Split SIEVE: Informatic Support for Differential Split LC-FTMS Experiments

M. Askenazi[1], J. Sutton[1], L. Bonilla[1], C. Ptak[2], F. M. Maroto[3]

[1]Thermo BRIMS Center, Cambridge, MA, USA; [2]Advion Biosciences, Ithaca, NY, USA; [3]Thermo Electron Corp., San Jose, CA, USA

Overview

Purpose: To demonstrate the automated analysis of differential proteomic expression profiling using a bottom-up, label-free integrated fractionation/nanoESI device coupled with LC-FTMS Detection (see Poster P071-M for more details on this unique workflow/setup).

Methods: Two standard protein mixtures were analyzed by, a semi-preparative, split workflow that coupled the TriVersa™ NanoMate™ (Advion Biosciences, Ithaca, NY) with a Finnigan™ LTQ FT™ (Thermo Electron, San Jose, CA) and simultaneously provided fractions for re-injection MS, MS/MS and MS^3 analyses. (Figure 1 below)

Results: This experiment demonstrates the effectiveness of a split workflow in supporting a thorough characterization of differential expression on the initial HPLC-MS/MS phase of the workflow as well as the usefulness of directed sample fraction re-injection in those instances where peaks have been under-characterized by the primary high-throughput analytic pass.

Introduction

At the Thermo BRIMS (Biomarkers Research Initiatives in Mass Spectrometry) Center, our goal is to develop, validate and share comprehensive MS-based solutions (from sample collection to informatics) for HT, robust biomarker discovery and development using Thermo Electron's broad portfolio of cutting-edge instrumentation. This goal is achieved by putting together a unique workflow providing tight integration of all major steps in the label-free analysis of a large scale biomarker experiment: from Sample Preparation through Sample MS-Analysis to Data Analysis. The SIEVE (Statistical Iterative Exploratory Visualization Environment) represents the final step in this workflow: a research-oriented informatic pipeline capable of reducing the large datasets produced by typical biomarker experimental designs into a small set of statistically significant observations. While the SIEVE already enables users to analyze data generated on Finnigan LTQ™, LTQ FT or Orbitrap™ instruments (using MALDI or LC ion sources) according to a predefined experimental design (typically, a two-group randomized controlled study or a single-group longitudinal study), the system has recently begun supporting experiments with a split design (where the high-flow branch is collected into 96 well plates for offline investigative work). By analyzing the standard samples in this fashion, we are able to show the power of the SIEVE in its primary high-throughput mode, in addition to the potential value of split-phase experiments.

FIGURE 1. Graphical representation of a typical bottom-up proteomic experiment within the context of the overall Thermo BRIMS Center analytical pipeline: five replicate injections of each sample (A,B) were run through the TriVersa, thereby generating a time-dependent fractionation of the sample (where each well in the 96-well plate represents approximately 36-seconds of chromatographic time).

