A New Software for Automated, High-Throughput Quantitative Proteomics

Bernard Delanghe[1], Rosa Viner[2], Torsten Ueckert[1]

[1]Thermo Fisher Scientific (Bremen) GmbH, Bremen, Germany; [2]Thermo Fisher Scientific, San Jose, CA, USA

Overview

Purpose: Quantification of precursor ion labeled peptides.

Methods: New algorithms.

Results: Fast, accurate and precise protein quantification for SILAC and isotopic labeling techniques.

Introduction

One of the major goals in proteomics research is the accurate quantification of protein expression in biological systems. Stable amino acid isotope labeling (SILAC) is widely used for relative protein quantification.[1,2] However, analyzing those large and complex datasets is still challenging and requires sophisticated data reduction algorithms and quantification schemes.

Here we describe a new workflow in Thermo Scientific Proteome Discoverer software for automated, fast and accurate SILAC quantification. With the introduction of high resolution and high mass accuracy Orbitrap™ analyzers it has now become possible to analyze a wide range of SILAC experiments, including experiments using heavy and light lysine, arginine, isoleucine , di-methyl lysine and other user defined labeling. The entire workflow from data acquisition to data analysis can be easily automated from within the acquisition sequence using Proteome Discoverer™ Daemon together with the Thermo Scientific Xcalibur sequence editor.

Methods

The complete workflow is shown in Figure 1 and consist of two parallel processes: identification and quantification.

Identification

The identification is done as usual and consists of 2 steps:

1. Selection of the spectra
2. Identification of the selected spectra using Mascot, Sequest or both of the search algorithms.

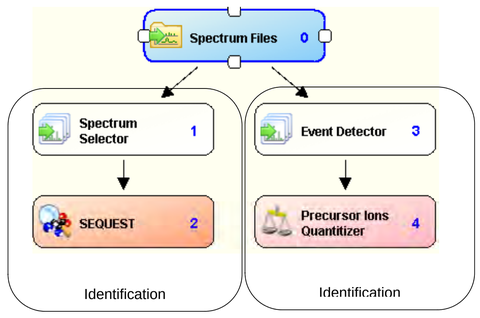
Quantification

Accurate detection and protein quantification consists of 3 steps:

1. Event detection: after noise removal all events (features) are detected and the individual peak areas are calculated.
2. Precursor quantification: The identified peptides are associated with the detected events based on mass accuracy and retention time. Using the information included in the quantification method, the theoretical mass(es) of the corresponding labeled pair (light, medium or heavy) are calculated. Those are then associated with the corresponding events the same way as the identified peptides. This information is also used to validate the results and errors are flagged. The peptide ratios are calculated using the same number of isotopes (Figure 3).
3. The protein ratio is calculated using the median peptide ratio (Figure 6B).

All raw files selected in the Spectrum Files step will be automatically processed. This can be further automated by adding the Proteome Discoverer Daemon as post processing method in the Sequence Editor of the Xcalibur™ software. In this configuration the raw files acquired on the instrument computer are automatically send to the data analysis computer and processed with the specified workflow.

FIGURE 1. Overview of the workflow.



Results

A large number of raw files from different SILAC experiments have been processed to evaluate the performance of the quantification algorithm and the speed of the software. Samples with different types of labels as well as with different fixed ratios were analyzed. All samples were analyzed using Thermo Scientific LTQ Orbitrap XL or LTQ Orbitrap Velos hybrid mass spectrometers.

Robustness and quality

To evaluate the quality and robustness of the software we processed the benchmark data set consisting of 72 LC/MS runs provided with the MaxQuant software.[3]

The result are summarized in Figure 2A. In total around 210,000 peptides have been identified with a false discovery rate (FDR) better then 1%. About 90% of those peptides have been quantified.

FIGURE 2. Summary of the quantification result of 72 raw files (A) and the main parameters used for the protein identification (B) .

A. Number of Peptides at 1% FDR	
Inconsistently labeled	437
No quantification values	24130
Not unique	12122
Redundant	81804
Unique	91484
Total	209977
Peptides quantified	89%

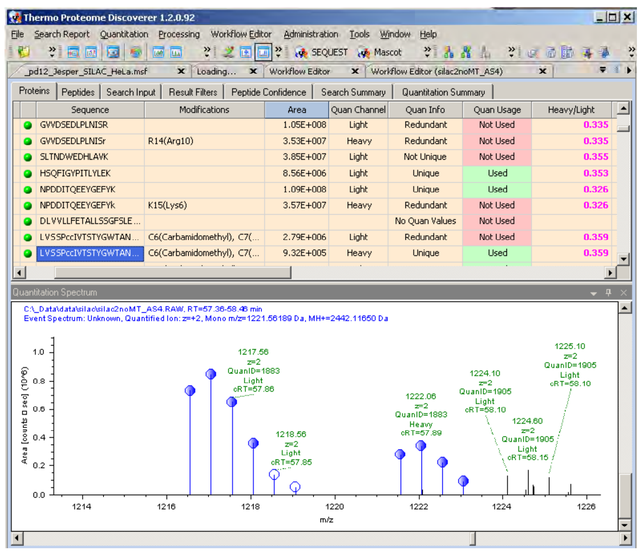
Number of raw files	72
Total file size (GB)	18.5
Total number of MS2 spectra	893636

