	41	VETGMIKPGMVVpTFAPTGLTTEVK	48	
171	gi|30687350	Elongation factor 1B alpha subunit 2	7	28	AVpTFSDLHTEEGVK	82	√
172	gi|15239877	Elongation factor 1B beta subunit	7	27	IPFVPIpSGFEGDNMIER	24	
188	gi|11277801	Ribosomal recycling factor, chloroplast precursor (RRF)	23	46	LpSEDNVKDLSSDLQK	41	√
195	gi|18398753	60S ribosomal protein L9	8	44	pTALSHVDNLISGVTR	78	
189	gi|11277801	RRF (Ribosomal recycling factor, chloroplast precursor)	23	46	LpSEDNVKDLSSDLQK	41	
196	gi|15242436	40S ribosomal protein S15	7	27	HGRPGVGATHSSR GVDLDALLDMSTDDLVK	45 34	
200	gi|15228098	60S ribosomal protein L12	10	48	VpTVVPSAAALVIK VpTGGEVGAASSLAPK HNGNIpSFDDVTEIAR	46 44 63	
207	gi|12322730	Putative elongation factor P (EF-P)	8	31	DGpSQFVFMDLTTYEETR	74	
220	gi|468773	Ribosomal protein L12	7	22	VTVVPSAAALVIK	60	
Signaling (12)							

153	gi|18399423	Protein phosphatase (PP2C)	17	59	pYLQTNLFDNILK	41	
159	gi|15242753	Adenosine kinase 1 (ADK1)	24	51	pSLIANLSAANCYK	32	
167	gi|15235363	Plasma membrane polypeptide family protein (DREPP)	17	29	pTFDESKETINK	71	√
					VpSVFLPEEVKTK	49	
					VVEpTYEATSAEVK	37	√
					AEEPAKpTEEPAK	35	√
					pTEEPAKTEGTSGEKEEIVEETK	89	√
					AVSEApSSSFGAGYVAGPVTFIEEK	155	
169	gi|1522987	14-3-3 protein GF14 chi isoform	10	38	EAAPAAAKPADEQQpS	44	
173	gi|18421762	14-3-3 protein GRF3 isoform	7	23	DNLTLWTpSDMTDEAGDEIK	38	
176	gi|166717	14-3-3 protein GF14 psi isoform	12	29	SASDIApTAELAPTHPIR	75	
178	gi|15218740	Extracellular dermal glycoprotein, putative	7	11	pTLPVPASVVFDLGGR	56	√
206	gi|7671458	14-3-3 like protein AFT1	12	42	LLDpSHLIPSAGASESK	46	
222	gi|15228674	Rotamase cyp 5 (ROC 5)	14	54	VFFDMSLpSGTPIGR	48	√
225	gi|16398	Nucleoside diphosphate kinase	7	37	NVIHGpSDSVESAR	31	√
229	gi|15227956	Parvulin 1 (PIN1AT)	10	57	VGDIpSDIVDTDSGVHIIIK	66	√
230	gi|15242724	Legume lectin family protein	19	44	KPLIVAHLDLpSK	65	√
Hormone and vitamin biosynthesis (2)							
161	gi|15239735	Thiazole biosynthetic enzyme (TH1)	17	29	HAALFTpSTIMSK	38	
					IVVpSSCGHDGPFGATGVK	75	
165	gi|15220770	ACC oxidase 2 (ACO2)	6	27	AHpTDAGGIILLFQDDK	76	√
					NASAVpTELNPTAAVETF	47	√
Nucleotide biosynthesis (3)							
119	gi|15235865	Pyrimidin 4 (PYR4)	8	15	EGpSVGATVTPQHLLLNR	61	√
120	gi|3850621	Putative RNA binding protein	19	40	GKLEpTESLLQSK	37	√
131	gi|1529384	Putative mRNA binding protein	10	27	SpTEQPPHVEGDAVK	32	
					AVPIPGpSGLQLTNISHVR	62	
Cell rescue and defense (7)							
113	gi|984052	Thioglucoside glucohydrolase 2 (TGG2)	15	25	DFLpSQGVRPSALK	39	
148	gi|15232129	PROHIBITIN 4	7	16	AEGEpSEAAQLISDATAK	67	√

