

Due Diligence Towards Building a Clinically useful SRM Experiment: The Uncommon Knowledge

Amol Prakash¹, Scott Peterman¹, Jennifer Sutton¹, Mary Lopez¹, Reiko Kiyonami², Karen Salomon², Tom McClure², and Andreas Huhmer²

Thermo Fisher Scientific, ¹Cambridge, MA, USA; ²San Jose, CA, USA

Overview

Purpose: To demonstrate prototype software that will automatically build SRM tables, instrument methods, and sequences as well as process targeted protein quantitation experiments and produce a consensus report.

Methods: Use a model system of milk proteins, and spike in known amounts of alpha and beta casein digestion standards to simulate biological and technical replicates to test the ruggedness of the prediction and processing software.

Results: The SRM Workflow prototype software predicted, grouped, and processed the resulting data for the biological and technical replicates and gave a quality scoring to quickly determine which samples, if any, fell out of the user-defined acceptance window.

Introduction

Recently, the proteomics community has begun shifting biomarker discovery from a pure shotgun approach to hypothesis-driven methods that integrate bioinformatics and utilize targeted SRM analysis. To satisfy the requirements of identifying and confirming biomarkers for clinical assays, key metrics need to be met. A biomarker has to be biologically relevant and yield unique peptides used to represent the targeted proteins. To increase throughput in method development, we introduce a software package designed to take advantage of the peptide and protein identifications determined using highly sensitive discovery experiments to drive targeted peptide quantitation. This prototype software handles all facets of the protein quantitation process from *in silico* programs designed to base proteotypic peptide selection off of library data or theoretical predictions, to processing the acquired data and correlating relationships between controls and sample results using a unique algorithm.

Data processing is accomplished by using the peptide information generated during the method building process to identify retention times, summed and individual peak areas, ion and area ratios, as well as relative and absolute quantitation for label-free and labeled quantitative approaches.

FIGURE 1. Screen capture of the prototype SRM Workflow homepage that guides the user through the method building, data collection, processing, and data review.

- Drop the protein or peptide sequence from any FASTA source.
- Define the proteotypic peptide candidates based on previous experiments or biostatistics.
- Pick the best SRM transitions for each peptide based on previous experiments or biostatistics.
- Make copies of proteotypic peptides for heavy labeled, iTRAQ™, SILAC™, or modified peptides and SRM transitions.
- Define the experiment being acquired, build the instrument method, and set the sequence being acquired.
- Load the files to be processed, define the quantitation approach, process the data, inspect the results, refine the experiment, create a report.

Methods

A 1 mL aliquot of milk was crashed using MeCN to produce proteins which were then dried down and re-dissolved in 6 M urea and digested using standard protocols. In addition, alpha and beta casein (SIGMA™, St. Louis, MO) were digested at 1 mg/mL. A constant aliquot of the milk protein digest was placed in separate vials. To each vial a known amount of alpha and beta casein were spiked in to simulate biological samples. A total of 5 biological samples were made and 5 technical replicates were acquired for each to generate statistics.

All data was collected on a TSQ Quantum Ultra™ (Thermo Fisher Scientific, San Jose, CA). A total of three SRM transitions were acquired for each targeted peptide to ensure the correct targeted peptide was identified. All experiments were separated on a Hypersil Gold™ 50 x 1 mm packed with 1.9 μm packing material. A binary solvent system consisting of A) 0.1% formic acid and B) MeCN (0.1% formic acid) were used in a fast gradient flowing at 70 μL/min.

FIGURE 2. Screen capture demonstrating the selection of proteotypic peptides. The inset shows user-defined filters for identifying proteotypic peptides and the capability of performing a BLAST search against a user-specified database.

Results

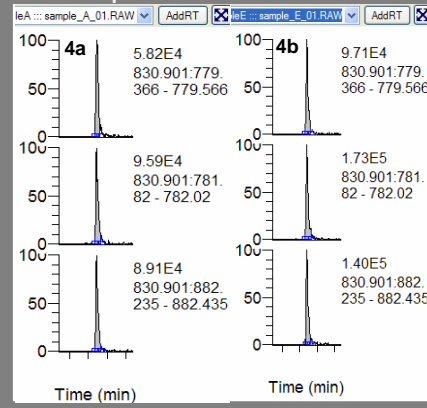
The key deciding factor in determining the success or failure of a targeted protein quantitation experiment is based on selection of the proteotypic peptides. Taking into account the results from discovery experiments dramatically reduces the amount of time needed in developing quantitation experiments. This SRM Workflow is designed to better integrate method-building with results from previous discovery experiments. Figure 1 shows the overall workflow of the SRM Workflow prototype software with each of the icons used to determine each step of the process. The key feature of each step is listed in Figure 1. Identified proteins from database search engines, software that identifies and links biological relationships, and proteins identified from SIEVE™ output are able to be dropped into SRM Workflow. Once in the SRM Workflow software, the user can define the proteotypic peptides generated for further method development.

