Comprehensive Quantitative Phosphoproteomic Approach using SWATH® Acquisition

Ryo Yokoyama[1], Takeshi Shibata[1], Masato Aoshima[1], Masaki Matsumoto[2], Keiichi I. Nakayama[2]
[1]SCIEX; Tokyo Japan, [2]Department of Molecular and Cellular Biology, Medical Institute of Bioregulation, Kyushu University; Fukuoka Japan

Protein phosphorylation plays a crucial role in many biological processes such as inter- and intracellular signaling, cell cycle, protein synthesis, degradation and apoptosis. To understand these biological processes, it is important to develop analytical strategies to identify and quantify a broad range of phosphorylated peptides and proteins. In the study of post-translational modifications, it is challenging to differentially quantify the different isoforms of modified peptides as they often possess the same modification at different locations, creating peptides of the same mass that can co-elute. Previous label free strategies have relied on MS1 based quantitation which cannot distinguish the modified forms unless they are completely chromatographically resolved.

Recently SWATH® Acquisition was reported as an alternative strategy to increase the comprehensiveness of data collection and processing[1]. SWATH acquisition is a data-independent acquisition (DIA) strategy that generates, in a single measurement, a complete recording of the parent ion and fragment ion spectra of all the detectable analytes in a biological sample. Here quantitation is performed using the MS/MS data which allows for differential quantitation of individual peptide isoforms (Figure 1).

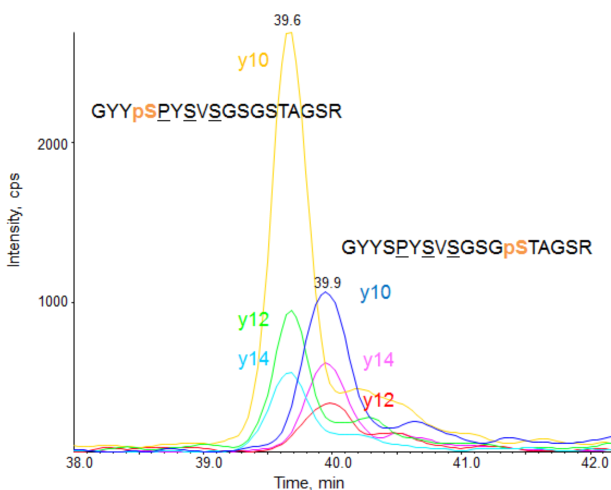


Figure 1. Fragment Ions Key for Elucidating Multiple Phosphopeptide Isoforms. Phosphopeptide isoforms have the same parent ion mass so strategies reliant on MS1 quantitation cannot distinguish the different forms. As quantitation is performed on the MS/MS data with SWATH Acquisition, site specific ions can be used to selectively quantitate the different phosphopeptide forms, even when they are co-eluting.

In this study, we used SWATH acquisition combined with Immobilized Metal Affinity Chromatography (IMAC) for quantitative phosphoproteomics. HeLa cells were stimulated with EGF (Epidermal Growth Factor) and the phosphorylation changes of many signaling proteins relating MAPK/ERK pathway were quantified.

Key Features of SWATH® Acquisition for Phosphoproteomics

- SWATH Acquisition on the TripleTOF® 5600+ System
  - Comprehensive high quality quantification with confirmation of identity on every detectable protein and peptide
  - Full scan MS/MS of all detectable ions is acquired, enabling post-acquisition interrogation of data
- Extraction of site specific fragment ions to distinguish multiple PTMs sites on the same peptide
  - Quantitation is performed on the fragment ion data, so even close eluting phosphopeptide isoforms can be differentially quantified (Figure 1)
  - No risk of missing sites due to dynamic exclusion in data dependent strategies

Methods

Sample Preparation: HeLa cells were stimulated with EGF or EGF + U0126 (MEK inhibitor). After stimulations, whole cell lysates were obtained at six time points for quantitative analysis by SWATH acquisition (Figure 2). Portions of these stimulated samples were also mixed for qualitative analysis by standard data dependent analysis. All samples were reduced, alkylated, digested with trypsin and phosphorylated peptides were enriched using IMAC. Phosphorylated peptides for qualitative analysis were separated to 12 fractions by using reverse phase chromatography at high pH.