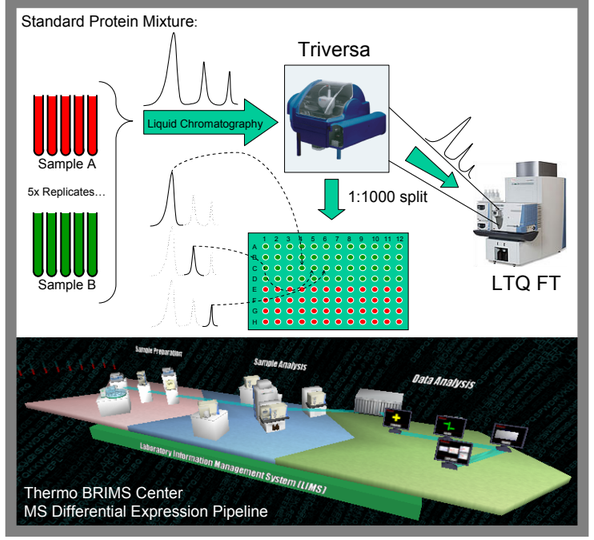
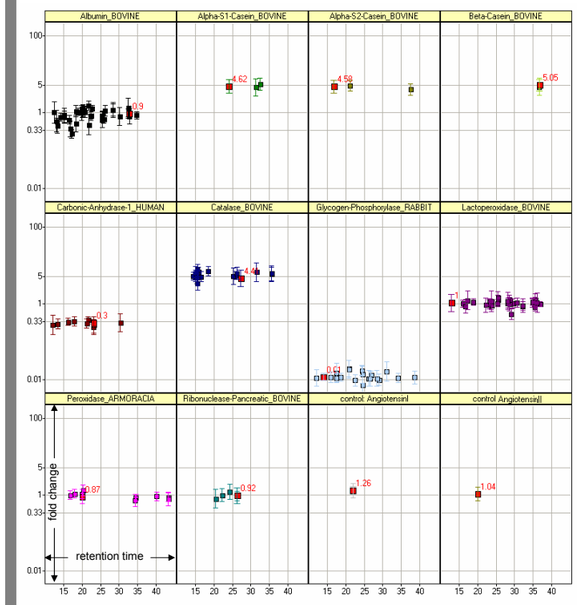


FIGURE 2. Protein summary of high-quality, high-precision peptide expression ratio estimates (with 95% confidence intervals) by retention time by protein. Every datapoint represents a C_{12} isotope with an identification which passes a stringent charge-state/XCorr filter (z={1,2,3},XCorr={1.8,2.5,3.8}) and has a fold-change confidence interval with a range that is less than four-fold. The red datapoints are the median per protein which is an appropriately robust measure of centrality for the protein ratio estimate.



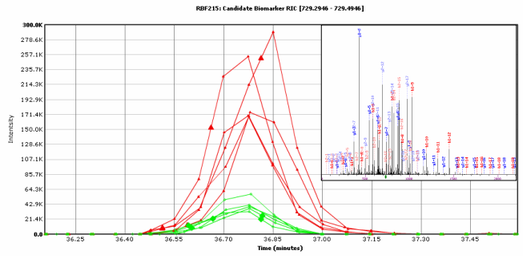
Methods

Enzymatic Digestion
Proteins were dissolved in 100mM NH_3HCO_3 pH 7.8, reduced with 100mM dithiothreitol for 1 hour at 56°C, then alkylated using 300mM iodoacetamide for 1 hour at room temperature in the dark. Proteolytic enzyme was added at ~ 1:20 w/w enzyme-to-protein ratio and incubated for 15 hours at 37°C. Reaction was quenched with acetic acid and stored at -80°C until analysis.

Sequential LC-MS/MS Analysis
Sample digests were analyzed by LC MS/MS using a Finnigan LTQ FT mass spectrometer (San Jose, CA) equipped with a TriVersa NanoMate (Advion BioSciences, Ithaca, NY). Peptides were captured and concentrated for on-line RP-HPLC using a C18 capillary column (2.1mm i.d. x 50cm), with 5.0 μm packed particles (Higgins Analytical Inc., Mountain View CA). A split of 1:1000 (NanoMate LTQ FT to plate) was achieved using MicroTee (Upchurch Scientific, Oak Harbor, WA). The low-flow branch of the split was directed into the mass spectrometer and full scan (MS) data was collected in the FT-ICR MS in profile mode between 400-1800 m/z with simultaneous Data Dependent™ MS/MS scanning of the top five most abundant ions performed in the ITMS. High-flow branch was collect into a 96 well plate using a time based fractionation of 36 sec/fraction.

Re-Infusion
Collection plates were lyophilized to dryness then reconstituted in 25 μL of 50% Methanol in 0.1% acetic acid. Infusion was achieved using infusion manual of the TriVersa NanoMate. A fifteen minute interactive analysis of an "orphan" differentially expressed peptide peak (one which was not subject to MS/MS during any of the initial, high-throughput phase injections) was performed using less than 5 μL of the samples.

