Poster Note 64441

Development of a High-Throughput Urine Analysis for Global Protein Profiling

Scott Peterman, David Sarracino, Amol Prakash, Bryan Krastins, Gregory Byram, Gouri Vadali, Maryann Vogelsang
[1]Thermo Fisher Scientific BRIMS, Cambridge, MA; [2]Optys Technology Inc. Philadelphia, PA;
[3]Nuclea Biosciences, Cambridge, MA

Overview

Purpose: Develop a comprehensive translational proteomics workflow for urine combining sample preparation, data acquisition, and processing.

Methods: Performed pSMART data acquisition to sample UHPLC peaks for increased protein/peptide content generation per unit time. Incorporate unique data processing strategies in Pinnacle software for automated qualitative and quantitative analysis.

Results: The experimental method using pSMART generated over 55% more peptides consistently sampled per injection. Unique data processing strategies in the Pinnacle software significantly increased the protein sequence coverage compared to Sequest searching for N- and C-terminal truncation analysis, which is key for urine proteomics.

Introduction

Translational proteomics has emerged as a powerful method to rapidly classify proteins/peptides as putative markers using fewer experimental steps. To succeed, global protein profiling studies place much greater emphasis on biological replicate analysis of well-defined, large number cohort (e.g. 20x20 to 100x100) analysis instead of fewer samples and greater numbers of technical replicates. This approach results in a greater understanding of biological variance to identify protein groups associated at defining biological states. The greater sample load requires more efficient acquisition methods to complete studies in a timely manner. To address this, we have utilized larger bore UHPLC columns with smaller particles to maximize peak and loading capacities. [1] Data acquisition schemes utilize pSMART [2] as it enabled reproducible peak detection, sequencing, and quantitation.

We have applied the combined method for urine proteomics studies due to its potential for disease proteomics. Urine samples contain numerous proteins localized into specific groups, accessible using various sample preparation methods. [3] Following sample preparation, LC-MS analysis is performed on either intact or bottom- up analysis.

Methods

Sample Preparation

Two second morning urine samples (approximately 400 mL each) was collected with informal consent from a healthy volunteer. Samples were treated with 40 mL of a 10% MeOH (1% acetic acid) solution and centrifuged at 4000 xg for 30 min. to eliminate cell debris. From the stock solution, four different samples were prepared by a two-step centrifugal separation using 150kDa and 9 kDa molecular weight cutoff filter (MWCO) (Thermo Fisher Scientific). Sample volumes of 20, 40, and 80 mL were used for each replicate. The high- and low-MW samples were digested and analyzed. A portion of the low-MW sample was kept for intact analysis. Each sample was spiked with the PRTC kit (Thermo Fisher Scientific) prior to LC-MS analysis.

Liquid Chromatography (or more generically Separations)

A Vanquish UHPLC system (Thermo Fisher Scientific) was used for all experiments. A binary solvent system consisting of A) 0.2% formic acid and B) MeCN (0.2% formic acid) was used for all separations on an Acclaim 120 column with dimensions of 250 x 2.1 mm with 2.2 µm particles. The analytical gradient was 0.6% per min. at a flow rate of 150 µL per min.

Mass Spectrometry

All experiments were performed using an Orbitrap Fusion™ mass spectrometer using standard DDA and modified DIA method. [2] Both methods utilized 60,000 and 15,000 resolution for MS and tandem MS spectral acquisition and loop counts of 7. To accommodate the narrow peak widths, a precursor m/z range of 450-1200 Da was used and an overall cycle consisting of 9 scan events was used.

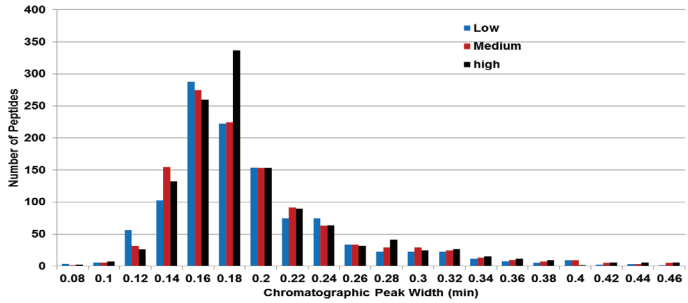
Data Analysis

All bottom-up data was processed using Proteome Discoverer (PD) 1.4.(Thermo Fisher Scientific) and used to generate spectral libraries. The Pinnacle software (Optys Technology, Inc) performed all comparative qualitative and quantitative data analysis across different levels and incorporated the spectral libraries. Pinnacle search strategies performed post-acquisition targeted data extraction using the peptide target parameters imported from the PD search results. In addition, the confident set of protein (67) sequences was used to extended data extraction specifically for PTMs and partial cleavage peptides not readily detected in the original PD search.

