Advantages of the LTQ Orbitrap for Protein Identification in Complex Digests

Rosa Viner, Terry Zhang, Scott Peterman, and Vlad Zabrouskov; Thermo Fisher Scientific, San Jose, CA

Overview

We have compared the performance of the QSTAR® Elite and the LTQ Orbitrap XL™ for identification and structural characterization of peptides in a highly complex sample mixture.

Introduction

Comprehensive, accurate identification of proteins in complex sample mixtures is an important fundamental capability for any proteomics research laboratory. Technology advancements in both hardware and software continue to expand and refine our view of any proteomic system in terms of protein identities and their posttranslational modifications (PTMs). It has been suggested that the very recent ability to routinely obtain accurate mass measurements (<5 ppm RMS) on precursor and MS/MS fragment ions in proteomic experiments should lead to unprecedented accuracy in the ability to identify and characterize proteins.[1] This paper compares alternative approaches to this challenging application using two high performance platforms for proteomics: a QqTOF instrument (QSTAR Elite from Applied Biosystems) and a hybrid linear ion trap-orbitrap instrument (LTQ Orbitrap XL from Thermo Fisher Scientific).

Methods

Sample Preparation: Ten µL of *E. coli* cell lysate diluted 20 fold with 6 M guanidine HCl in 50 mM ammonium bicarbonate (pH 8.0) was reduced with 5 mM DTT, alkylated with 25 mM iodoacetic acid and digested at 37 °C for 16 hours.
HPLC: Column: C18 Packed tip, 75 µm x 75 mm (QSTAR Elite); C18 column, 75 µm x 100 mm (LTQ Orbitrap XL); Mobile phase A: 0.1% Formic Acid in Water with 2% Acetonitrile; Mobile phase B: 0.1% Formic Acid in Acetonitrile; Flow Rate: 300 nL/min Gradient: 5% B to 35% B in 90 min.
Mass Spectrometry: LTQ Orbitrap XL with nanospray ion source was operated as follows: External calibration (weekly); Top 7 MS^2 (in the ion trap), 1 microscan; Resolution (FT full scan): 60 000 *m/Δm50%;* Monoisotopic toggle: ON; MS^2 threshold: 8000 (~S/N 2:1); MS target: 1e6; MS^2 target: 5e3; Dynamic Exclusion: 60 s; Repeat count: 1; Exclusion mass tolerance: 10 ppm; Charge state selection: 2+, 3+; Injection times (FTMS): 700 ms; Injection times ($ITMS^2$): 100 ms.
The LTQ Orbitrap XL consists of an LTQ XL™ linear ion trap mass spectrometer with an orbitrap high resolution, accurate mass detector. The instrument was operated in a parallel mode that concurrently provided 1) accurate precursor mass measurement (<5ppm) and 2) high sensitivity MS/MS peptide fragment ion spectra from the LTQ XL. QSTAR Elite analysis was performed by an independent 3rd party laboratory. Care was taken to select a lab with a certified, SCIEX®-trained operator to optimize results from the QSTAR Elite. To ensure reproducibility the sample was analyzed several times and the best results from both platforms were used for this study.
Data acquisition parameters for QSTAR Elite with nanospray ion source were as follows: External calibration (daily); Top 5 MS^2; Dynamic exclusion: 90 s; Repeat count: 1; Charge state selection: 2+, 3+; Max MS^2 time: 200 ms; "Smart Exit": ON
Dynamic Background Subtraction feature of Analyst® QS 2.0 software was found to be unreliable and was turned off. The acquisition was optimized for maximum protein IDs.
Database Searching: To facilitate comparison of results, all data was processed with Mascot™ software (Matrix Sciences) using identical databases and the following search parameters: Fixed modifications: Carboxymethyl (C); Variable modifications: Deamidation (NQ), N-Acetyl (Protein), N-Formyl (Protein), Oxidation (M), Methyl (KR), tri-Methylation (KR), di-Methylation (KR), Pyro-Glu (QE); Mass values: Monoisotopic; Protein Mass: Unrestricted; Peptide Mass Tolerance: ± 5 ppm (0.1 Da for QqTOF); Fragment Mass Tolerance: ± 0.8 Da (0.2 Da for QqTOF); Max Missed Cleavages: 2; Instrument type: ESI-FTICR, ESI-QTOF; Databases: Uniprot/reverse uniprot; Data format: mz Data.xml or .mgf (QqTOF); Report top hits: auto
To exploit benefits of vendor specific software, the QqTOF data was also searched with ProteinPilot™ using the following parameters: Modifications: Biological Group with amino acid substitution; Peptide Mass Tolerance: ±0.1 Da; Fragment Mass Tolerance: ±0.2 Da; Instrument type: QSTAR; Databases: Uniprot; Search: Thorough
The Orbitrap data was also searched using SEQUEST® in BioWorks™ 3.3 against an *E. coli* database.

