1000 Proteins per Hour [pph] Identified from an *E. coli* Digest

*Maximizing Protein Identification on the LTQ Orbitrap Velos*

Kai Scheffler, Thermo Fisher Scientific, Dreieich, Germany  
Eugen Damoc, Thomas Möhring, Thermo Fisher Scientific, Bremen, Germany

**Introduction**

Enormous improvements in mass spectrometry technology over the last few years have allowed for ever deepening analysis of complex proteomes. However, many low abundance proteins remain undetected in current large-scale proteomic analyses due to the limits on the rate at which MS/MS spectra can be acquired. To solve this undersampling problem, we developed the Thermo Scientific LTQ Orbitrap Velos, a new hybrid linear ion trap-Orbitrap mass spectrometer that provides faster MS/MS scanning in the ion trap (up to 10 peptide sequences per second), thereby increasing the depth of analysis of complex protein mixtures.

This application note demonstrates and discusses the ability of the newly developed hybrid instrument to identify more proteins, with increased sequence coverage and confidence. Different strategies regarding chromatographic run times and utilizing the instruments various dissociation techniques are compared.

**Instrument Innovation**

The instrument evaluated in this application note is the Thermo Scientific LTQ Orbitrap Velos mass spectrometer equipped with Electron Transfer Dissociation (ETD). It is a hybrid instrument which includes a novel dual-pressure linear ion trap combined with an Orbitrap™ detector as shown in Figure 1.

Significant technological improvements have been implemented in both parts of the hybrid instrument. Changes implemented in the front-end of the hybrid instrument, the Thermo Scientific LTQ Velos ion trap, include: (a) a progressive stacked-ring ion guide (S-Lens) with 5-10x increased ion transmission; (b) a dual-pressure differentially pumped linear ion trap with a higher-pressure cell for improved ion trapping, isolation, and CID efficiencies, and a lower pressure cell for improved resolution and/or scan speed and (c) predictive automatic gain control (AGC) for increased scan speed.

Improvements in the rear portion of the instrument, the Orbitrap mass analyzer section, are: (d) an HCD-collision cell with axial field combined with the C-trap and (e) improved vacuum in the Orbitrap detector chamber for improved analysis of intact proteins.

**Experimental**

**Sample Preparation**

The lyophilized soluble fraction of an *E. coli* whole cell lysate was solubilized in 50 mM ammonium bicarbonate buffer followed by reduction of disulfide bonds with dithiothreitol (DTT) at 56 °C for 1 hr, and alkylation of cysteines with iodoaceticamide for 30 min at room temperature in the dark. The sample was then enzymatically digested for 14 hours with a Lys/Arg specific protease. The digest was split in aliquots and stored at -20 °C until use. Samples were diluted in 0.1% formic acid prior to analysis by reversed-phase nanoHPLC-MS/MS.

**LC-MS Analysis**

The proteolytic digest of *E. coli* was separated by online reversed-phase chromatography for each run using a Thermo Scientific Surveyor MS Pump Plus and Micro AS autosampler with a reversed-phase peptide trap (100 μm inner diameter, 2 cm length) and a reversed-phase analytical column (75 μm inner diameter, 10 cm length, 3 μm particle size, both NanoSeparations, NL), at a flow rate of 250 nl/min. The chromatography system was coupled on-line with an LTQ Orbitrap Velos mass spectrometer equipped with an ETD source.

![Figure 1: Schematic of the Thermo Scientific LTQ Orbitrap Velos mass spectrometer with ETD.](image-url)
Details for the chromatographic and mass spectrometric settings are listed in Table 1 and 2. Dilutions of the injected sample amounts were performed to evaluate the effects of decreasing sample quantity in regards to sensitivity of the instrument as judged upon the total number of proteins identified as well as their sequence coverage.

