Application Note: 30179

Increasing Proteome Coverage with ETD and CID in One Run using Intelligent Data-Dependent Decision Tree Logic

Martin Zeller, Torsten Ueckert, Bernard Delanghe, Thomas Moehring,
Thermo Fisher Scientific (Bremen) GmbH, Hanna-Kunath-Str. 11, 28199, Bremen, Germany

Key Words

- Electron Transfer Dissociation, ETD
- Data-Dependent Decision Tree, DDDT
- Proteome Discoverer Biosoftware Suite
- Multiple Search Engines

Introduction

With the introduction of electron transfer dissociation (ETD) on the Thermo Scientific LTQ Orbitrap mass spectrometers, ETD is now a widely accepted alternative dissociation technique and used together with the more traditional collision-induced dissociation (CID)[1]. ETD has shown to be complementary to CID and is mainly used to increase proteome coverage and for analysis of labile post-translation modifications (PTM) which are preserved in ETD.

To make use of the complementary nature of CID and ETD, samples would have to be injected twice for each method or CID and ETD spectra have to be generated for each precursor ions which leads to an increase of cycle time.

Josh Coon and co-workers have found in a large-scale study that peptides can have a higher probability of identification in a database search using either CID or ETD depending on the peptide charge state *z* and mass-to-charge ratio (*m/z*)[2]. These rules have been implemented in a data-dependent decision tree (DDDT) logic in the instrument method setup. The aim of the directed fragmentation is to inject a sample only once and the dissociation method that most likely will lead to a positive identification of the peptide will be applied (see Figure 1).

Goal

In this work we describe the method setup of the data-dependent decision tree as well as the workflow for data-mining the mixed raw data containing CID and ETD spectra within the Thermo Scientific Proteome Discoverer software suite.

Data-Dependent Decision Tree – Method Setup

The setup of the DDDT method within the Thermo Scientific Xcalibur method editor in Tune 2.5.5 for LTQ Orbitrap XL ETD™ and Tune 2.6 for LTQ Orbitrap Velos ETD™ is straight-forward. Figure 2 shows the screenshot with the changes in the "Data Dependent Settings" compared to a standard "top N" method.

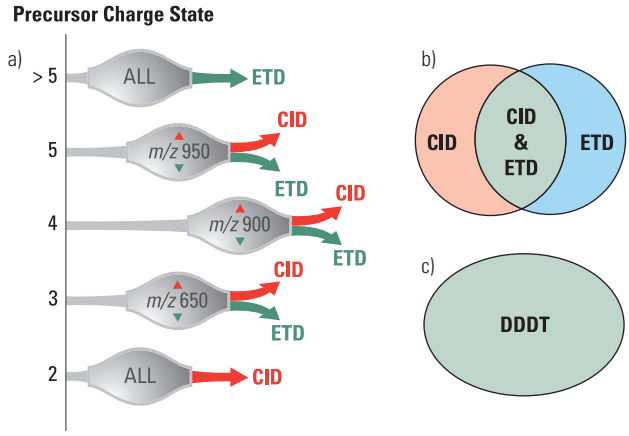


Figure 1:
a) Data-dependent decision tree rules to subject precursor ions either to CID or ETD based on charge state *z* and mass-to-charge ratio
b) Result space of two separate runs using CID and ETD
c) Result space of one run using DDDT.

When ticking the check-box "Use procedure" in figure 2a the new window "Procedures" opens as shown in figure 2b. The *m/z* ratios in the field "Value" are recommended values extracted from a large scale data evaluation by Coon and co-workers[2].

"Activation type" in *Scan Event – Activation* (Figure 2c) should be set to "CID"; this will subject all doubly charged precursor ions to CID. Doubly charged peptide precursor ions fragment very efficiently with CID whereas ETD fragmentation efficiency for doubly charged precursor ions is rather low. This is because one of the two charges is reduced and only one charge remains to charge the generated ETD fragments. These steps have to be repeated for all scan events of the "top N" method.