Results and Discussion

The base peak chromatograms from both analyses are qualitatively similar (Figure 1), lending confidence to the comparison. Data from the 100 ng and 500 ng samples were searched with Mascot. Results (p < 0.001, Protein False Positive Rate << 1%) are shown in Figure 2. The Orbitrap identified significantly more proteins in both samples, a total of 315 vs. 125 in the 500 ng sample and 150 vs. 60 in the 100 ng sample. Of all the proteins identified in this study, 31% were found by both instruments, 6% were seen only by QqTOF and 63% were seen only by the Orbitrap.

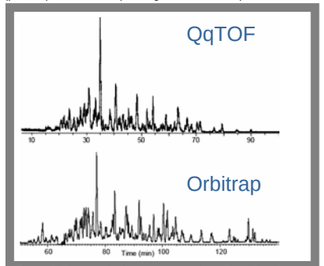


Figure 1: Base peak chromatograms from *E. coli* lysate analyses by QqTOF (top) and Orbitrap (bottom)

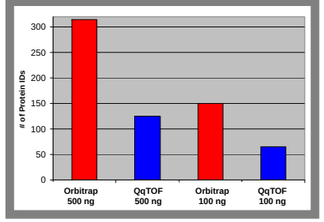


Figure 2: Number of proteins identified in 100 ng and 500 ng *E. coli* lysate samples analyzed by Orbitrap and QqTOF. Database searching using Mascot parameters outlined in the text and p < 0.001, resulted in a protein false positive rate (FPR) less than 1%.

We compared the redundancy rate of peptide sequencing for both instrumental platforms using 500 ng runs. The ratio of [number of distinct peptides/total number of identified peptides] was 43% for the QqTOF and 71% for Orbitrap confirming that the Orbitrap acquired significantly fewer redundant MS^2 spectra. In addition, only at the most stringent Mascot filter (p< 0.001, ion score >25), did the protein false positive rate approach zero (for both types of data). This filter setting was used throughout the experiments to ensure the highest confidence level of protein identification.

Of the highest scoring 125 proteins identified by the QqTOF in the 500 ng sample, the Orbitrap routinely achieved higher sequence coverage for the same proteins. Figure 3 shows that, for the top four and bottom four proteins on the QqTOF list, the Orbitrap identified more peptides to provide higher primary sequence coverage. It is important to note that, in addition to these 125 proteins, the Orbitrap confidently identified an additional 190 low abundance proteins that were not detected by the QqTOF. To assure that the QqTOF data was not disadvantaged by Mascot searching, it was also searched by ProteinPilot. Results from Mascot and ProteinPilot were approximately the same. To match the ProteinPilot results, Mascot result filtering was relaxed to significance threshold of 1% (p< 0.01), with a corresponding increase in the protein FPR to 2.4%.

Mass accuracy is expected to provide several significant benefits in proteomics experiments, including 1) increased confidence in protein database search results, 2) improved ability to assign post translational modifications, and 3) enhanced de novo sequencing capabilities. This study showed clear benefit for database searching and PTM assignment. Note that the benefit of accurate mass for de novo sequencing has been previously shown.[2] The Orbitrap showed a precursor mass error of 2 ppm RMS, compared to QqTOF error of 19 ppm RMS for the top protein hit (60 kDa chaperonin). The high mass accuracy achieved with the Orbitrap mass analyzer has a direct benefit on the quality of protein identifications. During a Mascot database search, the precursor mass accuracy is taken into account during calculation of the ion score. As shown in Figure 4, this dramatically increases the number of proteins that are confidently identified – at the same FPR – as mass accuracy increases.

The effect is similar for QqTOF data, but much less pronounced due to the lower overall mass accuracy achievable with the QqTOF mass analyzer and a larger spread of mass error. Mascot expectation values which reflect the significance of the hit are further improved by the characteristically robust peptide fragmentation seen in ion trap MS/MS spectra. As more predicted fragment ions are matched, Mascot ion scores increase and expectation values drop. Figure 5 shows a typical high sensitivity peptide MS/MS spectrum from the Orbitrap and an MS/MS spectrum of the same peptide from the QqTOF. The Orbitrap spectrum contains both b- and y-ion series, while the QqTOF spectrum has only a y-ion series. Many more predicted ions are matched in the Orbitrap spectrum which significantly improves Mascot scores for these peptides.

