JPP JOURNAL OF PROTEINS AND PROTEOMICS
7(4), 2016, pp. 305-309

CC BY SA

Research Communication

PHOSPHOPROTEOMIC PROFILING USING BIOPSY-SCALE PROTEIN AMOUNTS

Keshava K. Datta[1,2], Bharti Dhaka[1], Krishna Patel[1,3], Vinuth N. Puttamallesh[1,3], Harsha Gowda[1,4*]

[1]*Institute of Bioinformatics, International Tech Park, Bangalore-560066, India.*
[2]*School of Biotechnology, KIIT University, Bhubaneswar, Odisha-751024, India*
[3]*Amrita School of Biotechnology, Amrita University, Kollam-690525, India*
[4]*Center for Systems Biology and Molecular Medicine, Yenepoya University, Mangalore-575018, India*

Abstract: Protein phosphorylation is an important post-translational modification that regulates various signaling pathways in cells. Aberrant phospho-signaling due to gain-of-function or loss-of-function of kinases is associated with various diseases including cancers. Phosphorylation serves as a surrogate to monitor kinase activity.Advent of mass spectrometry based phosphoproteomics has enabled identification and quantitation of large number of phosphopeptides in a high-throughput manner. Due to low stoichiometry, phosphopeptides are often enriched from proteolytic digests of large amounts of protein lysate prior to mass spectrometry analysis. Conventionally, phosphoproteomics has been carried out using 2-5 mg protein as starting material. This is often not feasible in biomedical research studies where only biopsies are available. In this study, we evaluated the feasibility of carrying out phosphoproteomics studies with biopsy-scale protein amounts. We demonstrate that identification of 500-2,500 phosphopeptides is feasible with protein amounts as little as 20-100 µg.

Key words: Phosphoproteomics; clinical samples; mass spectrometry

Note : Coloured Figures and Supplementary Information available on Journal Website in "Archives" Section

Introduction

Reversible protein phosphorylation plays an important role in regulating various signaling pathways in cells. Proteins can be reversibly phosphorylated by enzymes known as kinases. Human genome encodes ~500 kinases (Fleuren *et al.*, 2016). Phosphorylation regulates numerous biological activities within the cell. Further, extracellular signals are usually transduced into the cell through a cascade of phosphorylation events on specific protein substrates. Aberrant regulation of kinase signaling pathways is associated with various diseases including cancers. Aberrantly activated kinases are used as therapeutic targets in various cancers. For example, a small molecule inhibitor called imatinib is used to target mutant EGFR in non-small cell lung cancer (Wu *et al.*, 2016).

Phosphoproteomics has emerged as a valuable approach to characterize kinase mediated signaling activity in biological specimens (Olsen *et al.*, 2006). It allows identification and quantitation of protein phosphorylation in a site specific manner. Owing to low stoichiometry, phosphoproteomic studies have often followed a standard framework where milligram amounts of protein lysate is used as a starting material from which phosphopeptides are enriched using either antibody or metal affinity chromatography based methods. Anti-

Corresponding Author: **Harsha Gowda**
E-mail: harsha@ibioinformatics.org

Received: September 22, 2016
Accepted: November 10, 2016
Published: December 14, 2016

phosphotyrosine monoclonal antibodies including 4G10, p-Tyr-100 and pY10 marketed by various vendors have been successfully used for enriching tyrosine phosphorylated peptides prior to mass spectrometry. Similarly, immobilized metal affinity chromatography (IMAC) has been extensively used for enrichment of serine/threonine phosphopeptides (Harsha and Pandey, 2010). Milligram amounts of protein is often not a constraint in cell culture based experiments. However, sample amounts become limiting in biomedical research investigations using clinical specimens. The feasibility of phosphoproteomic studies using protein amounts that could be obtained from biopsy tissue specimen is not well understood. Typical biopsy material available for research purposes would be in the range of 1 – 5 mg. This would yield protein amounts in the range of 25 – 250 micrograms. We evaluated the feasibility of carrying out phosphoproteomic studies using such low amounts of protein.

Materials and methods

Cell culture

Human esophageal cancer cell line TE-5 was grown in DMEM containing 10% fetal bovine serum (Clontech, Mountain View, CA) and 1% penicillin/streptomycin mixture at 37°C in a humidified 5% CO_2 atmosphere. The cells were grown to 70% confluency and post twelve hours of serum starvation, were harvested and lysed in SDS lysis buffer (2% SDS in 100mM TEABC).