Data-Dependent Decision Tree – Data Mining

The decision to subject precursor ions for dissociation by either CID or ETD is made "*on-the-fly*" during an LC run, thus resulting in a raw data file that contains both CID and ETD spectra.

a) Procedure b) Values c) Activation

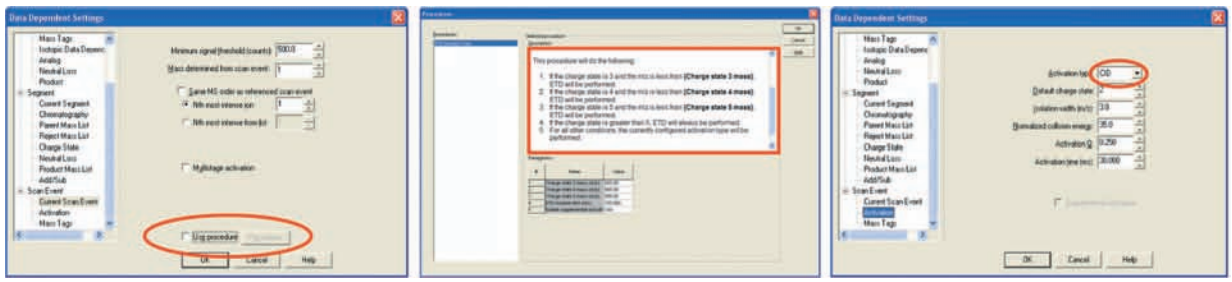


Figure 2: DDDT method in Xcalibur™ Instrument Setup.

CID and ETD spectra cannot be searched with the same search parameter. CID and ETD spectra have to be separated and searched with the appropriated search parameters and/or dedicated search algorithms.

Figure 3a shows the workflow for data analysis of the DDDT data within the Thermo Scientific Proteome Discoverer Biosoftware suite. Proteome Discoverer™ is a multi search engine, workflow data processing application targeting peptide and protein identification as well as quantitation. It is designed to process complex data sets with different search algorithms and/or dissociation techniques in one go. A false discovery rate (FDR) determination for each search engine is implemented via decoy database search.

CID and ETD spectra are separated with the "Scan Event Filter" (node #2 for CID and node #5 for ETD spectra) using the fragmentation type for differentiation (see Figure 3b and 3c). CID spectra are then searched using Mascot™ and SEQUEST® (node #3 and node #4).

ETD spectra are searched using also Mascot and SEQUEST and, in addition, ZCore, which is a dedicated ETD search algorithm. Mascot and SEQUEST are initially designed to search CID data but are recently adopted to cope with ETD data.

Characteristic for ETD fragment spectra are two types of ion peaks: true fragment ion peaks and peaks that are related to the precursor ions. The peaks that are related to the precursor are the peaks of the un-reacted precursor ions, peaks of the charge reduced species of the precursor ions and neutral losses thereof. Most database search algorithms such as SEQUEST and Mascot are designed for the analysis of CID spectra which typically do not contain precursor-related peaks to such extend. These search algorithms basically score the experimentally generated spectra versus calculated theoretical spectra and all those non-fragment ion peaks in ETD spectra can potentially lead to false positive identifications in database searches.

However, if the charge state of the precursor ions is determined, all 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 high resolving power such as the LTQ Orbitrap™ instrument family can unambiguously determine the charge state of precursor ions typically generated in LC-MS runs of enzymatically degraded proteins.

Figure 3a shows the "Non-Fragment Filter" (node #6) between the "Scan Event Filter" to filter for ETD spectra and Mascot (node #7) and SEQUEST (node #8). Figure 3d shows the parameter for the removal of the precursor-related peaks in the mass lists. The benefits of removing the precursor-related peaks are increased search confidence and reduction of false positive identifications[3-6].

Figure 4 shows the database search parameters for the ETD spectra for Mascot, SEQUEST and ZCore.