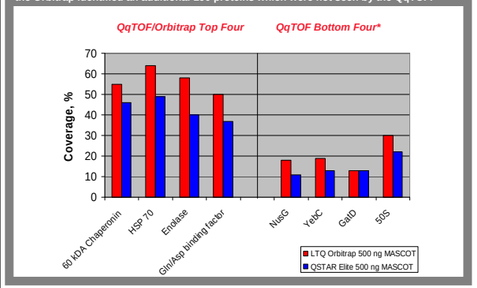


Figure 3: Primary sequence coverage for the top 4 proteins on the QqTOF list and for the bottom 4 proteins. The QqTOF identified 125 proteins in a 500 ng sample of *E. coli* lysate. In all cases, sequence coverage obtained from Orbitrap platform was higher. Importantly, the Orbitrap identified an additional 190 proteins which were not seen by the QqTOF.

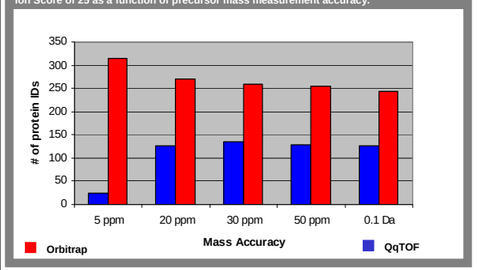


Figure 4: Number of proteins identified by Mascot search with p< 0.001 (FPR << 1%) and Ion Score of 25 as a function of precursor mass measurement accuracy.

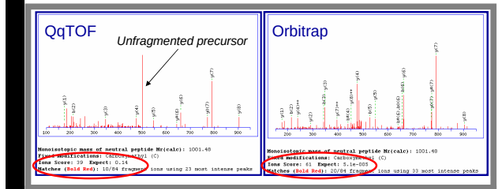


Figure 5: Comparison of MS/MS fragment ion spectra from the QqTOF and the Orbitrap for peptide VLENAEGDR from HSP 70 protein (500 ng runs). The Orbitrap spectrum contains a full b- and y-ion series while the QqTOF spectrum has only a y-ion series and retains some unfragmented precursor. This has a dramatic effect on the Mascot ion score.

A new Paragon™ algorithm (ProteinPilot software) enables database searches which account for many peptide modifications. Without true accurate mass capabilities, however, this routinely leads to mis-assignments and false positive identifications.

Without true accurate mass, however, this routinely leads to mis-assignments and false positive identifications. Fig. 6 shows the example of a confident assignment of an A to V amino acid substitution in EF-Tu protein using ProteinPilot. Upon inspection of the MS/MS spectrum used to make this determination, none of the ions required for actual confirmation of this assignment (y1, y2, b12 and b13) are present. The delta mass for this substitution is 28.0313 Da. Given the absence of confirming ions, this mass could also be explained by di-methylation or other amino acid substitutions, such as K to R. Interestingly, in a subsequent run the same peptide was identified (less confidently) as having a K12 to R12 substitution (Fig. 7) based on the presence of a weak y3 ion. The delta mass for this substitution is 28.0062 Da. The peptide mass difference resulting from A to V conversion (or di-methylation, [28.0313 Da]) and isobaric K to R conversion [28.0062 Da] is only 0.025 Da or 15 ppm at this mass (1630 Da). Therefore the mass analyzer must be capable of measuring with better than 5 ppm mass accuracy (RMS) to confidently distinguish between these two isobaric modifications. While it is not possible to confidently distinguish between K to R substitution and di-methylation of K with the mass accuracy of a QqTOF, this difference can easily be confirmed using an Orbitrap. Fig. 8 shows a Mascot search results of the Orbitrap data that confidently identified the presence of a di-methylated K12 in this peptide, supported by 1.4 ppm mass error for the precursor ion and the presence of b11, b12, and b13 peptide MS/MS ions. This modification in EF-Tu protein is well known and is annotated in the SwissProt database.

