Methods 54 (2011) 379–386

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

A solid phase extraction-based platform for rapid phosphoproteomic analysis

Noah Dephoure *, Steven P. Gygi *

Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA

ARTICLE INFO

Article history:
Available online 1 April 2011

Keywords:
Proteomics
Phosphorylation
Chromatography/methods
Mass spectrometry/methods
Phosphoproteomics

ABSTRACT

Protein phosphorylation is among the most common and intensely studied post-translational protein modification. It plays crucial roles in virtually all cellular processes and has been implicated in numerous human diseases, including cancer. Traditional biochemical and genetic methods for identifying and monitoring sites of phosphorylation are laborious and slow and in recent years have largely been replaced by mass spectrometric analysis. Improved methods for phosphopeptide enrichment coupled with faster and more sensitive mass spectrometers have led to an explosion in the size of phosphoproteomic datasets. However, wider application of these methods is limited by equipment costs and the resultant high demand for instrument time as well as by a technology gap between biologists and mass spectrometrists. Here we describe a modified two-step enrichment strategy that employs lysC digestion and step elution from self-packed strong cation exchange (SCX) solid phase extraction (SPE) columns followed by immobilized metal ion affinity chromatography (IMAC) and LC–MS/MS analysis using a hybrid LTQ Orbitrap Velos mass spectrometer. The SCX procedure does not require an HPLC system, demands little expertise, and because multiple samples can be processed in parallel, can provide a large savings of time and labor. We demonstrate this method in conjunction with stable isotope labeling to quantitate peptides harboring >8000 unique phosphorylation sites in yeast in 12 h of instrument analysis time and examine the impact of enzyme choice and instrument platform.

1. Introduction

Reversible protein phosphorylation modulates protein activity, localization, stability, and protein–protein interactions, and in so doing regulates a diverse array of cellular events that govern nearly every aspect of cellular growth and division [1]. Traditional methods for studying protein phosphorylation, such as in vitro kinase assays, metabolic labeling with radioactive phosphate, or the use of phospho-specific antibodies are laborious and slow and often require prior knowledge or educated guesses about the sites under study. In the wake of whole genome sequencing, mass spectrometric analysis has emerged as a powerful tool for unbiased protein analysis and has proven especially well-suited to the analysis of protein phosphorylation.

Search algorithms such as SEQUEST [2] and Mascot [3] match experimental peptide mass spectra with theoretic spectra generated from protein database sequences. In conjunction with a target-decoy searching strategy [4], they can accurately and reliably identify thousands of peptides and proteins from complex mixtures without prior knowledge of the sample content. Phosphopeptides are identified by considering the 79.9963 dalton mass difference of a phosphate group in each spectral match. In many cases, specific MS/MS fragment ions can uniquely localize the site of phosphorylation within the peptide. In some of the largest datasets, >10,000 unique phosphorylation sites have been identified [5–12].

The success of mass spectrometric phosphorylation analysis was only achieved after overcoming a number of complicating factors. Phosphate groups are particularly labile. Upon fragmentation for MS/MS, phosphopeptides often simply lose phosphoric acid, resulting in a neutral loss of 98 daltons, and may fail to produce sufficient fragmentation of the peptide backbone for identification. A number of strategies such as reselection of the neutral loss ion for additional fragmentation and MS^3 [13] analysis, additional activation and pseudo MS^3 [14] analysis, or alternate fragmentation methods, such as ETD [15], have been attempted with some success. More recent work suggests that this hurdle has largely been overcome by faster and more sensitive instrumentation [16].

A much more difficult problem is the low stoichiometry of most phosphorylation sites. It is still nearly impossible to detect phosphopeptides in complex mixtures without enrichment. One

Abbreviations: ACN, acetonitrile; ETD, electron transfer dissociation; IMAC, immobilized metal ion affinity chromatography; SCX, strong cation exchange chromatography; m/z, mass to charge ratio; CID, collision induced dissociation; S/N, signal to noise ratio; SPE, solid phase extraction; FA, formic acid; HPLC, high performance liquid chromatography; AGC, automatic gain control; FDR, false discovery rate; SILAC, stable isotope labeling with amino acids in cell culture.

* Corresponding authors.
E-mail addresses: noah_dephoure@hms.harvard.edu (N. Dephoure), sgygi@hms.harvard.edu (S.P. Gygi).

1046-2023/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymeth.2011.03.008

effective strategy for enrichment is the use of phospho-specific antibodies. This has been most successful in the analysis of tyrosine phosphorylation [17], due to the high selectivity of existing antibodies. Similar quality antibodies for phospho-serine and phospho-threonine have not been described.

The success of unbiased large-scale phosphoproteomic analysis has been enabled by two major enrichment techniques: (1) phospho-affinity purification, using immobilized metal ion affinity chromatography [18] (IMAC) or titanium dioxide [19] (TiO_2), and (2) charge-based separation using strong cation exchange chromatography (SCX). Early applications of each demonstrated their effectiveness, but resulted in quite modest datasets by current standards. In a breakthrough report in 2001 using IMAC, Ficarro et al. [20] reported 383 phosphorylation sites from *Saccharomyces cerevisiae*, the largest such dataset at the time. In 2004, 2002 sites enriched by SCX from HeLa cell nuclear extracts were reported [21], again a then unprecedented number.

