Incorporation of an Automated Scoring Metric for High-Resolution/Accurate-Mass Targeted Peptide Quantitation Routine

Scott Peterman[1], Amol Prakash[2], Mary Lopez[2], John Rogers[3]

[1]Thermo Fisher Scientific, San Jose, CA: [2]Thermo Fisher Scientific BRIMS Center, Cambridge, MA; [3]Thermo Fisher Scientific, Rockford, IL

Overview

Purpose: Incorporate a robust method to automate targeted peptide identification, and verification based on HR/AM MS-level data when used in conjunction with spectral library information.

Methods: LC-MS properties were determined by processing the peptide retention time calibration (PRTC) mixture peptides and applying these rules to the processing of targeted peptides. Establishment of correct retention times, charge state and isotopic *m/z* values to establish XIC plots which are then used to calculate isotope distribution correlation coefficients. HR/AM MS-level results were corroborated with spectral library data for confirmation of peptide presence.

Results: Automated verification scheme for targeted HR/AM MS-level data analysis demonstrated 99% success rate for identifying targeted peptides analyzed neat and in biological matrices.

Introduction

Targeted protein quantitation applications have expanded from solely being performed on triple quadrupole mass spectrometers to high-resolution/accurate-mass (HR/AM) instruments to better leverage discovery data. Full-scan HR/AM MS detection simultaneously detects 1000's of components and enables post-acquisition quantitation. The primary limitation is an efficient and reliable method to extract MS-level data for a quantitative analysis. We present a robust workflow that incorporates spectral library information created during unbiased database matching with a synthetic peptide QC kit to characterize the LC-MS parameters and enables corrections to be applied to targeted peptides resulting in an automated method for building, processing, verification, and reporting quantitation results.

Methods

Sample Preparation

The Thermo Scientific Pierce Peptide Retention Time Calibration mixture, a set of 15 carefully-selected synthetic peptides, was analyzed neat as well as spiked into each sample at a constant amount. Targeted peptide analysis was performed for an equal-molar 12-protein digest standard analyzed neat as well as in matrix. BSA and HeLa cell lysate digest samples were used for false positive analysis.

LC/MS

All experiments were performed on a Thermo Scientific LTQ Orbitrap XL hybrid mass spectrometer equipped with a nano-LC pump. A binary solvent system of A) 0.1% formic acid and B) 0.1% formic acid in acetonitrile was used for all separations. A flow rate of 500 nL/min was used with a 0.7% per minute gradient (5-45% B) for the initial discovery experiments as well as quantitation. All data were acquired using a full-scan MS event (@ 50,000 resolution) in the Orbitrap™ mass analyzer followed by 5 data-dependent MS/MS events acquired in the linear ion trap.

Data Analysis

Unbiased discovery data was processed using Thermo Scientific Proteome Discoverer software version 1.2 to identify targeted peptides from the 12-protein digest. The resulting .msf files were used to create spectral libraries containing measured retention times (RT) as well as precursor and product ion *m/z* values from the discovery experiments.

All targeted quantitation data was processed using Thermo Scientific Pinpoint software version 1.1. The initial list of targeted peptides were imported from the spectral libraries. Each peptide was quantified using the four most abundant isotopes per charge state. Data extraction was performed using 5 ppm mass windows. Verification was performed using retention time, accurate mass, and isotopic distribution analysis for targeted peptides.

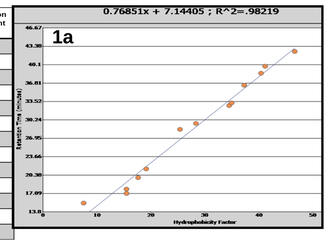
Results

The first step is to process the PRTC peptides to evaluate key LC-MS parameters such as mass accuracy and RT characterization. The table lists important key metrics that will be used to evaluate all targeted peptides. Figure 1a shows the linear relationship between hydrophobicity factors and measured RTs. The linear equation can be used to predict RTs for all targeted peptide as hydrophobicity factors are readily calculated in Pinpoint software based on the SSRCalc algorithm. Figure 1b shows the RT correlation between neat samples and different degrees of complexity. The linear equations are used to correlate spectral library retention times between samples to identify the most likely RTs for further verification schemes. In addition, the mass accuracy analysis for the PRTC peptides are used to evaluate the systematic error associated with detection in an Orbitrap mass spectrometer.

