A RAPID METHOD FOR IDENTIFICATION AND REMOVAL OF XENOBIOTIC RESPONSES FROM MASS SPECTROMETRY BASED METABOLOMIC DATA

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

John P. Shockcor, Jose M. Castro-Perez, and Kate Yu
Waters Corp. Milford, MA USA

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)

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

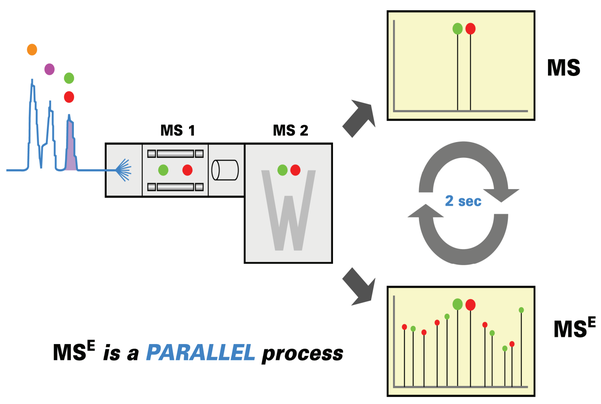

Figure 1. MS^E data acquisition mode

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
Injection volume: 10 µL

Dosing:

Human volunteers were given a single 400 mg dose of ibuprofen prior to sleep. Urine samples were obtained 8 hours after dosing. The control samples were obtained from each volunteer just prior to dosing.

DATA EXTRACTION AND EXACT MASS DATA 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 exact mass, retention time pairs (EMRT's) observed in all the samples and the measured area from their XICs listed.

Since we are interested in only those EMRT 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 ibuprofen m/z 205.1234, we are concerned with only 0.1234. If we hydroxylate ibuprofen we add oxygen which would change the fractional portion of the m/z to 0.1184 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 with decimal values which are not 5.1 mDa below our parent compound, the result being the elimination of chemical noise from our analysis. 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		Δ values for calculation of exact mass data filters
	Nominal mass	Accurate mass	
+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, created using **SIMCA-P** version 11.5 from **Umetrics**, shown in Figure 2. Here we can clearly see that the dose samples are, as we would expect, different from the control.

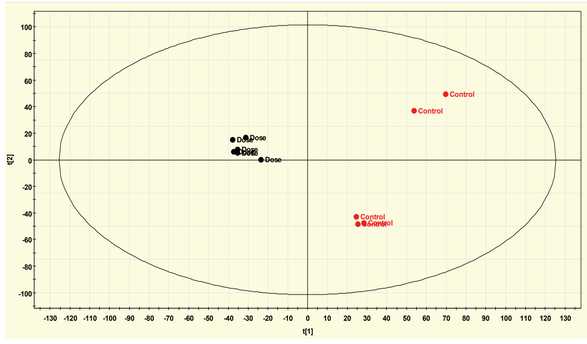

Figure 2. PCA scores Plot, showing dose and control groups.

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

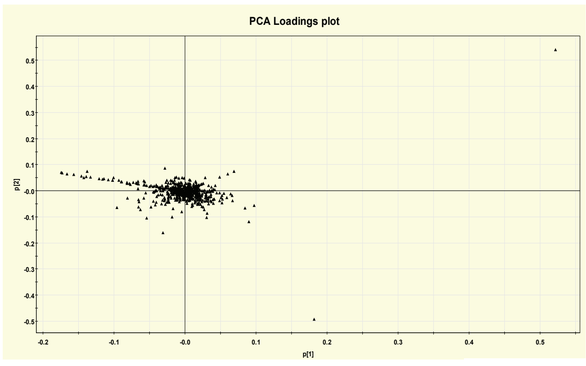

Figure 3. PCA Loadings plot.

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 loading can be more easily interpreted.

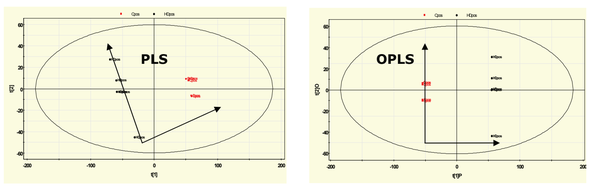

Figure 4. PLS plot compared to an OPLS plot showing the effect of orthoganlization.

If we plot the weight of the variable along the x-axis against a parameter which describes the analytical reliability 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 reliability of the measured variables (Figure 5 and 6).

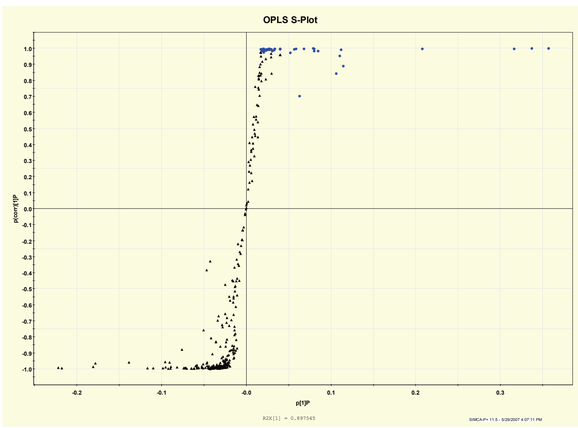

Figure 5. S-Plot from the OPLS analysis.

The S-Plot is very easily interpreted. The further along the x-axis from the origin the larger the contribution to the variance between the groups. The further along the y-axis from the origin the higher the reliability of the data. Thus the *blue dots* in the upper right quadrant are those *EMRT pairs* which are high in the dosed sample and potential metabolites of Ibuprofen .