Fig. 6: Unsubstantiated assignment of an amino acid substitution using QqTOF and ProteinPilot. ProteinPilot confidently identifies A to V amino acid substitution in EF-Tu peptide AFDQIDNAPEEKAR, even though the y1, y2, b12 or b13 ions required to confirm this assignment are not present in the MS/MS spectrum.

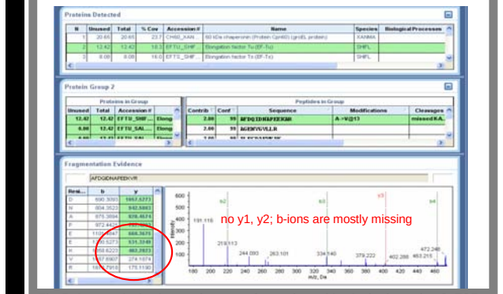
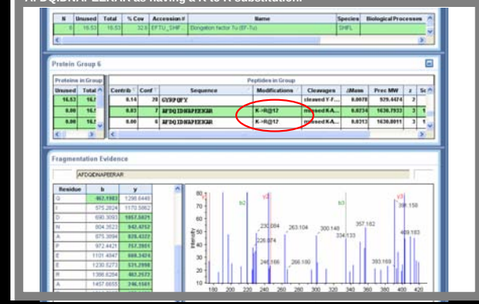


Fig. 7: A second mistaken assignment of an amino acid substitution using QqTOF and ProteinPilot. This MS/MS scan identifies the same EF-Tu peptide AFDQIDNAPEEKAR as having a K to R substitution.



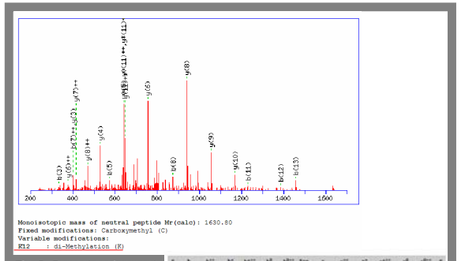


Fig. 8: Correct assignment of the di-methylation present on peptide AFDQIDNAPEEKAR from EF-Tu protein based on the Orbitrap data. Precursor mass error was 1.4 ppm. Presence of b11, b12, and b13 ions confirms (blue circle) the assignment.

To enhance Mascot search results of the Orbitrap data to account for PTMs and non-specific cleavages, it is possible to do an error-tolerant search. For example, error-tolerant Mascot search for the RS6 protein identified three more peptides (resulting from non-specific cleavages), and hence, increased the sequence coverage from 38% to 51%, but it also proposed two possible modifications in HAVTEASPMVK peptide (di-methylation of K and V3 to A3 conversion) based on the MS^2 ions and parent mass accuracy (data not shown due to space limitation). It appears that this is a much more conclusive approach to PTM assignment than the use of QqTOF and ProteinPilot.
Orbitrap data was also searched with SEQUEST against the *E. coli* database. Results were comparable to those achieved by Mascot.

Conclusions

Protein identification and PTM assignment remain critically important experiments in support of biological, drug research, biomarker, and clinical applications of proteomics. In this study, it has been shown that in identical experiments:

1. The Orbitrap is able to identify more than 2.5 times as many proteins as the QqTOF in a complex *E. coli* lysate. The Orbitrap data provided enhanced primary sequence coverage for almost all of the 125 proteins on the QqTOF list.

2. Protein database search results from Mascot, SEQUEST and ProteinPilot search engines appear to be highly similar at FPR << 1%.

3. Parallel acquisition – ability to perform Data Dependent™ MS/MS analyses in the ion trap while accurate mass precursor measurements are performed in the Orbitrap – provides ideal data for proteomics experiments. The combination of accurate precursor mass and robust peptide fragment ion information allow highly sensitive and confident peptide identification and PTM assignment.

4. ProteinPilot made spurious PTM assignments which were not supported by spectral information. In the example cited in this study, ProteinPilot provided two erroneous mis-assignments of the same modification in the same peptide. Using Orbitrap data for the same peptide, it was possible for Mascot to confidently assign the modification based on excellent precursor mass accuracy agreement (1.4 ppm error) and the presence of several confirming MS/MS fragments.

References: **1.** On The Proper Use of Mass Accuracy in Proteomics. Zubarev R. and Mann M. Mol. Cell Proteomics. (2007) 6(3): 377–381. **2.** In Depth Evaluation of Data Acquisition Modes for de novo Sequencing on High-Performance FT-MS Instruments. Scigelova M., Wolfendin G., and Rogers I. ABRF, Tampa, FL (2007).