Strategies for the efficient use of Fourier Transform Ion Cyclotron Mass Spectrometry

Andreas Römpp[1,2]; Marco Konijnenburg[2]; Ioana Taban[2]; Mark C. Duursma[2]; Erika Amstalden[2]; Vinh An Thieu[1];
Dieter Kirsch[1]; Theo Luider[3]; Bernhard Spengler[1]; Ron M. A. Heeren[2]

[1] Justus Liebig University Giessen (JLU), Giessen, Germany; [2] FOM-Institute for Atomic and Molecular Physics (AMOLF), Amsterdam, The Netherlands;
[3] ErasmusMC (EMC), Rotterdam, The Netherlands

Overview

Purpose:
• analyze large data sets
• use full information of mass spectra generated by FTICR
Methods:
• parallel processing in a distributed computing environment
• composition based sequencing (CBS), remote control
Results:
• significant increase in processing speed for biomarker identification
• identification of peptides/proteins that are not included in databases

Introduction

The high mass resolution and mass accuracy of modern Fourier Transform Ion (FTICR) mass spectrometers offer new possibilities for the analysis of complex samples. At the same time the resulting wealth of information poses new challenges for efficient data handling and analysis. Software tools have to be able to handle very large data sets and make full use of the mass spectral information offered by FTICR-MS measurements. We present strategies to use this promising technology in a flexible and efficient way.

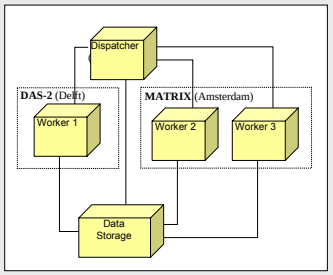
Methods

• CSF samples from breast cancer patients with/without leptomeningeal metastases
• nano-HPLC system from LC Packings
• heavily modified Bruker APEX FTICR mass spectrometer (7T magnet) at AMOLF
• Finnigan LTQ-FT (linear ion trap and FTICR) mass spectrometer at JLU
• several computer setups ranging from a single PC to supercomputers with up to 100 CPUs

Parallel Processing of nanoLC-FTICR MS data (CSF samples)

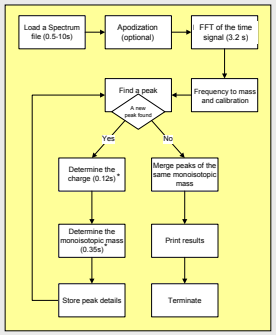
NanoLC-FTICR MS experiments result in about 800 spectra corresponding to up to 4 GB of mass spectral data. In order to speed up the analysis of this data the spectra are analyzed in parallel on several computers.

Scheme of distributed computing (GRID)



A central dispatcher distributes the jobs (one spectrum to be analyzed) to the workers (computer that processes the data). In contrast to a classical computer cluster the workers in a GRID environment can run different operating systems and can be distributed between different geographical locations.

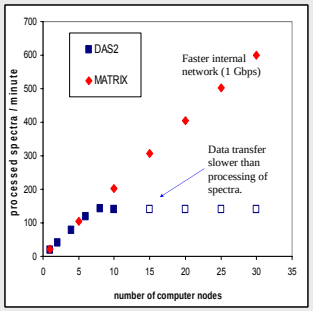
Flow-chart of worker program



* based on THRASH algorithm [1]

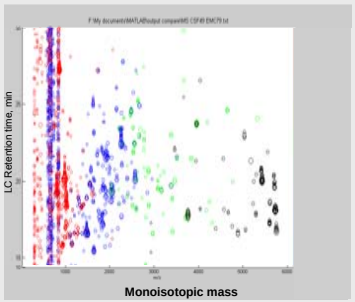
The program is written in Java and thus independent of the operating system. The modular setup allows for easy integration of new processes and adaptation to new data formats.

Increase in processing speed



Data transfer is the limiting step in distributed computing. This is expected to be less of a problem in the future as the network speed is projected to grow much faster than processing speed. Total processing time was reduced from more than one hour to less than 2 minutes.

CSF sample from breast cancer patient

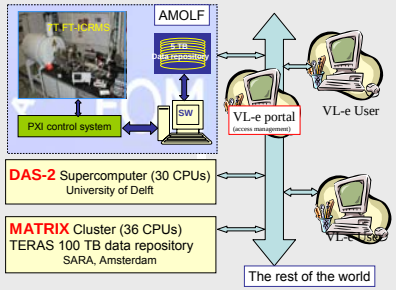


PEPTIDE SEQUENCE	Corresponding protein
ANRPFLVFIR	Anti-thrombin III
MEEVEAMLLPETLKR	α-1-Antichymotrypsin
VTSIQDWVQK	Haptoglobulin
WQEEMELYR	Apolipoprotein A1
EGYYGYTGAFR MYLGYEYVTAIR	Serotransferrin

The peptides indicated in the figure above were found to be upregulated in CSF samples from breast cancer patients with leptomeningeal metastasis (brain tumor). [2]. Identification is based on MSMS experiments and SEQUEST database search. Several others peptides have been tentatively identified.