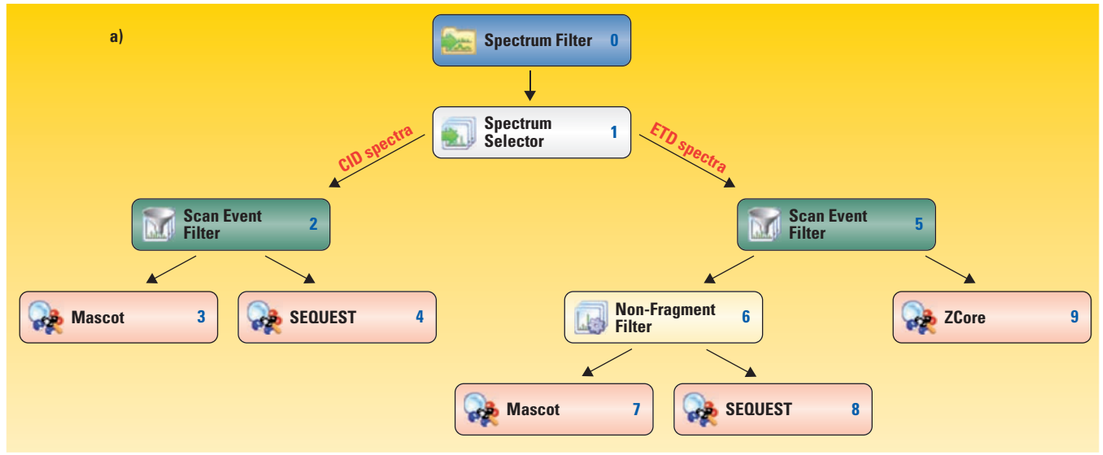
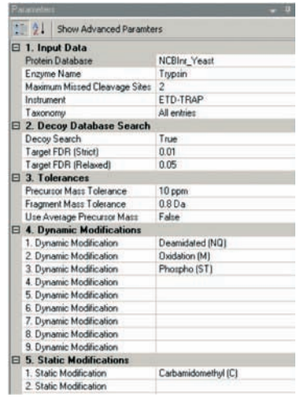
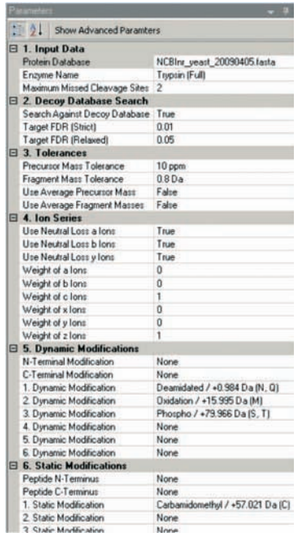


Figure 3: a) Workflow for database search of DDDT data in Proteome Discoverer using multiple search engines; b) "Scan Event Filter" node #2 filters only for CID spectra; c) "Scan Event Filter" node #5 filters only for ETD spectra ; d) Recommended set of parameters for "Non-Fragment Filter".

a) Mascot



b) SEQUEST



c) ZCore

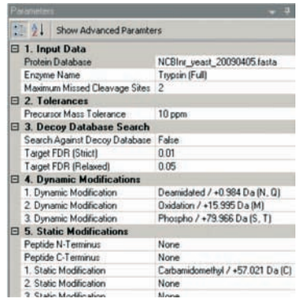


Figure 4: Set of parameter to search ETD spectra with Mascot (a), SEQUEST (b) and ZCore (c).

Data-Dependent Decision Tree – Result Discussion

In this report we compare the results of two runs of a "top 8" CID resp. ETD method with one run of a "top 8" DDDT method (survey scan detected in the Orbitrap™ mass analyzer and eight data-dependent CID resp. ETD MS/MS scans detected in the linear ion trap).

a) Run Time, Number of Spectra etc.

Table 1 summarizes the experimental details such injected sample amounts, run times, MS/MS spectra triggered and search results. For the two runs with CID and ETD the injected sample amounts, run times as well as triggered MS/MS double compared to the run with the DDDT logic.

	CID and ETD	DDDT	Change
Sample amount injected	1,000 ng	500 ng	100%
Total run time	360 min	180 min	100%
MS/MS spectra triggered	32,353	16,900	91%
CID / ETD total	19,951 / 12,402	11,553 / 5,347	73% / 132%
Success rate	14.6% / 17.7%	19.6%	
CID / ETD 2+ precursor ions	9,849 / 12,510	8,862 / -	
Identified peptides	3,847	3,311	16%
Identified proteins	892	810	10%

Table 1: Comparison of CID and ETD method with DDDT method.

