Published in final edited form as:

Anal Chem. 2009 June 1; 81(11): 4210–4219. doi:10.1021/ac802487q.

An integrated workflow for characterizing intact phosphoproteins from complex mixtures

Si Wu[#], Feng Yang[#], Rui Zhao, Nikola Toli , Errol W. Robinson, David Camp II, Richard D. Smith, and Ljiljana Paša-Toli

Pacific Northwest National Laboratory, Richland, WA, 99352

Abstract

The phosphorylation of any site on a given protein can affect its activity, degradation rate, ability to dock with other proteins or bind divalent cations, and/or its localization. These effects can operate within the same protein; in fact, multisite phosphorylation is a key mechanism for achieving signal integration in cells. Hence, knowing the overall phosphorylation signature of a protein is essential for understanding the "state" of a cell. However, current technologies to monitor the phosphorylation status of proteins are inefficient at determining the relative stoichiometries of phosphorylation at multiple sites. Here we report a new capability for comprehensive liquid chromatography mass spectrometry (LC/MS) analysis of intact phosphoproteins. The technology platform built upon integrated bottom-up and top-down approach that is facilitated by intact protein reversed-phase (RP)LC concurrently coupled with Fourier transform ion cyclotron resonance (FTICR) MS and fraction collection. As the use of conventional RPLC systems for phosphopeptide identification has proven challenging due to the formation of metal ion complexes at various metal surfaces during LC/MS and ESI-MS analysis, we have developed a "metal-free" RPLC-ESI-MS platform for phosphoprotein characterization. This platform demonstrated a significant sensitivity enhancement for phosphorylated casein proteins enriched from a standard protein mixture and revealed the presence of over 20 casein isoforms arising from genetic variants with varying numbers of phosphorylation sites. The integrated workflow was also applied to an enriched yeast phosphoproteome to evaluate the feasibility of this strategy for characterizing complex biological systems, and revealed ~16% of the detected yeast proteins to have multiple phosphorylation isoforms. Intact protein LC/MS platform for characterization of combinatorial posttranslational modifications (PTMs), with special emphasis on multisite phosphorylation, holds great promise to significantly extend our understanding of the roles of multiple PTMs on signaling components that control the cellular responses to various stimuli.

Introduction

Posttranslational modification (PTM) of proteins, such as phosphorylation, plays a critical role in cell signaling and other fundamental cellular functions in living organisms [1–3].

*Corresponding author: Ljiljana Paša-Toli , ljiljana.pasatolic@pnl.gov, EMSL, MSIN: K8-98, Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352.

[#]contributed equally

Studies aimed at analyzing signaling pathways require methods that can specifically detect, identify, and quantify phosphoproteins. While traditional methods [4] typically allow characterization of one phosphoprotein (often limited to a particular phosphorylation site) at a time, recent advancements in LC/MS technology now enable proteome-wide study of phosphorylation (i.e. phosphoproteomics) [5–13]. In spite of numerous technological advances, phosphoproteome analyses are still challenged by the fact that only a small percentage of all cellular proteins are phosphorylated at any given time. Consequently, enriching the phosphorylated fraction prior to MS analysis is a prerequisite for being able to detect rare and possibly novel phosphoproteins. Enrichment strategies (e.g. IMAC) [14] are typically applied in a two-step scheme; that is, selectively isolate phosphoproteins first, then isolate phosphorylated tryptic peptides since MS analysis is typically performed at the peptide level (i.e. from the bottom-up) [11, 12, 15]. Although the bottom-up approach has allowed identification of thousands of phosphorylation sites in a proteome [5–13, 16], many phosphorylation sites remain unidentified due to incomplete sequence coverage. Similarly, it is not possible to assess whether different phosphopeptides are derived from one or more forms of the parent protein or to determine the occupancy of a given phosphorylation site when there are multiple sites on different peptides. This inability to precisely characterize endogenous phosphoprotein forms for the array of gene products is a significant drawback of conventional phosphoproteomic approaches.

Multiple gene products or protein isoforms are common and phosphorylation often coexists with other PTMs and may occur on multiple distinct sites on the proteins. Since the phosphorylation of any site can act as an on/off specific switch for protein activity or localization [17], knowing the relative abundances of the overall phosphorylation signature of “intact” protein isoforms (i.e. the occupancy and coordination of all sites) is essential for understanding the “state” of a cell and characterization of the cellular pathways.

Top-down mass spectrometry [18, 19] measures intact proteins and facilitates the characterization of protein isoforms including posttranslationally modified proteins [18–22]. Further characterization of their primary structure to determine the specific PTM site can be achieved by different fragmentation techniques (CID [23], ECD [24], and ETD [25]) at the intact protein level. The relative abundance of protein isoforms can be retrieved from MS peak intensities or stable isotope labeling approaches [21, 26, 27].

The top-down approach has been successfully applied for the characterization of various protein PTMs including phosphorylation [19–21, 27, 28]. However, previous phosphorylation characterization has been often limited to the study of a single purified protein. Capabilities for the broad MS-based characterization of intact phosphorylated proteins would provide new insights into biological systems.

In response to these challenges, we have developed a capability for comprehensive high-throughput analysis of intact phosphoproteins using FTICR MS and micro-separations. The technology platform was built upon a novel integrated top-down and bottom-up approach that is facilitated by intact protein reversed-phase liquid chromatography (RPLC) concurrently coupled with Fourier transform ion cyclotron resonance (FTICR) MS and fraction collection. [29] [30] The integrated strategy can be readily applied to measure

differential protein abundances, and provides a platform for selection of biologically relevant targets for further characterization using offline tandem MS (i.e. MS/MS). We have developed and optimized a metal free RPLC-ESI-MS platform for intact phosphoprotein analyses to minimize losses of phosphorylated species due to the formation of metal ion complexes at various metal surfaces. In this work, we have coupled this platform to a 12 Tesla FTICR mass spectrometer, which offers sensitive intact protein mass measurements with high resolution and mass measurement accuracy (MMA). In the proof of principle experiment, we have enriched and identified about twenty isoforms of α 1-casein, α 2-casein, and β -casein from a standard protein mixture containing phosphorylated and non-phosphorylated proteins. We have also used the integrated strategy to characterize the yeast phosphoproteome. Top-down proteomics determined 16% of the yeast phosphoproteome has multiple phosphorylation isoforms (such information is unattainable using the traditional bottom-up approach). The strategy reported in this work builds a foundation for characterizing multisite phosphorylation and accurately quantifying changes in degree of phosphorylation to enable the characterization of changes in the state of phosphorylation in complex biological systems. Such analysis will enhance the analysis capabilities of proteomics for systems biology research.

