APPLICATION OF NOVEL MULTIVARIATE STATISTICAL METHODS TO THE IDENTIFICATION OF XENOBIOTIC METABOLITES IN BIOFLUIDS

John P. Shockcor[1,2] and Jose M. Castro-Perez[1]
[1]Waters Corp. Milford, MA USA
[2]University of Cambridge, Cambridge, UK

INTRODUCTION

Multivariate statistical methods have typically been used in metabonomic/metabolomic studies to study endogenous metabolite patterns and how they are altered after some perturbation of a biological system. In this work we present a method for identification of xenobiotic metabolites in a complex biological matrix using advanced statistical methods and exact mass data filters.

Data from control and dosed animals is first filtered using an exact mass data filter to remove much of the endogenous background. Exact mass data filters exploit the mass sufficiency/deficiency of the parent drug and its potential metabolites and require that the data be collected using accurate mass measurement.

The resulting filtered data is then analyzed by orthogonal projection on latent structure (OPLS). OPLS is a modification of the common partial least squares method which allows filtering of confounding variations from the analysis resulting in an easier to interpret dataset. This enhances our ability to detect metabolites (differences between control and dosed samples related to the parent drug) which might not be easily predicted.

METHODS

LC/MS Methodology:

Mass Spectrometer: Q-Tof Premier™
MS scan range: 70-900 Da
Mode of Operation: +/-ve ion mode ESI
V-mode, pDRE (dynamic range enhancement)
Lock Mass: Leucine Enkephalin at 200 pg/mL

MS^E Methodology:

The Q-Tof Premier was operated in a parallel data acquisition mode with a wide band RF mode in Q1(Figure 1). Thus, allowing all ions in the collision cell. This resulted in one single injection in which data was collected under one single data file with two functions. These were;

Function 1) Low energy acquisition (5eV) which contained the unfragmented compounds

Function 2) High energy or MS^E acquisition (15eV-30eV ramp) which contained all of the fragmented ions

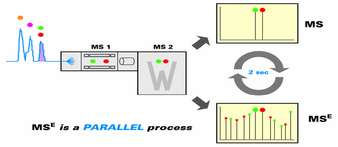


Figure 1. MS^E data acquisition mode with the Q-Tof Premier.

LC conditions:

ACQUITY UPLC™
ACQUITY UPLC BEH C18 Column
100 x 2.1 mm id, 1.7 µm
Mobile phase A: 0.1 % formic acid
Mobile phase B: Acetonitrile
Flow rate: 0.6 mL/min
Gradient: 0 min 98% A, 0-10 min 20% A, 10-11 min 0% A, 11-14 min 98%A
Injection volume: 10 µL

DATA EXTRACTION AND EXACT MASS DATA FILTERS FILTERS

The data from an LC/MS or GC/MS analysis must be in a tabular format before it can be mined using multivariate statistics. This is accomplished using the MarkerLynx™ Application Manager. MarkerLynx will process the samples and create a data matrix with all m/z, retention time pairs observed in all the samples and the measured area from their XICs listed.

Since we are interested in only those m/z, retention time pairs which are parent drug related we can apply an exact mass data filter to remove a significant amount of the chemical background from our matrix. Exact mass data filters use the decimal portion of the parent mass, in the case of acetaminophen m/z 152.0706, we are concerned with only 0.0706. If we hydroxylate acetaminophen we add oxygen which would change the fractional portion of the m/z to 0.06552 which is a decrease of 5.1 mDa. This decrease will be the same for the mono-hydroxylation of any compound. So it is possible for us to create an hydroxylation filter which would reject any m/z values with decimal values which are not 5.1 mDa below our parent compound, the result being the elimination of chemical noise from our analyte. In practice we would set a window to include the parent and other common metabolic transformations using our knowledge of the decimal portion of these transformations as a guide (see Table 1).