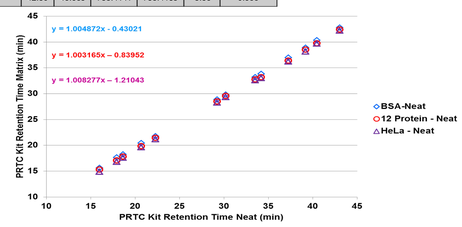
Following PRTC kit analysis, the sample to determine experimental RT values, mass accuracy, charge state determination, and isotopic distribution. The LC-MS values determined for the PRTC peptides were applied to targeted peptide analysis to establish RT windows as well as mass accuracy for data extraction and processing. Each targeted peptide sequence is used to determine the charge state and isotopic

FIGURE 1. PRTC kit peptide analysis used to evaluate all targeted peptides.

Peptides	RT	HF	Theo m/z	Exp m/z	Mass Error (ppm)	Correlation Coefficient
SSAAPPPPPR	15.34	7.5681	488.7642	488.7670	-5.73	0.999
HVLTSIGEK	17.18	15.522	492.2729	492.2755	-5.28	1.000
GISNEGQNASIK	17.83	15.500	609.3096	609.3124	-4.43	1.000
DIPVPKPK	19.89	17.651	447.2763	447.2785	-4.70	0.999
IGDYAGIK	21.57	19.157	418.7292	418.7313	-5.02	1.000
YYTITEVATR	28.61	25.406	608.8140	608.8140	-4.43	0.999
ELGQSGVDTYLQTK	29.44	28.368	676.3664	769.8885	-5.20	0.997
GILFVGSGVSGGEEGAR	32.77	34.52	796.4073	796.4105	-4.02	0.998
SFANQPLEVVYSK	33.23	34.965	741.3853	741.3896	-5.67	0.998
LTILEELR	36.4	37.303	493.7976	493.8001	-4.86	0.998
NGFILDGFPR	38.34	40.420	568.2983	568.3014	-5.28	0.999
ELASGLSFPVGFK	39.79	41.185	676.3664	676.3703	-5.62	0.998
LSSEAPALFQFDLK	42.30	46.663	783.4141	783.4183	-5.36	0.999

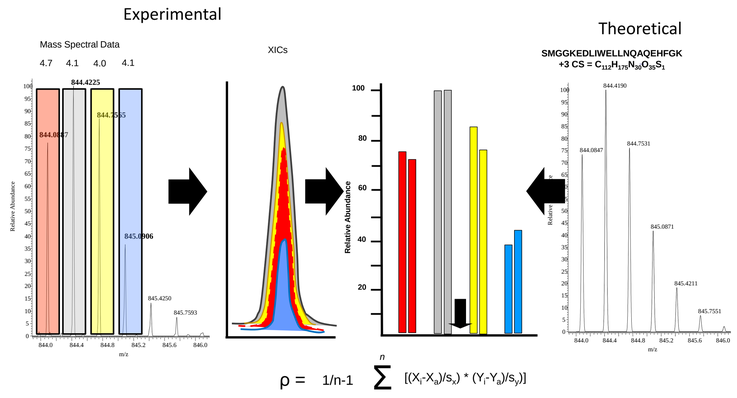


1b



m/z values that were used to perform extracted ion chromatograms (XICs). Figure 2 shows the method of data extraction. The experimental XICs were used to first identify retention times based on covariance analysis. The resulting XIC peaks were integrated and normalized to the most abundant isotope. The relative AUC values were compared to the theoretical values to determine the overlap for verification. Figure 3 shows a situation where there are multiple retention time points. Correlation coefficient analysis for each retention time point and charge state overlap identifies the correct RT as B and D both have coefficients greater than 0.99 but only the RT point at 24.2 min. has XIC traces for both the +2 and +3 charge states. The RT window was confirmed by the spectral library entry.

