

Spectral Merging to Improve Peptide Identification and Quantitation

Rosa Viner¹, Torsten Ueckert², Terry Zhang¹ and Vlad Zabrouskov¹

Thermo Fisher Scientific, ¹San Jose, CA, USA and ²Bremen, Germany

Overview

Purpose: Evaluation of the benefits of combining CID and collision cell spectra for simultaneous peptide identification and quantitation using iTRAQ reporter ions.

Methods: Datasets for analysis were generated from iTRAQ labeled A549 cells digests treated with 5 μ M camptothecin and harvested at 0, 2, 8 and 24 hours post treatment. Selected parent peptide ions were fragmented in the HCD collision cell and/or in the Ion Trap (IT) by CID. Spectra pre-processing was done by Proteome Discoverer 1.0 software.

Results: Using standard data analysis (no CID/HCD spectra merging); on average 400 proteins were identified and 85% of identified proteins were quantified in single LC run in whole cell lysates. With the new data analysis workflow that merges low resolution ion trap CID spectra with high mass accuracy and resolution collision cell spectra obtained on the same precursor, approximately the same number of proteins was identified and quantified. However, the number of the identified peptides increases substantially thus providing better quantitative statistics per protein.

Introduction

Quantitation of differentially expressed proteins is an important area of proteomics research. Both labeling and non-labeling approaches have been explored. Among the labeling techniques, iTRAQ is the most widely used. In this method, peptides are labeled with isobaric tags that produce low mass reporter ions during MS/MS fragmentation. The abundance of the reporter ions is the most important criteria for reliable and reproducible quantitation. Higher collision energies in HCD cell can be used to generate more abundant iTRAQ reporter ions. However, this can compromise the intensity of sequence ions and as a result decrease the confidence of peptide identifications from this high resolution fragmentation data. On the other hand ion trap fragmentation usually produces best quality spectra for peptide identification. Therefore the two dissociation techniques provide complementary fragmentation data. Combining CID ion trap and collision cell Orbitrap spectra can increase the confidence of peptide identification and provide quantitative information based on the iTRAQ reporter ions. In this study, we examined the benefits of merging of ion trap and collision cell MS/MS spectra for peptide identification and iTRAQ quantitation in complex mixtures.

Methods

Sample Preparation and Sequential LC-MS/MS analysis.

A549 cells (1×10^7) were treated with 5 μ M camptothecin (Sigma), harvested at 0, 2, 8 and 24 hours post treatment, and lysed with 2 mL of lysis/binding/wash buffer (whole cell lysate) containing Halt Protease and Halt Phosphatase inhibitors (Thermo Fisher Scientific). The enrichment of phosphorylated proteins was accomplished with the phosphoprotein enrichment kit (Thermo Fisher Scientific). Whole or PE cell lysates (50 μ g from each time point) were reduced with 5 mM DTT, alkylated with 25 mM iodoacetamide, acetone precipitated, digested and individually labeled with iTRAQTM reagents (Applied Biosystems) as follows: 0 hr - 114; 2 hr - 115; 8hr - 116 and 24 hr - 117. Labeled samples were combined in 1:1:1:1 ratio and applied to PepCleanTM C-18 spin columns (Thermo Fisher Scientific). Phosphopeptides were further purified from 15 μ g of iTRAQ labeled phospho-enriched fractions using a Pierce[®] phosphopeptide isolation kit (Thermo Fisher Scientific) according to manufacturer's instructions. The labeled samples were analyzed using nanoLC-ESI-MS/MS with LTQ Orbitrap XL. An Eksigent[®] NanoLC-2DTM HPLC with a C18 column 75 μ m ID x 15cm was used to separate peptides using a 5-30% gradient (A: water, 0.1% formic acid; B: acetonitrile, 0.1% formic acid) at 200 nL/min over 180 min. The following settings were used for the LTQ Orbitrap XTMLT: FT: MSⁿ Target = 2e5; MS/MS = 2 μ scans, 300 ms max ion time; MS = 400-1500 m/z, 30 000 resolution, MS Target = 5e5; MS/MS = Top Three Data DependentTM acquisition HCD/ Top Three Data Dependent acquisition CID IT; Dynamic Exclusion = Repeat count 2, Duration 30sec, Exclusion duration 120sec; HCD Parameters: Collision Energy = 35%, resolution 7500. MSⁿ Target Ion Trap = 1e4

Data Processing:

Proteome Discoverer 1.0 (Thermo Fisher Scientific) with SEQUEST[®] or MASCOT[™] search engines was used for spectra-preprocessing, protein ID and quantitation based on iTRAQ reporter ions. Mascot software 2.2 (Matrix Sciences) was also used for protein ID based on a probability of 99.9 ($p < 0.01$) and a precursor mass accuracy of 10 ppm. For iTRAQ quantitation significant (95% t-test) normalized median ratios were used. The human IPI database was used for both search engines

FIGURE 1. Workflows Used for the Data Analysis:

- Spectra were normalized and merged before the database search
- Spectra were merged before the database search
- No spectra- preprocessing before the database search

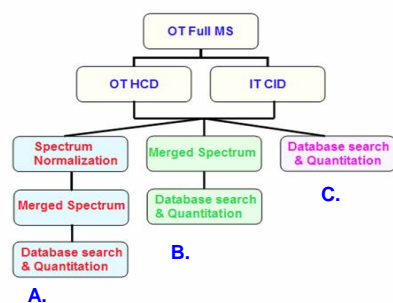


TABLE 1. Results of Data Analysis of A549 Cell Lysate Digest by Workflows Displayed in Figure 1 Using MASCOT Search Engine at FPR<2% (n=3)