174	gi|15233357	FIBRILLIN (FIB)	10	39	AElpSELITQLESK	42	√
175	gi|15228198	PYK-10-BINDING PROTEIN 1 (PBP1)	12	35	pSPEEVTGEEHGK VpYVGQAQDGISAVK GANLWDDGpSTHDAVTK TpSDVIGSDEGTHFTLQVK QTpSPPFGLEAGTVFELK	57 44 68 98 82	√ √ √
208	gi|18405982	Avirulence-responsive protein, putative	6	23	pTVEVVLTDTSEKK	43	
209	gi|15242451	Avirulence-responsive protein-related	4	23	IPEIVpSATLPGFK	36	
212	gi|18399630	Cysteine protenase inhibitor	8	42	SNpSLFPYELLEVVHAK	72	√
Amino acid biosynthesis and metabolism (10)							
101	gi|5238686	Cobalamin independent methionine synthase (ATCIMS)	32	41	pYGAGIGPGVYDIHSPR GMLpTGPVTILNWSFVR ALGVDpTVPVLVGPVSYLLLSK	102 37 48	
109	gi|14532594	Dihydroxyacid dehydratase, putative	17	21	VpSDAVPFLADLKPSGK	42	√
110	gi|4467128	Putative aminoacylase	15	19	pTSKPEIFPASTDAR	50	√
118	gi|15229033	S-adenosylmethionine synthetase 3 (MTO)	119	46	NGpTCPWLRPDGK VLVNIEQQpSPDIAQGVHGHLTK VHTVLIpSTQHDETVTNDEIAADLK	46 49 108	
121	gi|110740313	Cytosolic O-acetylserine (thiol)lyase	23	30	pYLSTVLFDDTR IGFpSMISDAEKK IDGFVpSGIGTGGTITGAGK	47 35 85	√ √
136	gi|531555	Aspartate aminotransferase	20	40	VVIpSSPTWGNHK IGAINVVCSpSADAATR	40 38	
141	gi|18404496	Binding/catalytic/coenzyme binding	22	46	ALDLApSKPEGTGTPTK ALDLApSKPEGTGTPTKDFK	47 48	√
154	gi|80490	Cysteine synthase	20	31	pYLSTVLFDATR IHYEpTTGPEIWK IDGFVpSGIGTGGTITGAGK	43 37 65	 √
155	gi|11131561	Cysteine synthase	12	33	IGFpSMISDAEQK	32	√
211	gi|15225026	Alanine glyoxylate aminotransferase (AGT)	10	22	ALpSLPTGLGIVCASPK	67	√

Cell wall biosynthesis (1)							
135	gi|15225605	Endo-xyloglucan transferase (EXGTA-1)	12	35	LpYSSLWNADDWATR	83	
Circadian rhythm (1)							
217	gi|15226605	ATGRP7	9	49	SGGGGGYSGGGGpSYGGGGGR	62	

1 – Total number of peptides (phosphorylated and non-phosphorylated) matched to the protein.

2 – Percentage (%) of the protein sequence covered by all matching peptides (phosphorylated and non-phosphorylated).

3 – Phosphorylated serine, threonine and tyrosine residues are marked pS, pT and pY, respectively.

4– Check marks indicate phosphopeptides that were also detected in F4 without using IMAC enrichment. The 11 phosphoproteins identified in the F4 that were not identified after IMAC are listed in Supplementary Table S2.

5 – Rubisco phosphopeptides for which MS/MS spectra are shown in Figure 5.

1031 Supplementary Table S2: *Arabidopsis thaliana* leaf phosphoproteins identified in the
1032 supernatant (F4) following PEG fractionation but only prior to IMAC enrichment.

Gene locus	Protein names	CV (%)	Phosphopeptides sequences	Ion score
gi|18396217	General regulatory factor 7 (GRF7)	56	EApSKHEPEEGKPAETGQ	44
gi|15221850	(S)-2-hydroxyacid oxidase	49	QLDYVPApTISALEEVVK	40
gi|15221850	F12M16.14	46	pTQDGGTEVVEAK	70
gi|30678347	Carbonic anhydrase 1 (CA1)	46	pYGGVGAAIEYAVLHLK VIpSELGDSAFEDQCGR EAVNVpSLANLLTYPFVR	72 94 49
gi|18408627	Haloacid dehydrogenase-like hydrolase family protein	47	AVpSAIVSCLLGPER KKPDPAIpYNLAAETLGVDPSK	51 46
gi|15242753	Adenosine kinase (ADK1)	54	GELVpSDDLVVGIIDEAMNKPK	45
gi|15228869	Copper chaperone (CCH)	57	VEpTVTETKTEAETK	52
gi|7658343	Unknown protein (13384-11892)	60	pSLGLDKDLSAALLGPR LAEGpTDITSAALLGPR KVVIFGLPGAYpTGVCSQQHVPSYK	82 72 32
gi|15233985	ATP synthase delta chain	40	QLEDIASQLELGEIQLApT SSSLQHSHTpSNFLNVLVDANR	85 48
gi|15236930	Vistitone reductase-related	36	LAGQDANVTTVPVpSVLR	51
gi|15241338	3-isopropylmalate dehydrogenase	31	ANPLApTILSAAMILLK	52