Chromatography: The samples were analyzed using the nanoLC™ 425 system combined with the cHiPLC® system (SCIEX) in trap-elute mode. The samples were first loaded on the cHiPLC trap (200 µm x 500 µm C18-CL ChromXP™ column, 3 µm, 300 Å) and washed for 10 mins at 2 µL/min. Then, an elution gradient of 2-4% acetonitrile (0.1% formic acid) in 4mins, 4-25% in 71mins, 25-32% in 15mins was used on a cHiPLC column (75 µm x 15 cm C18-CL ChromXP column, 3 µm, 300 Å). Trap and column were maintained at 40ºC for retention time stability.

Mass Spectrometry: Eluent from the column was sprayed using the NanoSpray® Source into a TripleTOF® 5600+ system. Data dependent acquisition was performed for protein identification where the MS was acquired from 400-1000 m/z (250 msec accumulation time). MS/MS was acquired from 100-1600 m/z (100 msec) on top 20 precursors passing the selection criteria. SWATH acquisition was performed for quantitative analysis were Q1 was stepped from 400-1000 m/z (12 Da wide steps) and high resolution MS/MS was acquired from 100-1600 m/z (40msec accumulation time) giving a total cycle time of 2.1 sec.

Data Processing: The protein identification data was searched with ProteinPilot™ Software 4.5 beta to create a spectral library of the proteins and peptides in the samples. All data were processed using the MS/MS^{ALL} with SWATH Acquisition MicroApp in PeakView® Software using the spectral library created. Quantitative data of the proteins and peptides were exported to MarkerView™ Software for statistical analysis using Principal Components Analysis - Discriminant Analysis (PCA-DA).

Sample Preparation Strategy

The preparation of the phoshopeptide samples for LC/MS analysis is shown in Figure 2. Proteins were extracted for each time point, for both the +EGF and +EGF with inhibitor experiments. After protein extraction, a portion of all the individual samples were mixed for protein ID experiments. For quantitative analysis, the remainder of the sample was subjected to small scale IMAC (~400µg each sample) to enrich the phosphorylated peptides. For qualitative analysis, the mixed sample was subjected to large scale IMAC (~1mg sample). After large scale IMAC, the sample was fractionated by using reverse phage chromatography on high pH condition before analysis to obtain great depth of coverage.

