

Novel Software Approaches to Building, Processing, and Scoring Large Targeted Peptide Assays using iSRM

S. Peterman¹, R. Kiyonami¹, A. Prakash², D. Sarracino², B. Krastins², T. Rezai², M. Athanas³, Mary Lopez², and Bruno Domon⁴ ¹Thermo Fisher Scientific, San Jose, CA, ²Thermo Fisher Scientific BRIMS, Cambridge, MA, ³Vast Scientific, Wayland, MA, ⁴ETH Zürich, Zürich, Switzerland

Overview

Purpose: Demonstrating an automated method workflow for combining discovery screening, verification, and maintaining a high level quantitation for large targeted peptide lists.

Methods: Integrated workflow involving Pinpoint software, the TSQ Vantage triple quadrupole mass spectrometer, and T-SRM and iSRM strategies. Candidate selection and verification schemes are based off of spectral library entries acquired on an LTQ-based mass spectral platform.

Results: In one 40 minute gradient experiment, 757 peptides from 492 proteins were quantified and verified using a combination of primary and secondary SRM acquisition. All data processing and verification steps were performed automatically using Pinpoint software.

Introduction

The role of SRM assays for peptide and protein verification and quantitation has expanded to monitor large numbers of candidates (500-1000 peptides) during initial screening phases. This expansion has created additional challenges to the biomarker verification workflows to effectively build intelligent methods, process data, and automatically classify proteins, peptides, and individual transitions. To address these concerns, we present a novel software package to build large SRM assays, acquire data, process and score the resulting processed data and to enable user-defined metrics to refine the SRM assay to include the optimal transitions for a final method. The software package is integrated into discovery data acquired in-house as well as public database repositories, to intelligently build large SRM assays based on experimental data.¹ In addition, the software enables user-defined criteria to score resulting processed data to automatically parse the best peptides and transitions per protein into a refined method.²

Methods

Sample Preparation

Yeast cell lysate (1 mg) was enzymatically digested. A total of 1 µg was loaded on column and analyzed in triplicate.

LC/MS/MS

All data acquisition was performed using a Thermo Scientific TSQ Vantage triple quadrupole equipped with a nanoLC pump (1D plus, Eksigent, Dublin, CA). A binary solvent system was used A) 0.1% formic acid and B) MeCN (0.1% formic acid) pumped at 300 nL/min for a 1% per min. gradient over 40 minutes from 5% B to 45% B. All the experiments were carried out using iSRM instrument control software.³ Each peptide was monitored using 8 SRM transitions, the 2 most intense product ions were designated as primaries, and the remaining 6 product ions classified as secondary SRM transitions. Product ions were determined from the spectral library read in from MRM Atlas. The time-based transition acquisition (T-SRM) was used for all primary transitions and the time window applied for each primary transition set was 4 minutes, with experimentally determined RT as the center. A cycle time of 2 seconds was used to monitor the primary SRM transitions and a 0.10 second cycle time was used to monitor the secondary transitions. Dynamic exclusion was set to trigger the data-dependent secondary SRM acquisition only once for each peptide. Scheme 1 shows the workflow used to analyze and verify the putative biomarkers from yeast cell lysate. The study originated with 1000 candidates, analyzed, and refined for the best 757 peptides to be monitored for qualitative and quantitative information in one experiment.

Data Processing

All data processing was performed using Thermo Scientific Pinpoint software in an automated fashion. The spectral libraries were read in from MRM Atlas which contained the protein, peptide sequence, *m/z* values for precursor and product ions as well as the measured product ion intensities. The SRM transitions table created to acquire the data was used to process the resulting RAW files keeping the hierarchical relationship intact.

Scheme 1. Automated workflow for building, acquiring, processing, and verifying large SRM-based assays. The user specifies the spectral library used to build methods as well as process the resulting data. The final step in the workflow enables automated method refinement to include only the biomarker candidates meeting the user-defined criteria.

