Optimizing Data Acquisition for Automated *de novo* Sequencing

Iain Rogers[1], Michaela Scigelova[2], Gary Woffendin[2]

[1]Bioinformatics Solutions, Canada; [2]Thermo Fisher Scientific, UK;

Overview

Purpose: We have evaluated various data acquisition methods on the LTQ Orbitrap to obtain data most suitable for automated de novo sequencing. We then reviewed the examples where an integrated approach using de novo sequencing and blast search indicated an amino acid substitution or sequence variation in analysed peptides.

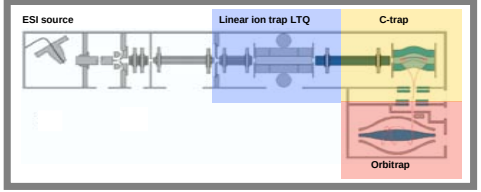
Methods: Medium complex peptide sample (6 proteins digest) was analysed by nano-LC-MS/MS using three different methods combining various fragmentation modes (linear ion trap vs C-trap) and levels of accuracy of fragment ion detection. De novo sequencing was performed with PEAKS Studio 4.2.

Results: Higher collisional energy dissociation in the C-trap provided the best quality data for automated de novo sequencing (no low mass cutoff, fragments measured with low ppm accuracy). Several sequence variations on known peptides were identified with the PEAKS sw, increasing protein coverage.

Introduction

De novo sequencing enables identification of peptides and proteins from unsequenced genomes [1,2] or validation of the results of a database search [3]. To be of practical use this process must be automated, with a throughput matching that of data acquisition.
The LTQ Orbitrap™ delivers routine mass measurements with deviations of less than 3 ppm (external calibration). In the context of proteomics experiments measuring the precursor ion highly accurately means fewer false positive identifications [4, 5].
There is, however, no clear consensus regarding the benefit of MS/MS mass accuracy. This is because the accurate mass detection in the Orbitrap analyser takes longer than the fragment detection in a linear ion trap, resulting in potentially less peptides being fragmented and identified. Also, the LTQ Orbitrap can fragment peptides in the linear ion trap or in the C-trap, each method being characterised by particular spectra qualities (Figure 1).
We performed a detailed comparison of data acquisition methods on LTQ Orbitrap with respect to their suitability for automated *de novo* sequencing with PEAKS™ Studio 4.2 software. As this package combines de novo sequencing with BLAST searches in databases we were also interested in indications of amino acid substitutions or unexpected modifications.

FIGURE 1. The diagram of LTQ Orbitrap mass spectrometer showing the key components: the linear ion trap LTQ, C-trap, and Orbitrap mass analyser. These mass analysers can operate independently of each other or in a convenient parallel mode, maximising the duty cycle. CID fragmentation is carried out in the linear ion trap, and the fragments can be detected either in the linear ion trap or Orbitrap. HCD fragmentation is performed in the C-trap and the fragments are detected in the Orbitrap (inset).

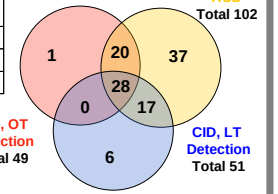


Methods

Peptides in a six protein digest mix (Dionex; 100 fmol) were separated via Surveyor™ LC equipped with MicroAS™ autosampler (Thermo Fisher Scientific) using a peptide trap column (PSDVB, Michrom Bioresources) and a PicoFrit™ nanocolumn packed with 5 µm BioBasic™ C18, 300 Å pore size (New Objective), flow rate 300 nL/min. A gradient of 5 - 40% acetonitrile in 40 minutes was used. The tandem mass spectrometric analysis was performed on the LTQ Orbitrap (Thermo Fisher Scientific). The so called 'Top3' acquisition routine was used, comprising a full MS scan followed by 3 Data Dependent™ MS/MS scans. All full scan spectra were recorded at resolving power 30,000 in the Orbitrap analyzer. The MS/MS spectra (fragmentation and detection) were obtained as detailed in Figure 2 (inset).
The database search (BioWorks™ 3.3.1) was performed against the Uniprot-SwissProt database (containing 216,381 entries), considering fully tryptic peptides and carboxyamidation on Cys (+ 58.0055 Da) as a fixed modification. The precursor mass tolerances for all experiments were 10 ppm; the fragment tolerances considered for Orbitrap-measured MS/MS spectra were 10 mmu, whilst ion trap spectra were searched with 1 Da tolerances.

FIGURE 2. An overview of the acquisition methods used (inset). HCD spectra generated in the C-trap delivered the highest number of **complete correct *de novo* sequences** from among the three methods tested.

Precursor Scan	Fragmentation Method	MS/MS Detection
Orbitrap	CID (LTQ)	LTQ
Orbitrap	CID (LTQ)	Orbitrap
Orbitrap	HCD (CTrap)	Orbitrap



The results were filtered with 'Xcorr vs charge state' filter so as to ensure less than 3% false positive rate (FPR) of identification established using reverse database searches.
These results were used to benchmark the outcome of *de novo* sequencing with PEAKS Studio 4.2. The mass tolerance set for Orbitrap-measured spectra (both precursor and fragment ions) was 0.01 Da while 1 Da was used for the ion trap-measured fragments. The result was considered correct if the complete sequence matched the one obtained in the database search, while allowing for the following ambiguities: Leu = Ile; GlyGly = Asn. In the particular case of fragment masses measured with the LTQ: Gln = Lys, and one possibility of an amino acid positional interchange was still admissible, i.e. NA = AN.

FIGURE 3: Qualitative MS/MS spectral differences. CID spectra contain prominent y- and b-ion series, while HCD spectra tend to show a dominant y-ion series and presence of immonium ions.