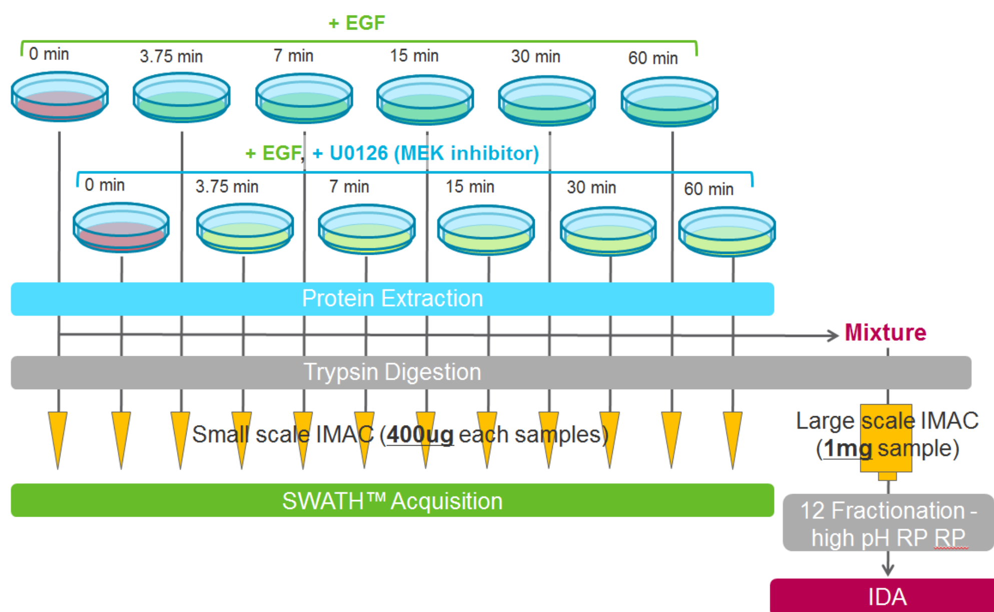


Figure 2. Workflow for Comprehensive Quantitative Phosphoproteomic Approach. Time courses for the changes in protein expression in EGF or EGF + U0126 treated HeLa cells were generated. After protein extraction, these samples were mixed for protein ID experiments. For quantitative analysis, small scale IMAC was conducted using 400 µg each sample to enrich phosphorylated peptides. These samples were analyzed by SWATH Acquisition. For qualitative analysis, large scale IMAC was conducted using 1mg sample. After large scale IMAC, the sample was fractionated by using reverse phage chromatography under high pH conditions, then analyzed by low pH analytical LCMS.

Identification of Phosphorylation Sites

The enriched and fractionated sample was analyzed by nanoLC/MS/MS to identify as many proteins and peptides as possible. As shown in Table 1, over 3700 proteins were identified from over 20000 peptides using ProteinPilot™ software. Good enrichment of phosphorylated peptides was obtained as over 85% of detected peptides were phosphorylated. Further characterization of the phosphopeptide indicated the majority of the sites were phosphoserine and the great majority were singly-phosphorylated.

Table 1. Identification and Phosphorylated Proteins and Peptides.

	Number
# of Proteins (1% global FDR)	3777
# of Peptides (1% Global FDR)	20530
# of Phosphopeptides (1% Global FDR)	17678
# of Phosphorylation Sites (% Ser : Thr : Tyr)	18368 (85.4 : 14.1 : 9.1)
Distribution of Phosphorylation Sites (Single : Double : Triple)	(16929 : 740 : 9)

SWATH® Acquisition Data Processing

Approximately 3600 proteins from the protein identification result were imported into the SWATH Acquisition MicroApp in PeakView Software (Figure 3, top left) and all confidently identified phosphopeptides per protein (>95% confidence) were selected for processing. Top right panel below shows an overlay of the 12 SWATH acquisition experiments, the total ion chromatograms (TICs) overlay extremely well showing high reproducibility across the sample set. In this experiment, smaller Q1 isolation windows (12 Da) were used for increased quantitative specificity with no loss in quantitative quality.

An example of a phosphorylated peptide from the Serine/Arginine repetitive matrix protein is shown in the bottom left panel. Four extracted ion chromatograms (XICs) were generated for the sequence ions across the peptide sequence VSGRT**pS**PPLLDR. A peak group with good confidence was detected at 31.78 mins and relative quantitation information on that peptide across the samples can be determined from the XIC peak areas.

Finally, the ion library spectrum (pink) can be compared to the MS/MS spectra obtained at 31.7min in the Q1 m/z window for the 3+ parent m/z of the peptide.

