Improving LC-MS throughput for phosphoproteomics research

T. W. Clauss, R. J. Moore, M. Gritsenko, Y. Wang, R. Zhao, S. O. Purvine, W. B. Chrisler, A. Shukla, M. E. Monroe, F. Yang, and R. D. Smith
Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA

Pacific Northwest NATIONAL LABORATORY

Overview

- **High resolution FTMS instruments coupled to custom built metal-free automated nanoLC systems are used in PNNL's high throughput facility for phosphoproteomic analysis of samples [1].**
- **Data acquisition times range from 200 to 450 min for each phosphoproteomic sample analysis.**
- **Use of two columns and overlapping LC cycles reduce total LC cycle times of 4.5 - 8.5 h to approximately the data acquisition (gradient) time [1,2].**
- **Compromise between data acquisition time and phosphopeptide identifications is explored.**

Phosphoproteomics nano-LC platform coupled to a Thermo Scientific LTQ Orbitrap Velos ETD mass spectrometer. Sample flow-path is metal free.

Introduction

The generally low level of protein phosphorylation in biological samples requires initial larger amounts of proteins (up to10 mg) for off-line fractionation prior to phosphopeptide enrichment [3]. This presents significant challenges in phosphoproteomics research when limited amounts of material are available for comprehensive phosphopeptide coverage.

Without the ability to reduce sample complexity, MS/MS undersampling significantly impacts the number of phosphopeptides identified in a single LC-MS run. We find that our standard 200 min gradient separation time is adequate for sample fractions, but longer times are required for more complex samples (e.g., enriched HMEC).

Here we explore the correlation between separation times and phosphopeptide identifications in order to better understand future directions that should be taken to improve throughput while mitigating the effects of MS/MS undersampling.

Methods

Sample analysis

- Tryptic digests of human mammary epithelial cells (HMEC) were enriched using Ni-NTA-agarose magnetic beads [4]. Three 20 µL replicate sample aliquots were tested at each condition.
- Automated metal-free nano-LC system was used for all analyses. Pump pressures were kept constant for mobile phase A (0.1 M HOAc, 100% water) and B (0.1 M HOAc in 70:30 ACN:water) at 1500 psi and 3000 psi, respectively.
- A 360 µm o.d. x 50 µm i.d. x 55 cm long column packed in-house with Jupiter 3 µm C18 particles (Phenomenex) was used for all separations. The initial flow rate under starting conditions (100% mobile phase A) was ~70 nL/min.
- Gradient times were adjusted by changing the split flow rate (pre-injection) to obtain desired elution windows.
- Data acquisition was performed using an LTQ Orbitrap Velos ETD mass spectrometer with a100k precursor scan from *m/z* 400-2000 followed by the top10 data-dependent ion trap MS/MS scans where activation (CID or ETD) was determined using the data dependent decision tree algorithm option in Xcalibur [5].

Data analysis

- CID and ETD spectra were extracted with isotope correction using DeconMSn [6] and filtered for subsequent SEQUEST analysis.
- Tryptic peptides within 50 ppm were dynamically searched for phosphorylation on S,T and Y residues.
- 0.5 Da fragment ion tolerance was chosen due to low mass precision detection. Decoy IPI human protein file (ipi.HUMAN.v3.54) was used to assess appropriate cutoff values to achieve ~1% FDR.
- A program developed in-house (PhoPAD2) was used to recalculate the delCn of phosphopeptides [7].
- Cutoff values applied to forward-only sequences to obtain final peptide lists.

Results

Elution window vs. gradient time

a) Elution window for 200 min gradient **b) Elution window for 450 min gradient**

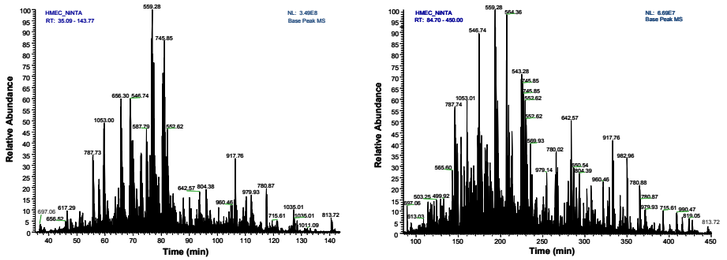

Figure 1. Base peak chromatographic profile from the elution window of a phosphopeptide enriched HMEC sample using: **a)** 200 min gradient and **b)** 450 min gradient. The earliest (*m/z* 697.06) and latest (*m/z* 813.72) LC peaks observed at both gradient times were used to establish elution windows. The 3x increase in the elution window is roughly proportional to the increase phosphopeptide identifications shown in Table 1.