B. **Search Conditions**	
Search Engine:	Mascot
FASTA:	IPI_Human
Peptide Tolerance:	10 ppm
Fragment Tolerance:	0.8 Da
	Oxidation (M); Lys8 (K); Arg10 (R);
Dynamic Modifications:	Acetyl (Protein N-Term)
Fixed Modifications:	Carbamidomethylation (C)

Speed

It takes about 8 minutes to quantify an average raw file (~ 300MB, 12,500 MS2 spectra) of a SILAC experiment on a typical desktop computer. Computation time varies with file size, number of full scans, sample concentration and complexity

FIGURE 3. Example of SILAC pair quantification. All detected isotopes are indicated by a blue circle (spectrum in lower panel). Isotopic peaks included in the quantification calculations are highlighted with filled blue circles, ensuring consistency and precision of peak area determination.



Accuracy and precision of the algorithms used for quantification.

To test the accuracy and the precision of the quantification algorithms we analyzed heavy and light SILAC labeled samples mixed in known ratios. In case of chemical labeling (dimethyl-Lysine labeling or 18O labeling) a digested protein mixture, non labeled and labeled, were mixed together at a certain ratio and then analyzed by LC-MS/MS.

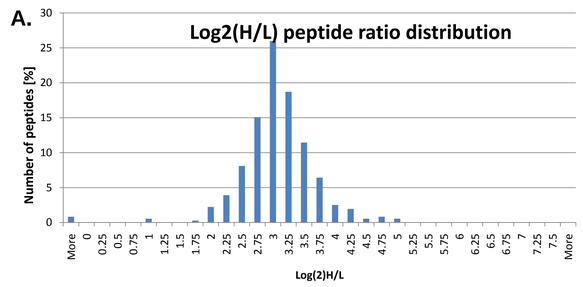
Figure 4 shows the result of all peptide ratios with FDR better then 1% of a classic SILAC experiment where treated (labeled with arginine 10 and lysine 6) and untreated cells are mixed 10 to 1.

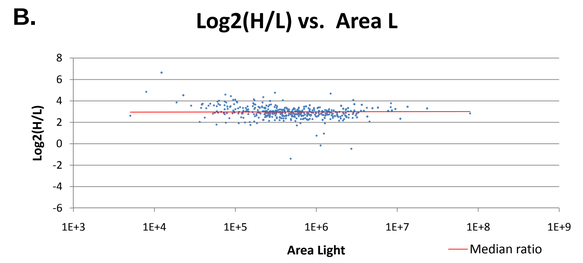
The results shows that the median peptide ratio is 7.6 with a standard deviation (SD) of 1.5 for this example. As the heavy labels are not completely incorporated, the labeling efficiency is ~96% for this example and quantification is resulting in an overall lower peptide ratio.

This example demonstrates that accurate peptide quantification can be obtained over 4 orders of magnitude of precursors ion abundances at 10:1 ratios (Figure 4B).

FIGURE 4 . Result of a "classic" SILAC experiment (Arginine 10, Lysine 6) with fixed 10/1 ratio. (A) displays the distribution of the log2(heavy/light) ratio of the peptides. (B) displays the same ratio versus the area of the light peptide.

Median ratio	Mean ratio	SD
7.6	7.7	1.6
Log2(Median Ratio)	Log2(Mean Ratio)	Log2(SD)
2.9	2.9	0.7





Similar results have been obtained using isoleucine 6 as an isotopic label. As not all peptides contain isoleucine as part of the peptide sequence , fewer peptides in the sample are quantified (data not shown).

Dimethyl-lysine labeling

As "classic" SILAC is not generally applicable other techniques, as dimethyl-lysine labeling have recently been developed. [4]

In this case peptides are chemically labeled.

The software has been tested on a light, medium, and heavy dimethylated labeled proteins. Also here the protein result have good precision and accuracy for both heavy to light and medium to light ratios (data not shown).

Conclusion

Proteome Discoverer software provides fast, accurate and precise protein quantification for SILAC and isotopic labeling techniques.

Robustness and speed of the quantification algorithm enable quantification of the majority of peptides from large datasets in an economical fashion.

The entire workflow from data acquisition to data analysis can be easily automated from within the acquisition sequence using Xcalibur Sequence Editor together with the Proteome Discoverer Daemon.

References

1. Everly, P.A., et al. (2004). Quantitative cancer proteomics: Stable isotope labeling with amino acids (SILAC) as a tool for prostate cancer research. Mol & Cell Proteomics. 3.7: 729-735.
2. Mann, M. (2006). Functional and quantitative proteomics using SILAC. Nature Reviews. 7:952-959.
3. Cox, J. et al. (2009). A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. Nature Protoc 4, 698-705
4. Boersema, P.J, et al. (2009). Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nature Protoc , Published online 19 March 2009; doi:10.1038/nprot.2009.21