Trypsin digestion and titanium dioxide-based phosphopeptide enrichment

Protein concentration was determined using bicinchoninic acid (BCA) assay. Aliquots of 20 µg and 100 µg samples were reduced with dithiothreitol at 60°C and alkylated using iodoacetamide at room temperature. Protein precipitation was done using 6 volumes of ice cold acetone and digested overnight at 37°C using modified trypsin (Promega). Peptides from each aliquot were enriched for phosphopeptides using TiO_2 based enrichment method. Briefly, TiO_2 beads were heated for 10 minutes at 95°C and washed in 5% 2, 5-dihydroxybenzoic acid (DHB) for 20 minutes on the rotator at room temperature. The peptides were dissolved in 5% DHB and incubated with TiO_2 beads for 30 minutes on a rotator at room temperature. The TiO_2 beads were washed 3 times and eluted twice with 2% ammonia. The enriched peptides were concentrated using vacuum concentrator and desalted using C_{18}StageTips. The desalted peptide samples were subjected to mass spectrometry analysis.

LC-MS/MS analysis

The samples were analyzed on a Thermo Scientific™ Orbitrap Fusion™ Tribrid™ Mass Spectrometer (Thermo Scientific, Bremen, Germany) interfaced with Easy-nLC II nanoflow liquid chromatography system (Thermo Scientific, Odense, Denmark). Peptide samples were reconstituted in 0.1% formic acid and loaded onto a trap column (75 µm x 2 cm) packed in-house with Magic C_{18} AQ (Michrom Bioresources Inc., Auburn, CA, USA). Peptides were resolved on an analytical column (75 µm x 25 cm) at a flow rate of 280 nL/min using a linear gradient of 3-30% solvent B (0.1% formic acid in 95% acetonitrile) over 105 minutes and a total run time of 120 minutes. MS data acquisition was carried out from 400-1600 m/z range using Orbitrap mass analyzer at a resolution of 120,000 at 400 m/z. The AGC target was set to 200,000 with ion injection time of 55 ms and dynamic exclusion was set to 30 seconds. Precursor ions acquired with top speed mode were fragmented using HCD fragmentation with 34% normalized collision energy. Fragment ions were detected from 110-2000 m/z range at a resolution of 30,000 at 400 m/z. The AGC target for MS/MS scan was set to 100,000 with ion injection time of 200 ms. Internal calibration was carried out using lock mass option (m/z 445.1200025) from ambient air.

Data Analysis

Mass spectrometry data was searched using Mascot and SEQUEST search algorithms against Human RefSeq75 protein database supplemented with frequently observed contaminants through the Proteome Discoverer platform (v1.4, Thermo Scientific, Bremen, Germany). For both algorithms, search parameters included a maximum of 2 missed cleavages; carbamido methylation at cysteine as a fixed modification,

oxidation at methionine, phosphorylation at serine, threonine and tyrosine as variable modifications. The MS error tolerance was set at 20 ppm and MS/MS error tolerance at 0.1 Da. The data was searched against a decoy database and the results from both searches were used to estimate q values using the Percolator algorithm within the Proteome Discoverer suite. Peptides were considered identified at a q value of <0.01. The probability of phosphorylation for each S/T/Y site on each peptide was calculated by the PhosphoRS node (Taus *et al.*, 2011) (Version 3.0) in the Proteome Discoverer. Motif analysis was performed using motif-x tool.(Schwartz and Gygi, 2005).

Results and Discussion

By carrying out single pot enrichment without any pre-fractionation, we identified a total of 2,442 phosphopeptides from 1,085 proteins in this study. Phosphoproteomic analysis of peptides enriched from 20 µg protein lysate resulted in identification of 483 phosphopeptides and 100 µg lysate resulted in identification of 2,374 phosphopeptides. This included identification of 32 protein kinases. As anticipated, serine phosphorylation had the highest representation in our dataset followed by threonine and tyrosine (Table 1). Using phosphoRS probability score, we could confidently assign 550 phosphosites for 483 phosphopeptides identified from 20 µg protein lysate and 2,523 phosphosites for 2,374 phosphopeptides identified from 100 µg protein lysate. All the phosphoPSMs identified from 20 µg and 100 µg of protein lysate can be found in Supplementary Table S1 and S2, respectively.

