MsXelerator: A Software Suite for LC- MS based Metabolomics

Marco M.A. Ruijken†
†: MsMetrix, Maarssen, the Netherlands
info@msmetrix.com www.MsMetrix.com

MsMetrix
Accelerating Data Analysis in LC/MS

MsXelerator Introduction

In Metabolomics, LC-MS based experiments are often used to compare and find differences between complex biological samples across multiple conditions. Fast, powerful computational tools are needed to explore and detect significant differences between sample/control, treated non-treated or time-related states. The comparative analysis might be restricted to two samples or can be applied to large series of samples belonging to different classes or treatments.

During LC/MS data processing complicating factors are often encountered: strong and complex chromatographic sample shifts, how to handle normalization of samples, the use of low-resolution MS versus high-resolution MS, processing the enormous amounts of peaks into significant results, identification of relevant peaks and the interpretation of (multivariate) models.

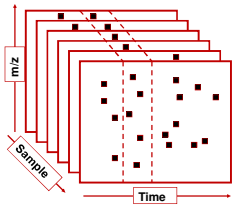
MsXelerator is dedicated to all of the above tasks and contains the following modules:

Mbrowser:

- Interactive Exploration of LC-MS data
- Import data formats of all major vendors: Thermo, Waters, Bruker, Agilent, Sciex

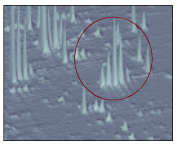
MPeaks:

- Fast Peak-Picking & Peak Filtering at any resolution
- Charge States, Dual mode Differential Analysis
- Targeted and non-targeted Analysis
- Identification



MS Compare:

- Biomarker Discovery
- Statistical Comparison of series or classes of samples
- Dedicated alignment algorithms
- Peak Matching
- Profiling, Multivariate Analysis tools like PCA, Clustering, Correlation Maps, Regression Analysis, etc.



IPeaks:

- Isotope Search Engine
- Peak detection based on isotopic signatures
- Performs any type of isotope ratio calculation for quantitation or recognition.

Alignment Algorithms

MsCompare contains five alignment algorithms that can be run individually or sequentially. Very difficult data sets might need a combination of the available algorithms.

- **Offset Correction:** apply fixed time shift based on single selected peak. Select the peak maximum of a reference peak to correct the full chromatogram.
- **Cross Correlation:** fixed offset correction based on correlation in selected region. Shape more important compared to size in recognition of peaks.
- **Correlation Optimized Warping (COW):** handles linear and non-linear shifts. Needs relatively well defined Total Ion Currents or Base Peak Chromatograms and can take some time to compute.
- **Manual Correction:** apply shifting, stretching and shrinking in specific regions. Graphical implementation. To be applied in very difficult situations or as a first step in a sequential alignment procedure.
- **Peak Reference Warping:** If TICs and BPCs are not well defined, reconstruct a chromatogram based on automatically or user selected m/z values (low or high resolution), leaving out high background ions. The algorithm automatically selects reference peaks that are well spread across the chromatogram and present in all samples. Next, apply a non-linear shift correction based on the position of these reference peaks. Extremely fast.

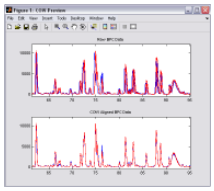
Results and Examples

Alignment 1: COW

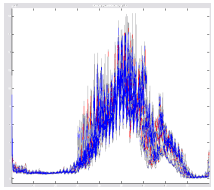
Water Qtof LC/MS: high quality TIC/BPC ➞ easily aligned using COW

Alignment 2: Reference Peak Warping

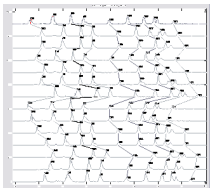
Thermo LTQ: too many peaks show up in TIC ➞ Select reference peaks and apply Reference Peak Warping. Selected peaks are checked on equal spread and Intensity



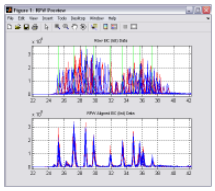
Original and aligned base peak chromatograms



Original TICs in overlay (14 samples)

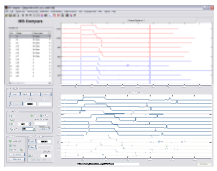


Automatically reconstructed TICs based on 9 selected reference m/z values



Original and aligned TICs after RPW

Alignment 3: Manual Approach



Raw (bottom) and aligned chromatograms (top), using manual alignment, Process Chromatography Data

Alignment Summary

COW often performs well on TIC or BPC chromatograms having not too many peaks that are nicely separated. If the data get more crowded or complex, Peak Reference Warping on selected reference peaks is preferred. Optionally one might use COW after RPW based on the full TICs or BPCs. In cases in which the shifts are really complicated (example 3), it is preferred to start with a manual alignment, followed by either COW or RPW. In the literature we find more than 10 different alignment algorithms, all having their pro's and con's. We believe that the combination of the 3 algorithms implemented in MsCompare can solve any alignment problem.