Experimental Procedures

Phosphoprotein enrichment from a standard protein mixture

Proteins for the standard protein mixture were purchased from Sigma: ubiquitin, β -lactoglobulin A, β -lactoglobulin B, β -casein, α -casein, carbonic anhydrase II, bovine serum albumin, lysozyme, and ribonuclease A. A 1 mg/mL stock solution of each chosen standard protein was prepared in water.

Phosphorylated proteins (α - and β -casein) were enriched using TALON PMAC phosphoprotein enrichment kit (Clontech, Mountain View, CA) according to the manufacturer's instructions. Briefly, equal amounts of the proteins listed above were mixed and then diluted with buffer A from the kit to a final concentration of 0.1 mg/mL for each protein. A 4.5 mL aliquot of this mixture was loaded onto the phosphoprotein enrichment column. After washes with buffer A, 1 mL of buffer B (20 mM sodium phosphate in 500 mM KCl) was used to elute the phosphoproteins off the column, and this elution step was repeated 4 more times. Each buffer B elution step was collected as a separate fraction. To verify phosphoprotein enrichment, an aliquot from each fraction was loaded onto a Bio-Rad pre-cast 4–12% SDS-polyacrylamide gel (Hercules, CA) and then stained with GelCode Blue Stain Reagent (Pierce, Rockford, IL).

For MS analysis, purified proteins were first buffer-exchanged (in 25 mM NH_4HCO_3) using Microcon centrifugal filter units (YM3, 3 kDa mass cutoff, Millipore, Billerica, MA) to remove the high non-volatile salt in buffer B. Due to notable leakage of iron from the enrichment kit, extensive buffer-exchange steps (3 times) were used.

Intact phosphoprotein LC-FTICR MS with/without on-line fractionation

The LC-FTICR MS with/without on-line fractionations were accomplished using Triversa NanoMate 100 (Advion BioSciences, Inc., Ithaca, NY). The RPLC system used for online intact protein separations was similar to previously reported system [31]. Briefly, for LC/MS analysis without on-line fractionation, a 75 μm i.d. $\times$ 70 cm column was packed in-house with Phenomenex Jupiter particles (C5 stationary phase, 5 μm particle diameter, 300 Å pore size). For LC/MS analysis with on-line fractionation, a 200 μm i.d. column was used due to the higher solvent flow rate through the column required to collect up to 96 fractions of sufficient volume for further analysis. Mobile phase A was composed of 0.01% trifluoroacetic acid (TFA), 0.6% acetic acid, 5% isopropanol, 25% acetonitrile (ACN), and the balance water, while mobile phase B consisted of 0.01% TFA, 0.6% acetic acid, 9.39% water, 45% isopropanol and 45% ACN. The operating pressure was 10,000 psi, and the flow rate was $\sim$ 300 nL/min and 5.5 $\mu\text{L}/\text{min}$ for the 75 μm i.d. and 200 μm i.d. column, respectively.

During on-line fractionation, $\sim$ 300 nL/min of the flow was directed to an nanoESI chip (Advion BioSciences, Inc., Ithaca, NY) for ionization and introduction into a modified Bruker 12 T APEX-Q FTICR mass spectrometer [31]. The remaining $\sim$ 5.2 $\mu\text{L}/\text{min}$ was collected into a 96-well plate. A back pressure of 0.25 to 0.35 psi and a voltage of 1.45 to 1.7 kV was used for the nanoESI employing an Advion NanoMate 100 system. A novel compensated trapped-ion cell with improved DC potential harmonicity was employed to enhance mass measurement accuracy and sensitivity [32, 33]. During the LC/MS analysis, a single mass spectrum was recorded using 512k data points, and the average of two mass spectra was used for data analysis.

On-plate fraction digestion and MS/MS analysis using RPLC-Ion Trap MS

To obtain bottom-up protein identifications for collected LC/MS fractions of interest, a 20 μL solution of 10 $\mu\text{g}/\text{mL}$ trypsin (Promega, Madison, WI) in 30% (v/v) aqueous ACN with 50 mM ammonium bicarbonate buffer (pH=8.2) was added to each well. Digestion was performed overnight at 37 °C. Organic solvent in the sample was removed by vacuum centrifugation. The final sample volume was then adjusted to 15 μL with 0.1 M acetic acid in water and analyzed using standard LC-MS/MS procedures [5] on a ThermoFisher LTQ linear ion trap (San Jose, CA).

Offline protein FTICR-MS/MS

ESI of the collected fractions was performed using the Advion NanoMate 100 autosampler and a nanoESI chip with previously stated conditions [32]. Precursor ions were manually selected using Xmass (Bruker Daltonics) to determine the proper quadrupole settings to transmit the ion of interest into the external collision hexapole. Selected ions were accumulated for 0.2 sec in the collision hexapole and N_2 gas was pulsed into the collision cell to increase trapping and fragmentation efficiency. The CID-MS/MS analyses were accomplished by reducing the DC offset on the collision hexapole from 0 V to -25 V during ion accumulation, and 10 to 50 mass spectra were averaged to get the fragment ion information of sufficient quality.