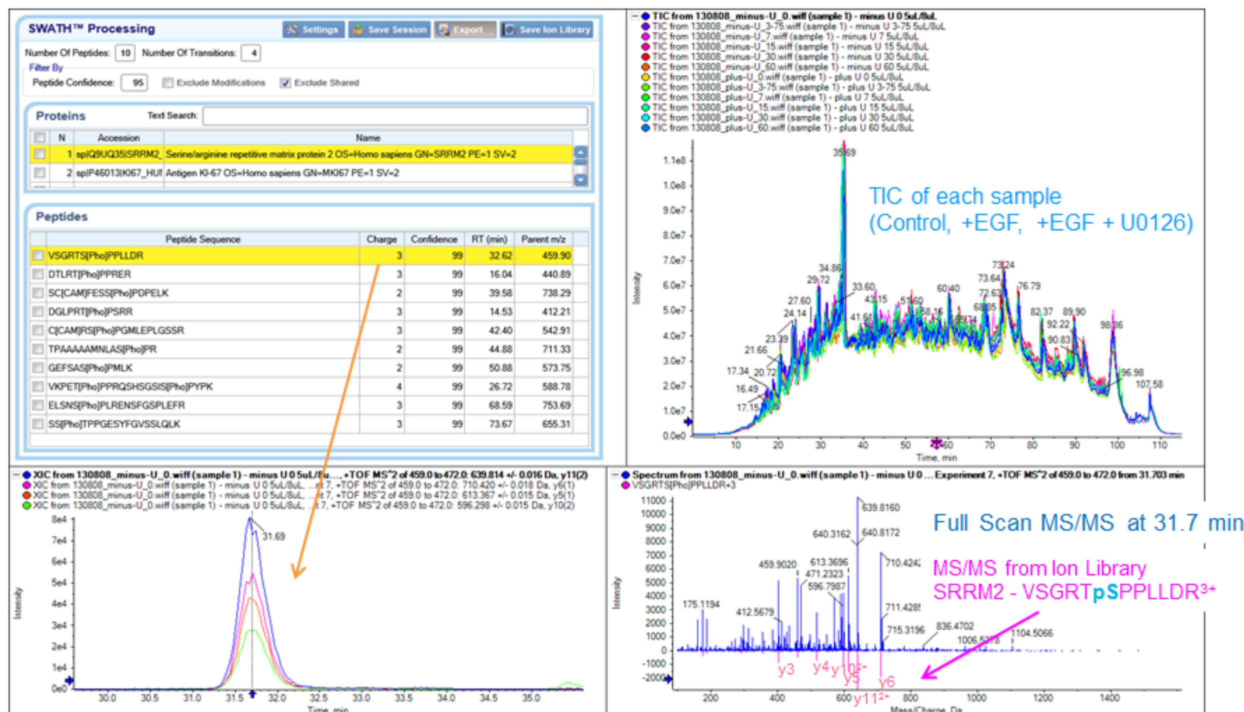


Figure 3. Processing of SWATH Acquisition Data. The 12 SWATH acquisition experiments were processed using an ion library of ~3600 proteins. Extracted ion chromatograms for sequence specific ions were generated for the phosphopeptides identified.

Changes in Phosphorylation During the Time Course of EGF Stimulation

Many proteins have multiple phosphorylation sites that change during signaling and often these sites are on the same tryptic peptide. SWATH acquisition is a powerful strategy for the quantitation of phosphopeptide isoforms due to the use of MS/MS data. This advantage is highlighted in Figure 1, where the two forms are closely eluting but can be easily quantified by using sequence ions that differentiate the two forms. Figure 4 shows another case of multiple phosphorylation sites on a peptide from the EGFR. Both sites were identified and extracted across the time course samples and each show a different behavior after EGF stimulation.

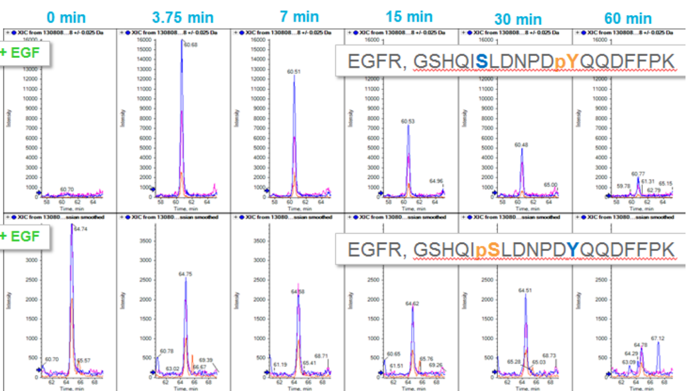


Figure 4. Quantification of Differentially Phosphorylated Peptides. A peptide from EGFR was found to have two different sites of phosphorylation, GSHQISLDNPDpYQQDFFPK and GSHQIpSLDNPDYQQDFFPK as shown in the database search results (top). XICs were generated from the SWATH™ Acquisition data to both peptides using site specific fragment ions (bottom). The time course of phosphorylation was found to be different for each phosphosite.