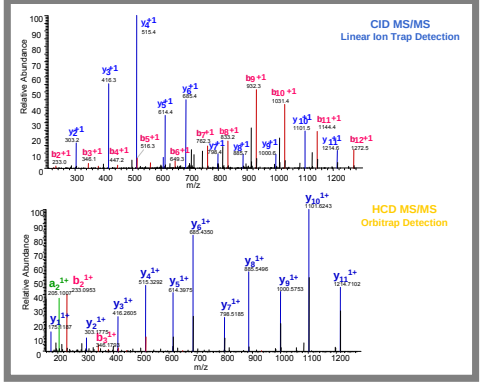
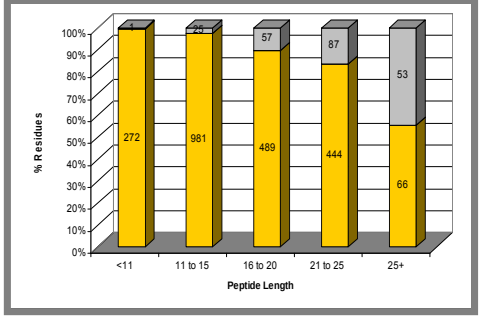


FIGURE 4. Effect of peptide length on the outcome of automated *de novo* sequencing. 98% residues were correctly assigned for peptides 15 or fewer amino acids long.



Results

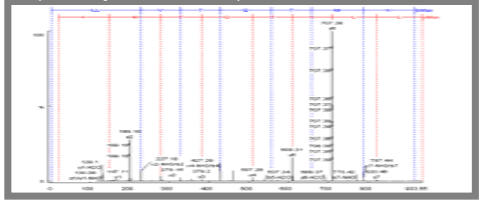
The database search identified confidently (less than 3% FPR) 158, 108, and 167 spectra from the CID LTQ Detection, CID Orbitrap Detection, and HCD Orbitrap Detection experiments, respectively. These spectra were *de novo* sequenced with PEAKS.

The MS/MS spectra obtained with higher energy collisional dissociation (HCD) in the C-trap differed from those generated by standard collision-induced dissociation (CID) in the ion trap (Figure 3).

The results were first evaluated considering the number of complete correct sequences (Figure 2).

HCD data (167 MS/MS spectra) were further evaluated considering the assignment of individual amino acids to a correct position within the peptide. In total, 2252 of 2475 amino acids were correctly assigned which represents 91% of all residues. The successful outcome of *de novo* sequencing was clearly dependent on the length of the peptide. The program delivered a correct assignment for 98% of residues in peptides with 15 or less amino acids (Figure 4).

FIGURE 5. HCD spectra with fragments measured accurately in the orbitrap enable a confident assignment of amino acids. This example shows peptide LLVTQTMK with internal glutamine residue that can not be mistaken for a lysine even though the mass difference would be just 38 mDa.

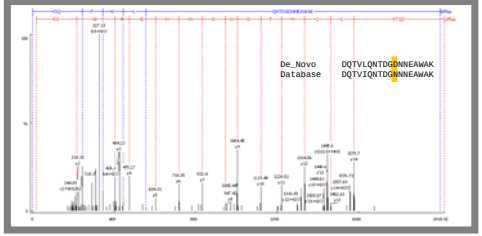


The high mass accuracy of the HCD fragmentation spectra eliminates de novo sequencing errors that result from the fact that masses of certain combinations of amino acid residues are very similar in mass. For instance, it is easy to resolve the difference between K and Q (128.172 x 128.130 Da; Figure 5), or MM and YV (262.081 and 262.131 Da).

While the standard search with BioWorks was able to identify all six of the proteins in the sample, there remained a considerable number of spectra (335 out of 638 total good quality MS/MS spectra) unassigned after a database search. The identified proteins' sequences and reversed sequences were used as the reference database for SPIDER algorithm which is part of PEAKS software package.

SPIDER homology search enabled us to assign further 85 MS/MS spectra which belonged to the 6 proteins analysed. The variations in the sequence included modifications like deamidation (Figure 6).

FIGURE 6. Homology search with SPIDER highlighted many cases of peptide modification or amino acid substitution, like this example of deamidation.



Conclusions

De novo sequencing requires the best data quality available.

The LTQ Orbitrap method employing higher energy collisional dissociation (HCD) performed in the C-trap generates data which is well suited for automated *de novo* sequencing.

The C-trap MS/MS spectra are detected in the Orbitrap analyser. The high resolution is very useful for confident charge state assignment of fragment ions from higher charge state precursors (we observed correct sequence assignment for several 4+ peptides).

HCD data are characterised by 3 ppm accurate mass measurement of fragment ions, prominent and mostly complete y-ion fragment series, presence of low m/z fragments and immonium ions.

Success rate of automated *de novo* sequencing expressed as correct amino acid residue assignment was 91% and increased to 98% for peptides up to 15 amino acid long.

Homology search using SPIDER in PEAKS helped increase protein sequence coverage and identified large number of modifications/amino acid substitutions.

The data of this study were used for further improving the performance of PEAKS Studio 4.2 software with ion trap MS/MS data. See poster xxx at this meeting.

References

Sunyaev et al., **Anal. Chem.**, 2003, 75, 1307-15
Frank et al., **J. Proteome Res.**, 2005, 4, 1287-95.
Wielsch et al., **J. Proteome Res.**, 2006, 5, 2448-56.
Yates, J.R. et al., **Anal. Chem.**, 2006, 78, 493-500.
Zubarev, R., Mann, M., **Mol. Cell. Proteomics**, 2007, *in press*.