Data Analysis

Peptide MS/MS data obtained from the digested fractions were processed with SEQUEST (Thermo Scientific, San Jose, CA) [33] using the sequences of the standard protein mixture added to the Human international protein index (IPI) 2008 sequence database. During the SEQUEST data analysis no enzyme rules were applied. The identified peptides were then filtered according to the criteria suggested by Washburn et al. [34] and to include only fully tryptic peptides (no missed cleavages). The identified phosphopeptides were then manually confirmed. A list of identified proteins was created for each fraction that contained 5–40 kDa proteins supported by at least two distinct peptide identifications.

Intact protein RPLC-FTICR mass spectra were de-isotoped and clustered to calculate a mono-isotopic mass for each observed LC/MS feature using in-house developed software (ICR-2LS and Viper [37], available for download at <http://ncrr.pnl.gov/software/>) as described in Sharma et al. [31] and Wu et al.[34]. All the spectra were externally calibrated using myoglobin and ubiquitin spectra acquired in a standard protein LC/MS analysis.

The LC/MS feature clustering was based on the neutral mass, charge state, abundance, isotopic fit, and spectrum number (relating to RPLC retention time). Spectra that corresponded to a particular feature were summed, and the resulting spectrum reprocessed as described above. Next, all charge states in the m/z range were collapsed into a zero charge state spectrum (i.e. neutral mass). Accurate mono-isotopic masses from the intact protein LC/MS analysis were then searched against the corresponding provisional protein lists assembled from bottom-up data for tentative intact protein identifications. Discrepancies between the measured intact protein mass and the predicted mass for proteins in the provisional protein list were used to identify target proteins with potential PTMs (discussed in following section). Here, when reporting the molecular mass of a protein, we report the relative molecular mass (M_r) of the most abundant isotopic composition.

Intact protein MS/MS spectra were analyzed using ICR-2LS and/or ProSightPTM (<https://prosightptm.scs.uiuc.edu/>) combined with protein sequences identified in bottom-up analyses.

Results and discussion

A metal free HPLC platform for intact phosphoprotein analysis

An ideal LC/MS platform requires both high-resolution separation and high-sensitivity detection of the proteins. Previously, we have established an LC platform using high pressure which is critical for better intact protein separations using long small microparticle packed capillary columns [31]. However, since the major parts in this high-pressure system were made of stainless steel, which is well known to trap the phosphorylated species [38, 39], it was necessary to modify the system by maximally eliminating the metal parts to increase phosphoprotein sensitivity.

Therefore in this work, we have modified the high pressure LC/MS system for phosphoprotein analysis, as illustrated in Figure 1. The Triversa NanoMate silicon based nanoESI chip and conductive plastic tip were used for applying ESI voltage while collecting

fractions. The sample injection loop, column frit, and all metal unions were replaced with non-metal equivalents. Only the metal exposure in the high pressure switching valves remains (as non-metal alternatives are currently unavailable). Reducing the exposed metal surface area improved phosphoprotein sensitivity.

Figure 2 shows the LC/MS analysis results of 1 μ g β -casein using either the metal free interface or the standard interface. It should be pointed out that, for this set of experiments, both the metal free and standard interfaces include a metal column frit which was replaced with a Kasil frit for later experiments. Higher total ion chromatogram (TIC) intensity was obtained with the metal free configuration and more protein species were identified (Figure 2a). The spectra for β -casein variant B, the most abundant species in both cases had a 3.5 fold improvement in ion intensity using the metal free interface (Figure 2b). In addition, a much lower metal adduct peak (M+Fe, +53 Da peak from neutral mass spectrum) was also noticed with the metal free interface (Figure 2b). After exchanging the metal column frit with a non-metal equivalent, the metal adduct peak was further reduced (data not shown). These results confirm the sensitivity of phosphoprotein detection is increased by eliminating metal surfaces in the platform. To further improve RPLC performance, various chromatography conditions such as stationary phase, capillary column, ion-pairing agent, and solvent system were also optimized to improve the sensitivity and resolution of intact proteins (data not shown).

Enrichment of phosphorylated proteins

To test the efficiency of the phosphoprotein enrichment method, we applied a mixture of standard proteins to the PMAC phosphoprotein enrichment column, as described in the experimental section. The phosphoproteins were eluted of the enrichment column into 5 fractions. High enrichment of β -casein and α -casein, known phosphorylated proteins, was confirmed by SDS-PAGE (Figure 3) and intact LC/MS data. Non-phosphorylated proteins with a molecular mass less than 25 kDa were only detected in the flow through, and not detected in any of the five enriched fractions. For BSA, a 66 kDa protein, the majority of the protein eluted in the column flow through. A small portion was also detected in the enriched fractions as shown by the faint BSA band in Figure 3 likely due to the non-specific binding of the negatively charged groups in the protein to the PMAC column (the larger protein the greater degree of non-specific binding). Additional wash steps may reduce or eliminate this non-specific binding, but may also cause increased loss of the bound phosphoproteins. Overall, phosphoproteins constituted the majority of the proteins in the enriched fractions. For RPLC/MS of the enriched phosphoproteins, we combined fractions 2 through 5, dialyzed the sample against 25 mM NH_4HCO_3 , and concentrated them to an appropriate volume for analysis.

The phosphoprotein enrichment efficiency was further demonstrated by applying intact LC/MS analysis to the protein mixtures before and after phosphoprotein enrichment. The base peak chromatograms from FTICR mass spectra acquired during an RPLC separation of the standard proteins (Figure 4b) and enriched phosphoproteins (Figure 4a) show that most of the non-phosphorylated proteins were not detected after enrichment and that the phosphorylated proteins were effectively enriched. For carbonic anhydrase, only a small

portion, ~ 5% of the original intensity, was still detected in phosphoprotein enriched fraction. BSA was not detected in either LC/MS analysis. However, BSA was detected in an LC/MS analysis when other proteins were excluded from the sample solution [33]. The observed LC peak for BSA was broader than peaks observed for other proteins suggesting the reduced chromatographic resolution and matrix effects or ionization suppression due to co-elution with the casein proteins as the major contributing factors for the lower apparent sensitivity for BSA. It is unclear to what extent other factors contribute to the effect, though in this regard the LC/MS carbonic anhydrase results correlated as expected with the gel based results.