FIGURE 2. Data processing scheme for targeted peptide quantitation and verification. The Pearson correlation coefficient listed below was used to calculate isotopic distribution overlap.



$$\rho = \frac{1}{n-1} \sum_{i}^{n} [(X_i - X_a)/s_x) * (Y_i - Y_a)/s_y)]$$

FIGURE 3. XIC overlap analysis to identify true retention time for the targeted peptide KDSGFQMNQLR. The plot at the bottom of the figure shows comparative relative AUC values for the 4 retention time points.

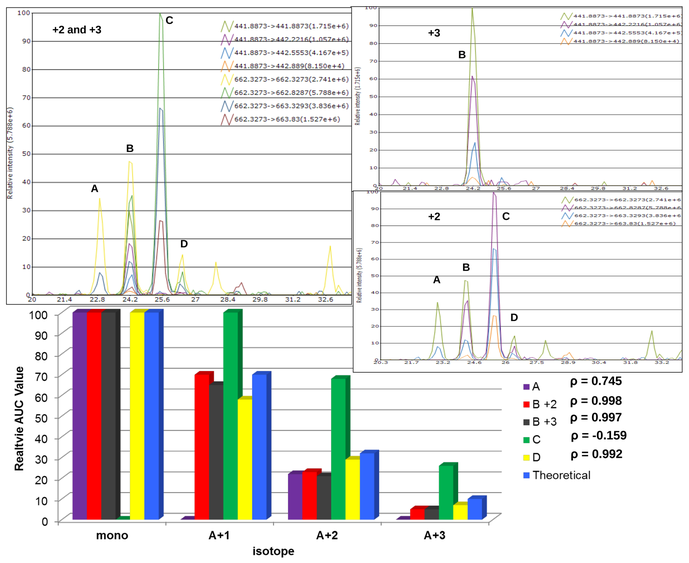
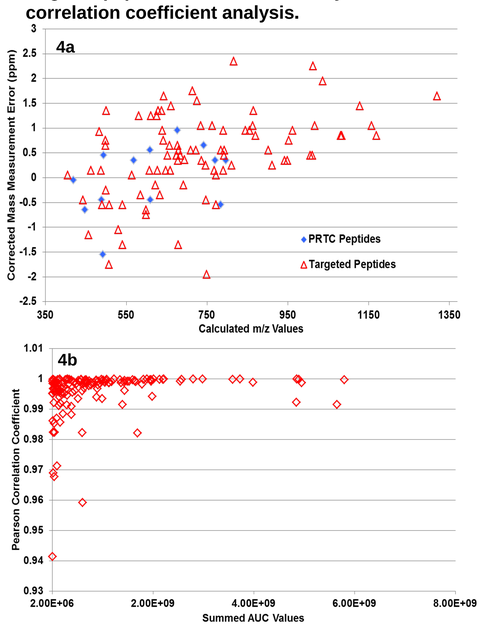


Figure 4 shows consolidated results following data treatment for the 200 targeted peptides. Due to the systematic mass error, mass values were corrected prior to performing XIC. The increased mass tolerance enabled tighter mass windows to be used for analysis. The resulting XIC values per peptide were compared against the theoretical distribution to determine the correlation coefficient. Figure 4b shows a plot of the correlation as a function of summed AUC value. Even low-abundance peptides still registered exceptionally high overlap with theoretical values.

Final confirmation was performed using RT correlation between spectral library entries and targeted analysis for the true positive data extraction and the false positive in a HeLa cell lysate digest. Figure 5 shows the correlation for the true values between the two methods. Each of the spectral library RT values were corroborated by unbiased product ion spectral matching.

FIGURE 4. Scoring results for the PRTC and targeted peptides for mass accuracy and correlation coefficient analysis.



