MsCompare: An Untargeted LC & GC/MS Metabolomics Platform for Quality Control, Precise Deconvolution and Data Analysis

Marco M.A. Ruijken:
MsMetrix, Maarssen, the Netherlands

MsMetrix
Accelerating Data Analysis in LC/MS

MsCompare Introduction

A GC/MS or LC/MS workflow for Metabolomics applications includes a number of distinct steps: experimental design, sampling, sample preparation, data acquisition, data processing, deconvolution, identification and data interpretation.

The MsXelerator software includes Quality Control procedures and Analysis of Variance (ANOVA) to properly control each step in this workflow and to obtain reliable results for both quantitation and identification. MsXelerator supports all major vendors and is dedicated to all of the above tasks. It contains the following modules:

QC Quant:

- Quality Control for GC/LC MS Projects
- Include: Samples, Blanks, Controls: Instrument Performance checks
- ANOVA workflow : Sample types, Sample preparation, Origin, etc.
- Fast and sensitive GC/MS Peak Picking including Multivariate Analysis

GC Analyzer:

- Compare highly similar samples: Sample/Control
- Accurate sensitive Peak Picking and Advanced Deconvolution
- Identification: Link to NIST MS Search, includes Ret. Index Search

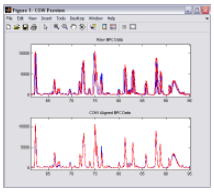
MsCompare:

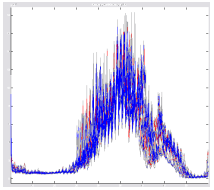
- Metabolomics & Biomarker Discovery
- Statistical Comparison of multi class samples
- Dedicated alignment algorithms (COW, RPW)
- Targeted & Un-Targeted Peak Picking
- Profiling, Multivariate Analysis tools like PCA, PLS, Library Search.

Alignment Algorithms

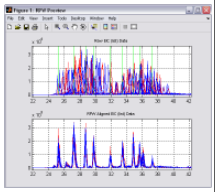
MsCompare contains five alignment algorithms that can be run individually or sequentially. Complex data sets might need a sequential combination of the available algorithms.

- **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 ions. The algorithm selects reference peaks that are well spread across the chromatogram and present in all samples. Next, apply a Fast non-linear shift correction based on the position of these reference peaks.

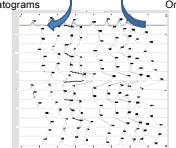




Original and aligned base peak chromatograms



Original and aligned TICs after RPW



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

GC/MS QC Quant

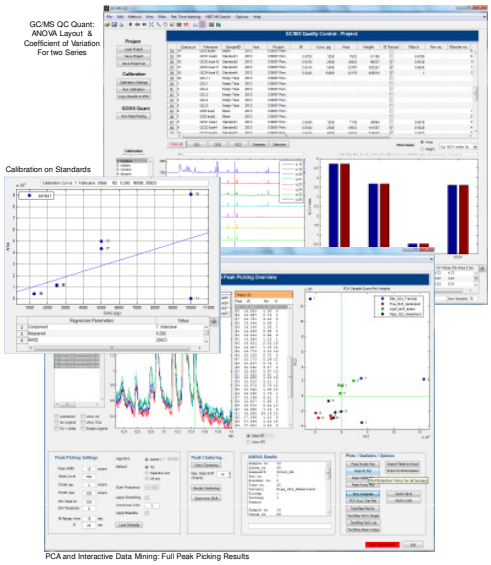
Sample Lists: in QC Quant the following setup of samples was applied: empty tubes (11), blanks (4), quality control standards of a mixture of 9 n-alkanes at 4 levels (8) and 16 samples from 4 tomato varieties split into technical and biological replicates.

ANOVA Layout: a total of 10 ANOVA fields may be defined. These fields can be used to specify sample, instrument or sample preparation conditions (e.g. blender type, tube number, analyst). The ANOVA fields can be used in the analysis, e.g. PCA to link certain groupings or outliers with experimental conditions or sample types.

Quality Control: based on the QC samples and blanks, instrument performance is controlled during the total experiment by examining calibration plots and residuals plots, coefficient of variation between QC replicates. Projects from different time periods can be merged and combined results can be analyzed.

Peak Picking and Multivariate Analysis: QC Quant contains very sensitive and fast peak picking algorithms for detection of all chromatographic peaks in all individual GC/MS ions traces.

Advanced Analysis: Results from QC Quant can be analyzed in MsCompare using more advanced techniques like PLS-DA, ECVA or to find the discriminating peaks from different sample groups.



