# MS-Xelerator™: Advanced Algorithms for LC/MS Data Processing applied to Biomarker Discovery, Differential Analysis and Quantitative Proteomics

**Marco M.A. Ruijken, MsMetrix, Maarssen, The Netherlands**

mr@msmetrix.com

## MS-XELERATOR OVERVIEW

LC-MS based proteomic experiments are used to compare complex biological samples across multiple conditions. Fast, powerful computational tools are needed to explore and detect differences in the areas of **Expression Proteomics** and **Biomarker Discovery**.

In general, specialized steps are necessary to solve these difficult problems (binning, alignment & normalization, peak picking, relative quantitation, classification etc.). MS-Xelerator is a collection of software tools dedicated to all of the above tasks.

**MBrowser:** Interactive graphical environment for LC-MS data mining.

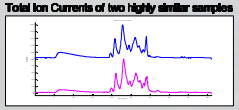
**MPeaks:** Fast Peak-Picking & Peak Filtering, Charge State Calculations, Differential Analysis and linking of results to Mascot.

**IPeaks:** Quantitative Proteomics based on labelling experiments: SILAC, $^{16}O/^{18}O$, $^{14}N/^{15}N$ labelling or user defined isotopic labelled peaks.

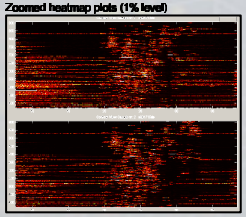
**MS Compare:** Biomarker Discovery and Statistical Comparison of series or classes of samples.

## DIFFERENTIAL ANALYSIS

Total Ion Currents of two highly similar samples

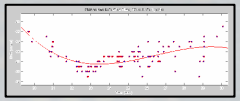


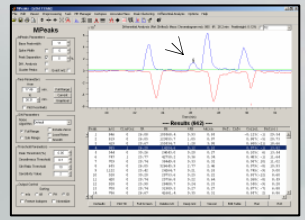
Zoomed heatmap plots (1% level)



**Differential Analysis** between two samples is performed after detection of all chromatographic peaks in both samples. The applied algorithm uses an auto-alignment procedure to correct for chromatographic shifts. Auto-alignment is locally applied to pairwise detected peaks.

The figure at right shows the results of **Auto Alignment** for a number of selected peaks (retention time versus optimal shift). Although a general time-related trend can be observed, many peaks show large deviations from this overall trend.





**MPeaks Differential Result Screen.** Change folds for all peaks are calculated and visualized. The mass chromatogram plotted displays a differential peak at a level of 0.12%.

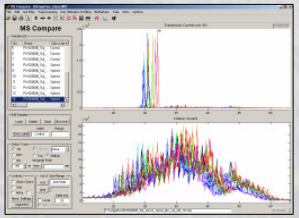
Peak-Picking for both files (FT-MS), following differential analysis was completed in 45 seconds.

## BIOMARKER DISCOVERY

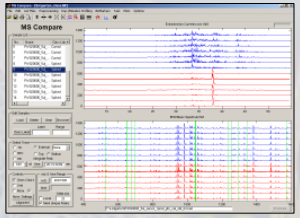
The MS Compare module offers multiple tools (graphically and computationally) for the detection and evaluation of unique or discriminating peaks in Biomarker Discovery. The search can be performed directly, based on full spectrum methods or by using peak detection/matching algorithms.

Identification of unique peaks by direct comparison of mass spectra offers the advantage that no alignment will be necessary. In all other cases, alignment and possibly other pre-processing steps will be part of the workflow. The data set used for this study consists of 14 serum samples: 7 controls and 7 samples spiked with a peptide calibration standard. Computation time was in the order of 2 minutes. The average number of detected chromatographic peaks for a single sample was in the order of 10.000.

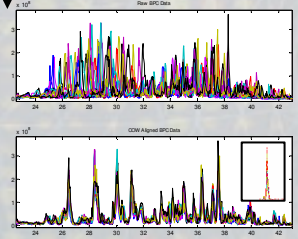
**1. Exploration & Data Mining** based on TIC's / BPC's. Bottom, overlay of TIC's. Top, extracted ion currents showing severe alignment problem.



