

High Confident 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: Evolution 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: LTQ Orbitrap XL ETD™, LTQ FT Ultra™ equipped with ECD, and Proteome Discoverer Software

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

Introduction

Electron Transfer Dissociation (ETD) and Electron Capture Dissociation (ECD) fragment spectra typically show two types of ion peaks: True fragment ion peaks and peaks that are related to the precursor ions. Those precursor related peaks are the peaks of the un-reacted precursor ions, peaks of the 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 and up to the 6th harmonic peaks can be detected, depending on the *m/z* of the precursor ions and the start *m/z* of the spectrum.

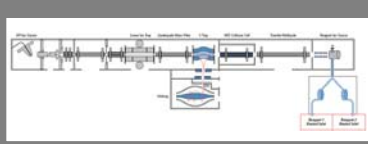
All those non-fragment ion peaks can lead to false positive identifications in database searches that use search algorithms that basically search the experimentally generated spectrum versus a calculated theoretical spectrum (such as SEQUEST® and Mascot™ for example).

However, with the knowledge of the charge state of the precursor ions, most those non-fragment ion peaks can be easily calculated and removed from the mass lists that are generated before submission for database searches.

Instruments with a high resolving power such as the 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 degraded proteins.

We evaluate in this work the effect of the removal of the non-fragment ion peaks in ECD and especially ETD spectra on the database search results.

FIGURE 1. Schematic of the LTQ Orbitrap XL ETD.



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 a Surveyor™ LC equipped with MicroAS™ autosampler (all Thermo Fisher Scientific) using a peptide trap (NS-MP-10 C18 100 µm, 2 cm, NanoSeparations) and a C18 analytical column (NS-AC-07 C18 75 µm, 10 cm), 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 (RP 60,000) followed by five data-dependent™ ETD MS/MS scans with detection of the ETD fragment ions in the linear ion trap. Target values were 565 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.

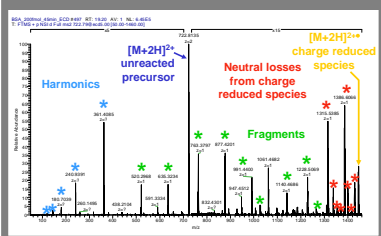
The LTQ FT Ultra performed a full MS scan (RP 50,000) followed by 3 data-dependent ECD MS/MS scans (RP 25,000) Target values were 565 for full FTMS scans, 2e5 for FT MSn scans. ECD duration was 70 msec. Energy was set to 5. Charge-state dependent ECD duration was used for all ECD MSn scans.

Data analysis was done using Proteome Discoverer software (Thermo Fisher Scientific) with an additional mass list preprocessor node.

Results

Figure 1 shows the schematic of the LTQ Orbitrap XL ETD. 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. The fragmentation principle for electron transfer in ECD and ETD spectra are the same, however ETD spectra do not show harmonic peaks in the spectrum. Figure 2 shows a typical ECD spectrum.

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



With the knowledge of the charge state of the precursor ion ions, most non-fragment ion peaks can be easily calculated and removed.

$$m/z_{\text{unreacted precursor}} = m/z_{\text{precursor}} + z$$

$$m/z_{\text{charge reduced}} = \frac{m/z_{\text{precursor}} + z}{2}$$

$$m/z_{\text{neutral loss}} = m/z_{\text{precursor}} - z$$

In addition, the neutral losses from the charge reduced species are removed as well.

Neutral loss masses *k* are taken from Ref [1]; the following masses are currently used:

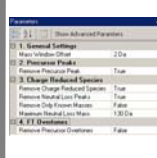
17.027 Da	NH ₃	44.037 Da	CH ₃ N ₃	74.019 Da	C ₂ H ₅ S
18.011 Da	H ₂ O	45.021 Da	CH ₃ NO	82.053 Da	C ₂ H ₅ N
27.896 Da	CO	46.006 Da	CH ₃ O ₂	86.072 Da	C ₂ H ₅ N ₂
32.028 Da	CH ₃ OH	46.042 Da	C ₂ H ₃ O	99.089 Da	C ₂ H ₇ N ₃
34.063 Da	N ₂ H ₄ (2+NH ₂)	59.037 Da	C ₂ H ₃ NO	101.095 Da	C ₂ H ₇ N ₂
35.037 Da	H ₂ NO	59.048 Da	CH ₃ N ₂	108.059 Da	C ₂ H ₇ O
36.021 Da	H ₂ O ₂ (2+H ₂ O)	73.089 Da	C ₂ H ₅ N	1131.074 Da	C ₂ H ₅ N

Table 1 shows that the number of isotope peaks used depending on the (uncharged) molecular mass of the precursor ions. The table was taken from Ref [2].