More than 70% of the phosphosites identified from 20 µg protein lysate were also identified from 100 µg protein sample (Figure 1A). Evaluation of phosphopeptides identified in 100 µg samples alone revealed that many of these peptides were below the detection limit of mass spec in 20 µg samples. We determined some of the most enriched phosphomotifs in our dataset using motif-x software tool (Figure 1B). This tool looks for overrepresented sequence residues flanking the phosphosite. Kinases use some of these motifs as recognition sequences on their substrates to phosphorylate specific sites. Most of them were proline directed motifs where serine is followed by proline in +1 position. These motifs are known to be targets of cyclin dependent kinases, mitogen-activated protein kinases and glycogen synthase kinases (Luo *et al.*, 2005). We have identified several kinases belonging to these groups in our dataset suggesting some of the phosphoproteins we have identified are likely to be substrates of these kinases. Establishing kinase-substrate relationships is an active area of research. Various computational tools have been developed to build networks to predict upstream kinases and associated pathways based on phosphoproteomics datasets. These strategies are valuable in clinical samples as they provide insights into aberrantly activated pathways that could be potentially targeted. Kinase inhibitors are being used to treat various cancers. They are also some of the most actively pursued class of molecules in oncology. The described strategy could be potentially useful to systematically investigate activated kinase signaling pathways across cancers even when sample amounts are limiting.

In the past, various groups have shown that large amounts of starting material paired with extensive fractionation is required for carrying out phosphoproteomic analysis. For example, using 4 mg of protein as starting material, Zarei *et. al.* reported 3607, 1899, 1334 phosphosites with SCX, ERLIC and HILIC respectively (Zarei *et al.*, 2011). For each chromatographic technique, they pre-fractionated samples into 19 fractions that were subsequently analyzed on mass spectrometer. Using 500 µg of starting material and various chromatographic techniques, Nie *et. al.* (Nie *et al.*, 2010) identified 920 phosphosites using a runtime of 4 hours and 2,706 phosphosites using a runtime of 76 hours. In this study, we used 2 different sample amounts that are feasible from biopsy-scale tissue specimen as starting material and carried out phosphoproteomic analysis without any pre-fractionation. We demonstrate that it is possible to identify 500-2500 phosphopeptides using such low sample amounts.

Phosphoproteomics studies are often carried out using large amounts of protein as starting material. In most biomedical research studies,