Results

Translational proteomics requires new workflows that maximizes content in the form of qual/quan analysis covering as much of the proteome as possible in the shortest period of time to facilitate larger cohorts. Thus we employ unique experimental strategy to address the stated requirements for translational proteomics. The first aspect is dramatically increasing chromatographic resulting through wide-bore UHPLC columns and higher flow rates. Figure 1 shows the resolution afforded using higher flow rates resulting in peak capacities ca. 230 for a 38 min. gradient. Increased chromatographic resolution results in greater selectivity and sensitivity in less time. A gradient length 3- to 5-times longer would be needed to match the peak capacity. Wide bore columns also increase loading capacity without broadening peaks. Due to using higher flow rates, the delay in matching the solvent composition delivered by the pumps in the column is minimized resulting in extremely reproducible elution times from run to run.

FIGURE 1. Distribution of chromatographic peak widths for detected peptides. The range of peak widths represent the three different loading amounts.

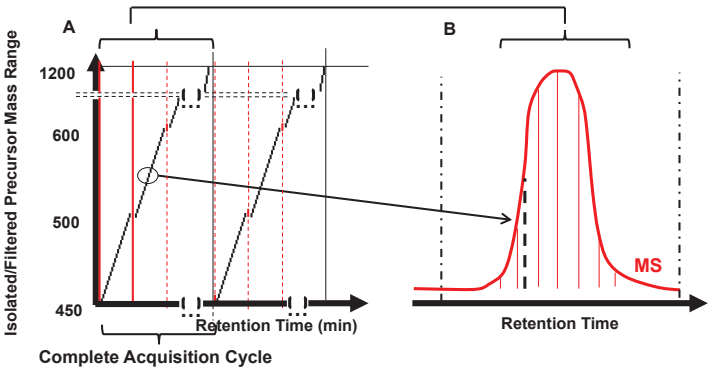


The second aspect is data acquisition strategies. UHPLC requires short cycle times to enable robust quantitation. The pSMART acquisition method provides flexibility in maintaining sensitivity and selectivity across short cycle times. (Figure 2) The method leverages the trapping schemes of the Orbitrap as well as the higher resolution/accurate mass analysis for the entire precursor m/z range enabling only ca. 350 msec for global quantitation. The independent DIA scan events acquired in between MS scan events are customized to address the average peak widths but still enable routine sampling across the targeted precursor m/z range for post-acquisition data processing. This departure from standard DIA or all ion fragmentation results in an increased coverage of the proteome while minimizing experimental time.

Thermo SCIENTIFIC

The second aspect is data acquisition strategies. UHPLC requires short cycle times to enable robust quantitation. The pSMART acquisition method provides flexibility in maintaining sensitivity and selectivity across short cycle times. (Figure 2) The method leverages the trapping schemes of the Orbitrap as well as the higher resolution/accurate mass analysis for the entire precursor m/z range enabling only ca. 350 msec for global quantitation. The independent DIA scan events acquired in between MS scan events are customized to address the average peak widths but still enable routine sampling across the targeted precursor m/z range for post-acquisition data processing. This departure from standard DIA or all ion fragmentation results in an increased coverage of the proteome while minimizing experimental time.