PCA and Interactive Data Mining: Full Peak Picking Results

GC Analyzer

Detecting Differences between Sample/Control: GC Analyzer performs very sensitive peak picking using all GC/MS ion traces. Results from peak picking are compared against a control / reference / target sample. The high sensitivity allows for the detection of very small differences even when these are totally buried under large peaks.

Retention Indices: GC Analyzer uses a mix of methyl esters or n-alkanes for RI calibration.

Deconvolution: GC Analyzer performs deconvolution based on all ion traces (if no target present) or using only the subset of differential peaks. In this way even perfectly co-eluting component can be deconvoluted.

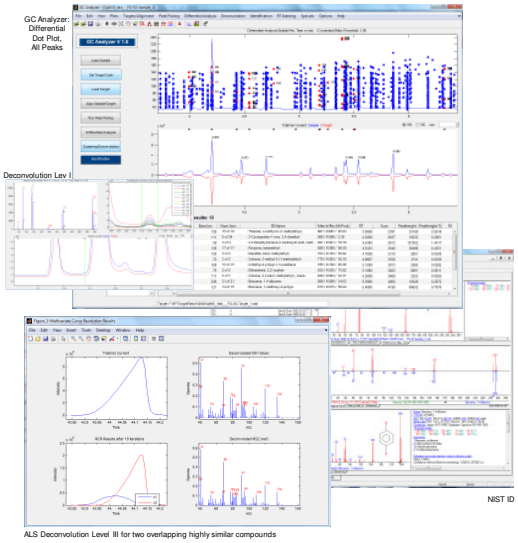
Deconvolution can be applied on three levels:

Level I: Closely eluting peaks having all unique ions.

Level II: Closely eluting peaks with similar ions but some unique ➔ apply Regression Analysis.

Level III: Closely eluting peaks with no unique ions ➔ apply Alternating Least Squares (ALS)

Identification: Deconvoluted spectra are automatically identified using NIST MS Search software. Using the latest NIST version you can apply Retention Indices to obtain more reliable results for very similar spectra.



ALS Deconvolution Level III for two overlapping highly similar compounds

MsCompare

Peak Picking and Peak Matching: MsCompare is the preferred module for multi sample and multi group comparisons. Run targeted or un-targeted peak picking on hundreds of samples from different groups.

Alignment: use correlation optimized warping or reference peak warping to quickly align all chromatograms.

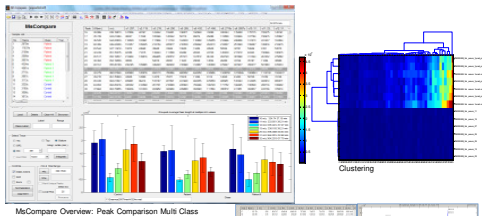
Univariate Analysis Tools: Many problems are not very complex and do not require multivariate tools. MsCompare includes many powerful univariate statistics, including multi group statistics.

Multivariate Analysis Tools: MsCompare contains the following Multivariate Analysis Techniques: Hierarchical Clustering, Principal Component Analysis, PLS Regression, Correlation Maps and Extended Canonical Variates Analysis (ECVA).

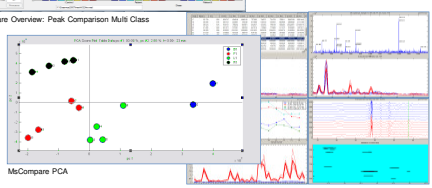
Interactive Graphics: Stay in touch with your data and directly link all results with EIC and MS plots of your samples.

Identification: MsCompare allows direct linking with user created data bases or internet linked data bases for identification. For GC/MS the identification is controlled using NIST Library Searching.

Multi Vendor Data Import: MsCompare allows processing of data from the following vendors: Thermo Xcalibur, Waters Masslynx, Bruker, Agilent MassHunter and AB Sciex. Import is also supported for NetCDF and mzXML data formats.

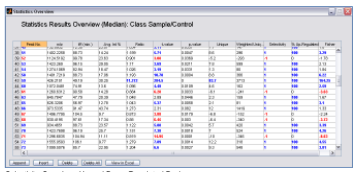


MsCompare Overview: Peak Comparison Multi Class



MsCompare PCA

MS & EIC Overviews



Selectivity Overview; Up and Down Regulated Peaks

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 and GC/MS Data Processing and Profiling studies. Advanced Alignment algorithms in combination with Multivariate techniques and Quality Control solve most of the problems encountered in Metabolomics projects.