1033

1034

1035

1036

1037

1038

Figure 1

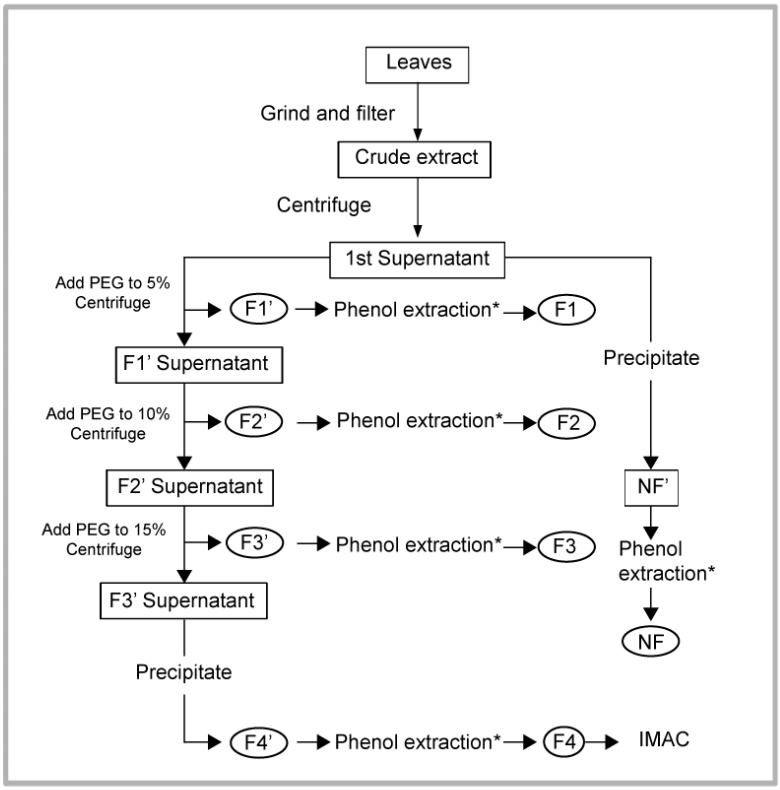

Leaves
Grind and filter
Crude extract
Centrifuge
1st Supernatant
Add PEG to 5%
Centrifuge
F1'
Phenol extraction*
F1
F1' Supernatant
Precipitate
Add PEG to 10%
Centrifuge
F2'
Phenol extraction*
F2
F2' Supernatant
NF'
Add PEG to 15%
Centrifuge
F3'
Phenol extraction*
F3
Phenol
extraction*
F3' Supernatant
NF
Precipitate
F4'
Phenol extraction*
F4
IMAC