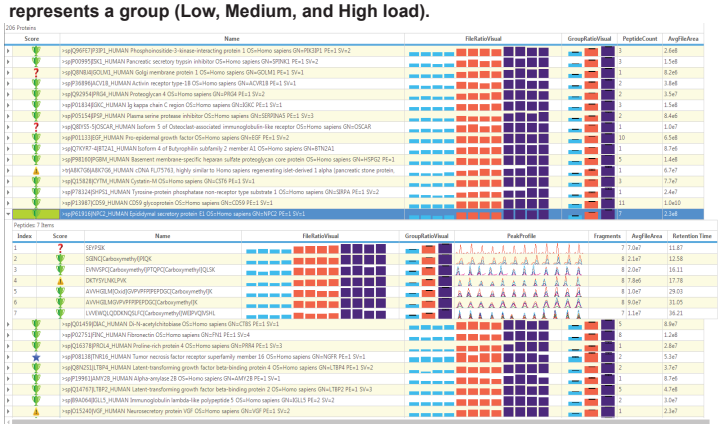
FIGURE 2. Schematic of pSMART data acquisition enabling a 9 sec. cycle time for the acquisition of at least 9 HR/AM MS events and at least 1 narrow DIA window covering a precursor m/z range of 450-1100 Da.



The last aspect necessary for robust analysis is automated data processing. The Pinnacle software was used for all qual/quan data analysis. The inclusion of the DDA results provided an introductory library from which more exhaustive and confident searching was performed. Pinnacle provides a unique data reviewing strategy moving from top-down or bottom-up data review. (Figure 4) The weighted peptide results are used to define the protein response across each group. The columns provide fast and comprehensive data review through the area under the curve (AUC) ratios for each file

As well as the grouped results. The global peptide analysis can be displayed based on protein-peptide review or flat peptide lists. Figure 3 shows the view at the peptide level showing not only the measured AUC values per file and group, but the integrated peak shape for rapid and confident evaluation. The Pinnacle software enables users to score and filter data based on the experimental design.

FIGURE 3. Comprehensive data review from the Pinnacle software showing the processed protein response as well as the expanded peptide view. Each color represents a group (Low, Medium, and High load).



The comprehensive data displayed in Figure 3 is based on peptide detection and quantitation. For maximum coverage, the Pinnacle software utilizes a unique feature of combining spectral library content with in silico predictions to increase peptide-specific feature detection. The spectral library created from the DDA experiments contains peptide specific information such as protein origin, retention time, precursor and product ion m/z values, as well as product ion distribution profiles. The DDA-derived spectral libraries are limited in peptide-specific content due to the nature of DDA methods. Figure 4 shows the benefit of the Pinnacle searching approach by expanding the precursor charge states associated with a targeted peptide. The increased sampling results in repetitive product ion spectral acquisition using two charge states as opposed to only one. As shown in Fig. 1, the experimental method was set up for a 9 sec. cycle time, but a large section of peptides had narrower peak widths. In addition to increasing the sampling rate per peptide, acquiring DIA spectra on multiple charge states increases the confidence in peptide scoring. Figure 5 shows the narrow DIA spectra for a long peptide. The combined product ion spectra generated 20 fragment ions almost covering the entire sequence.

The comprehensive data displayed in Figure 3 is based on peptide detection and quantitation. For maximum coverage, the Pinnacle software utilizes a unique feature of combining spectral library content with in silico predictions to increase peptide-specific feature detection. The spectral library created from the DDA experiments contains peptide specific information such as protein origin, retention time, precursor and product ion m/z values, as well as product ion distribution profiles. The DDA-derived spectral libraries are limited in peptide-specific content due to the nature of DDA methods. Figure 4 shows the benefit of the Pinnacle searching approach by expanding the precursor charge states associated with a targeted peptide. The increased sampling results in repetitive product ion spectral acquisition using two charge states as opposed to only one. As shown in Fig. 1, the experimental method was set up for a 9 sec. cycle time, but a large section of peptides had narrower peak widths. In addition to increasing the sampling rate per peptide, acquiring DIA spectra on multiple charge states increases the confidence in peptide scoring. Figure 5 shows the narrow DIA spectra for a long peptide. The combined product ion spectra generated 20 fragment ions almost covering the entire sequence.

FIGURE 4. Comparative peptide analysis using only spectral library information (Fig. 4A) vs. expanded peptide content (Fig. 4B). Precursor isotopic XICs are used for quantitation and the triangles represent narrow DIA scan events isolating the precursor m/z values. Open triangles represent DIA spectra lacking peptide specific product ion information, which is represented by the blue triangles.

