

The MarkerLynx Application Manager: Informatics for Mass Spectrometric Metabonomic Discovery

Chris L. Stumpf, Jeff Goshawk

Waters Corporation

For research use only. Not for use in diagnostic procedures.

Abstract

The Waters MarkerLynx Application Manager for MassLynx Software addresses the problem of identifying multiple interrelated markers by looking for patterns in related LC-MS data sets.

Benefits

- The MarkerLynx Application Manager for MassLynx Software accurately detects ions of interest from an MS sample set by extracting the relevant data
- The seamless integration of the data reduction, integration, and statistical analysis of the LC-MS raw data allows for the rapid identification of biomarkers of disease, toxicity, or genetic modification

Introduction

Metabonomic studies interrogate biological fluids looking for changes in an organism's metabolic profile induced by genetic modifications, toxicological insults, or environmental stimuli. These changes can often be subtle and are often hidden by other metabolic variations such as age, diurnal effects, and gender. In such studies, very rarely does a system perturbation produce concentration changes in just one metabolite, but rather there is a cascade of metabolic changes and these changes would be difficult to visualize. The Waters MarkerLynx Application Manager for MassLynx Software addresses the problem of identifying multiple interrelated markers by looking for patterns in related LC-MS data sets. The samples are categorized based on the similarity of their LC-MS signal. Based on this classification, interrelated markers can be quickly identified. These markers of interest can then be further studied by MS/MS or other analyses of choice.

The MarkerLynx Application Manager

The MarkerLynx Application Manager is an integrated software package that provides automated peak detection followed by principal components analysis (PCA) with direct links to the original chromatograms and ion intensity trend plots. In brief, PCA is a mathematical process to reduce multi-dimensional data sets down to coordinates that can be readily displayed in a graphical view. PCA plots reveal intersamplerelationships via their spatial proximity to one another, the trajectory of metabolic changes occurring, and the compounds responsible for such changes. The advantage of using such a statistical tool is that similar chromatographic and/or mass spectrometric data are revealed in clusters so that the marker ions that contribute to the variance can be readily discerned.

As an example, Figure 1 illustrates the use of MarkerLynx to distinguish between male and female white mice sample based on the metabolites present in urine.