Figure 2

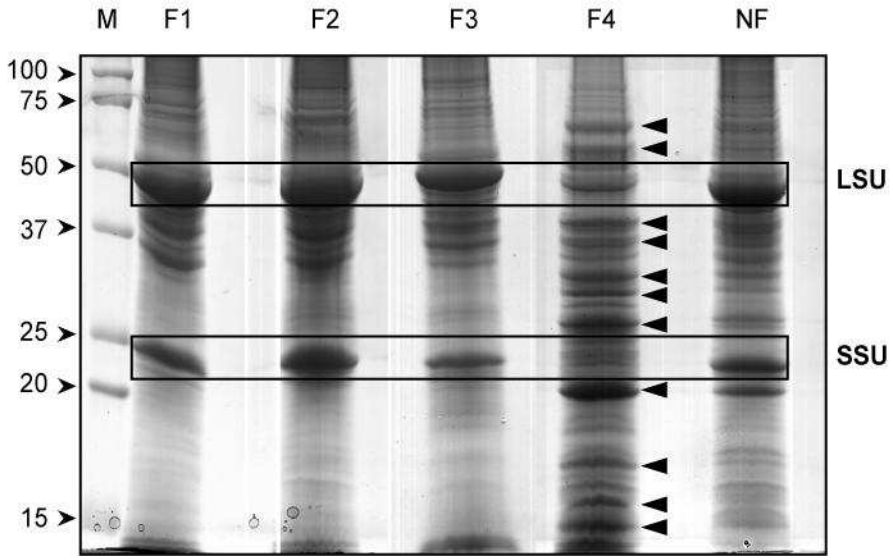


Figure 3

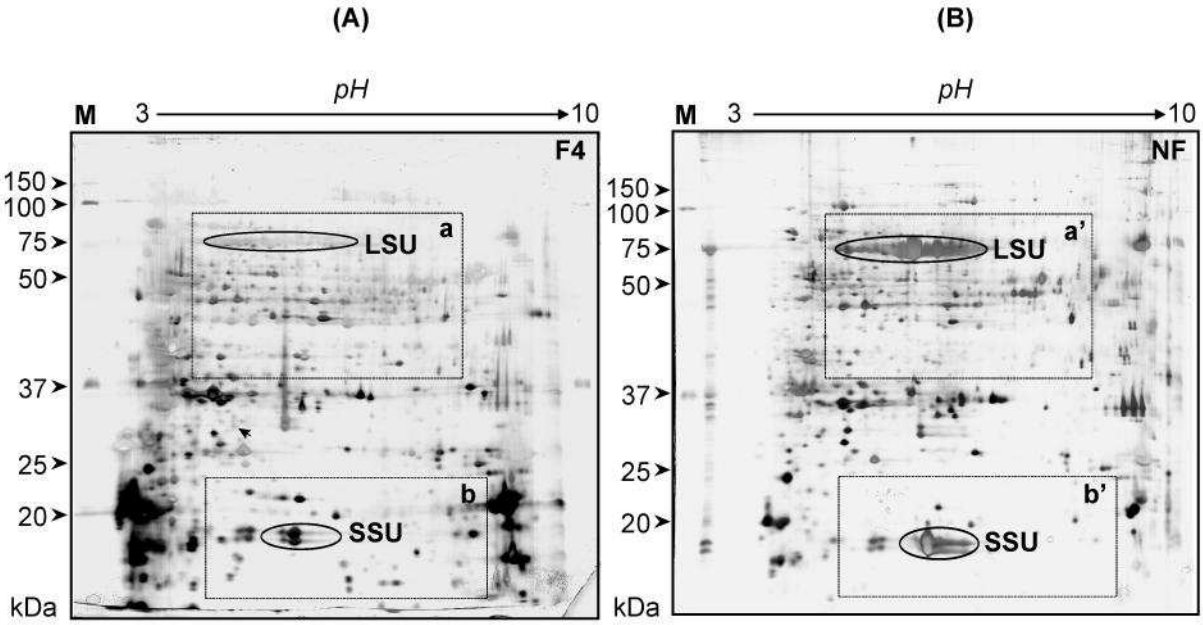
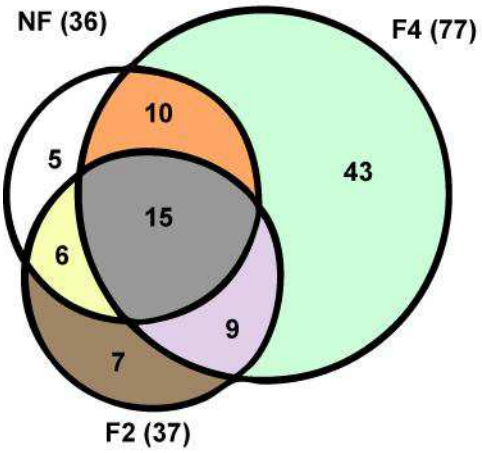
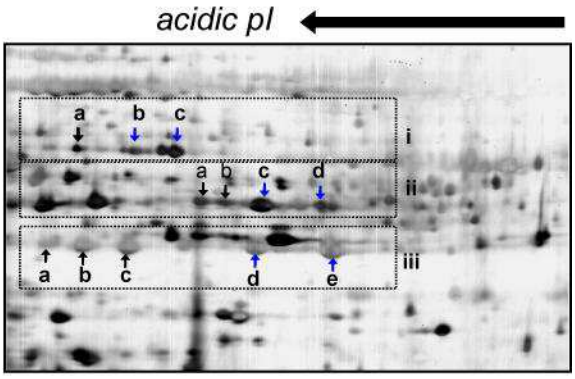


Figure 4

(A)



(B)



Spot	ID	PP	Spot	ID	PP
ia	PGK	3	iid	FBA	0
ib	PGK	0	iiia	ATLFNR1	3
ic	PGK	0	iiib	ATLFNR1	3
iia	FBA	5	iiic	ATLFNR1	2
iib	FBA	3	iiid	ATLFNR1	0
iic	FBA	0	iiie	ATLFNR1	0