For the direct comparison of the CID and ETD run it is important to note that in the "top 8" ETD run also doubly charged precursor ions are fragmented by ETD whereas with DDDT logic they are fragmented by CID. ETD fragmentation efficiency for doubly charged precursor ions is comparably lower and also the optimum ETD reaction time for doubly charged precursor ions is about 50% longer than for triply charged precursor ions which increases cycle time.

b) Database Search Results

Although run time and the number of MS/MS scans double for two runs with CID and ETD, the number of identified peptides and proteins increase only by 16 and 10% (all at 1% FDR) compared to the one run with DDDT logic.

Figure 5 shows the distribution of peptides identified by CID and/or ETD. The number of peptides identified by CID is almost the same for the CID only run compared with the DDDT run although a lot less CID spectra (42% less) have been triggered. This reflects that doubly charged peptides are fragmented very efficiently with CID and the DDDT methods subjects all doubly charged precursor ions to CID.

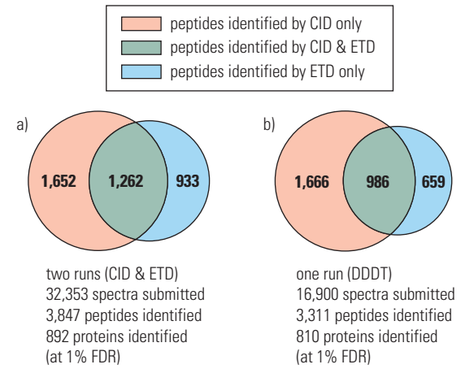


Figure 5: Venn diagrams of the identified peptides using CID and ETD method (a) and DDDT (b).

Also the higher success rate (ratio of identified peptides to number of MS/MS spectra triggered) reflects the higher efficiency of the DDDT logic.

An advantageous strategy to maximize the number of identified peptides/proteins is to search the raw data with multiple search engines. The workflow setup of the Proteome Discoverer software allows searching the raw data with the search engines Mascot and SEQUEST for CID data and Mascot, SEQUEST and ZCore for ETD data.

Figure 6 illustrates the distribution of peptides identified by each search engines. For CID data of tryptic digests, Mascot appears to be slightly superior to SEQUEST since Mascot has a higher identification rate for rather small peptides. For ETD data all three search engines perform equally with ZCore showing similarities to SEQUEST based on the high number of overlapping peptides.

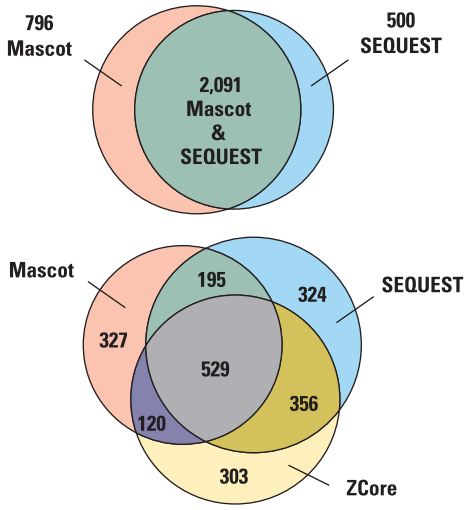


Figure 6: Venn diagram for the number of peptides identified with Mascot and SEQUEST for the activation type CID (top). Venn diagram for the number of peptides identified with Mascot, SEQUEST and ZCore for the activation type ETD (bottom).

Conclusion

We have shown that the data-dependent decision tree method improves peptide and protein identifications compared to separate runs using CID and ETD. This is especially useful when low sample amounts and/or limited instrument time are available. Furthermore, we show that Proteome Discoverer has all the tools that are necessary for data mining of mixed raw files containing CID and ETD spectra.

Acknowledgements

We would like to thank Dr. David M. Good (Karolinska Institute Stockholm, Sweden) for the kind gift of the yeast digests.

Experimental

All spectra were acquired on the LTQ Orbitrap XL ETD. 500 ng of the complex Saccharomyces cerevisiae samples were separated for each experiment via Thermo Scientific Surveyor MS Pump Plus LC equipped with Thermo Scientific MicroAS Autosampler using a peptide trap (C18 Biobasic, 100 µm inner diameter, 2 cm length) and a C18 analytical column (C18 Biobasic, 75 µm inner diameter, 10 cm length, both NanoSeparations, NL), at a flow rate of 250 nl/min.