Data-dependent tandem MS acquisition methods were used for all experiments. Different fragmentation techniques, namely CID, ETD and HCD and combinations thereof by using the data dependent decision tree (DDDT) acquisition method were compared. All runs were performed at least in duplicates. In all experiments the Full MS scans were acquired over a mass range of 380-2000 \( m/z \) with detection in the Orbitrap mass analyzer at a resolution setting of 30,000. MS/MS spectra were acquired in the ion trap except for fragment ions produced using higher-energy collision-induced dissociation (HCD) taking place in the collision cell positioned in axial position next to the C-trap.

Fragment ion spectra produced via HCD were acquired in the Orbitrap mass analyzer with a resolution setting of 7,500. For data-dependent acquisition, the method was set to analyze the top twenty most intense ions when fragmentation was performed using CID (Table 3).
The top ten most intense ions per scan cycle were fragmented in case of ETD spectra and also when the DDDT method was applied. Parameters for the CID/ETD data dependent decision tree were set to perform ETD when the precursor charge state was $z=3$ and $m/z$ of the precursor $< 650$ u, with charge state $z=4$ and $m/z < 900$ u and charge state $z=5$ with $m/z < 900$ u. In all other cases peptide fragmentation was triggered with CID.

For data-dependent acquisition of HCD spectra, the top ten most intense ions were selected for fragmentation in each scan cycle and Full MS as well as fragment ion spectra were detected in the Orbitrap mass analyzer. Exclusion conditions were optimized according to observed chromatographic peak width.

Due to the greater ion transmission of the ion source, a lower injection time was selected for all MS/MS spectra detected in the ion trap: 50 ms compared to 100-200 ms on previous instruments. The activation time for resonance CID fragmentation was 10 ms compared to 30 ms used on the Thermo Scientific LTQ XL linear ion trap, taking advantage of the increased fragmentation efficiency in the higher-pressure cell of the ion trap. For HCD experiments a lower target value of $5e4$ ions was used compared to $2e5$ on previous instruments, taking the advantage of better transmission of fragment ions from the axial field collision cell.

Database Search

The raw data files were searched using Thermo Scientific Proteome Discoverer software v. 1.2 with Mascot™ v. 2.2.1 search engine (Matrix Sciences Ltd., London, UK). The search parameters applied in the database searches are listed in Table 4. The reverse database search option was enabled in Mascot, and all data were filtered to satisfy a false discovery rate (FDR) of 1% or better. All datasets were processed using the workflow feature in Proteome Discoverer™ software. The workflow used to process datasets containing ETD spectra included the node called non-fragment ion filter which aims at the removal of intact precursors and its charged reduced species since they are very often the most abundant peaks in the fragment ion spectra und thus compromise the score.

Protein Annotation

All identified proteins from the duplicate runs from CID, ETD, HCD and DDDT methods were exported from the Proteome Discoverer software as ProtXML format and submitted to the Thermo Scientific ProteinCenter data interpretation software to retrieve relevant biological information from publicly available protein databases including gene ontology references for function, biological process and subcellular location.

Results and Discussion

The LTQ Orbitrap Velos mass spectrometer identifies more than 1,000 proteins per hour [pph]

Figure 2 illustrates the base peak chromatograms from analyses of 1,000 ng, 100 ng and 20 ng of a proteolytic digest of *E. coli* whole cell lysate, separated over 60 min via reversed-phase chromatography on the LTQ Orbitrap Velos mass spectrometer with a split flow from a Surveyor MS pump. The peak intensities in the chromatograms correlate very well with the sample amount analyzed and also demonstrate the moderate complexity of the sample compared to more complex lysates from other organisms such as *C. elegans*² or human HeLa cells.³
For a typical and reproducible 60-minute separation the LTQ Orbitrap Velos mass spectrometer is able to identify over 1,000 proteins from 1 μg of a proteolytic digest of *E. coli* whole cell lysate using a ddTOP20 CID method with acquisition of the full scan in the Orbitrap detector and the fragment ion spectra in the ion trap. Reducing the sample amount by a factor of 10 to a final 100 ng applied on column, the number of identified proteins is only reduced by about 10%. Reducing the sample amount again by a factor of 5 to 20 ng, the number of identified proteins is again reduced by about 10%, resulting in 844 proteins on average as depicted in Figure 3a.