1050

1051

Figure 5

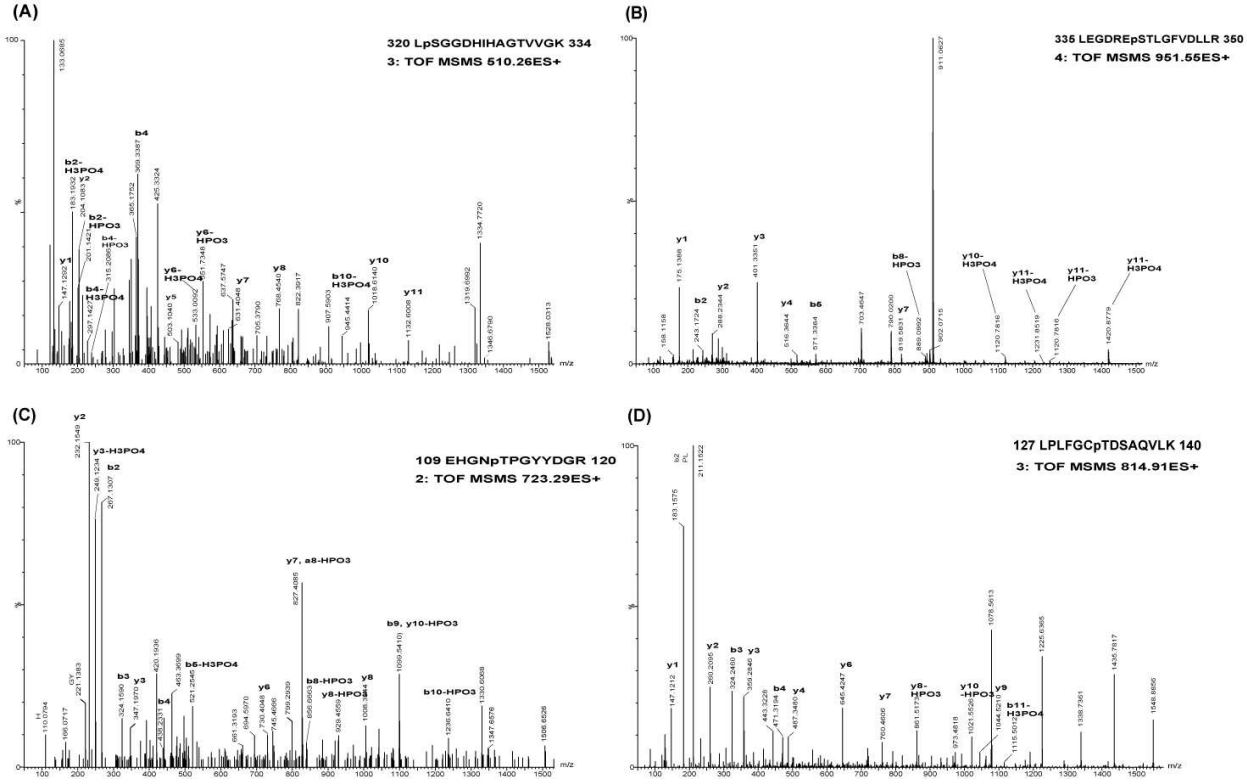
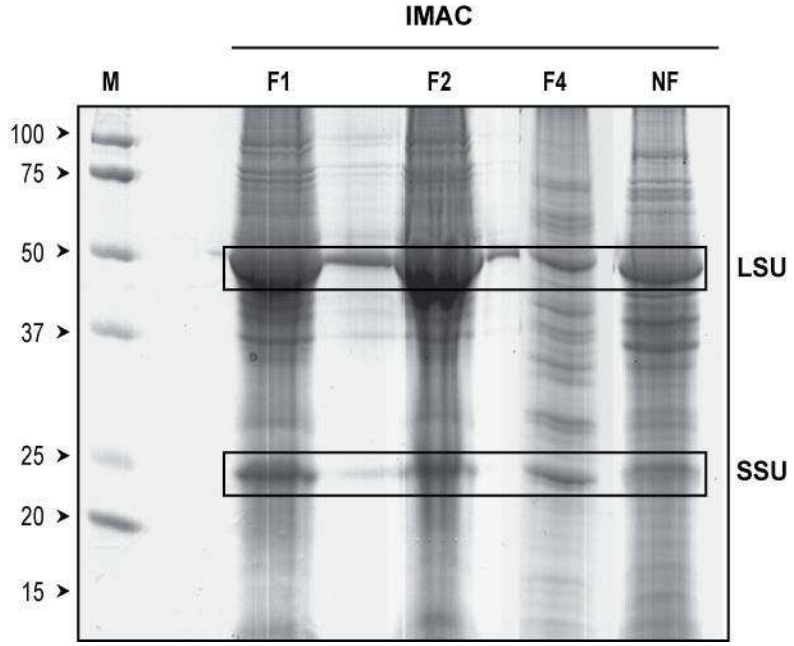

(A)
320 LpSGGDHIHAGTVVGK 334
3: TOF MSMS 510.26ES+
(B)
335 LEGDREpSTLGFVDLLR 350
4: TOF MSMS 951.55ES+
(C)
109 EHGNpTPGYYDGR 120
2: TOF MSMS 723.29ES+
(D)
127 LPLFGCpTDSAQVLK 140
3: TOF MSMS 814.91ES+