A gradient of 5 - 30% acetonitrile containing 0.1% formic acid in 130 minutes was used. The LTQ Orbitrap XL ETD performed a full MS scan (RP 60,000) followed by eight Data-Dependent™ MS/MS scans with detection of the fragment ions in the linear ion trap. Target values were 1e6 for full FTMS scans and 1e4 for ion trap MS^n scans. Anion target value was 3e5. ETD activation time was set to 90 msec for charge state 2+. Charge state dependent ETD reaction time was used which reduces the specified ETD reaction time for higher charged peptide precursor ions. Supplemental activation was used for all ETD MSn scans. For the CID method, CID was used as the activation type and ETD analogous for the ETD method. For the DDDT method the "procedure" was used as described in Figure 2.

Data analysis was done using a pre-release version of Proteome Discoverer 1.1 software suite. For all three search engines SEQUEST, Mascot and ZCore, the peptide precursor mass tolerance was set to 10 ppm and fragment ion mass tolerance to 0.8 Da. Carbamidomethylation on cysteine residues was used as fixed modification and oxidation of methionine, deamidation of asparagine and glutamine as well as phosphorylation of serine and threonine as variable modifications. For the Mascot search, "ETD trap" was used as "instrument" to search ETD spectra and "ESI trap" to search CID spectra. The SEQUEST search used the value "1" as weight for c and z ions to search ETD spectra and "1" as weight for b and y ions to search CID spectra. All spectra were searched against NCBI database filtered for yeast proteins containing 26918 entries. The spectra were also search against decoy database using a target false discovery rate (FDR) of 1% for strict and 5% for relaxed conditions. Venn diagrams were generated using PNNL Venn Diagram Plotter v1.3.3250.

References

1 McAlister, G.C., Berggren, W.T., Griep-Raming, J., Horning, S., Makarov, A., Phanstiel, D., Stafford, G., Swaney, D.L., Syka, J.E.P., Zabrouskov, V., Coon, J.J., A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-orbitrap mass spectrometer. *J. Proteome Res.* (**2008**) *7*, 3127-3136.

2 Swaney, D.L., McAlister, G.C., Coon, J.J., Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat. Methods* (**2008**) *5*, 959-964.

3 Good, D.M., Wenger, G.D., McAlister, G.C., Coon, J.J., Post-acquisition ETD spectral processing for increased peptide identifications. *J. Am. Soc. Mass Spectrom.* (**2009**) *20*, 1435–1440.

4 Good, D.M., Wenger, G.D., Coon, J.J., The effect of interfering ions on search algorithm performance for ETD data. *Proteomics, in press.*

5 Zeller, M., Delanghe, B., Ueckert, T., Moehring, T., Workflow for maximizing proteome coverage using CID and ETD. *Proceedings of the 57th ASMS Conference*; Philadelphia, PA, June **2009**, M150.

6 Sweet, S.M., Jones, A.W., Cunningham, D.L., Heath, J.K., Creese, A.J., Cooper, H.J., Database Search Strategies for Proteomic Data Sets Generated by Electron Capture Dissociation Mass Spectrometry. *J. Proteome Res.*, in press

In addition to these offices, Thermo Fisher Scientific maintains a network of representative organizations throughout the world.

Africa-Other
+27 11 570 1840
Australia
+61 2 8844 9500
Austria
+43 1 333 50 34 0
Belgium
+32 53 73 42 41
Canada
+1 800 530 8447
China
+86 10 8419 3588
Denmark
+45 70 23 62 60
Europe-Other
+43 1 333 50 34 0
Finland/Norway/Sweden
+46 8 556 468 00
France
+33 1 60 92 48 00
Germany
+49 6103 408 1014
India
+91 22 6742 9434
Italy
+39 02 950 591
Japan
+81 45 453 9100
Latin America
+1 608 276 5659
Middle East
+43 1 333 50 34 0
Netherlands
+31 76 579 55 55
South Africa
+27 11 570 1840
Spain
+34 914 845 965
Switzerland
+41 61 716 77 00
UK
+44 1442 233555
USA
+1 800 532 4752

www.thermo.com

AN30179_E 08/16C