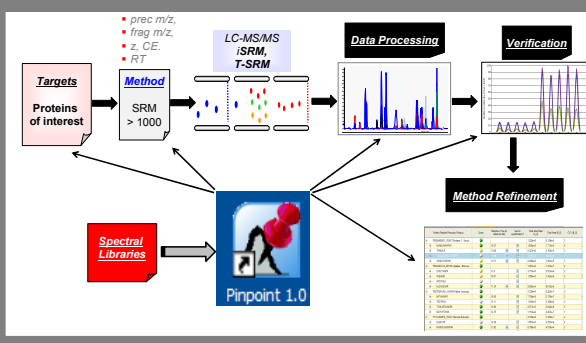
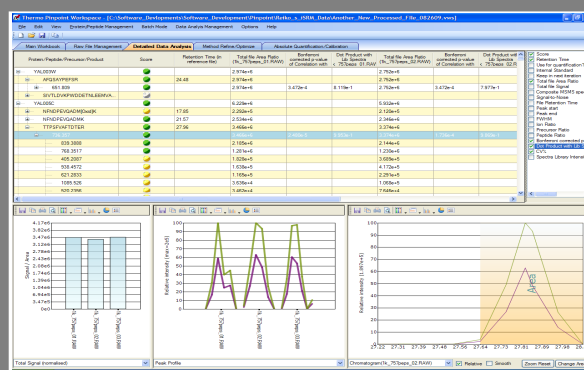


FIGURE 1. Screen capture of the automated processed results for the 757 targeted yeast peptides. The page is divided into the results table layered at the protein, peptide, precursor, and individual SRM transition levels. The interactive graphs at the bottom of the page display the results for the highlighted row. The columns shown are controlled by the check boxes in the upper right hand corner. The screen capture below shows the user-defined settings to automatically classify each of the targeted peptides monitored in the experiment.



Results

Pinpoint™ software was used to create a targeted SRM assay to simultaneously quantify and verify 757 peptides from yeast cell lysate. The novel method utilized T-SRM and iSRM acquisition strategies distributing the 6056 SRM transitions into primary (1514) transitions used for quantitation and secondary (4542) transitions for verification. Three technical replicates were acquired to assess quantitative precision as well as verification strategies based on the proper triggering and acquisition of composite MS/MS spectra comprised of primary and secondary SRM transitions for each of the targeted peptides. Figure 1 shows the cumulative results for the experiment. The reporting page is broken into two sections, the table listing the measured values and a score for each level measures as well as interactive graphs updated upon the user-defined row that is highlighted. The values that can be used to classify each level displayed in the table are shown in Figure 1b. The current values used highlight strength of the iSRM method, %CV (based on primary SRM acquisition) and library matched p-value (based on secondary SRM acquisition). The strategy provides a high degree of confidence in the scoring and enables an automated refinement step within the Pinpoint software to create more robust methods targeting the best peptides and SRM transitions while dramatically reducing manual interrogation.

An example of the scoring function is shown in Figure 2. The relative abundance of the composite MS/MS spectrum is shown as it compared to the spectral library entry (red bars). The eight product ions monitored provides sequence coverage as well as verification. The %CV for each product ion is shown with excellent reproducibility which is key for performing a dot-product correlation coefficient analysis and the subsequent p-value. The dot-product correlation coefficient equation uses the relative abundance values of each product ion obtained from the SRM measurement and the spectral library entry as opposed to the rank order. Using this approach enables a better distinction between real matches and false hits. The p-value is the probability that a random arrangement of the spectral library abundance values result in a greater coefficient than the real spectral library entry. For the example, in Figure 2,

FIGURE 2. Comparative analysis of the product ion abundance of the composite MS/MS spectrum and the spectral library entry for the targeted peptide TTPSFVAFDTER. The composite MS/MS spectrum was acquired using iSRM.

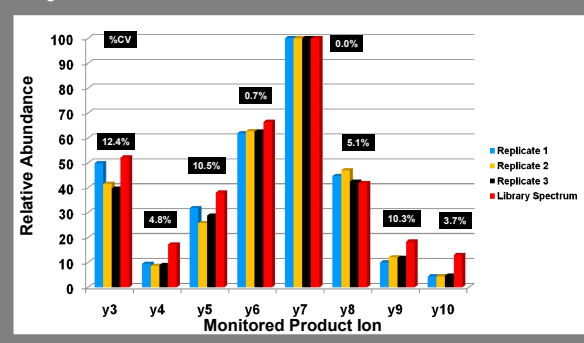


FIGURE 3. Distribution of the %CV measurements for the 757 targeted peptides monitored in one experiment. The results were from automated data processing performed in Pinpoint.