Workflow	# ID peptides	# ID proteins	# Quantified Proteins
A	1595	389	329
B	1641	419	326
C	1017	434	341

Results

To evaluate the approach of combining sequence fragment rich IT CID spectra with collision cell spectra containing iTRAQ reporter ions for protein identification and quantitation in complex mixtures, we performed three types of data analysis for iTRAQ labeled A549 whole cell or phosphoprotein enriched fraction digests as shown in Figure 1. Proteome Discoverer software enables spectral merge of fragments from multiple dissociation techniques such as ETD, HCD and CID and different mass analyzers. The same filters/search criteria were used across all data analysis: protein identification data were from FTMS/ITCID runs using MASCOT search engine ($p < 0.01$) and quantitation data were significant normalized median ratios (95% t-test, five replicates LC runs) for proteins identified at FPR<2% with 2 peptides minimum

Table 1 shows results of all employed workflows for iTRAQ labeled A549 cell lysate digest. So while the total number of identified and quantified proteins slightly decreases (from 434(C) to 389(A)), the number of the identified peptides increases substantially (from 1017 to 1641) thus providing better quantitative statistics per protein. Interestingly that spectrum normalization prior merging did not improve the number of protein IDs or quantitation results (workflow A) probably due to differences in signal/noise ratio in LTQ versus Orbitrap spectra. Detailed analysis of the obtained results has shown that spectra merging provides best value when one or both of the original spectra are of insufficient quality to provide confident identification. In this case as demonstrated in Figure 2 for phosphorylated peptide by merging scans the quality of the resulting spectrum increases thus providing more confident peptide IDs and more peptides can be quantified. This observation was confirmed by data analysis of phosphoprotein enriched fraction digest (Table 2). PE fraction contains more low abundant proteins and modified peptides than whole cell digest and as result quantitation is significantly improved after spectral merging and gap in total number of protein IDs is narrow. And as expected no effect of spectral merging was observed for simple nine protein mixture as shown in Table 3. However, in certain cases lower ion scores were obtained for combined spectra than for individual ones that probably explains lower number of total protein IDs in Table 1 (Figure 3). These results demonstrate that improvement is required in the way how we merge the scans from two different analyzers.

FIGURE 2. Comparison MASCOT Search Results for Peptide TAPSFSESR

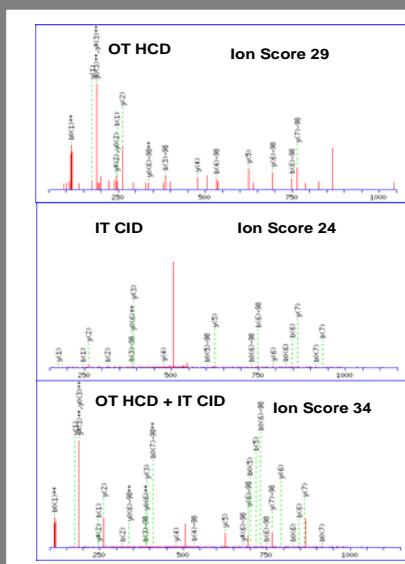


FIGURE 3. Comparison MASCOT Search Results for Peptide HQGVoXMGVoxMGQK

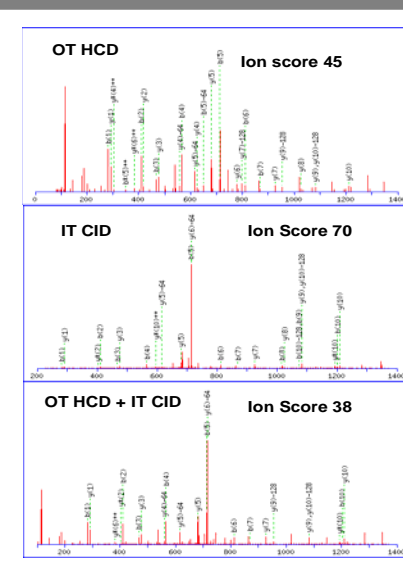


TABLE 2. Results of Data Analysis of PE Fraction of A549 Cell Lysate Digest by Workflows Displayed in Figure 1 Using MASCOT Search Engine at FPR<2% (n=3)

Workflow	# ID peptides	# ID proteins	# Quantified Proteins
B	1334	229	205
C	1074	249	207

TABLE 3. Results of Data Analysis of Nine Protein Mixture Digest by Workflows Displayed in Figure 1 Using SEQUEST Search Engine at FPR<2% (n=3)

Workflow	# ID peptides	# ID proteins	# Quantified Proteins
B	73	9	9
C	71	9	9

Conclusions

- Spectral merging of IT CID and OT HCD increases the number of the identified peptides in average by 30-50% thus significantly improves protein quantitation in complex mixtures.
- Spectral merging does not provide any benefits for general protein identification
- Normalization of absolute spectral intensities before the merge doesn't improve the data quality.
- Scan merging provides complementary fragment ion information and improves both peptide identification and iTRAQ quantitation for low abundant or modified peptides

SIGMA is a trademark of Sigma-Aldrich Co. and its affiliate Sigma-Aldrich Biotechnology. iTRAQ is a trademark of Applied Biosystems. NanoLC-2D is a trademark and Eksigent is a registered trademark of Eksigent Technologies. SEQUEST is a registered trademark of the University of Washington. Mascot is a trademark of Matrix Science Ltd. All other trademarks are the property of Thermo Fisher Scientific Inc. and its subsidiaries.