Elemental Composition shift	Mass shift Nominal mass	Mass shift Accurate mass	Δ values for calculation of exact mass data filters
+O	+16	+ 15.9949	- 0.0051
$+O_2$	+32	+ 31.9898	- 0.0102
$-H_2$	-2	- 2.0157	- 0.0157
$-CH_2$	-14	- 14.0157	- 0.0157
-Cl+O	-18	- 17.9662	+0.0338
$+C_2H_2O$	+42	+ 42.0106	+0.0106
$+SO_3$	+80	+ 79.9568	- 0.0432
$+C_6H_8O_6$	+176	+176.0321	+0.0321
$+C_6H_8O_7$	+192	+192.0270	+0.0270
$+C_2H_5NO_2S$	+107	+107.0042	+0.0042
$+C_{10}H_{15}N_3O_6S$	+305	+305.0682	+0.0682

Table 1. Some common biotransformation and their associated Δ values for calculation of exact mass data filters.

DATA ANALYSIS

Multivariate Data Analysis (MVA) extracts the information from large data sets and presents the results as interpretable plots based on the mathematical principle of projection. Even data characterized by thousands of variables can be reduced to just a few information rich plots. Basically we are separating signal from noise in data with many variables and presenting the results in a graphical format. Any large complex table of data can be easily transformed into intuitive plots summarizing the essential information. The First step in analyzing samples from a xenobiotic metabolism study is to determine if the data show a difference between the control and dosed groups. This is accomplished by examination of the scores plot shown in Figure 2. Here we can easily see that the dose samples are, as we would expect, different from the control.

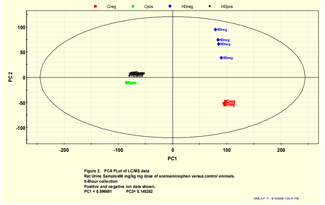


Figure 2.

The second step is to examine the loadings plot (Figure 3) and from it determine which compounds, described as m/z, retention time pairs, are responsible for the difference between the dosed animals and the controls. The m/z, retention time pairs which are the most significant in separating the groups, are those furthest from the origin of the loading plot. Acetaminophen and/or its metabolites will be among these. Using this approach we can extract our signal from the chemical noise in the samples.

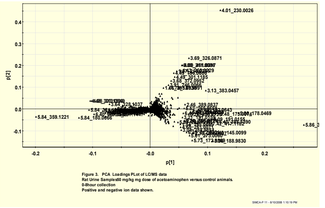


Figure 3.

ADVANCED STATISTICAL METHODS

One of the problems with this approach is that often the loading plot is difficult to interpret and it does not provide us with information about the confidence limits of our measurements. Orthogonal projection on latent structure (OPLS) is a novel new technique which allows us to compare two groups and orthognalizes the results such that the variance between the groups is aligned with the x- axis (Figures 4 and 5) and loading can be more easily interpreted.

PLS

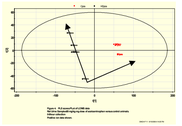


Figure 4.

OPLS

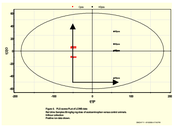


Figure 5.

If we then plot the weight of the variable along the x-axis against a parameter which describes the variance of the loading we obtain the S-Plot. The S-Plot is superior to the normal loadings scatter plot because of its two dimensional nature. It allows us to easily extract even subtle differences in data sets and provides an easy way to visualize both weight and confidence limits for the variables (Figure 6 and 7).

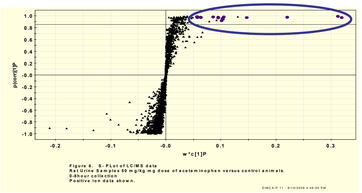


Figure 6.

The blue circled loadings shown in Figure 6 are m/z, retention time values which are elevated in the dose samples and not present in the control samples. Figure 7 shows these values in a bar graph format. From an examination of these results we observe responses for all the metabolites of acetaminophen in or dosed animals urine samples. The masses of these metabolites and their structures are shown below.

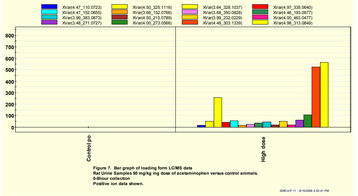
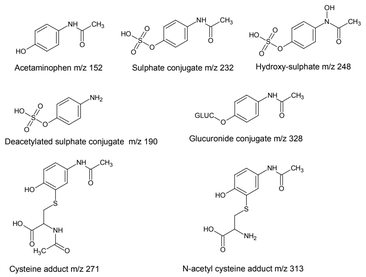


Figure 7.



TO DOWNLOAD A COPY OF THIS POSTER VISIT WWW.WATERS.COM/POSTERS

720001853EN