FIGURE 3. Example of a detailed individual frame report (the median for Beta-Casein as seen in figure 2). Reconstructed Ion Chromatograms are represented for every individual re-injection decorated with the location of the location in time where a representative MS/MS was taken for the peak's precursor in every run. The peptide sequence is a hyperlink to a visualization of the annotated MS/MS data for the given run.



Estimate of intensity ratio: **5.05** [3.57 - 7.15] (pval = 6.00e-006)

File	ProteinID	Peptide	XCorr	Charge	Time	Description
ABRF_A1	gi|115660	R.DMPIQAFLLYQEPVLGPVR.G	6.384900	3	36.66	Beta-Casein_BOVINE
ABRF_A2	gi|115660	R.DMPIQAFLLYQEPVLGPVR.G	6.063600	3	36.59	Beta-Casein_BOVINE
ABRF_A3	gi|115660	R.DMPIQAFLLYQEPVLGPVR.G	5.137100	3	36.51	Beta-Casein_BOVINE
ABRF_A4	gi|115660	R.DMPIQAFLLYQEPVLGPVR.G	5.236200	3	36.51	Beta-Casein_BOVINE
ABRF_A5	gi|115660	R.DMPIQAFLLYQEPVLGPVR.G	6.411200	3	36.81	Beta-Casein_BOVINE
ABRF_B1	gi|115660	R.DMPIQAFLLYQEPVLGPVR.G	4.260500	3	36.60	Beta-Casein_BOVINE
ABRF_B2	gi|115660	R.DMPIQAFLLYQEPVLGPVR.G	5.720800	3	36.66	Beta-Casein_BOVINE
ABRF_B3	gi|115660	R.DMPIQAFLLYQEPVLGPVR.G	5.131500	3	36.59	Beta-Casein_BOVINE
ABRF_B4	gi|115660	R.DMPIQAFLLYQEPVLGPVR.G	5.925500	3	36.81	Beta-Casein_BOVINE
ABRF_B5	gi|115660	R.DMPIQAFLLYQEPVLGPVR.G	6.253200	3	36.81	Beta-Casein_BOVINE