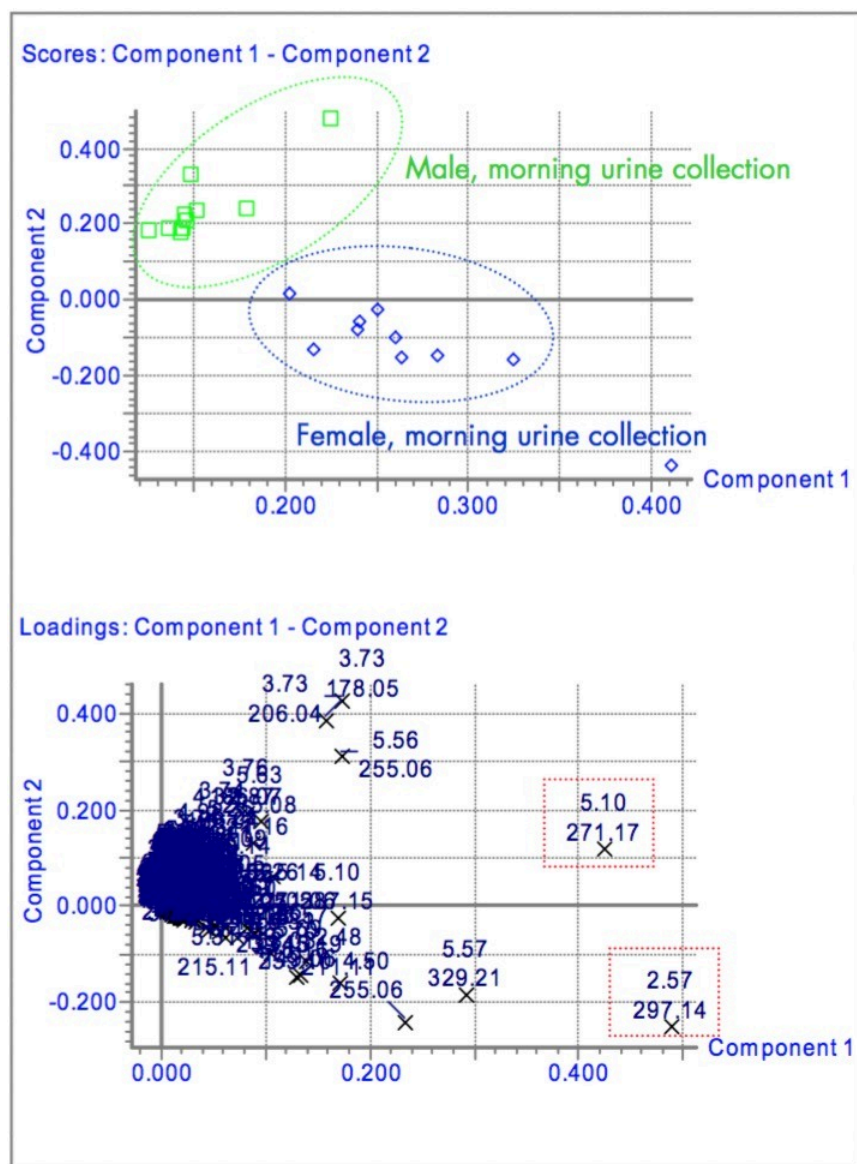


Figure 1. PCA scores and loadings plots of mouse urine samples. The scores plot (top) displays the natural clustering or separation within the data set. The loadings plot (bottom) then identifies those variables that contribute most significantly to the variance observed within the data set. Metabolite ions indicated in red are most influential to the observed gender-based group clustering.

Experimental

Three Synthetic Mixtures: An Example

A simple data set consisting of three sample mixtures was created to illustrate the processing features of MarkerLynx. The first two mixtures were either aqueous (Mixture 1) or rat urine-based (Mixture 2) and were spiked at varying concentrations (sample-to-sample) with theophylline, caffeine, hippuric acid, benzoic acid, chlorpheniramine, nortriptyline, tolbutamide, and reserpine. The final mixture (Mixture 3) contained a constant concentration of theophylline, caffeine, hippuric acid, nortriptyline, and 4-nitrobenzoic acid in water. According to the experimental design, MarkerLynx should classify three distinct group clusters and identify those marker ions responsible for the group separation. Inspection of the MarkerLynx browser below demonstrates that indeed three sample groups clusters were produced. Nortriptyline (m/z 264), theophylline (m/z 181), and reserpine (m/z 609) were correctly identified as marker ions.