Virtual Laboratory for e-Science (VL-e)

Set-up of remote experimentation

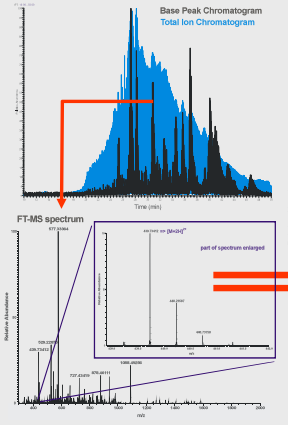


The Virtual Laboratory for e-Science provides a platform for cooperation between laboratories at different locations (www.vl-e.nl). Access to unique/expensive analytical instruments, software tools and computing resources are shared through the internet. The FTICR MS at AMOLF can be controlled from any computer that has access to the world wide web (provided that the user has the necessary access rights). Only the graphical representation of the data is shown at the remote site, thus greatly reducing the data transfer rate and processing time.

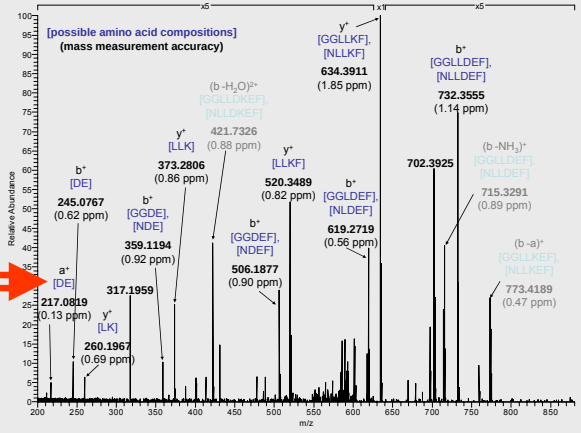
Composition-based sequencing (CBS) for nanoLC-FTICR data

The following example shows a peptide from a human renal cancer cell line that could not be identified unequivocally in a database search.

NanoLC-FTICR MS measurement



MS/MS spectrum of m/z 439.7



Calculate possible amino acid composition

The number of possible amino acid compositions **(AAC)** is gradually reduced by combinatorial calculations [3]. More details about this procedure can be found on Poster ThP 22.

m/z (observed)		number of possible AACs (at +/- 2.5 ppm)
439 $[M+2H]^{2+}$		274
217 / 245	[DE]	8
260	[LK]	4
373	[LLK]	2
		2

Determine peptide sequence

The sequence is determined by comparing observed and theoretical fragment ion masses of possible parent ion sequences.

score	sequence	# matched ions
1.6153	DENFLLK	33
1.3837	DEFNLLK	31
1.3365	DEGGFLLK	31
...		

This peptide was identified as a partial sequence of peptidyl-prolyl cis-trans isomerase A.

Conclusions

• Distributed processing allows for substantial reduction of the processing time for large data sets from FTICR measurements if
• CBS can be used to identify peptide sequences that were not found in an automatic database search
• Remote experimentation facilitates the cooperation between laboratories at different geographic locations

Acknowledgements

This work has been supported by the Stichting voor Fundamenteel Onderzoek der Materie (FOM), the Dutch Ministry of Education, Culture and Science (OC&W), the Dutch Ministry of Economic Affairs (EZ) and the German Federal Ministry of Education, Research (grant 0312834A) and the German Research Council (DFG Sp314/3-1,2). CSF samples have been provided by Theo Luider (ErasmusMC, Rotterdam). The acid strip samples have been provided by Thomas Flad (Center for Medical Research, University of Tuebingen Medical Clinic, Germany).

References

[1] Horn, et al., *JASMS* (2000) **11**(4): 320.
[2] Dekker, L. J., W. Boogerd, G. Stockhammer, J. C. Dalebout, I. Siccama, P. Zheng, J. M. Bonfrer, J. J. Verschuuren, G. Jenster, M. M. Verbeek, T. M. Luider, P. A. Sillevis Smitt, *Journal of Molecular and Cellular Proteomics*, submitted.
[3] Spengler, B.: *JASMS* (2004), **15**, 703.