TABLE 1. Number of Isotope peaks used

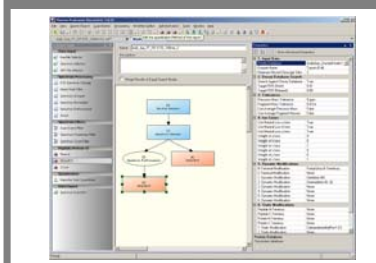
M (uncharged)	Number of Isotope Peaks
< 1000	3
< 2000	4
< 3000	5
< 4000	6
< 5000	7
< 6000	8
< 7000	9
≥ 7000	10

FIGURE 3. Implementation of the Spectrum Preprocessor node in Proteome Discoverer



Figures 3 and 4 show the implementation of the mass spectrum preprocessor node in Proteome Discoverer software.

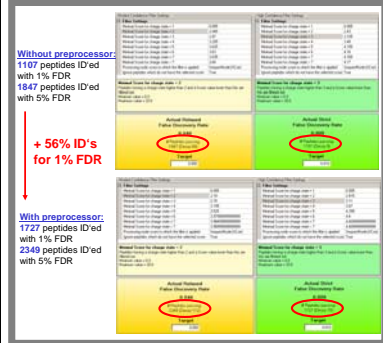
FIGURE 4. Proteome Discoverer workflow to search ETD spectra acquired with the LTQ Orbitrap XL ETD with and without the Spectrum Preprocessor node



To evaluate the performance and the results with the mass list preprocessor we have subjected a complex Arabidopsis thaliana digest to LC-MS analysis using ETD as the fragmentation technique. A workflow was generated in Proteome Discoverer to search the data with and without the Preprocessor node to compare the results (see Figure 4).

Proteome Discoverer Software 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 in a single analysis step (e.g. Data Dependent Decision Tree). It has a false discovery rate determination for each search engine via decoy database search.

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



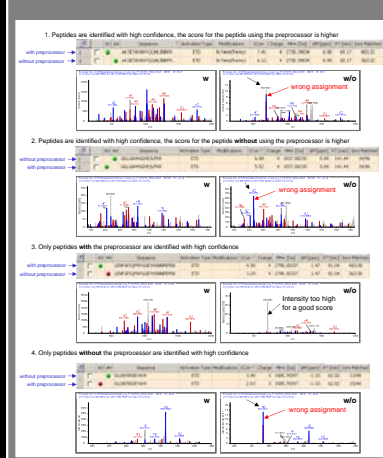
In-depth analysis of the peptide pairs in the search result with and without the spectrum preprocessor showed that the results can be divided into basically four different categories:

- Both identifications of the peptide pairs are with high confidence, the score for the peptide **with** the preprocessor is higher
- Both identifications of the peptide pairs are with high confidence, the score for the peptide **without** using the preprocessor is higher
- Only the identification of the peptide **with** the preprocessor is with high confidence
- Only the identification of the peptide **without** the preprocessor is with high confidence

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

The average increase of the Xcorr value was 0.34. This number does include the peptide pairs where spectra were not evaluated with the spectrum preprocessor received a higher score due to the false assignment of the precursor peak or the charge reduced species.

FIGURE 6. Comparison of the search results of different peptide pairs



Conclusions

We have shown that it is possible to increase the database search confidence of ETD and ECD data using the Spectrum Preprocessor node in Proteome Discoverer software. The prerequisite for application 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-fragment ion peaks for ETD and ECD analysis.

References

[1] Cooper, H. J., Hakansson, K., Marshall, A.G., Hudgins, R. R., Haselmann, K. F., Kjeldsen, F., Budnik, B. A., Poffler, 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).
 [2] http://www.matrixscience.com/help/search_field_help.html

Acknowledgements

We would like to thank Dr. Bernd Müller from the LMU Munich, Germany, for supplying the Arabidopsis thaliana digests.

SEQUEST is a registered trademark of the University of Washington. Mascot is a trademark of Matrix Sciences Ltd. All other trademarks are the property of Thermo Fisher Scientific Inc. and its subsidiaries.