

High confidence protein identification of ETD and ECD spectra with a new mass list preprocessor

Martin Zeller, Torsten Ueckert, Bernard Delanghe

Thermo Fisher Scientific, Bremen, Germany

Overview

Purpose: Evaluation of a mass list processing tool to remove non-fragment ion entries from mass lists of ETD and ECD spectra before submission to database searches

Methods: Thermo Scientific LTQ Orbitrap XL ETD™ and LTQ FT Ultra™ equipped with ECD mass spectrometers, and Proteome Discoverer software suite were utilised

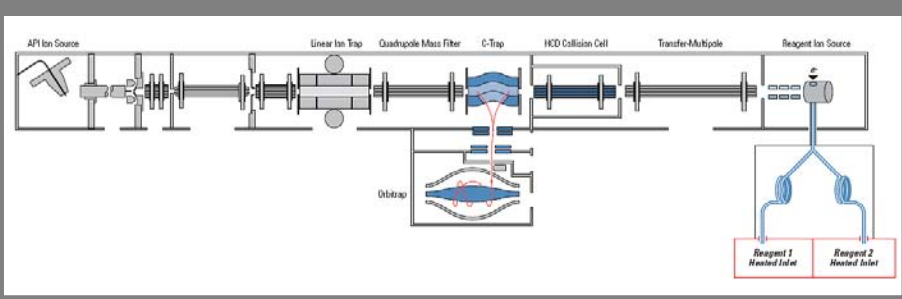
Results: The search results show higher scores, less false positive identifications and more true positive identifications, resulting in more reliable databases search for ETD and ECD spectra after use of the mass list preprocessor

Introduction

Electron Transfer Dissociation (ETD) and Electron Capture Dissociation (ECD) MS/MS spectra typically show two types of ion peaks: true sequence fragment ion peaks and peaks that are less useful and related to the precursor ions. These are typically the un-reacted precursor ions, charge reduced species of the precursor ions and neutral losses thereof. In addition, ECD spectra show peaks for the overtones of the precursor which are in about the same intensity range as the fragment ion peaks. Up to the 6th harmonic peaks can be detected. These non-sequence ion peaks can lead to false positive identifications in database searches that use search algorithms which basically score the experimentally generated spectrum versus a calculated theoretical spectrum (such as Sequest® and Mascot™). With knowledge of the charge state of the precursor ions, these non-sequence ions can be calculated and removed from the mass lists that are generated before submission to database search engines. Instruments with a high resolving power such as the Thermo Scientific LTQ Orbitrap XL ETD and the LTQ FT Ultra can unambiguously determine the charge state of precursor ions typically generated in LC-MS runs of enzymatically digested proteins and thus confidently predict non-sequence ion masses.

We have evaluated in this work the effect of the removal of the non-sequence ion peaks in ECD and ETD spectra from mass lists on the subsequent database search results.

FIGURE 1. Schematic of the LTQ Orbitrap XL ETD.



Results

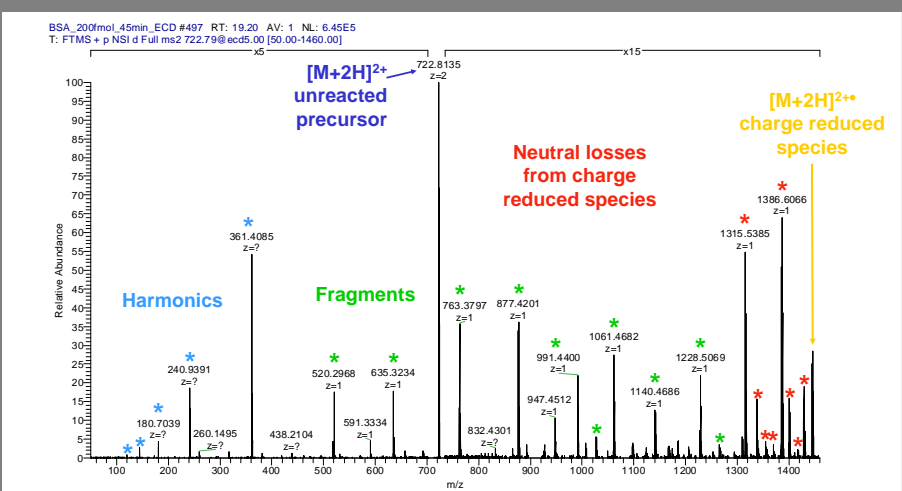
Figure 1 shows a schematic diagram of the LTQ Orbitrap XL ETD instrument. Analyte cations are accumulated in the linear ion trap and precursor ions are accumulated and isolated in the linear ion trap. Fluoranthene anions are produced via chemical ionization (CI) and transferred through the HCD collision cell and the C-trap into the linear ion trap. The ETD reaction takes place in the linear ion trap allowing the detection of the ETD fragment ions either in the linear ion trap or in the orbitrap detector. The fragmentation principle is the same for ECD and ETD spectra, however ETD spectra do not show harmonic peaks in the spectrum. Figure 2 shows a typical ECD spectrum.