The relative percentage of detected phosphoprotein in each LC/MS analysis was estimated by dividing the ion intensity of phosphoproteins by the total ion intensity. Before enrichment, about 15% of the detected ion intensity was due to phosphorylated proteins and after enrichment 96%, a significant shift in the proportion of the signal from phosphoproteins. The actual enrichment efficiency might be slightly different due to some uncertainty of the contribution from BSA, which was not detected in either LC/MS analysis.

Analysis of intact enriched phosphorylated proteins

We now focus on the LC/MS analysis of the enriched phosphoprotein based on our recently introduced integrated top-down and bottom-up approach [30] which featured concurrent LC/MS analysis and fraction collection, allowing in-depth characterization of protein PTMs and genetic variants. Here, the collected fractions were separated into separate aliquots for bottom-up and intact MS/MS analysis. The relative abundance of each protein in a fraction was estimated by the ratio of identified peptides from that protein in that fraction to the number of peptides identified from that protein in all fractions [40–42]. This method is sufficient to determine the relative amount of each protein between fractions and protein elution profiles generated from the intact LC/MS data have the expected temporal correspondence to the abundance of proteins identified in bottom-up analysis (Figure 4c). A total of 4 phosphoproteins were identified by bottom-up analysis, and each of them displayed distinctive elution patterns. For instance, κ -casein, which is a known contaminant in typical β -casein preparations, mainly eluted in fraction 1, whereas α -casein S2 primarily eluted in fractions 2 to 4, α -casein S1 in fractions 5 to 8, and β -casein in fractions 9 to 13. Due to the differing elution patterns, the bottom-up results were used to constrain the identity of the intact protein masses observed in LC/MS analysis to a particular combination of PTMs and single nucleotide polymorphisms (SNPs).

Table 1 and Figure 5 illustrate a sample of the variety of protein isoforms, as detected using top-down proteomics, that co-exist for a single phosphoprotein identified using the integrated approach. Note that the identification of protein isoforms detected in the top-down analysis was facilitated by constraints obtained from the bottom-up results. Overall, the integrated strategy confirmed the presence of over 20 casein isoforms, arising from genetic variants (SNPs) and varying numbers of phosphorylation sites. For instance, five β -casein isoforms were assigned as β -casein genetic variants A2, A1, and B with 4 or 5 phosphorylation sites based on their accurate masses and corresponding bottom-up results (Figure 5). The three genetic variants A2 (M_r = 23983.23 Da), A1 (24023.22 Da), and B

(24092.34 Da) with 5 phosphorylation sites were further confirmed by top-down MS/MS (as illustrated in our previous study, [30]).

The bottom-up results (Figure 4c) revealed a majority of α -casein S1 eluted in fractions 5–8. Based on the intact protein LC/MS data we also found that three distinct proteins with M_r = 24092.34, 24023.22, and 23983.23 Da, eluted sequentially in this region, which were matched to the major genetic variant C of α -casein S-1 with different degrees of phosphorylation (i.e., 7 to 9 phosphorylation sites) (Figure 5). Intact protein CID data, obtained using reconstituted fractions, confirmed that these three proteins are in fact genetic variant C of α -casein S-1. For example, in the CID spectrum of putative α -casein S-1 with 7 phosphorylation sites, about 20 fragment ions that matched α -casein residues up to residue 40 were assigned as unmodified b ions, with the last identified unmodified y ion being y68. This indicates the absence of phosphorylation in this region (Figure 6a and 6b).

We also confirmed 2 phosphorylations in the region between residues 140 and 146 (QELAYFY) based on the internal fragment ions b72-140 to b72-146. Thus, intact protein MS/MS data constrained the phosphorylation sites between residue 41 and 205. Among them, 5 phosphorylated sites are between residues 41 to residue 71, and 2 additional phosphorylated sites are between residues 72 and 146. We were not able to identify any differentially phosphorylated fragment ions between the three phosphorylated isoforms. However, all intact protein MS/MS spectra contained the internal fragments b72-140 to b72-146 with 2 phosphorylations (Figure 6c). Therefore, the different degree of phosphorylation should occur between residues 41 and 71. From the bottom-up data, a peptide K.DIGSESTEDQAMEDIK.Q presented a single phosphorylation (S48) as well as double phosphorylation (S46 and S48), as illustrated in Figure 7. The monophosphorylated peptide primarily eluted in fraction 5, while the peptide with 2 phosphorylation sites primarily eluted in fraction 7. This is consistent with the sequential elution of intact variant C proteins with 7 to 9 phosphorylations from 65 min to 72.5 min (corresponding to fractions 5 to 8), as shown in Figure 5.

In addition to characterizing specific protein isoforms, the intact protein LC/MS data were also used to obtain information on the relative quantity of each protein isoform. As different isoforms have differing elution profiles, the relative quantitation was obtained from the average of isoform intensities across the elution window (Figure 8). It is important to note that using conventional bottom-up proteomics approaches it is not possible to assess whether different phosphopeptides are derived from the same protein molecule or to determine a percent occupancy of a given phosphorylation site. Since multisite phosphorylation serves as a common mechanism for increasing the regulatory potential of proteins, our protein-centric strategy overcomes a significant pitfall inherent with the conventional bottom-up approach. The integrated strategy greatly benefits from the protein identifications made in the bottom-up analysis to constrain the possible modified proteins which may be present in the sample. This combined approach facilitates reliable identifications of intact proteins and characterization of the protein isoforms.