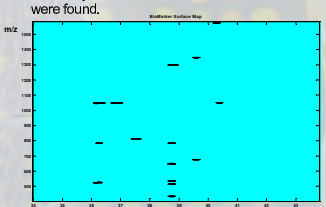
**Direct Mass Spectral Comparison:** bottom, "base chromatogram mass spectra" m/z 400-1600, unique features automatically marked. Top, EIC's shown for m/z 649.



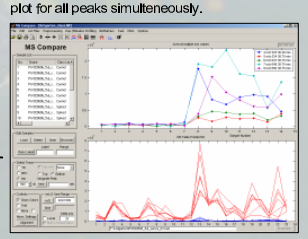
**2. Alignment:** top, raw BPC data, bottom aligned BPC's

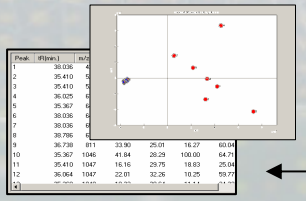


**3. Biomarker Surface Map**, map shows unique features. All 14 full 2D surfaces can be searched based on user defined criteria. In the example full selectivity was used. A total of 23 unique marker were found.




**4. MS Compare Overview:** top, profile plot for 5 discriminating peaks. Bottom, class overview plot for all peaks simulteneously.





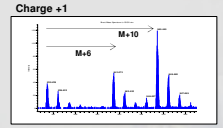
**5. Multivariate Analysis:** PCA or Clustering provides overview of class separation.

## QUANTITATIVE PROTEOMICS

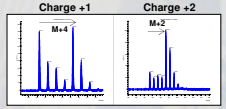
Recently, there has been significant interest in using isotopic labelling for the quantification of peptides and proteins in biological samples. Among the many formats for quantitative proteomics, stable-isotope labelling by amino acids in cell culture (SILAC) has emerged as a simple and powerful one. Other (metabolic) labelling strategies use $^{14}N/^{15}N$ labelling, $^{16}O/^{18}O$ labelling or specific labelling reagents. In general, labelling is facile and inexpensive, but the interpretation of the resulting data has been difficult and manual.

The IPeaks module contains 4 very fast algorithms for searching labelled peaks based on mass spectra or mass chromatograms. The user may select from a number of predefined labelling strategies ($^{16}O/^{18}O$, $^{14}N/^{15}N$, Arg0/Arg6/Arg10, Binomial Distribution) or may create a user specified labelling pattern based on expected mass differences and ratio's.

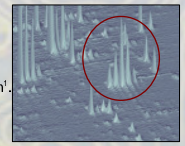
**SILAC labelling: Arg0 / Arg6 / Arg10**
Simultaneous analysis of three cell populations. Since the isotopic patterns will depend on charge state, a fast algorithm (peak based) automatically determines charge states and subsequently calculates the ratio's (0/6, 0/10 and 6/10). Complete analysis can be done in about 1 minute.

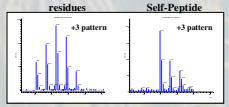



**$^{16}O/^{18}O$ labelling:**
Comparative analysis of peptide lysates of $^{18}O$ untreated cells with treated cells labelled with $^{16}O$. Due to the incorporation of two oxygen atoms, the algorithm will search for peaks having mass differences of 4 Dalton (singly charged peptides).



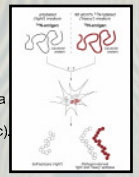

**SITE: Stable Isotope Tagging of Epitopes:**
Site is based on metabolic labelling of endogenously synthesized proteins during infection[1]. The technique involves special pooling of cell batches. Infection induced peptides show up having a very specific binomial isotopic pattern.



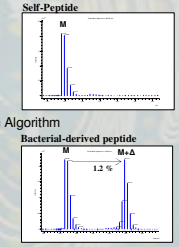



**$^{14}N/^{15}N$ labelling:**
Mass chromatograms are searched for characteristic $^{14}N/^{15}N$ ion doublets for OMV[1] derived peptides using a very fast iterative table based algorithm (30 sec).

Search Criteria:
- Ratio 1:1
- Mass Difference ~ 1.2%




Search Algorithm




[1] Meiring, H.D., Targeted identification of disease-related MHC-presented epitopes, Thesis 2005, Utrecht, The Netherlands

MsMetrix
www.MsMetrix.com

MS-Xelerator 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.