The total number of proteins identified in the duplicate runs for the three sample concentrations are very close and the overlap of the identified proteins was 90% across all concentrations indicating high reproducibility. Identification rate on the peptide levels demonstrates a success rate of identified peptides relative to the total number of CID spectra acquired of ~53% for all concentrations. The percentage of unique peptides identified versus all identified peptides varies between 70% (1 μg load) and 90% (20 ng load) as displayed in Figure 3b.

Figure 2: Base peak chromatograms of 1 μg (top trace), 100 ng (middle trace) and 20 ng (bottom trace) of *E. coli* proteolytic digest loaded on a 75 μm C18 reversed phase column and separated via a 60 min gradient (2-30% ACN).

Figure 3: Identification of unique proteins (a) and identification of unique peptides (b) from 1 μg/100 ng/20 ng of *E. coli* proteolytic digest using the 60 min ddTOP20 CID method. For a discussion on: "False discovery rates of protein identification: A strike against the two-peptide rule." see [1].
The LTQ Orbitrap Velos offers great acquisition speed

The high numbers of identified peptides and proteins using the LTQ Orbitrap Velos system is strongly based on the high acquisition speed that the instrument offers. Figure 4 highlights the speed at which spectra are acquired. Every vertical line in the diagrams represents a scan resp. a spectrum in the chromatogram. The green lines represent the Full MS scans, whereas the orange lines represent the MS/MS spectra. In the ddTOP20 CID method (Figure 4, top) the time elapsed for the acquisition of a Full MS spectrum is 1.13 sec due to the acquisition with 2 μscans and a resolution setting of 30k. The time elapsed to acquire 20 CID MS/MS spectra is 2.36 sec resulting in 3.49 sec for a full scan cycle comprising of 1 Full MS spectrum (Orbitrap detector) and 20 MS/MS spectra (ion trap detector).

In higher energy collision induced dissociation (HCD), a dissociation technique that has proven to be especially useful for de novo sequencing as well as for quantitation using chemical labeling, all scans are detected in the Orbitrap mass analyzer at high mass accuracy and a high resolution setting of choice between 7,500 and 100,000. The cycle time of a ddTOP10 HCD (Figure 4, bottom) is 2.93 sec, comprising of 0.64 sec for the Full MS acquired with 1 μscan and 2.29 sec for 10 MS/MS HCD spectra recorded at a resolution setting of 7,500. These two examples show the higher speed at which CID and HCD spectra can be acquired compared to the LTQ Orbitrap XL mass spectrometer, which is due to the implemented hardware improvements.

Comparing different fragmentation techniques for maximizing protein ID

The LTQ Orbitrap system offers various different fragmentation techniques respectively methods. Whereas CID, PQD (pulsed Q Dissociation), ETD and HCD are fragmentation techniques, DDDT is an instrument method.

The choice which of the available fragmentation methods is most suitable for the analysis very much depends on the complexity and the amount of sample available for repetitive runs, the protease used for digestion, and last but not least the aim of the analysis such as purely protein ID or identification combined with quantitation, de novo sequencing or the analysis of post-translation modifications. In this current study the aim was to identify the maximum number of proteins using a standard digest with a K,R-specific endoprotease.

Figure 5 summarizes the results regarding peptide and protein identification resulting from four of the possible fragmentation modes run as ddTOP20 CID, ddTOP10 DDDT, ddTOP10 ETD and ddTOP10 HCD. All methods were run in duplicate runs with 1 μg and 100 ng sample amount each with a separation of the peptide mixture over the course of 90 min. The numbers presented in the diagram are the average of the duplicate runs. All methods applied show a very high number of unique peptides as well as unique proteins identified with CID leading the field with an additional ~5% of proteins identified compared to the DDDT method. Overlap of proteins identified with all four methods is about 80% (Figure 6).
The method using HCD for fragmentation also provided more than 1,000 proteins for 1 µg of sample, demonstrating a much improved success rate of identifying peptides and proteins compared with the LTQ Orbitrap XL mass spectrometer, which is also due to an improved HCD cell resulting in a higher spectral quality.