Application for enriched yeast phosphoproteins

The integrated workflow was applied to characterize an enriched yeast phosphoproteome to demonstrate the feasibility of this strategy for complex biological systems. More than 500 yeast proteins were identified in a single LC/MS analysis using bottom-up only method, 70% of which have been previously reported as phosphoproteins indicating the high efficiency of phosphoprotein enrichment. The total number of identified yeast proteins is somewhat lower, relative to typical bottom-up LC/MS analysis, due in part to the selective enrichment of phosphoproteins (which represent a subset of all proteins), the relatively small amount of initial material, and a single LC/MS analysis as opposed to typical bottom-up analysis employing strong cation exchange fractionation followed by RPLC/MS. (Supplemental Table 1 lists all proteins identified in any of the three bottom-up LC-MS/MS analyses of the yeast phosphoproteins.) From a 2 µg enriched yeast protein sample (Figure 9), over 1,000 putative proteins (with M_r greater than 5 kDa) were detected in the intact protein RPLC-FTICR MS analysis.

We observed that a large portion of the detected proteins have multiple protein isoforms with a characteristic mass difference of ~80 Da indicating the presence of multiple phosphorylation sites, as illustrated in Figure 9. With phosphoprotein enrichment, a total of 16% of the detected proteins showed multiple protein isoforms with the characteristic mass difference of 79.95 ± 0.05 Da. Without phosphoprotein enrichment, only 3% of the detected proteins had multiple protein isoforms with the characteristic mass difference of 79.95 ± 0.05 Da. A subset of these proteins has been tentatively identified (Table 2) by matching bottom-up derived protein identifications with the intact protein accurate masses. All of the putative proteins are identified with at least one phosphorylated site. Preliminary results show various classes of PTMs in addition to phosphorylation, including methylation, acetylation, as well as proteolytic processing events.

In the next stages of this investigation, comparative proteomics will be employed for high throughput selection of biologically interesting candidates as targets for gas-phase MS/MS characterization. Differential abundance of (modified) intact proteins can readily be obtained using intact protein LC/MS profiles and proteins of interest can be readily selected for further investigation by targeted approaches, such as MS/MS or Western blots. The integrated top-down bottom-up strategy brings together the strengths of each method for a more complete characterization of expressed protein isoforms and is well suited for enrichment strategies to enhance the analysis of specific PTMs.

Conclusion

The integrated top-down and bottom-up approach facilitated by concurrent LC/MS analysis and fraction collection was combined with phosphoprotein enrichment for the targeted analysis of phosphorylated proteins. Increased sensitivity for phosphoprotein was obtained with a metal free, high-pressure (up to 10,000 psi) nanoRPLC-ESI-MS platform optimized for intact phosphoprotein separations. In a proof-of-principle experiment, metal free LC/MS and the integrated workflow were applied to analyze casein proteins enriched from a standard protein mixture. High enrichment efficiency was confirmed by both SDS-gel and intact LC/MS data and the integrated strategy revealed the presence of over 20 casein

isoforms, arising from genetic variants (SNPs) and varying degree of phosphorylation. This integrated approach is an efficient strategy for characterization of combinatorial PTMs, with special emphasis on multisite phosphorylation, for measuring differential protein abundances, and provides a means to select and further characterize biologically relevant targets by targeted MS/MS (or Western blots). These technological developments were also applied to analyze the yeast phosphoproteome in an initial attempt to characterize phosphoproteome at the intact protein level. The enriched phosphoprotein integrated top-down bottom-up approach holds great promise to significantly extend our understanding of the roles of multiple PTMs on signaling components that control the cellular responses to various stimuli.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Drs. Natacha Lourette, Keqi Tang, Anil Shukla, and Rui Zhang for contributing to the improvement of instrumental capabilities and performance. Portions of this work were supported by the National Center for Research Resources (RR 018522), the National Institute of Allergy and Infectious Diseases (NIH/DHHS through interagency agreement Y1-AI-4894-01), the National Institute of General Medical Sciences (NIGMS, R01 GM063883), and the U. S. Department of Energy (DOE) Office of Biological and Environmental Research. Work was performed in the Environmental Molecular Science Laboratory, a DOE national scientific user facility located on the campus of Pacific Northwest National Laboratory (PNNL) in Richland, Washington. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830.

References

1. Hubbard MJ, Cohen P. On target with a new mechanism for the regulation of protein-phosphorylation. *Trends Biochem. Sci.* 1993; 18:172–177. [PubMed: 8392229]
2. Hunter T. Signaling - 2000 and beyond. *Cell.* 2000; 100:113–127. [PubMed: 10647936]
3. Cohen P. The origins of protein phosphorylation. *Nat. Cell Biol.* 2002; 4:E127–E130. [PubMed: 11988757]
4. Yan JP, Garrus JE, Giebler HA, Stargell LA, Nyborg JK. Molecular interactions between the coactivator CBP and the human T-cell leukemia virus tax protein. *J. Mol. Biol.* 1998; 281:395–400. [PubMed: 9698555]
5. Yang F, Stenoien DL, Strittmatter EF, Wang JH, Ding LH, Lipton MS, Monroe ME, Nicora CD, Gristenko MA, Tang KQ, Fang RH, Adkins JN, Camp DG, Chen DJ, Smith RD. Phosphoproteome profiling of human skin fibroblast cells in response to low- and high-dose irradiation. *J. Proteome Res.* 2006; 5:1252–1260. [PubMed: 16674116]
6. Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villen J, Li JX, Cohn MA, Cantley LC, Gygi SP. Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. U. S. A.* 2004; 101:12130–12135. [PubMed: 15302935]
7. Chi A, Huttenhower C, Geer LY, Coon JJ, Syka JEP, Bai DL, Shabanowitz J, Burke DJ, Troyanskaya OG, Hunt DF. Analysis of phosphorylation sites on proteins from *Saccharomyces cerevisiae* by electron transfer dissociation (ETD) mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* 2007; 104:2193–2198. [PubMed: 17287358]
8. Ficarro SB, McClelland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM. Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat. Biotechnol.* 2002; 20:301–305. [PubMed: 11875433]
9. Ndassa YM, Orsi C, Marto JA, Chen S, Ross MM. Improved immobilized metal affinity chromatography for large-scale phosphoproteomics applications. *J. Proteome Res.* 2006; 5:2789–2799. [PubMed: 17022650]