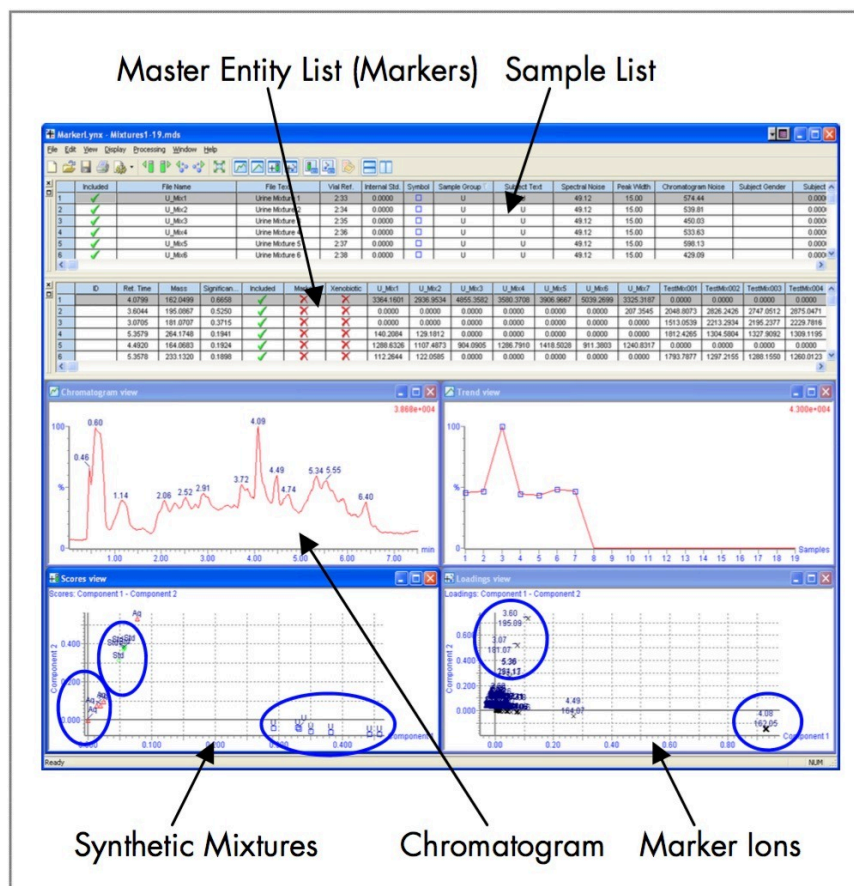


Figure 2. MarkerLynx browser results of the synthetic mixtures data set.

Results and Discussion

How Markerlynx Works

There are four distinct steps that MarkerLynx carries out to conduct a sample set analysis: (1) chromatographic peaks are detected, (2) ion intensities are identified, (3) ions from detected peaks are retained in a data matrix, and (4) PCA is performed on the resulting data.

1. Chromatographic Peak Detection

Multivariate analysis techniques are highly reliant upon correctly aligned data from both chromatographic and mass spectrometric peaks from sample to sample. LC-MS data that is coarsely binned and aligned (a common practice) will produce some potential marker metabolites, though much of the fine detail produced by a high mass accuracy, high resolution LC-MS system will be lost. In order to maximize the information generated from such a system (as with the Waters Metabonomics MS System), MarkerLynx utilizes the patented ApexTrack algorithm to perform accurate peak detection. ApexTrack takes the second derivative of a chromatogram (Figure 3 and 4) and locates the inflection points, local minima, and peak apex. After the apex is found for each peak, a retention time is assigned, and the data is correctly aligned.

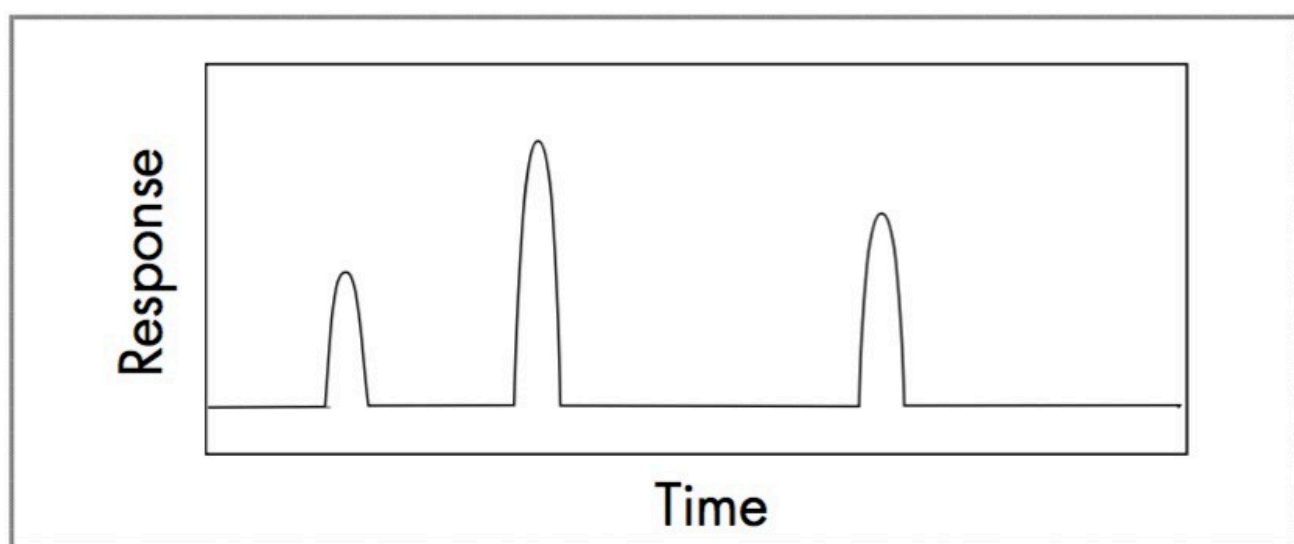


Figure 3. A simplified chromatogram.