Figure 2 shows the second step of the workflow in determining the proteotypic peptides for the casein proteins being monitored. The inset in Figure 2 shows the user-defined parameters that can be changed to refine the putative peptide list by selecting the enzyme being used, length of peptide, residues contained, and most importantly, providing a means of a BLAST search against the matrix to eliminate conserved sequences that may lead to false positives. The resulting peptides can then be copied and modified to build the resulting SRM table.

FIGURE 3. Screen capture of the data processing aspect of SRM Workflow. The report generator can process data from numerous biological and technical replicates and groups the results in proteins, peptides, and individual SRM transitions.

Once the experimental data has been acquired, automated processing can be performed. Figure 3 shows the results for samples A, B, C, and E. The results table is broken down into the protein, peptide, and individual SRM being reported. Note that the automated identification of the correct retention time is reported for each peptide based on detecting all of the SRM transitions at the same retention time. In addition, the area ratios for each sample are calculated and shown. Note the consistency of the integrations for each peptide and individual SRM transition as well as the variance.

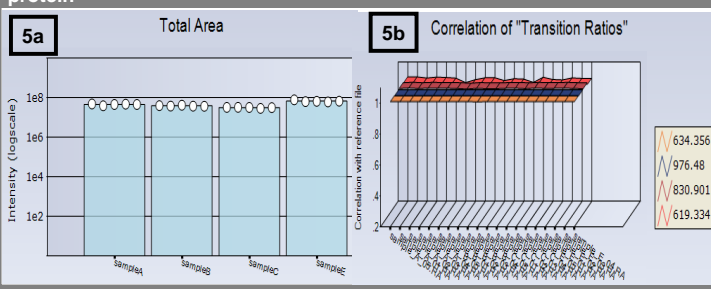
FIGURE 4. Individual SRM transition comparison for VPQLEIVPNSAEER where 4a is the transitions for the control and 4b is for Sample E.



The results for the phosphopeptide VPQLEIVPNSAEER are presented in Figure 4. A comparison of the individual SRM traces for the b₇, y₇, and the neutral loss product ion are shown for Sample A and E. Note that using three SRM transitions that all line up in RT confirms the correct retention time and more confidently enable the automated algorithm to identify the correct peak and integrate it. The intensity ratios for the two samples show excellent agreement for further confirmation.

In addition to the graph and individual retention times being displayed, two composite graphs are used to evaluate the peptide results across the entire sample set. Figure 5a shows the integrated peak areas for each biological and technical replicate within the experiment, and the overall sample set. In addition to the area plot, Figure 5b shows a correlation result for each peptide for a targeted protein. The scoring considers RT, ion ratios, and areas in relationship to the identified control file. The display in Figure 5b shows the correlation result for all proteotypic peptides associated with the targeted protein. The benefit of having these two additional graphical representations of the overall sample set is that it immediately enables the user to identify particular technical or biological injections that do not follow trend lines and can be more closely inspected.

FIGURE 5. Composite results for one targeted peptide. Figure 5a shows the integrated peak area as a function of biological groups and Figure 5b shows the Quality scoring across the different samples for a targeted protein



Conclusions

Building a clinically useful SRM experiment requires extreme care, due diligence and yet has to be high throughput. The SRM Workflow software enables the user to do so with an intuitive interface and a very powerful back end:

- Easy interface to allow progression from discovery experiments to targeted SRM experiments.
- Pre-defined and user-defined parameters to dramatically increase the ability to create a proteotypic peptide list.
- Automated BLAST search for increased confidence in its uniqueness of target peptides.
- Suggestion of the most optimal product ions, as well as enabling the user to select any SRM transitions and as many as necessary. Emphasis on multiple transitions per peptide.
- Enables labelling strategies like SILAC and iTRAQ, as well as monitoring PTMs.
- Intuitive and smart questionnaire to help build sequence files and instrument methods in an automated fashion. The resulting sequence can be tailored for quantitation curves, pure discovery, or biological and technical replicates with insertion of blanks or system washes in between each sample.
- Automated data processing directed by the original protein, peptide, and sequence file created during the method building steps.
- Automated and highly efficient computation of retention time, signal intensity, ratios, CV scores, alignment, noise levels and quality scoring. All computations based on strong statistical basis.
- Calibration experiments handled in an automated and smooth workflow.
- Visualization by means of graphs to assist quick inspection of data quality.

© 2007 Thermo Fisher Scientific Inc. All rights reserved. ITRAQ is a trademark of Applera Corporation. SILAC is a trademark of Invitrogen Corporation. SIGMA is a trademark of Sigma-Aldrich Co. and its affiliate Sigma-Aldrich Biotechnology. All other trademarks are the property of Thermo Fisher Scientific Inc. and its subsidiaries.