With the knowledge of the charge state of the precursor ion, all non-sequence ion peaks can be calculated and removed from mass lists submitted to database search engines. In addition, the neutral losses from the charge reduced species are removed as well.

Neutral loss masses were taken from Ref [1]; the following masses were used for preprocessing.

17.027 Da	NH ₃	44.037 Da	CH ₄ N ₂	74.019 Da	C ₃ H ₆ S
18.011 Da	H ₂ O	45.021 Da	CH ₃ NO	82.053 Da	C ₄ H ₆ N ₂
27.995 Da	CO	46.006 Da	CH ₂ O ₂	86.072 Da	C ₃ H ₈ N ₃
32.026 Da	CH ₂ OH	46.042 Da	C ₂ H ₆ O	99.068 Da	C ₄ H ₉ N ₃
34.053 Da	N ₂ H ₆ (2xNH ₃)	59.037 Da	C ₂ H ₅ NO	101.095 Da	C ₄ H ₁₁ N ₃
35.037 Da	H ₄ NO	59.048 Da	CH ₅ N ₃	108.058 Da	C ₇ H ₈ O
36.021 Da	H ₄ O ₂ (2xH ₂ O)	73.089 Da	C ₄ H ₁₁ N	(131.074 Da)	C ₉ H ₉ N

FIGURE 2. Example of a typical ECD spectrum (BSA, peptide YICDNQDTISSK). Besides the "real" fragment ion peaks the spectrum contains many peaks that are related to the precursor.



To evaluate the performance and the results with the mass list preprocessor we subjected a complex Arabidopsis thaliana digest to LC-MS analysis using ETD. A workflow was generated in Proteome Discoverer to search the data with and without the preprocessor node to compare the results. The Proteome Discoverer software suite is a multi search engine workflow data processing application, targeting peptide and protein identification. It is designed to process complex data sets with different search algorithms and/or dissociation techniques at the same time. It has a false discovery rate determination for each search engine via decoy database searches.

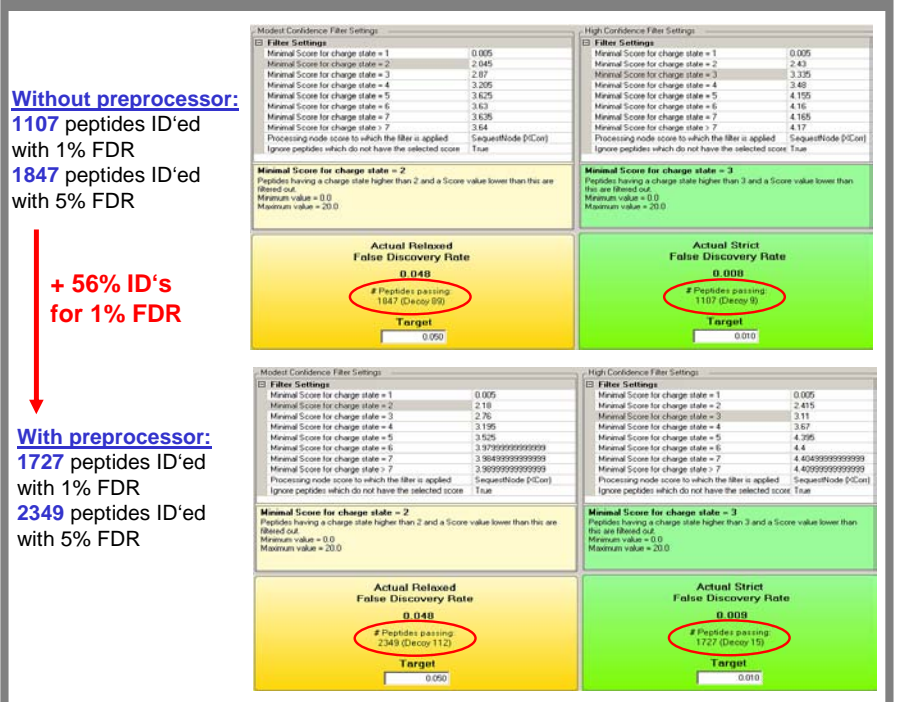
In-depth analysis of the peptide pairs in the search result with and without the spectrum preprocessor showed that the results can be divided basically into four different categories:

- Both ID's of the peptide pairs have high confidence, the score for the peptide with the preprocessor is higher → confident peptide ID
- Both ID's of the peptide pairs have high confidence, the score for the peptide without using the preprocessor is higher → no fragment matched to the precursor-related peaks
- Only the ID of the peptide with the preprocessor is with high confidence → confident peptide ID
- Only the ID of the peptide without the preprocessor is with high confidence → false positive

In almost every case where the score for the peptide without the spectrum preprocessor was higher it was either because the left-over un-reacted precursor ions or the charge reduced species were scored leading to a dubious score or in the worst case to false positive identifications.

The average increase of the Xcorr value using the preprocessor was 0.34. This number does include the peptide pairs where data without the Spectrum Preprocessor received a higher score due to the false assignment of the precursor peak or the charge reduced species.

FIGURE 3. Decoy database searches for 1% and 5% false discovery rate (FDR) with and without the preprocessor node.



Methods

All spectra were acquired on the LTQ Orbitrap XL ETD or LTQ FT Ultra equipped with an ECD cathode. The complex Arabidopsis thaliana samples were separated via Surveyor™ LC equipped with MicroAS™ autosampler (all Thermo Fisher Scientific) using a peptide trap (C18, 100 μm, 2 cm, NanoSeparations) and a C18 analytical column (C18, 75 μm, 10 cm, NanoSeparations), at a flow rate of 250 nL/min. A gradient of 2 – 30% acetonitrile containing 0.1% formic acid in 135 minutes was used.

The LTQ Orbitrap XL ETD performed a full MS scan followed by five Data Dependent™ ETD MS/MS scans with detection of the ETD fragment ions in the linear ion trap. Target values were 5e5 for full FTMS scans, 2e4 for ion trap MSⁿ scans. Anion target value was 1e6. ETD activation time was 90 msec. Supplemental activation was used for all ETD MSⁿ scans.

The LTQ FT Ultra performed a full MS scan followed by three Data Dependent ECD MS/MS scans. Target values were 5e5 for full FTMS scans, 2e5 for FT MSⁿ scans. ECD duration was 70 msec, Energy was set to 5. Charge-state dependent ECD duration was used for all ECD MSⁿ scans. Data analysis was done using the Proteome Discoverer software suite with an additional mass list preprocessor node.

Conclusions

We have shown that we could increase the database search confidence of ETD and ECD data using the Spectrum Preprocessor node in the Proteome Discoverer software suite. The prerequisite of the Spectrum Preprocessor is the ability to unambiguously determine the charge state of the precursor ions. Only instruments with high resolving power such as the LTQ Orbitrap XL ETD and the LTQ FT Ultra are able to determine the charge state of the precursor ions and therefore can benefit from the removal of all non-sequence ion peaks for ETD and ECD analysis.

References

- Cooper, H. J., Hakansson, K., Marshall, A.G., Hudgins, R. R., Haselmann, K. F., Kjeldsen, F., Budnik, B. A. Polfer, N. C., Zubarev, R. A., Letter: the diagnostic value of amino acid side-chain losses in electron capture dissociation of polypeptides. Comment on: "Can the (M(-)X) region in electron capture dissociation provide reliable information on amino acid composition of polypeptides?", Eur. J. Mass Spectrom. 8, 461-469 (2002).

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 and its subsidiaries.