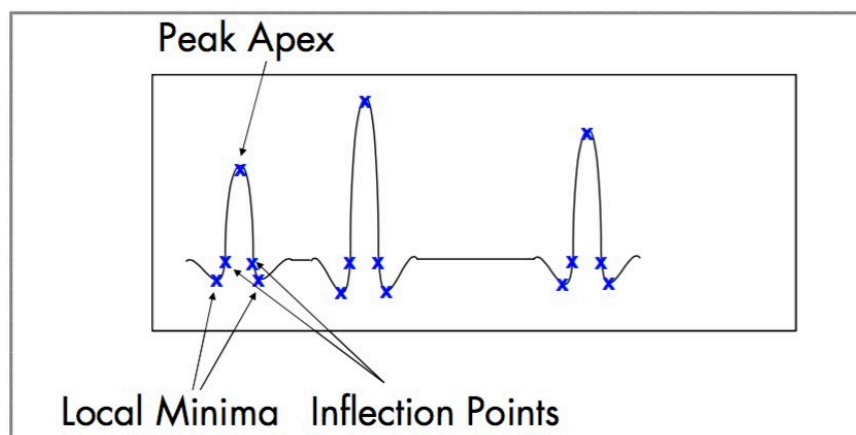


Figure 4. The second derivative is applied to the chromatogram shown in Figure 3. The inflection points, local minima, and peak apex are identified.

2. Locate Maximizing Ions

After the chromatographic peaks are located, the ions associated with these peaks are analyzed. When an ion's maximum intensity is found under the chromatographic peak, its retention time, ion intensity, and exact mass are captured (Figure 5). This process is repeated for all samples in the data set.

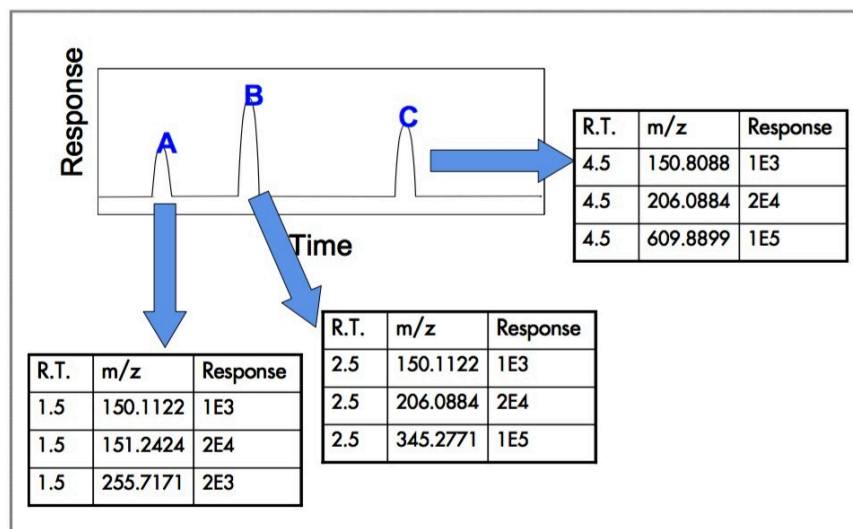


Figure 5. Ions found to maximize under chromatographic peaks are retained.