Phosphopeptide count vs. gradient time

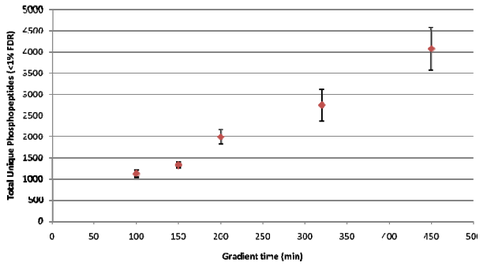

Figure 2. Average number of phosphopeptide identifications for 3 analyses at each gradient time. The number of unique phosphopeptides increased with longer gradient times. The number of identifications do not level off up to the 450 min gradient time.

The data in Figure 2 show a near-linear relationship between unique phosphopeptide identifications and gradient length out to 450 min, indicating phosphopeptide identifications are still largely limited by the mass spectrometer duty cycle.

Gradient times beyond 450 min have not yet been explored with this particular LC configuration and start becoming unmanageable in a high-throughput laboratory where biological and technical replicates are often required to identify proteins of interest.

Technical replicates, a common strategy for increasing proteome coverage, were analyzed using 150 and 450 min gradients (Table 1).

Table 1. Total unique phosphopeptides

	Gradient: 150 min	Gradient: 450 min
Replicate 1	1410	3534
Replicate 2	1299	4542
Replicate 3	1283	4164
Combined unique phosphopeptides	**1918**	**5952**

Combined results from triplicate analyses show ~45% more phosphopeptide identifications over a single measurement for the same gradient time. However, filtered peptide results for three combined measurements made at the shorter gradient time failed to produce the same number of unique phosphopeptide identifications from any single 450 minute analysis.

The ~45% increase in identifications for replicates at both gradient times further illustrates that undersampling is a limitation to coverage of a complex, enriched phosphopeptide sample.

Replicates do not effectively address undersampling

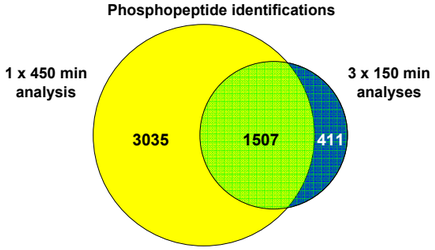

Figure 3. Venn diagram showing overlap of phosphopeptide identifications obtained from the combined results of three 150 min analyses and a single 450 min analysis of the same HMEC sample. More than twice the number of phosphopeptide identifications were obtained from a single 450 min analysis and included ~80% of the identifications from the 3 combined 150 min analyses.

Conclusions

- **Enrichment of phosphopeptides using Ni-NTA is useful for tracking phosphopeptide identifications under different LC-MS conditions.**
- **Replicate analyses do not compensate for MS/MS undersampling with complex samples. Further, a single analysis over a longer LC elution window will generally provide better coverage than technical replicates over a shorter elution window.**
- **This study demonstrates the need for alternative LC-MS approaches for dealing with complex samples where sample quantity is very limited (e.g., phosphoproteomics).**
- **Future Work: Development of on-line 2D nano-LC approaches to phosphopeptide analyses**

Acknowledgements

This research was funded by the NIH National Center for Research Resources (RR18522). Work was performed in the Environmental Molecular Science Laboratory, a U. S. Department of Energy (DOE)/BER national scientific user facility at Pacific Northwest National Laboratory (PNNL) in Richland, Washington. PNNL is operated for the DOE by Battelle under contract DE-AC05-76RLO-1830. We also would like to thank Dr. Steven Wiley for providing the HMEC samples used in this study.

References

1. R Zhao, et al. *J Chromatogr B Analyt Technol Biomed Life Sci* 877:663-670 (2009).
2. EA Livesay, et al. *Anal Chem* 80:294-302 (2008).
3. DLSwaney, et al. Proc Natl Acad Sci USA, 106:995-1000 (2009).
4. SB Fiacre, et al. *Anal Chem* 81:4566-4575 (2009).
5. M Zeller, et al. Thermo Fisher Scientific (Bremen) GmbH; *Thermo Scientific Application Note*: 30179 (2009).
6. AM Mayampurath, et al. *Bioinformatics* 24:1021-1023 (2008).
7. X Du, et al. *J Proteome Res* 7:2195–2203 (2008).

CONTACT: Therese Clauss
Biological Sciences Division, K8-98
Pacific Northwest National Laboratory
P.O. Box 999, Richland, WA 99352
E-mail: thereserw@pnl.gov