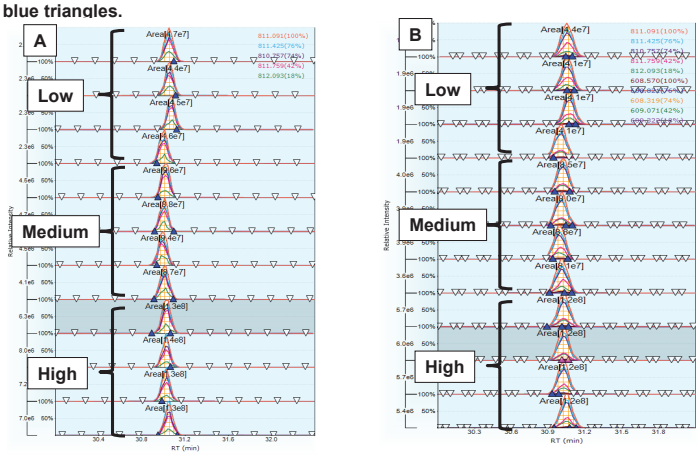
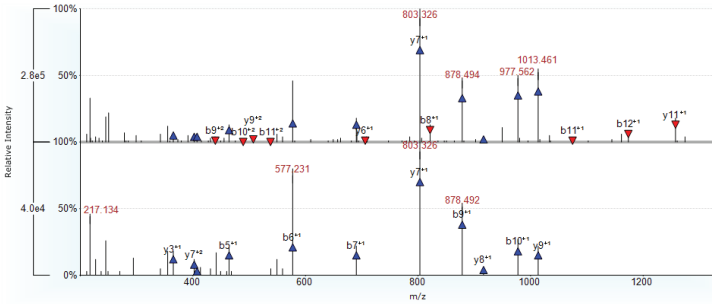


FIGURE 4. Narrow DIA spectra for the (A) +3 and (B) +4 precursor charge states for the peptide AVVHGILMGVPVPFPIPEPDGC*K. The blue triangles represent commonly measured product ions between the two spectra and the red triangles represent unique fragment ion m/z values.



The exhaustive searching strategy dramatically increased the number of peptides confidently identified, sequenced, and quantified across the different load amounts. Figure 5 shows ca. 50% more peptides identified per displayed proteins, which is representative for the overall set. Each accepted peptide must have accurate mass components (≤ 7 ppm), precursor isotopic and product ion distribution profiles meeting a set dot-product correlation coefficient (0.98 for precursors and 0.6 for products), and have a calculated AUC ratios between groups similar to the values calculated for peptides identified using only the spectral library information.

FIGURE 5. Comparative analysis of peptide coverage per targeted protein using the two described methods of data processing. The red bars represent data searching using only the spectral library information generated from DDA experiments and the blue bars represent the exhaustive search performed in the Pinnacle software. All peptides met stated requirements.

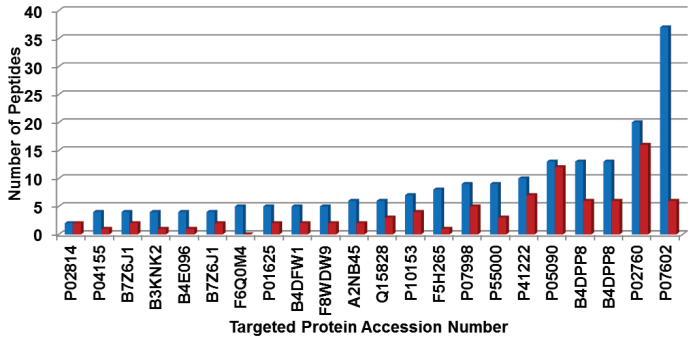
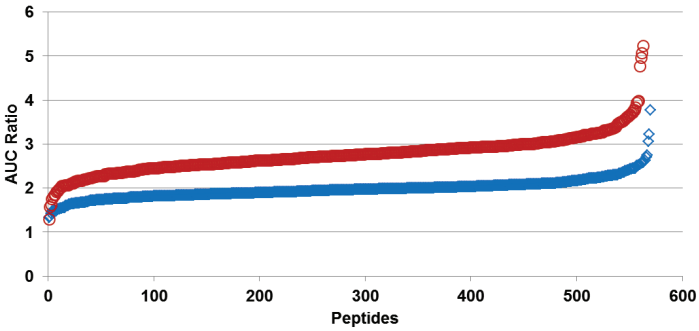
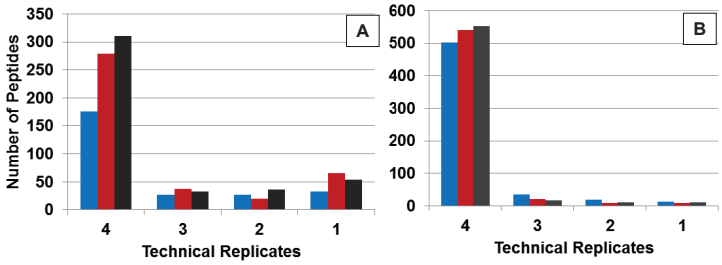