3. Assemble Ion Data into a Matrix

The ions found to maximize under the chromatographic peaks are assembled into a single data matrix (Figure 6). In order to be retained in the data matrix, an ion has to be present in at least two samples.

m/z	R.T.	Sample 1	Sample 2	Sample 3	Sample N
150.1122	1.5	1E3	1E3	1E3	
150.1122	2.5	1E3	1E3	1E3	
150.1255	2.5	0	0	1E4	
150.8088	4.5	1E3	1E3	1E2	
151.2424	1.5	2E4	0	0	
151.2425	1.5	0	2E4	1E3	
206.0884	2.5	2E4	0	2E4	
206.0886	2.5	0	2E4	2E4	
206.0884	4.5	2E4	2E4	2E4	
255.7171	1.5	2E3	2E3	1E3	
345.2771	2.5	1E5	1E5	1E4	
609.8899	4.5	1E5	1E5	1E3	

The matrix is filled with m/z and intensity information from sample 1 through sample N. Whenever an ion is detected, the ion list is first interrogated to determine whether it already exists. If the ion already exists in the matrix, the ion intensity is added to the corresponding location. If the ion does not exist it is then appended to the matrix, and a zero is inserted for samples in which the ion was previously undetected. Hence, retaining the validity of the matrix. A typical data matrix is shown in Figure 7.

4. Perform PCA on the Resulting Data Matrix

The resulting data matrix is subjected to PCA by MarkerLynx. During PCA processing, the data matrix is examined to find combinations of the ion intensities that best describe the maximum variance in the data. PCA produces two plots: a scores plot and a loadings plot. The scores plot effectively displays the intersamplerelationships in multi-dimensional hyperspace; with more similar samples clustering together and dissimilar samples separated. The loadings plot describes the relationship between the measured variables; those contributing most to the variance are those furthest from the origin of the plot. A typical LC-MS data matrix is shown below in Figure 7.

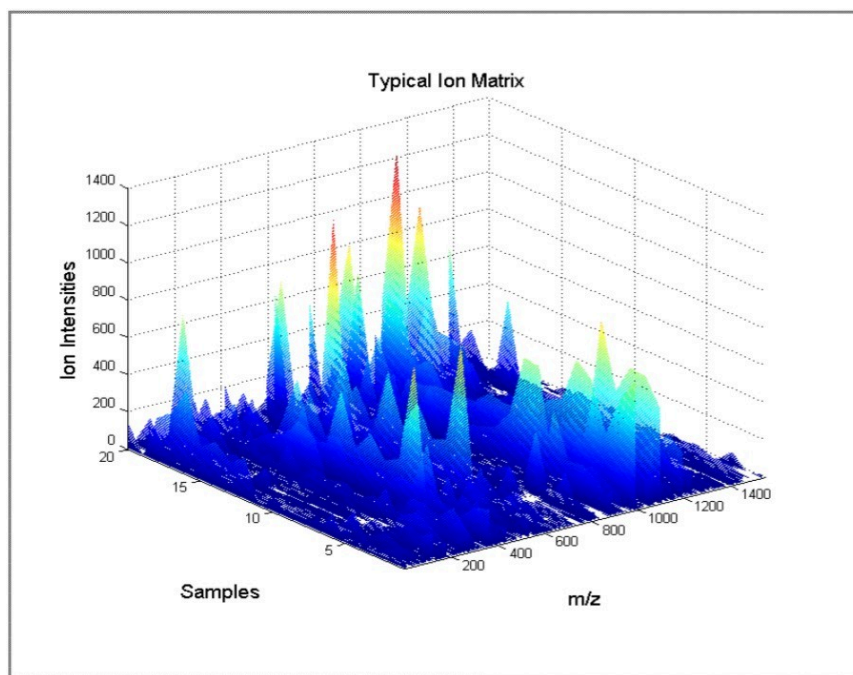


Figure 7. A typical data matrix consisting of mouse urine samples collected from the AM and PM time points from male and female white mice.

Conclusion

The MarkerLynx Application Manager for MassLynx Software accurately detects ions of interest from an MS sample set by extracting the relevant data in four automated processing steps: (1) chromatographic peak detection, (2) ion intensity identification, (3) data matrix construction, and (4) PCA on the resulting data set. The seamless integration of the data reduction, integration, and statistical analysis of the LC-MS raw data allows for the rapid identification of biomarkers of disease, toxicity, or genetic modification. MarkerLynx eliminates the need for simplistic data binning and retains the information-rich nature of the data produced by LC-MS, without the need for significant (and often error-prone) time-consuming manual data processing steps.

Featured Products

Progenesis QI <<https://www.waters.com/134790652>>

720001056, November 2004

© 2022 Waters Corporation. All Rights Reserved.