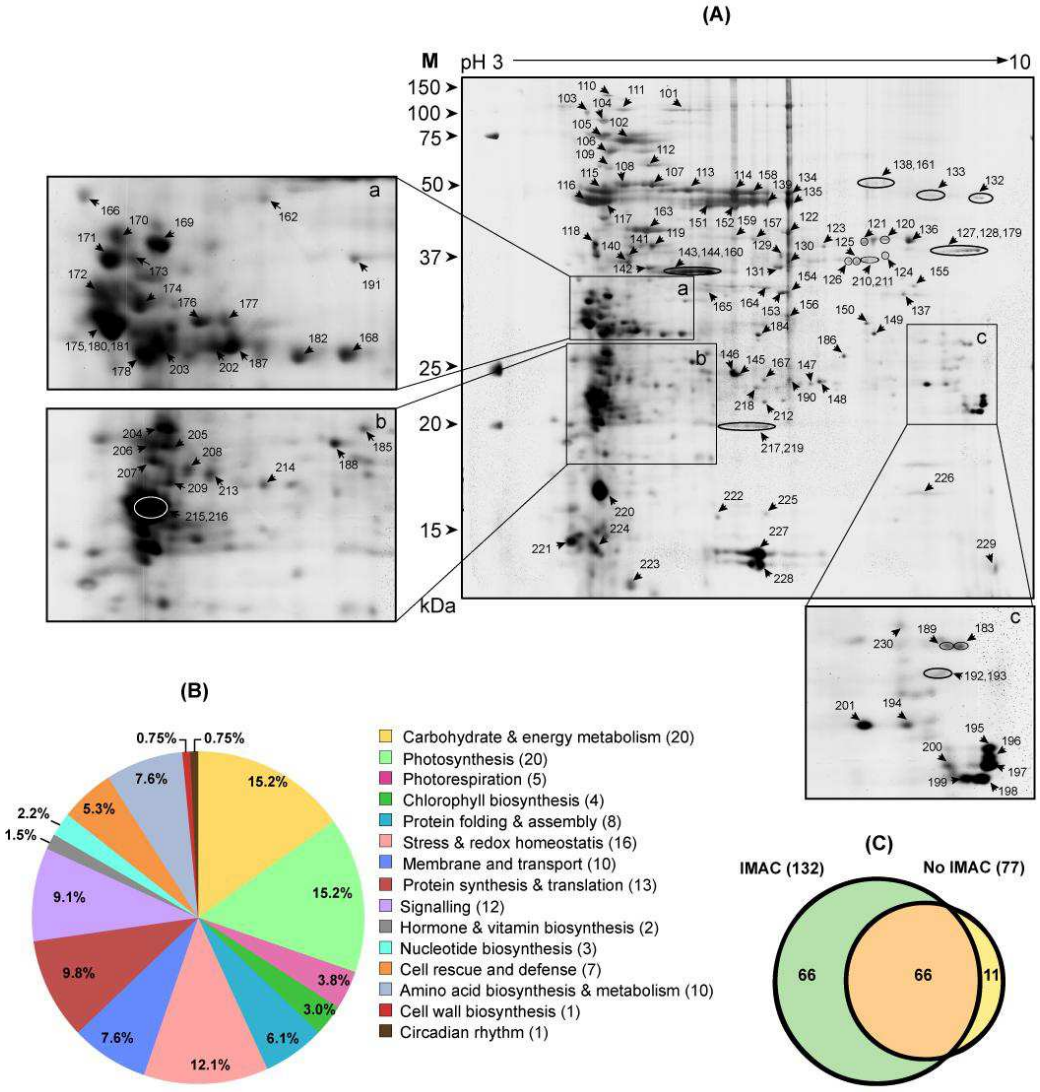
Figure 6



1054

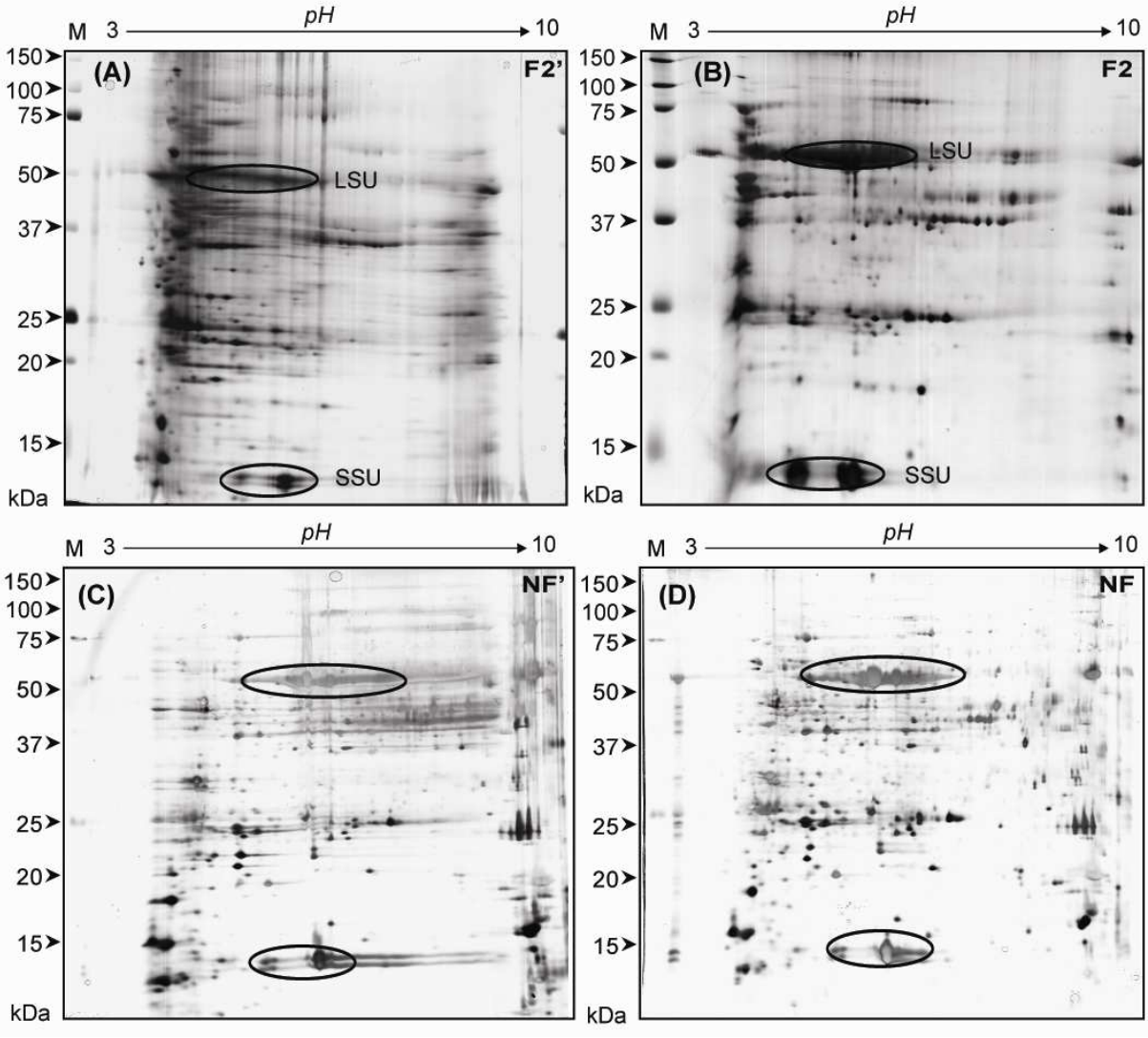
1055

Figure 7


(A)
M
pH 3
10
150
100
75
50
37
25
20
15
kDa
a
b
c
(B)
Carbohydrate & energy metabolism (20)
Photosynthesis (20)
Photorespiration (5)
Chlorophyll biosynthesis (4)
Protein folding & assembly (8)
Stress & redox homeostatis (16)
Membrane and transport (10)
Protein synthesis & translation (13)
Signalling (12)
Hormone & vitamin biosynthesis (2)
Nucleotide biosynthesis (3)
Cell rescue and defense (7)
Amino acid biosynthesis & metabolism (10)
Cell wall biosynthesis (1)
Circadian rhythm (1)
15.2%
15.2%
3.8%
3.0%
6.1%
12.1%
7.6%
9.8%
9.1%
1.5%
2.2%
5.3%
7.6%
0.75%
0.75%
(C)
IMAC (132)
No IMAC (77)
66
66
11