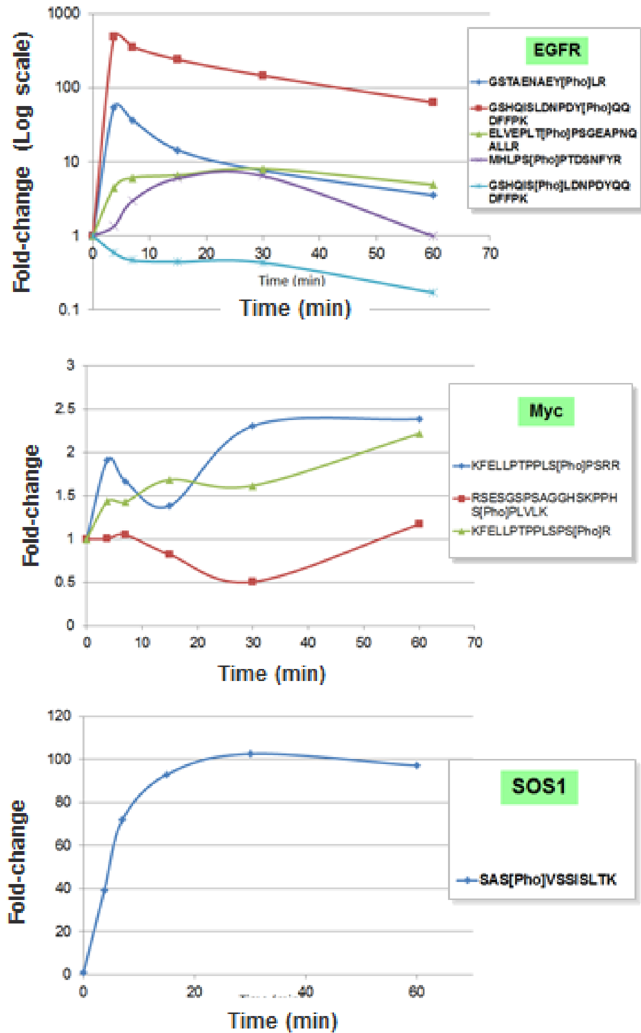


Figure 5. Quantitative Changes in Phosphorylation of Representative Signaling Molecules Related to EGF Stimulation. Multiple peptides from EGFR show a similar behavior of rapid phosphorylation followed by slow decay (top). Alternatively SOS1 protein shows a rapid phosphorylation but that is sustained for an extended time (bottom). Only small changes were observed for phosphopeptide from the Myc protein (middle).

Specific phosphopeptide sites can be tracked across the time course to understand perturbations in the signaling network (Figure 5). A number of phosphosites on EGFR show a rapid increase in phosphorylation upon EGF stimulation followed by a slow decay (top pane). Much smaller changes were observed for the phosphopeptides from Myc (middle pane). In contrast the serine phosphosite on SOS1 showed a slower and steadier increase in phosphorylation after EGF stimulation which persists for longer.

Statistical analysis was performed using MarkerView Software to find phosphopeptide features that distinguished the stimulated samples from the time zero controls as well as the stimulated vs inhibited samples (Figure 6).

Conclusions

Comprehensive phosphopeptide profiling using SWATH® Acquisition is powerful tool for large-scale non-targeted phosphorylation studies such as network biology. Because of the comprehensive nature of this quantitative strategy, large portions of signaling networks can be monitored in a single experiment, such as the example shown here for the EGF signal transduction network (Figure 7).