10. Zhou HL, Watts JD, Aebersold R. A systematic approach to the analysis of protein phosphorylation. *Nat. Biotechnol.* 2001; 19:375–378. [PubMed: 11283598]
11. Collins MO, Yu L, Choudhary JS. Analysis of protein phosphorylation on a proteome-scale. *Proteomics.* 2007; 7:2751–2768. [PubMed: 17703509]
12. Delom F, Chevet E. Phosphoprotein analysis: from proteins to proteomes. *Proteome Sci.* 2006; 4:15. [PubMed: 16854217]
13. Dephoure N, Zhou C, Villen J, Beausoleil SA, Bakalarski CE, Elledge SJ, Gygi SP. A quantitative atlas of mitotic phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* 2008; 105:10762–10767. [PubMed: 18669648]
14. Andersson L, Porath J. Isolation of phosphoproteins by immobilized Metal (Fe-3+) affinity-chromatography. *Anal. Biochem.* 1986; 154:250–254. [PubMed: 3085541]
15. Kjeldsen F, Savitski MM, Nielsen ML, Shi L, Zubarev RA. On studying protein phosphorylation patterns using bottom-up LC-MS/MS: the case of human alphacasein. *Analyst.* 2007; 132:768–776. [PubMed: 17646876]
16. Ham BM, Yang F, Jayachandran H, Jaitly N, Monroe ME, Gritsenko MA, Livesay EA, Zhao R, Purvine SO, Orton D, Adkins JN, Camp DG, Rossie S, Smith RD. The influence of sample preparation and replicate analyses on HeLa cell phosphoproteome coverage. *J. Proteome Res.* 2008; 7:2215–2221. [PubMed: 18412383]
17. Cohen P. The regulation of protein function by multisite phosphorylation - a 25 year update. *Trends Biochem. Sci.* 2000; 25:596–601. [PubMed: 11116185]
18. McLafferty FW, Breuker K, Jin M, Han XM, Infusini G, Jiang H, Kong XL, Begley TP. Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics. *FEBS J.* 2007; 274:6256–6268. [PubMed: 18021240]
19. Siuti N, Kelleher NL. Decoding protein modifications using top-down mass spectrometry. *Nat. Methods.* 2007; 4:817–821. [PubMed: 17901871]
20. Du Y, Parks BA, Sohn S, Kwast KE, Kelleher NL. Top-down approaches for measuring expression ratios of intact yeast proteins using Fourier transform mass spectrometry. *Anal. Chem.* 2006; 78:686–694. [PubMed: 16448040]
21. Roth MJ, Forbes AJ, Boyne MT, Kim YB, Robinson DE, Kelleher NL. Precise and parallel characterization of coding polymorphisms, alternative splicing, and modifications in human proteins by mass spectrometry. *Mol. Cell. Proteomics.* 2005; 4:1002–1008. [PubMed: 15863400]
22. Thomas CE, Kelleher NL, Mizzen CA. Mass spectrometric characterization of human histone H3: A bird's eye view. *J. Proteome Res.* 2006; 5:240–247. [PubMed: 16457588]
23. Loo JA, Edmonds CG, Smith RD. Primary Sequence Information from intact proteins by electrospray ionization tandem mass-spectrometry. *Science.* 1990; 248:201–204. [PubMed: 2326633]
24. Zubarev RA, Horn DM, Fridriksson EK, Kelleher NL, Kruger NA, Lewis MA, Carpenter BK, McLafferty FW. Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal. Chem.* 2000; 72:563–573. [PubMed: 10695143]
25. Coon JJ, Ueberheide B, Syka JEP, Dryhurst DD, Ausio J, Shabanowitz J, Hunt DF. Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* 2005; 102:9463–9468. [PubMed: 15983376]
26. Pesavento JJ, Mizzen CA, Kelleher NL. Quantitative analysis of modified proteins and their positional isomers by tandem mass spectrometry: Human histone H4. *Anal. Chem.* 2006; 78:4271–4280. [PubMed: 16808433]
27. Roth MJ, Parks BA, Ferguson JT, Boyne MT, Kelleher NL. "Proteotyping": Population proteomics of human leukocytes using top down mass spectrometry. *Anal. Chem.* 2008; 80:2857–2866. [PubMed: 18351787]
28. Zabrouskov V, Ge Y, Schwartz J, Walker JW. Unraveling molecular complexity of phosphorylated human cardiac troponin I by top down electron capture dissociation/electron transfer dissociation mass spectrometry. *Mol Cell Proteomics.* 2008; 7:1838–1849. [PubMed: 18445579]
29. Wenger CD, Boyne MT 2nd, Ferguson JT, Robinson DE, Kelleher NL. Versatile online-offline engine for automated acquisition of high-resolution tandem mass spectra. *Anal Chem.* 2008; 80:8055–8063. [PubMed: 18841935]