1057

Supplementary Figure S1

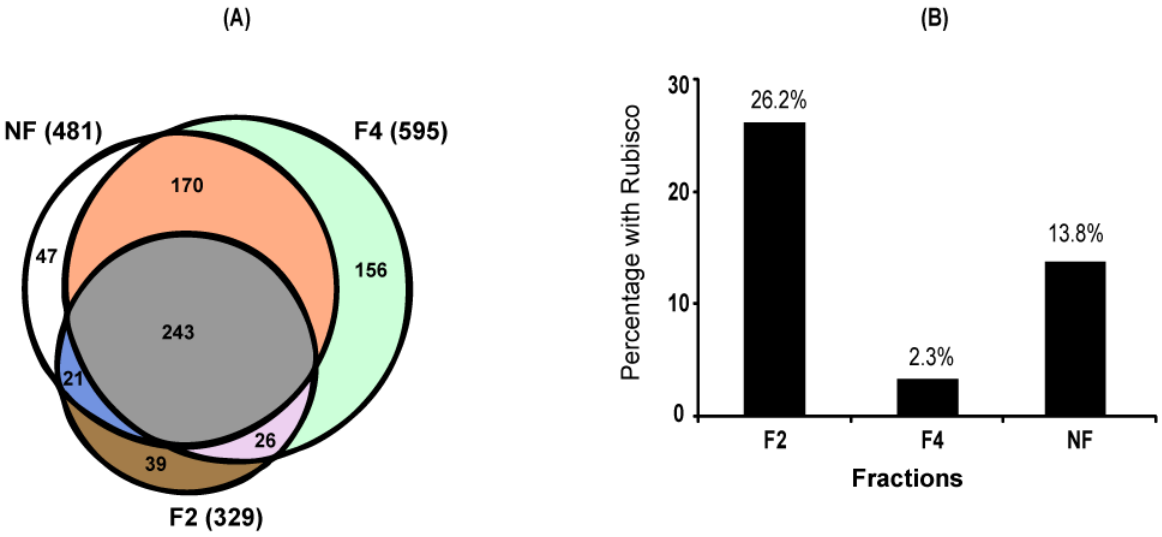


1058

1059

1060

Supplementary Figure S2

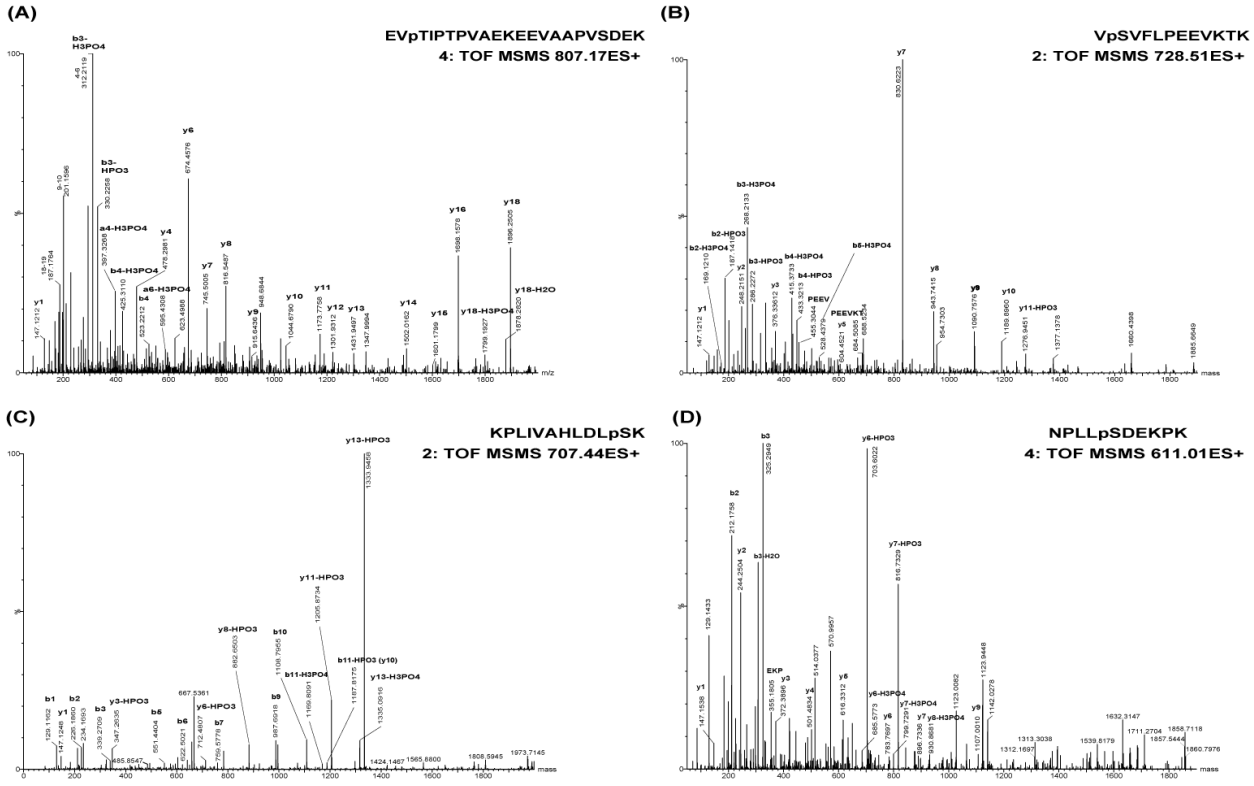