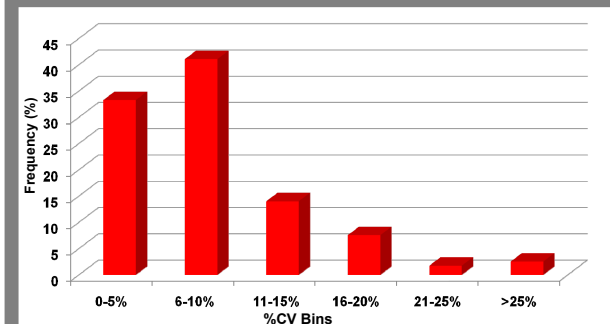
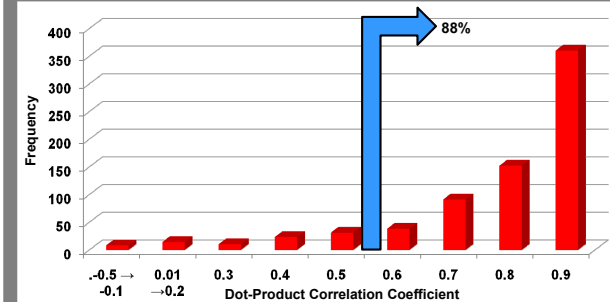


FIGURE 4. Dot-product correlation coefficient distribution for the 727 peptides that triggered secondary SRM transition acquisition. Each of the peptides triggered the acquisition of the secondary SRM transitions all three technical replicates.



the dot-product coefficient is 0.869 and the p-value is 2.48e⁻⁵ meaning there is one random match better than the experimentally determined value in 35,088 possibilities. An acceptable dot-product coefficient is 0.6 considering all targeted peptides were verified by monitoring 8 product ions.

The iSRM detection strategy enable accurate quantification as demonstrated by over 85% of the targeted peptides have %CV's less than 16% despite being monitored in one 60 minute gradient. Of the 85%, a majority of the peptides had %CV's less than 10%. A total of 727 peptides (96%) triggered secondary SRM acquisition in the 60 minute analysis. Of the 727, 88% had dot-product correlation coefficients greater than 0.6 which would provide high confidence in identification and verification that not only was the targeted peptide detected, but the signal measured in each SRM transition could be attributed to the peptide itself and less to the background matrix.

Conclusion

The workflow presented shows an integrated and robust method to predict, measure, verify, and perform method refinement in an automated manner.

- Spectral libraries provide information needed to build robust SRM transitions – proteins, peptides, charge states, *m/z* values for precursors and product ions, relative abundance, and RT information.
- Pinpoint software integrates discovery-based experiments with targeted protein quantitation experiments, building complex but effect methods providing quantitation and verification.
- Using T-SRM and iSRM acquisition strategies enables the a larger number of putative biomarkers to be interrogated without sacrificing quantitation or verification capabilities. Methods can be built automatically in Pinpoint software.
- Dot-product correlation coefficient analysis between spectral libraries and composite MS/MS data provides a high level of confidence on targeted product ion identity enabling automated scoring that can be used to refine methods and minimize manual interrogation.

References

- Prakash, A. et al. Expediting the development of targeted SRM assays: using data from shotgun proteomics to automate method development. *J. Proteome Research* 2008, 8(6), 2733-2739
- Gayen, A. K. The frequency distribution of the product moment correlation coefficient in random samples of any size draw from non-normal universes. *Biometrika*, 1951, 38, 219-247
- Kiyonami, R. et al. Increased Quantitative Throughput and Selectivity for Triple Quadrupole Mass Spectrometer-Based Assays Using Intelligent SRM (iSRM). 2009 ASMS poster # w313.

All trademarks are the property of Thermo Fisher Scientific and its subsidiaries.