The number of protein IDs with ETD is also very high yet lower compared to all other methods applied. This was as expected considering the nature of the sample consisting mainly of smaller peptides resulting in predominantly doubly charged peptides which provide better fragment ion spectra quality under CID and HCD conditions.

The location and biological function of all proteins identified from the proteolytic digest of \textit{E. coli} whole cell lysate in this study is detailed in Figure 7.
(a) Molecular Functions

- Antioxidant activity: 9
- Catalytic activity: 1021
- DNA binding: 211
- Enzyme regulator activity: 26
- Metal ion binding: 382
- Motor activity: 1
- Nucleotide binding: 394
- Protein binding: 330
- Receptor activity: 16
- RNA binding: 132
- Signal transducer activity: 44
- Structural molecule activity: 72
- Transcription regulator activity: 121
- Translation regulator activity: 19
- Transporter activity: 175
- Unannotated: 161

(b) Cellular Component

- Cell surface: 1
- Chromosome: 13
- Cytoplasm: 657
- Cytoskeleton: 7
- Cytosol: 53
- Endoplasmic reticulum: 6
- Extracellular: 43
- Membrane: 312
- Mitochondrion: 23
- Nucleus: 32
- Organelle lumen: 3
- Proteasome: 1
- Ribosome: 67
- Unannotated: 558

(c) Biological Processes

- Cell communication: 87
- Cell death: 7
- Cell differentiation: 8
- Cell division: 42
- Cell growth: 2
- Cell motility: 2
- Cell organization and biogenesis: 158
- Cell proliferation: 3
- Cellular homeostasis: 33
- Coagulation: 5
- Conjugation: 4
- Defense response: 6
- Development: 1
- Metabolic process: 1131
- Regulation of biological process: 225
- Reproduction: 20
- Response to stimulus: 178
- Transport: 254
- Unannotated: 205

Figure 7: Gene Ontology annotation of all proteins identified from *E. coli* digest by CID, ETD, HCD, and DDDT: (a) molecular functions, (b) cellular component, (c) biological processes.
Conclusion
The identification of proteins from complex mixtures requires high sensitivity as well as high acquisition speed. The technological advancements implemented in the LTQ Orbitrap Velos mass spectrometer with more than twice the scan speed of an LTQ Orbitrap XL mass spectrometer, help to overcome undersampling, increasing the depth of the analysis of complex proteomes. The enhanced sensitivity and improved duty cycle of the new instrument enables a dramatic increase of the number of unique peptide IDs, which translates into an increased number of proteins identified with higher sequence coverage and confidence. The LTQ Orbitrap Velos mass spectrometer
- provides an increased scan speed in MS and MS/MS mode due to its higher transmission ion source, dual pressure ion trap and improved HCD cell.
- identifies more than 1,000 proteins per hour from a proteolytic digest of E. coli whole cell lysate.
- provides a higher HCD scan speed and much improved spectral quality opening up new perspectives for de novo sequencing, quantitation of chemically labeled samples, and analysis of modifications which require high mass accuracy.
- offers an increase in experimental throughput, confidently identifying more peptides and proteins than any other mass spectrometer currently on the market.

Acknowledgements
We would like to thank Christian Ravnsborg for providing the results from the ProteinCenter™ data interpretation software and Eduard Denisov for his support and helpful discussions.

References

www.thermoscientific.com
Legal Notices: © 2010 Thermo Fisher Scientific Inc. All rights reserved. MAESTRO is a trademark of Matrix Sciences Ltd., UK. All other trademarks are the property of Thermo Fisher Scientific Inc. and its subsidiaries. This information is presented as an example of the capabilities of Thermo Fisher Scientific Inc. products. It is not intended to encourage use of these products in any manners that might infringe the intellectual property rights of others. Specifications, terms and pricing are subject to change. Not all products are available in all countries. Please consult your local sales representative for details.