1061

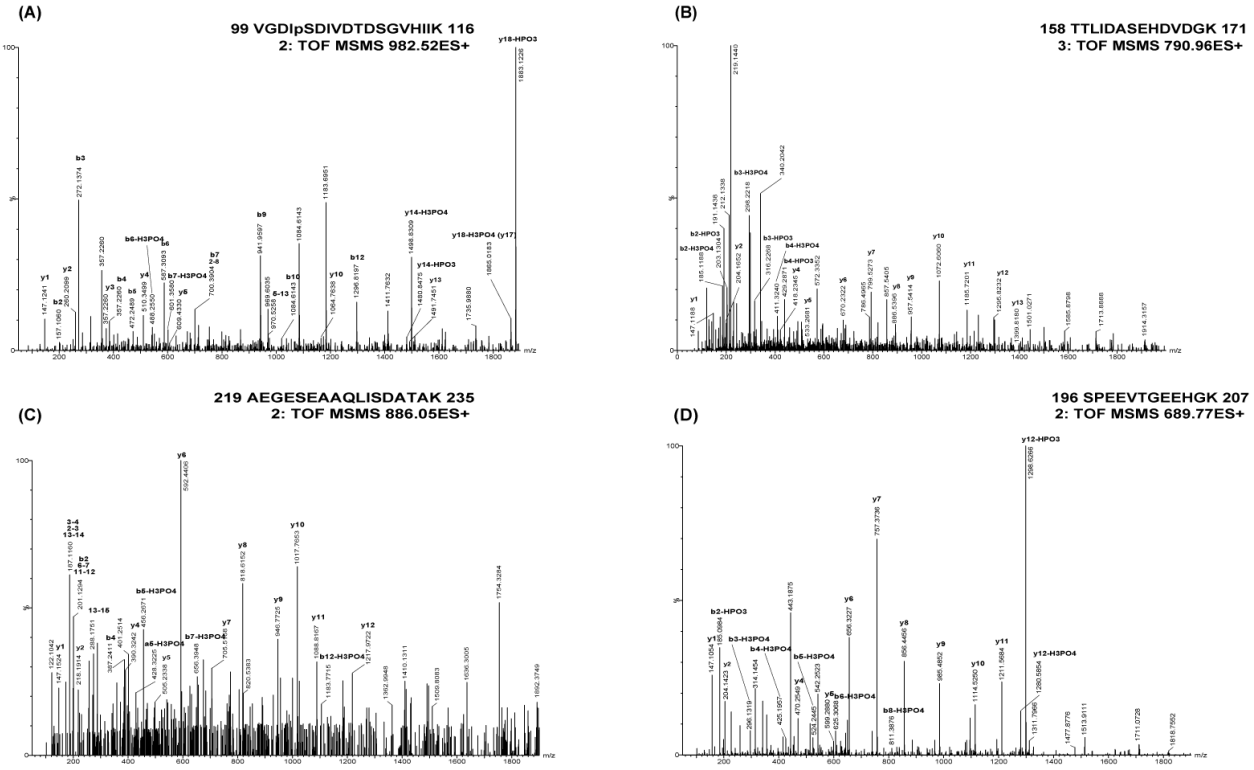
1062

1063

Supplementary Figure S3



Supplementary Figure S4



1067

1068

1069