30. Wu S, Lourette NM, Tolic N, Zhao R, Robinson EW, Tolmachev AV, Smith RD, Pasa-Tolic L. An Integrated Top-Down and Bottom-Up Strategy for Broadly Characterizing Protein Isoforms and Modifications. *J Proteome Res.* 2009
31. Sharma S, Simpson DC, Toli N, Jaitly N, Mayampurath AM, Smith RD, Pasa-Tolic L. Proteomic profiling of intact proteins using WAX-RPLC 2-D separations and FTICR mass spectrometry. *J. Proteome Res.* 2007; 6:602–610. [PubMed: 17269717]
32. Tolmachev AV, Robinson EW, Wu S, Kang H, Lourette NM, Paša-Toli L, Smith RD. Trapped-ion cell with improved DC potential harmonicity for FT-ICR MS. *J. Am. Soc. Mass. Spectrom.* 2008; 19:586–597. [PubMed: 18296061]
33. Tolmachev AV, Robinson EW, Wu S, Paša-Toli L, Smith RD. FT-ICR MS optimization for the analysis of intact proteins. *Int. J. Mass Spectrom.* 2008 accepted for publication.
34. Wu S, Lourette N, Toli N, Zhao R, Robinson EW, Tolmachev AV, Smith RD, Pasa-Tolic L. An integrated top-down and bottom-up strategy for characterizing intact proteins and their modifications. 2007 submitted.
35. Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. *J. Am. Soc. Mass. Spectrom.* 1994; 5:976–989. [PubMed: 24226387]
36. Washburn MP, Wolters D, Yates JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* 2001; 19:242–247. [PubMed: 11231557]
37. Monroe ME, Toli N, Jaitly N, Shaw JL, Adkins JN, Smith RD. VIPER: an advanced software package to support high-throughput LC-MS peptide identification. *Bioinformatics.* 2007; 23:2021–2023. [PubMed: 17545182]
38. Tuytten R, Lemiere F, Witters E, Van Dongen W, Slegers H, Newton RP, Van Onckelen H, Esmans EL. Stainless steel electrospray probe: A dead end for phosphorylated organic compounds? *J. Chromatogr. A.* 2006; 1104:209–221. [PubMed: 16378618]
39. Asakawa Y, Tokida N, Ozawa C, Ishiba M, Tagaya O, Asakawa N. Suppression effects of carbonate on the interaction between stainless steel and phosphate groups of phosphate compounds in high-performance liquid chromatography and electrospray ionization mass spectrometry. *J. Chromatogr. A.* 2008; 1198:80–86. [PubMed: 18541253]
40. Gao J, Opiteck GJ, Friedrichs MS, Dongre AR, Hefta SA. Changes in the protein expression of yeast as a function of carbon source. *J. Proteome Res.* 2003; 2:643–649. [PubMed: 14692458]
41. Liu HB, Sadygov RG, Yates JR. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* 2004; 76:4193–4201. [PubMed: 15253663]
42. Pang JX, Ginanni N, Dongre AR, Hefta SA, Opiteck GJ. Biomarker discovery in urine by proteomics. *J. Proteome Res.* 2002; 1:161–169. [PubMed: 12643536]

Figure 1.

Metal free HPLC system coupled with 12 Tesla FTICR-MS. Valve arrangement for (a) 75 μm ID and (b) 200 μm columns. Note the use of a solid phase extraction (SPE) column for loading the narrower ID columns. The standard configuration uses a stainless steel injection loop while the metal free design uses a fused silica sample loop. (c) Standard setup, which uses a metal screen as a column frit and a metal union to apply the ESI high voltage. (d) Metal free design which uses a Kasil frit and a PicoClear union for coupling to Triversa NanoMate (metal free LC-ESI interface). (e) Metal free design as in (d) with a larger ID column to facilitate fraction collection using Triversa Nanomate. Collected fractions were

analyzed using both conventional bottom-up proteomics and by tandem MS analysis of the intact proteins with CID in the external accumulation hexapole.

Figure 2.

Improved sensitivity using metal free interface. (a) The total ion chromatogram reconstructed from the FTICR spectra acquired during the RPLC separation of 1 μg of β -casein with “reduced-metal” (top) and standard (bottom) interfaces. (b) FTICR spectra and neutral mass spectra (insets) of β -casein variant A2 with reduced-metal (top) and standard (bottom) interfaces. The injector (needle, loop and valve) and the stainless steel unions employed in the conventional RPLC system are capable of mimicking the Fe(III)-IMAC behavior and

hence trap phosphorylated proteins/peptides. Stars in (a) indicate elution times of mass spectra in (b). Asterisks indicate noise peaks in (b).

Figure 3.

Enrichment of phosphoproteins from a mixture of standard proteins with a PMAC column was demonstrated by the differences in observed bands on a Bio-Rad 4–12% SDS polyacrylamide gel. The lanes of the gel were loaded with aliquots from either a molecular weight reference (Marker), the mixture of standard proteins before enrichment (Sample), column flow through (FT), or sequential fractions of collected column elution steps (E1-E5).

Figure 4.

Total ion chromatogram reconstructed from the FTICR spectra acquired during the RPLC separation before (a) and after (b) phosphoprotein enrichments. (c) 2D display reconstructed from the intact LC/MS data with the elution profile pattern obtained from bottom-up results. Heat map representation of protein elution patterns generated for the later portion of the LC/MS analysis using tryptic peptides identified in each fraction. The observation peptide counts were normalized with the total peptide counts for the protein from all the fractions

(row), with the scale ranging from 0 (i.e. least abundant, green) to 1 (i.e. most abundant, red). Each column in the heat map represent results obtained from the same RPLC fraction.

Figure 5.

Integrated analysis of a standard phosphoprotein mixture. Tentative identifications of proteins and modified proteins were accomplished by matching bottom-up data with the measured intact protein masses. Here are shown a sample of identified isoforms for α2-casein (a–d), α1- casein (e–g), and β-casein (h–j). The LC elution time corresponding to spectra (a–j) is indicated by corresponding letters on the ion chromatogram shown in (k).

Figure 6.

Bovine α -casein S1 variant C isoforms with 7, 8 or-9 phosphorylated sites were identified using the integrated approach. a) Intact MS/MS spectrum of bovine α -casein S1 variant C with 7 phosphorylated sites. b) The assigned fragment ions in Figure 6a with identified internal fragments highlighted using red fonts. c) A portion of the m/z spectrum to highlight some of the identified internal fragment ions indicating 2 phosphorylated sites between residue 70 and residue 140 for all three isoforms.