MsCompare: Data from the example below is from Listgarten et al. In total 14 samples (low level spiked) in two groups were measured on LTQ. Each of the samples contains more than 5000 peaks (Peak Picking results from MPeaks). The Peak Matching routine of MsCompare (figure 1) was used to calculate accurate areas and peak heights of all peaks. No separation of the two groups can be obtained on the full data matrix, or on the raw TICs/BPCs (Figure 2). PCA and PLS-DA did not give satisfactory results as the relevant peaks are very small and not based on a latent factor model. A simple straightforward ratio analysis picks out 15 relevant peaks. After removal of ^{13}C isotopes 10 peaks remain, giving a perfect separation (figure 3, 4)

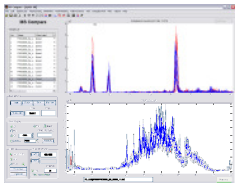


Fig 1. MsCompare Overview: run Peak Matching on all samples, view Extracted Ion Currents and spectra. Easily find relevant differences separating both groups using Multivariate Methods

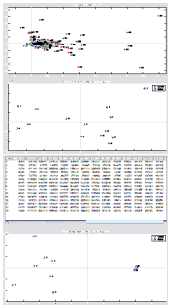


Fig 2,3 Top – PCA on full data matrix, showing now separation. Bottom – PCA applied to most discrimination peaks from a ratio analysis.

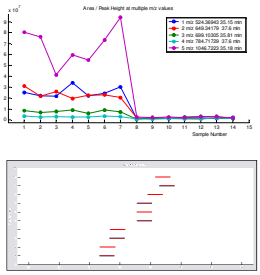


Fig 4,5 Top – Area Profile Plots for a number of highly discriminating peaks. Bottom, some of the selected peaks showing co-elution, which requires additional checks on the peaks and their MS spectra.

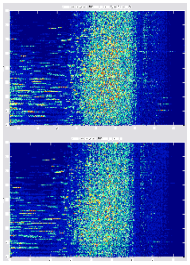
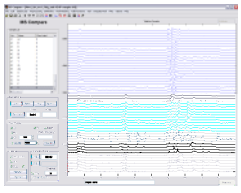


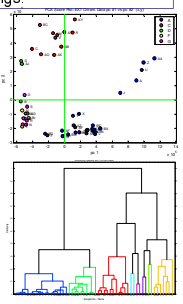
Fig 6 Contour plots of sample and control, showing complexity of LC/MS data sets

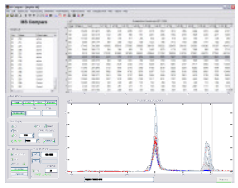
LC/MS Profiling versus Genetic Profiling: task was to compare genetic lineage of 80 organisms with the chromatographic profiles measured by LC/MS. There are about 6-9 clusters based on genetic lineage. LC/MS data are not properly aligned. Difficulty in alignment is that the chromatograms are quite different from group to group. It is not possible to use one single reference chromatogram for alignment. In this case, MsCompare was used to extract Average Mass Spectra from 0.1 Da. wide bins. High background ions were omitted in the analysis. Multivariate analysis was performed on Average Mass Spectra, a direction in which no alignment problems exist. Groupings obtained from PCA and Hierarchical Clustering were almost identical compared to genetic groupings.

Plant Metabolism: Natural Product Activity Screening: A topic of great interest concerns the screening for active natural products from higher plants in the search for new lead anti-inflammatory agents. LC/MS was used to screen for active components that correlate well to a bio-activity assay. In total 25 samples were run in positive and negative mode in triplicate (Waters QTOF premier). Regression analysis between the peak areas of about 1200 extracted peaks and the results from the bio-assay revealed a large number of correlating peaks. Future research will be focused on reduction of the number of interesting peaks using more dedicated analysis methods (sub-fractionation).

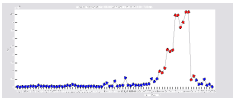


MsCompare showing stacked plot of 54 TICs. Unsupervised PCA and Clustering was applied on Average MS Spectra, showing clear and comparable groupings between both techniques.





MsCompare showing overlay of extracted ion currents for one of the interesting peaks (right peak).



Area Intensity of one of the highly correlating peaks versus sample number. Bio-active samples are marked red.

CONCLUSIONS

MsXelerator has been designed to be extremely fast, easy-to-use and is independent of instrument vendor. The software offers a multitude of algorithms and modules to solve a large number of complex problems in LC/MS Data Processing and Profiling studies. Advanced Alignment algorithms in combination with Multivariate techniques solves most of the problems encountered in Metabolomics. New algorithms, procedures ore user requirements are implemented on request.