To test the false positive rate, the m/z values used to assess the true positive analysis by correlation coefficient analysis of the isotopic distribution for the targeted peptides were used to perform data extraction for the HeLa cell lysate which was not spiked with the 12 protein mix. Figure 5b shows the RT correlation plot between the spectral library values and the FPR analysis. The comparisons between true and false positive analysis show very little overlap. To expand on the benefit of FPR analysis based on RT overlap, the distribution of correlation coefficients was plotted with and without RT constraints. The RT windows used in the study were determined from the 12 protein mix (Fig. 5a) and a 1 min. processing window was used for XIC extraction and coefficient determination. Figure 6 shows that factoring in expected RT windows dramatically reduces false positives by 33%. Additional interpretations could be implemented such as covariance analysis (peak shapes for isotopes) to further eliminate random noise from true isotopic ion signal being considered. In addition to the HeLa cell digest analysis, the true positive correlation distribution is presented. All but one of the isotopic groups had coefficients greater than 0.9. The inset in Figure 6 shows the distribution for the isotopic groups for the true positive analysis. The breakdown shown in Figure 6 in conjunction with the distribution shown in Figure 4b show that even low level signal still correlates.

FIGURE 5. Retention time evaluation and correction based on measured RT and spectral library RT values for the 12-protein digest mix and HeLa false positive analysis.

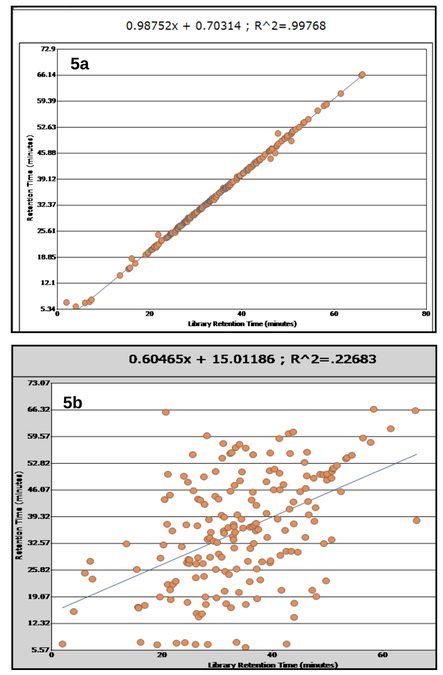
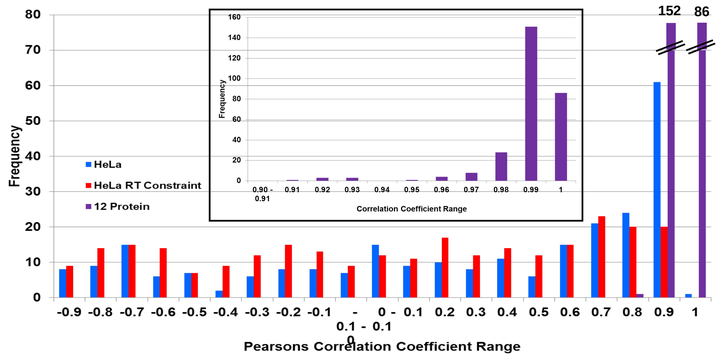


FIGURE 6. Correlation coefficient distribution for 200 targeted peptide isotopic *m/z* values analyzed in the 12-protein mix (true positive) and the HeLa (false positive) with and without RT constraints based on the true RT values.



Conclusion

The workflow presented here demonstrates and robust and highly confident method of automating target peptide sequence identification as well as determining abundant charge states, isotopic distributions and retention times. The resulting target peptide information can be incorporated into additional scheduled data acquisition methods.

The presented workflow demonstrated:

- Incorporation of a well-characterized synthetic peptide (PRTC) kit enabled LC-MS parameter characterization.
- Spiking the PTRC peptides into each sample enabled a robust method of RT determination across samples and instrumental methods.
- Use of accurate *m/z* values and tight mass tolerance windows provided sufficient mass discrimination capabilities to reduce background interference resulting in positive detection.
- Incorporation of Pearson correlation coefficient provided a robust means of determining isotopic distribution overlap between experimental and theoretical values.
- Full-scan HR/AM MS-level detection in the LTQ Orbitrap XL mass spectrometer maintained isotopic integrity over the intra-scan dynamic range, which is necessary for target verification.
- Pinpoint software incorporates a novel method for target verification, data processing and validation.

Thermo SCIENTIFIC