Figure 7.

Tandem mass spectra of the same peptide sequence with different degrees of phosphorylation. The proteins from which the peptides originated eluted in different LC fractions (a). MS/MS spectra of K.DIGSEpSTEDQAMEDIK.Q (b) and K.DIGpSEpSTEDQAMEDIK.Q (c) have been annotated to identify some of the observed MS/MS peaks. Peaks labels with asterisks are Fe adducts.

Figure 8.

Integrated analysis of a standard phosphoprotein mixture. “Bird’s eye view” offered by LC/MS data facilitates relative quantitation of $\alpha 2$ -casein (a), $\alpha 1$ -casein (b), and β -casein (c) phosphoprotein isoforms. The TIC of the LC/MS analysis (d) indicates which LC peaks were integrated to obtain spectra (a–c). The peaks labeled with asterisks are Fe adduct peaks.

Figure 9.

Example elution profiles and mass spectra of putative phosphorylated proteins (a–d) and a 2D display obtained for LC/MS analysis with a 75 μ m ID column of phosphoprotein enriched yeast lysate (e). The spots in the 2D display corresponding to protein isoforms (a–d) have been indicated. Tentative identifications of proteins and modified proteins were derived by matching bottom-up data (without fractionation) with the measured intact protein masses.

Table 1

Casein Isoforms Identified by using Phosphoprotein Enrichment Combined with Top-Down and Bottom-Up Proteomics

Protein ID	Protein	Measured Mr (Da)	Theoretical M _r (Da)	MMA (ppm)	Modifications
1	β-casein, variant B	24092.2150	24092.2661	-2.12	5 phosphorylations, P67H(SNP) and S122R (SNP) to variant A2
2	β-casein, variant A1	24023.1573	24023.1971	-1.66	5 phosphorylations, P67H(SNP) to variant A2
3	β-casein, variant A2	23983.1426	23983.1910	-2.02	5 phosphorylations
4	β-casein, variant A1	23943.1404	23943.2290	-3.70	4 phosphorylations, P67H(SNP) to variant A2
5	β-casein, variant A2	23903.1304	23903.2229	-3.87	4 phosphorylations
6	α-casein2, variant B	25066.1316	25066.0832	1.93	9 phosphorylations
7	α-casein2, variant B	25146.0688	25146.0513	0.70	10 phosphorylations
8	α-casein2, variant B	25225.9789	25226.0194	-1.61	11 phosphorylations
9	α-casein2, variant B	25305.9335	25305.9875	-2.13	12 phosphorylations
10	α-casein2, variant B	25385.9395	25385.9556	-0.63	13 phosphorylations
11	α-casein2, variant B	25465.9964	25465.9237	2.85	14 phosphorylations
12	α-casein1, variant B	23454.3416	23454.3163	1.08	6 phosphorylations
13	α-casein1, variant B	23534.3109	23534.2826	1.20	7 phosphorylations
14	α-casein1, variant B	23614.2199	23614.2489	-1.23	8 phosphorylations
15	α-casein1, variant B	23694.1793	23694.2152	-1.52	9 phosphorylations
16	α-casein1, variant C	23622.2894	23622.1941	4.03	9 phosphorylations
17	α-casein1, variant C	23542.3093	23542.2377	3.04	8 phosphorylations
18	α-casein1, variant C	23462.3269	23462.2615	2.79	7 phosphorylations

Table 2
Example Yeast Proteins Identified by Phosphoprotein Enrichment Followed by by Top-Down and Bottom-Up Proteomics

UniProt #	Protein	Measured Mr (Da)	Theoretical Mr (Da)	MMA (ppm)	Modifications
YBR085C-A	Uncharacterized protein YBR085C-A	9023.2383	9023.2117	2.95	5 phosphorylation, loss of N-term 7 residue
YBR085C-A	Uncharacterized protein YBR085C-A	9714.5522	9714.5658	-1.40	5 phosphorylation, 1 acetylation, loss of N-term Met
YCR031C	40S ribosomal protein S14-A (RP59A)	14700.7195	14700.7173	0.15	1 phosphorylation, 2 acetylation
YDR382W	60S acidic ribosomal protein P2-beta (P2B)	11209.3462	11209.3256	1.84	2 phosphorylation
YDR424C	Dynein light chain 1, cytoplasmic	10347.2519	10347.2927	-3.94	3 phosphorylation, loss of N-term 3 residues
YDR424C	Dynein light chain 1, cytoplasmic	10629.2439	10629.318	-6.97	4 phosphorylation, loss of N-term Met
YFL014W	12 kDa heat shock protein	11682.6032	11682.6114	-0.70	1 phosphorylation, 1 acetylation, loss of N-term Met
YFL014W	12 kDa heat shock protein	11602.6226	11602.6451	-1.94	1 acetylation, loss of N-term Met
YGR035C	Uncharacterized protein YGR035C	13137.3846	13137.3944	-0.75	4 phosphorylation, 1 acetylation, loss of N-term Met
YHL015W	40S ribosomal protein S20	13986.5802	13986.5342	3.29	1 phosphorylation
YHR132W-A	Protein IGO2	13510.6029	13510.6100	-0.53	5 phosphorylation, 1 methylation, loss of N-term 12 residue
YIL138C	Tropomyosin-2	19101.1719	19101.1965	-1.29	3 phosphorylation, 2 methylation, loss of N-term 2 residue
YLL050C	Cofilin (Actin-depolymerizing factor 1)	16276.7906	16276.8377	-2.89	4 phosphorylation, 1 acetylation, 1 methylation
YLR390W-A	Covalently-linked cell wall protein 14 (Inner cell wall protein)	23484.6962	23484.8575	-6.87	2 phosphorylation, 1 acetylation, 1 methylation