- Many phosphorylated proteins and peptides were identified using IMAC combined with data dependent acquisition.
- Phosphorylation of many signaling proteins relating EGF stimulation and MEK inhibitor were quantified using SWATH acquisition.
- Because SWATH acquisition is a full scan MS/MS based technique, specific fragment ions can be used to quantify site-specific phosphorylation changes on a peptide.
  - Smaller Q1 isolation windows are easily used for increased specificity
- PCA-DA provided visualization of quantitative trends of phosphorylation.

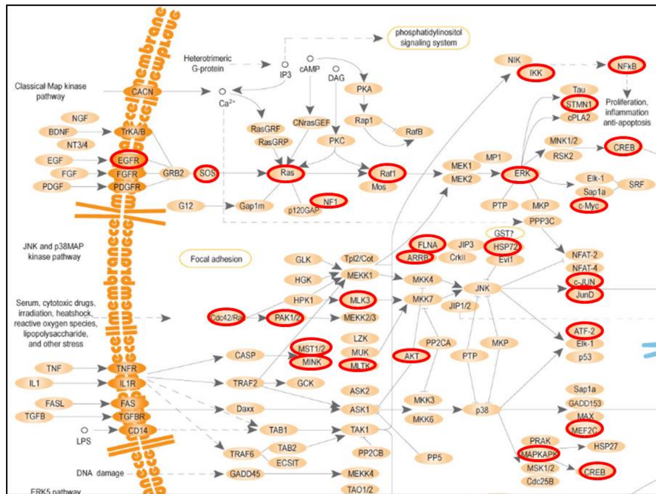
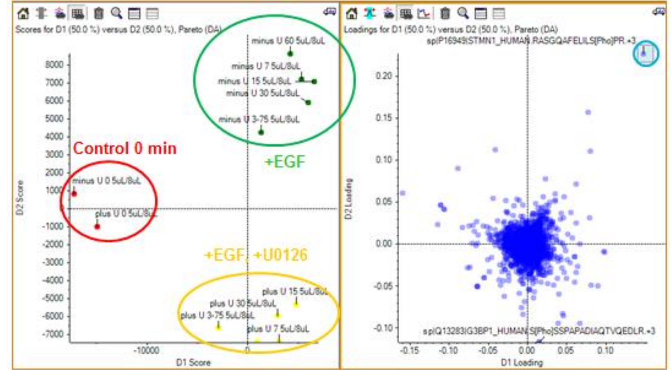


Figure 7. EGF Signal Transduction Pathway. Red circles indicate the proteins that had phosphorylation sites and were quantified in the SWATH acquisition experiment.



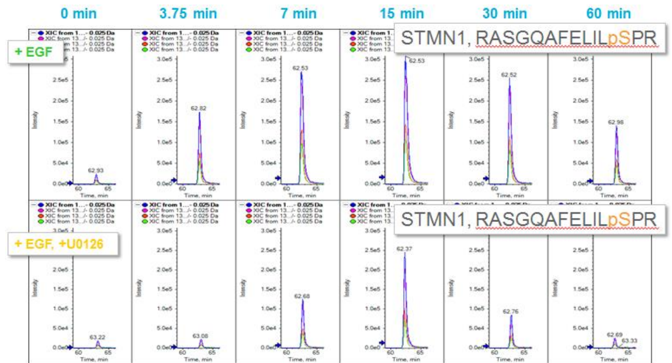


Figure 6. PCA-DA using MarkerView™ Software. (Top) Three groups were defined for the PCA-DA analysis, as shown in the scores plot. The variables that characterize the differences between the groups can be found on the Loadings plot (right). The feature selected from the loadings lot shows a serine phosphopeptide from STMN1 protein that shows similar behavior in both the EGF stimulation and EGF + inhibitor time courses.

References

1. Gillet LC et al (2012) *Mol. Cell. Prot.* **11**, 1-17.

© 2015 AB Sciex. For Research Use Only. Not for use in diagnostic procedures.

The trademarks mentioned herein are the property of AB Sciex Pte. Ltd. or their respective owners. AB SCIEX™ is being used under license.

Publication number: RUO-MKT-02-4296-A