FIGURE 6. Measured AUC ratios per peptide between the Low and Medium group (blue) and Low and High group (red). The AUC ratios represent peptides that first passed qualitative scoring methods described above.



The final consideration was determining the experimental success in repetitively sampling peptides across all injections. To satisfy the throughput requirements, technical replicates must be sacrificed in favor of greater biological replicate analysis for statistical purposes. Thus, the data acquisition method must provide exceptionally high confidence in product ion acquisition capabilities, even for UHPLC peak shapes. Figure 7 shows the dramatic increase in peptides routinely sampled across all 12 injections as compared to DDA analysis. In addition, there is must less of a difference in the number of peptides detected across the volume of urine prepared per group.

FIGURE 7. Comparative analysis of repetitive peptide product ion sampling across each technical replicate using A) DDA and B) pSMART. Successful product ion spectral matches resulted in a dot-product correlation coefficient of 0.6 or greater. For the exhaustive search containing DIA spectra covering precursor charge states not originally contained in the DDA library, the detection of consensus fragment ions (7ppm) and relative distribution was used.



Conclusion

The workflow presented comprises a unique approach to performing translational proteomics. The experimental method maximizes chromatographic and mass spectral performances while significantly reducing acquisition time per sample. The results demonstrate analytical strengths in the following:

- Incorporation of UHPLC pumps and columns resulted in extremely high peak capacity per unit time resulting in greater sensitivity and selection during LC-MS analysis.
- The pSMART method facilitated robust MS and DIA acquisition for reproducible qualitative and quantitative analysis on more peptides as compared to DDA.
- The automated data processing strategies in the Pinnacle software resulted in greater data extraction with high confidence without requiring manual integration.
- The combined method increased protein coverage by greater than 50%.

References

1. Gilar, M., Daly, A. E., Kele, M., Neue, U. D., Gebler, J. C. Implications of column peak capacity on the separation of complex peptide mixtures in single- and two-dimensional high-performance liquid chromatography, J. Chrom. A **2004**, *1061,* 183-192
2. Prakash, A., Peterman, S., Ahmad, S., Sarracino, D., Frewen, B., Vogelsang, M., Byram, G., Krastins, B., Vadali, G., Lopez, M. Hybrid data acquisition and processing strategies with increased throughput and selectivity: pSMART analysis for global qualitative and quantitative analysis J. Proteome Res. **2014,** *13(12)*, 5415-5430
3. Santucci, L., Candiano, G., Petretto, A., Bruschi, M., Lavarello, C., Inglese, E., Righetti, P. G., Ghiggeri, G. M. From hundreds to thousands: widening the normal human urinome, J. of Proteomics, **2015,** *112*, 53-62

www.thermofisher.com

Africa +43 1 333 50 34 0
Australia +61 3 9757 4300
Austria +43 810 282 206
Belgium +32 53 73 42 41
Canada +1 800 530 8447
China 800 810 5118 (free call domestic)
400 650 5118
Denmark +45 70 23 62 60
Europe-Other +43 1 333 50 34 0
Finland +358 10 3292 200
France +33 1 60 92 48 00
Germany +49 6103 408 1014
India +91 22 6742 9494
Italy +39 02 950 591
Japan +81 45 453 9100
Korea +82 2 3420 8600
Latin America +1 561 688 8700
Middle East +43 1 333 50 34 0
Netherlands +31 76 579 55 55
New Zealand +64 9 980 6700
Norway +46 8 556 468 00
Russia/CIS +43 1 333 50 34 0
Singapore +65 6289 1190
Spain +34 914 845 965
Sweden +46 8 556 468 00
Switzerland +41 61 716 77 00
UK +44 1442 233555
USA +1 800 532 4752

PN64441-EN 0616S

Thermo SCIENTIFIC
A Thermo Fisher Scientific Brand