Strong cation exchange chromatography separates peptides by their solution charge state. When performed at a pH below the pKa for the acidic amino acids, only phosphorylated amino acids can contribute a negative charge, and a given phosphopeptide will elute earlier than its unphosphorylated form, concentrating phosphopeptides in the earliest fractions. It was only when IMAC or TiO_2 was combined with SCX, that effective phosphopeptide enrichment for all but the earliest eluting SCX fractions was achieved. This two-step enrichment strategy, along with the introduction of faster mass spectrometers such as the LTQ, LTQ Orbitrap, and now the LTQ Orbitrap Velos, has led to an explosion in the number of reported sites. At the time of publication, the Phosphosite Database [22] (www.phosphosite.org) contained >95,000 unique protein phosphorylation sites from 21 different organisms, up from 13,018 in early 2008.

The first successful application of the two-step phosphopeptide enrichment was published in 2005 [13], describing the identification of 729 sites from yeast. By early 2007, a number of additional reports had surfaced [23–25], the largest of which identified 6600 sites from HeLa cells and 5635 sites from mouse liver. Since these early reports, this strategy and variations of it have been successfully applied in different experimental systems [26].

Despite the apparent robustness of these methods, almost all of these reports come from only a handful of experienced mass spectrometry labs. In the current report we have ported the HPLC-based SCX separation to a solid-phase extraction system using vacuum flow and step elution, providing a platform for processing many samples simultaneously. In addition to streamlining the procedure, our method relieves the need for access to, and expertise in, running an HPLC system, allowing nearly any laboratory to prepare highly enriched phosphopeptide samples ready for mass spectrometric analysis. We describe the method and demonstrate its power by performing a quantitative phosphoproteomic analysis of a 1:1 mixture of two metabolically labeled wild-type budding yeast cultures. We identified more than 8000 unique phosphorylation sites using only seven 95-min runs on a LTQ Orbitrap Velos mass spectrometer. We examine two additional factors in phosphoproteomic analysis, instrument platform and enzyme choice. We show that analysis of the same samples on the Velos yielded 27% more unique sites in 25% less time than analysis on a LTQ Orbitrap Discovery and demonstrate that lysC digestion generated 21% more unique phosphorylation sites than a parallel analysis using trypsin.

2. Rationale

As the phosphoproteomic platform in our lab matured, increases in the number of samples led to worsening bottle-necks at both the SCX peptide separation and at the mass spectrometer. To increase sample preparation throughput, we began experimenting with solid phase extraction separation, allowing us to process many samples in parallel. We reasoned that because we were collecting relatively broad 4-min fractions from the HPLC, that sufficient separation could be achieved with discrete salt elution steps from a SPE column of the same material. Instrumentation costs preclude any similar increase in the number of samples that can be simultaneously analyzed in the mass spectrometer. However, the introduction of a new instrument, the LTQ Orbitrap Velos, with improvements in both sensitivity and speed [27], allowed us to reduce both the number of samples analyzed and the length of each analysis without sacrifices in the size or quality of the dataset.

3. Materials and methods

The work-flow of a SILAC-based quantitative phosphoproteomic experiment is depicted in Fig. 1. We have previously described a detailed protocol for SCX-IMAC phosphopeptide enrichment and analysis [28]. Though the overall work-flow is unchanged, we describe a number of modifications to that method; namely – the replacement of (1) trypsin with lysC, (2) HPLC based SCX fractionation with a hand-packed SPE column and step elution, and (3) the LTQ Orbitrap with the LTQ Orbitrap Velos – and demonstrate the advantages of these changes. We address each step, and considerations for each, in turn, providing step-by-step protocols at select points, and detail how the described dataset was collected and analyzed.

3.1. Starting material

Most protein phosphorylation is sub-stoichiometric [29], i.e. for every unique amino acid sequence there are on average far more unphosphorylated molecules than phosphorylated. Successful phosphopeptide detection thus requires the enrichment of phosphopeptides and removal of the much more abundant unphosphorylated peptides. Because phosphopeptides make up such a small fraction of the total peptide pool, it is also necessary to start with a large amount of material in order to obtain detectable amounts of phosphopeptides from less abundant proteins. We routinely work with ~10 mg of total protein, using columns packed with 1 g SCX resin. Larger amounts of peptide can be processed by parallel separation on multiple columns. Where sample material is limited, valuable, though smaller datasets may still be achieved. We have identified nearly 5000 phosphorylation sites from 3 mg of total yeast protein. While we anticipate that even smaller amounts of starting material can still provide thousands of phosphorylation sites, this protocol was designed for large-scale analysis, and may not be suitable for smaller amounts of starting material.

3.2. SILAC labeling and cell growth

The included quantitative data were collected using stable isotope labeling of amino acids in cell culture [30] (SILAC). The method is, however, compatible with almost any quantitative approach as well as non-quantitative analysis.

SILAC is commonly performed by dual labeling with isotopically labeled lysine and arginine, followed by tryptic digestion. We favor a SILAC protocol that omits labeled arginine, using only isotopically labeled lysine, because it reduces the cost and complexity of experiments. Labeled arginine can be metabolized to proline, visible as satellite peaks in the mass spectra of proline containing peptides, impairing the accuracy of quantitation [31]. A lysine-only protocol also reduces the need for genetic manipulation required