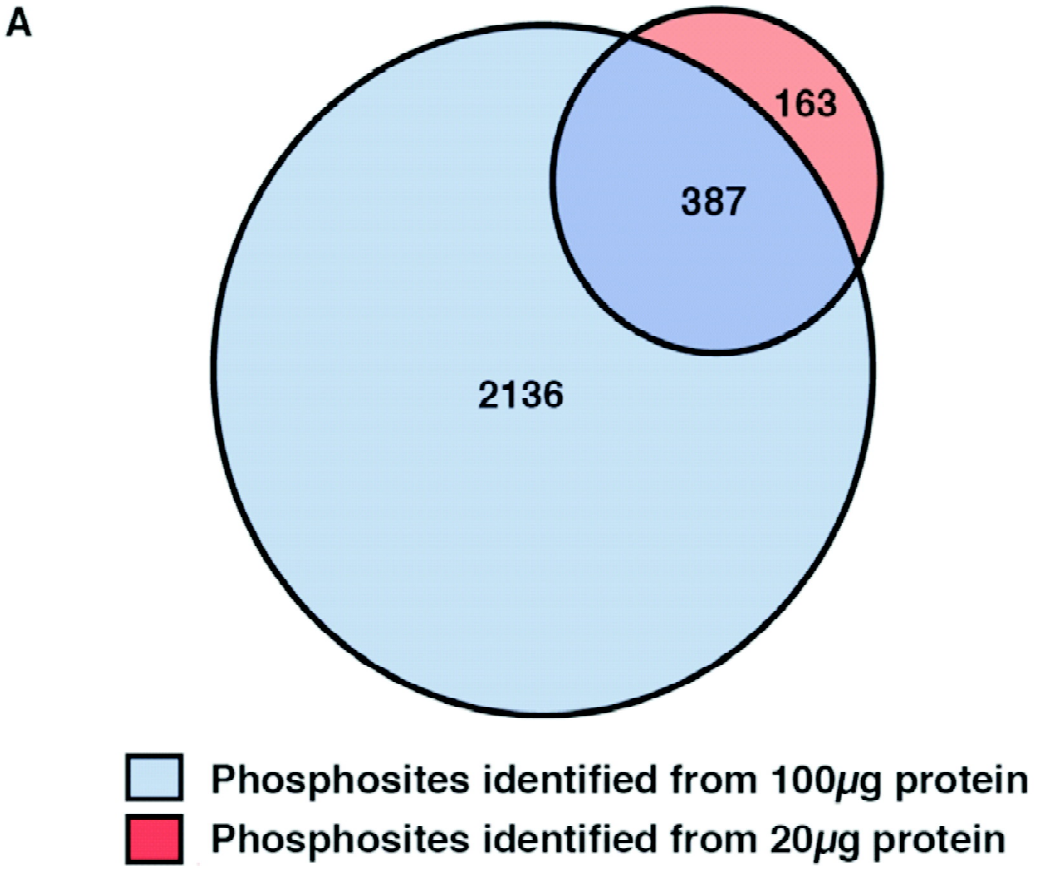

Figure 1 (A): **Venn diagram showing number of phosphosites identified from 20 μg and 100 μg protein lysate. (B) Top four most overrepresented phosphorylation motifs in our dataset**

Table 1
Summary of phosphosites identified in the study

Starting material (μg)	*No. of unique phosphopeptides*	*No. of unique phospho-sites*	*No. of phospho-serine*	*No. of phospho-threonine*	*No. of phospho-Tyrosine*
20	451 corresponding to 268 proteins	550	483	65	2
100	2265 corresponding to 944 proteins	2523	2168	328	27

biopsy specimen is readily available as opposed to surgically resected tissue specimen. In this study, we have shown that phosphoproteomics analysis is possible with low protein amounts available from biopsy specimens.

Acknowledgements

We thank the Department of Biotechnology (DBT), Government of India for research support to the Institute of Bioinformatics (IOB), Bangalore. Keshava K. Datta is a recipient of Senior Research Fellowship from University Grants Commission (UGC), Government of India.

Conflict of Interest

The authors declare no conflict of interest

References

Fleuren, E. D., L. Zhang, J. Wu and R. J. Daly (2016). The kinome 'at large' in cancer. Nat Rev Cancer 16, 83-98.

Harsha, H. C. and A. Pandey (2010). Phosphoproteomics in cancer. Mol Oncol 4, 482-495.

Luo, M., S. Reyna, L. Wang, Z. Yi, C. Carroll, L. Q. Dong, P. Langlais, S. T. Weintraub and L. J. Mandarino (2005). Identification of insulin receptor substrate 1 serine/ threonine phosphorylation sites using mass spectrometry analysis: regulatory role of serine 1223. Endocrinology 146, 4410-4416.

Nie, S., J. Dai, Z. B. Ning, X. J. Cao, Q. H. Sheng and R. Zeng (2010). Comprehensive profiling of phosphopeptides based on anion exchange followed by flow-through enrichment with titanium dioxide (AFET). J Proteome Res 9, 4585-4594.

Olsen, J. V., B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen and M. Mann (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127, 635-648.

Schwartz, D. and S. P. Gygi (2005). An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. Nat Biotechnol 23, 1391-1398.

Taus, T., T. Kocher, P. Pichler, C. Paschke, A. Schmidt, C. Henrich and K. Mechtler (2011). Universal and confident phosphorylation site localization using phosphoRS. J Proteome Res 10, 5354-5362.

Wu, P., T. E. Nielsen and M. H. Clausen (2016). Small-molecule kinase inhibitors: an analysis of FDA-approved drugs. Drug Discov Today 21, 5-10.

Zarei, M., A. Sprenger, F. Metzger, C. Gretzmeier and J. Dengjel (2011). Comparison of ERLIC-TiO2, HILIC-TiO2, and SCX-TiO2 for global phosphoproteomics approaches. J Proteome Res 10, 3474-3483.