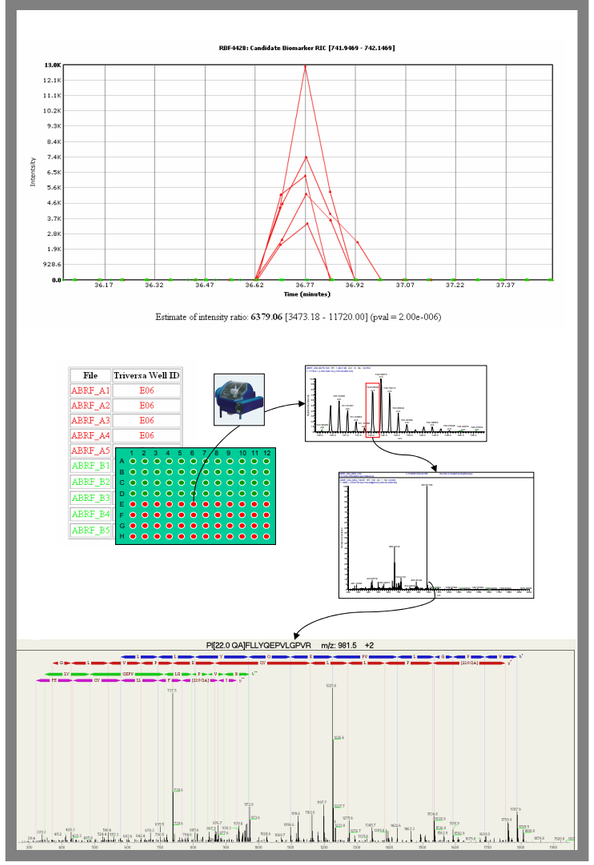
Results

The initial dataset consisting of 10 RAW files representing the two data samples re-injected five times, were run through the SIEVE, which identified the top 7000 most intense frames in the combined experimental dataset. These peaks were then filtered by a common XCorr vs. Charge State filter – essentially requiring that frames containing peptide identifications having charge state 1, 2 or 3 have a minimum XCorr of 1.8, 2.5 or 3.8 respectively (an exception was made for the internal controls which being very short peptides did not yield very high XCorr value despite their retention time and MS/MS being consistent with Angiotensin I/II). Of the surviving 818 frames, those yielding a 95% confidence interval for the expression ratio of the A to B samples (based on a t-test for a difference in the mean log-expression) ranging over more than 4 fold change (i.e. the upperbound was greater than 4 times the lowerbound) were filtered out. The remaining 518 frames passed an additional filter which involved removing frames whose m/z value was inconsistent with the peak being a monoisotopic mass (essentially a C_{12} filter). Finally, of the remaining 164, 6 frames which contained signal in only one of the two sample types (present/absent frames) were removed from the dataset – as well as 17 frames which were removed after manual inspection detected multiple peaks or incorrect identifications. These 141 frames yielded between 1 - 36 estimates for the A to B ratio per protein, and the median (a reasonable measure of centrality in this case) frame per protein is represented as red datapoints in Figure 2. Thanks to the split nature of the workflow, even frames which contained no MS/MS-based identification can be still be investigated as shown in Figure 3.

The results of this experiment demonstrate the effectiveness of a split workflow in supporting a thorough characterization of differential expression. This unique workflow is currently implemented at the Thermo BRIMS Center in Cambridge where it is being used routinely in our plasma-based label-free biomarker discovery program. The significant advantages of this workflow stem from our ability to use higher capacity LC columns AND to perform nanoESI, either in-line or off-line, for high-sensitivity FTMS detection.

Our ability to use larger capacity columns results in improved LC-MS reproducibility and ruggedness, both of which are desirable in a successful label-free semi-quantitative approach. Furthermore, the ability to load more sample onto the column is critical in biomarker discovery applications where low-level markers are usually pursued in the midst of complex matrices with large dynamic ranges of analyte concentration (e.g., plasma). This enables a higher success rate of identification of low-level components during the initial on-line HPLC-MS/MS phase of the workflow, as well as more intelligent and directed sample fraction re-injection in those instances where peaks have been undercharacterized by the on-line high-throughput analytic pass.

FIGURE 4. Example of an "orphan frame" which does not contain any MS/MS data – making it an ideal candidate for re-infusion. The appropriate well was re-infused and MS/MS followed by MS^3 data was collected. The MS^3 data was analyzed using the prototype DeNovoX V2.0 application to yield the modified peptide: DM*PIQ[22]AFLLYQEPVLGPVR



Estimate of intensity ratio: **6379.06** [3473.18 - 11720.00] (pval = 2.00e-006)

File	Triversa Well ID
ABRF_A1	E06
ABRF_A2	E06
ABRF_A3	E06
ABRF_A4	E06
ABRF_A5	
ABRF_B1	
ABRF_B2	
ABRF_B3	
ABRF_B4	
ABRF_B5	

Conclusions

In addition to increasingly comprehensive support for differential expression experiments the SIEVE framework has begun to be applied to split workflow experiments, enabling e.g. the post-acquisition analysis of "orphan peaks". Given the promising initial results achieved by split-workflow experiments, more aggressive automation is being pursued, e.g. in the robotic automation and control of the re-infusion process.

References

1. Jennifer Listgarten and Andrew Emili (2005), Statistical and Computational Methods for Comparative Proteomic Profiling Using Liquid Chromatograph-Tandem Mass Spectrometry. Molecular & Cellular Proteomics, 4.4, Review pp.419-434
2. Spotfire, DecisionSite 8.1™, http://www.spotfire.com

Thermo ELECTRON CORPORATION