

Nanoelectrospray and real-time database searches

Juri Rappsilber, Jens S. Andersen, Peter Mortensen*, Kenneth Budin*, Carsten Pedersen*, Matthias Mann

Protein Interaction Laboratory, Dept. of Biochemistry and Molecular Biology
University of Southern Denmark/Odense, Campusvej 55, DK-5230 Odense M, Denmark
*Protana, Staermosegaardsvej 16, DK-5230 Odense M, Denmark

Introduction

Interpretation of mass spectrometric data and identification of proteins are normally done subsequent to the acquisition of data. Information obtained at this stage needs to be sufficient to search databases as the analysis is finished. Yet, sequences from EST databases are short and error prone whereas genome sequences are of high quality but interrupted by non-coding sequences. In both cases often only single peptides can be found in the initial analysis. This gives one example where real time identification can be of advantage to direct sequencing and to increase the information obtained from a peptide sample. Directed sequencing means that already during an MS experiment spectra are analyzed and the obtained information is used to search sequence databases. Prediction tools are applied to analyze retrieved entries. This information is used in turn to guide the ongoing MS experiment towards particular masses for MS/MS. This allows to confirm the initial identification, distinguish which of several possible sequences belongs to the sample and find biologically crucial peptides.

Application of directed sequencing

Confirm	Distinguish	Find
<ul style="list-style-type: none">• identification of minor component (figure 1)• identification of EST• identification of genomic entry• prediction of overlapping ESTs found by sequence search (BLAST)• prediction of gene exons• prediction of exon boundaries• prediction of post-translational modification	<ul style="list-style-type: none">• very close homologes• sequence conflict in the database• alternative splice forms	<ul style="list-style-type: none">• N-terminal peptide• C-terminal peptide• modifications

Methods and Instrumentation

Proteins were separated by one and two dimensional gel electrophoresis. Bands were excised and in-gel digested. Peptide mixtures were analysed in a two-layered approach. First MALDI TOF spectra were acquired on a Bruker Reflex III with delayed extraction. Then fragmentation spectra were recorded on a PE Sciex QSTAR equipped with a nanoelectrospray source from MDS-Protana. Data interpretation and database searches were done within the Protein and Peptide Software Suite (PPSS) from MDS-Protana using a computer cluster.

Conclusions

- During the ongoing acquisition of MS data spectrum interpretation and database searches allow to direct the analysis to find, confirm and distinguish information on the sample.

- Nanoelectrospray of peptide mixtures has unique advantages for such a directed sequencing approach because any peptide can be selected for sequencing at any time during the experiment - even if the peptide signal was not visible in the primary MS spectrum.
- The interpretation time has to be used efficiently as the analysis time is limited. Full program integration is required.
- Searches have to be done in NRDB first but if no protein can be identified also in EST and genomic databases to use efficiently the available resources. This requires powerful software to deal with the massive databases and the fragmented information.

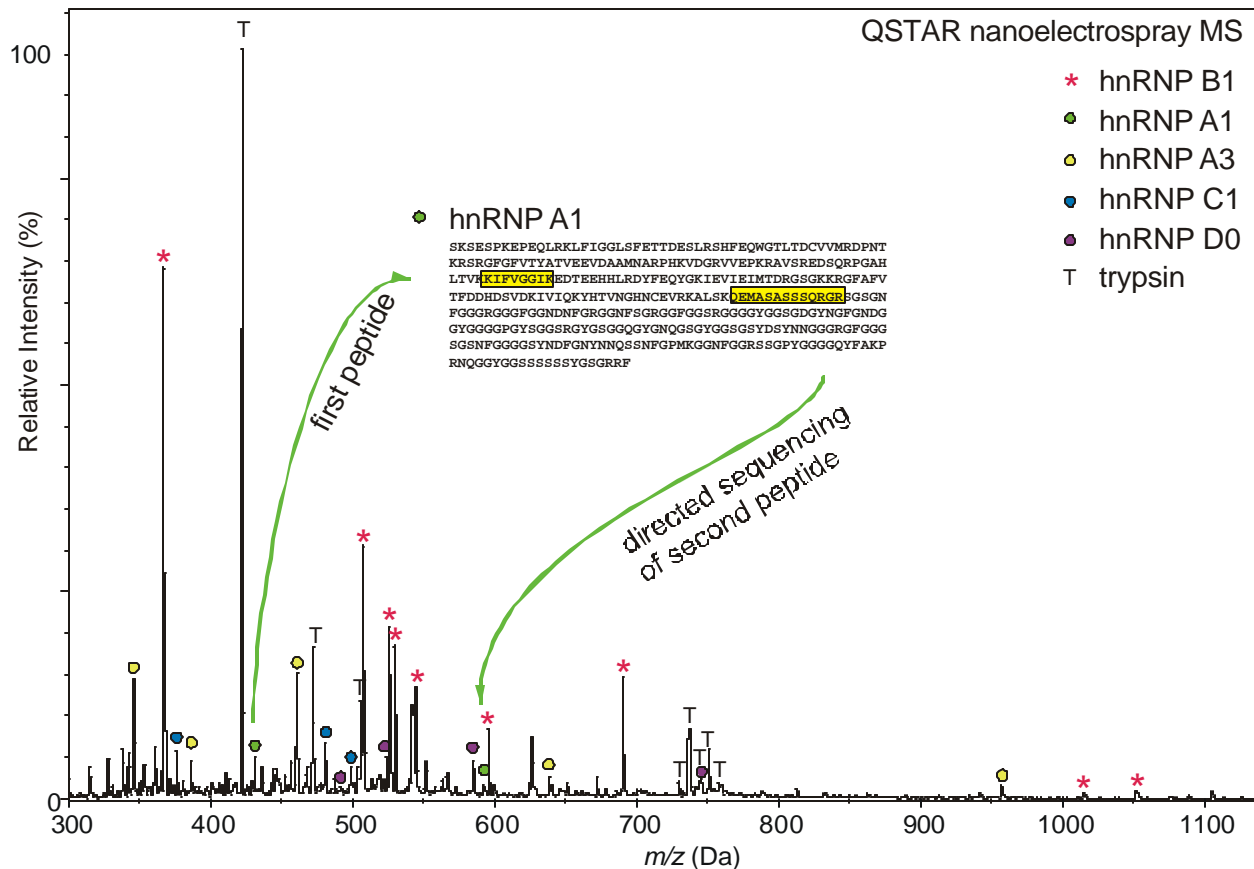


Figure 1: Nanoelectrospray mass spectrum of in-gel digests of a protein band containing a mixture of proteins.

The mixture analysis is done in two steps

Step one: identification and exclusion of major components

MS/MS data obtained from the most abundant peptides are used to identify the main components of the mixture. Tryptic peptides of those proteins are predicted and excluded to focus the analysis on other peptides.

Step two: identification and confirmation of minor components

Peaks in the primary mass spectrum left after step one are selected for MS/MS. If a new component of the mixture can be found, peptides are predicted and displayed in the primary mass spectrum to select a second peak for confirmation of the identification.