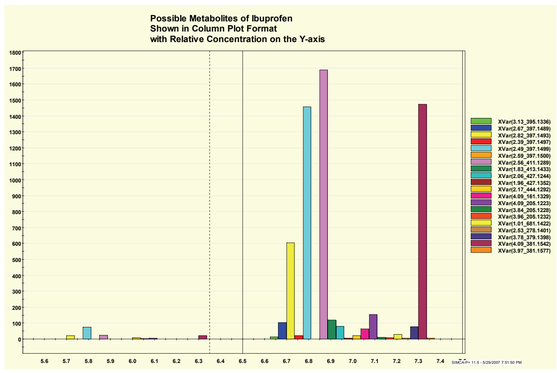

Figure 6. Possible metabolites of Ibuprofen and their relative concentrations.

IDENTIFICATION OF METABOLITES

Taking the list of potential metabolite EMRT's back to MarkerLynx and searching against a database of known and proposed ibuprofen metabolites we can quickly identify the biotransformations that have occurred (Table 2).

	ID	Ret. Time	Mass	Formula	Database	SIMCA1	SIMCA2
1	oxidized-tetra-hydroxy ibuprofen glucuronide	2.1445	443.1352	C19H24O12	ibuprofen (ChemFolder)	0.0237	0.9953
2	oxidized-tri-hydroxy ibuprofen glucuronide	1.9647	427.1352	C19H24O11	ibuprofen (ChemFolder)	0.0186	0.9827
3	oxidized-tri-hydroxy ibuprofen glucuronide	2.0577	427.1244	C19H24O11	ibuprofen (ChemFolder)	0.0786	0.9997
4	di-hydroxy-ibuprofen glucuronide	1.8342	413.1433	C19H26O10	ibuprofen (ChemFolder)	0.0845	0.9843
5	oxidized-di hydroxy ibuprofen glucuronide	2.5553	411.1289	C19H24O10	ibuprofen (ChemFolder)	0.3563	0.9996
6	hydroxy-ibuprofen glucuronide	2.4949	397.1499	C19H26O9	ibuprofen (ChemFolder)	0.3157	0.9966
7	hydroxy-ibuprofen glucuronide	2.3861	397.1497	C19H26O9	ibuprofen (ChemFolder)	0.0397	0.9971
8	hydroxy-ibuprofen glucuronide	2.8239	397.1493	C19H26O9	ibuprofen (ChemFolder)	0.2076	0.9990
9	oxidized-mono-hydroxy ibuprofen glucuronide	3.1313	395.1336	C19H24O9	ibuprofen (ChemFolder)	0.0305	0.9843
10	oxidized-mono-hydroxy ibuprofen glucuronide	3.0704	395.1330	C19H24O9	ibuprofen (ChemFolder)	0.0205	0.9759
11	ibuprofen acyl glucuronide	3.9652	381.1577	C19H26O8	ibuprofen (ChemFolder)	0.0170	0.9955
12	ibuprofen acyl glucuronide	4.0935	381.1542	C19H26O8	ibuprofen (ChemFolder)	0.3369	0.9993
13	oxidized-tetrahydroxy-ibuprofen	2.4393	267.1346	C13H18O6	ibuprofen (ChemFolder)	0.0274	0.9921
14	Ibuprofen	3.9582	205.1232	C13H18O2	ibuprofen (ChemFolder)	0.0262	0.9988
15	Ibuprofen	3.8377	205.1228	C13H18O2	ibuprofen (ChemFolder)	0.0275	0.9986
16	Ibuprofen	4.0943	205.1223	C13H18O2	ibuprofen (ChemFolder)	0.1117	0.9912

Table 2. Proposed metabolites of Ibuprofen.

If we desire to examine the effect of dosing ibuprofen on endogenous metabolism we need only highlight and exclude the EMRT pairs from our table and repeat the statistical analysis. However, this approach is also an effective way to conduct a xenobiotic metabolism study. From our list of metabolites in Table 2 we propose the structures shown Table 3.

#List	#	#ID	Structure	Name	Formula	Monoisotopic...
	1	1		Ibuprofen	$C_{13}H_{18}O_2$	206.13068 Da
	2	9		ibuprofen acyl glucuronide	$C_{19}H_{26}O_8$	382.162768 Da
	3	10		tetra-hydroxy-ibuprofen glucuronide	$C_{19}H_{26}O_{12}$	446.142426 Da
	4	11		tri-hydroxy-ibuprofen glucuronide	$C_{19}H_{26}O_{11}$	430.147512 Da
	5	12		di-hydroxy-ibuprofen glucuronide	$C_{19}H_{26}O_{10}$	414.152597 Da
	6	13		hydroxy-ibuprofen glucuronide	$C_{19}H_{26}O_9$	398.157682 Da
	7	14		oxidized-di hydroxy ibuprofen glucuronide	$C_{19}H_{24}O_{10}$	412.136947 Da
	8	15		oxidized-tri-hydroxy ibuprofen glucuronide	$C_{19}H_{24}O_{11}$	428.131862 Da
	9	16		oxidized-tetra-hydroxy ibuprofen glucuronide	$C_{19}H_{24}O_{12}$	444.126776 Da
	10	17		oxidized-mono-hydroxy ibuprofen glucuronide	$C_{19}H_{24}O_9$	396.142032 Da

Table 3. Structures of proposed metabolites of Ibuprofen.

Of interest is structure *9 in tTable 3,* an oxidized –tetra-hydroxy ibuprofen glucuronide. This metabolite has not been previously reported in the literature, yet it was found and identified using this approach.

Conclusions

The ability to mine a complex matrix like urine and generate a list of 40 to 50 EMRT pairs, with no user bias, and then from that list further mine for potential metabolites using a combination of analytical tools, like neutral losses and common fragment ions along with knowledge of metabolism, comprises a practical and effective approach to identify and/or remove xenobiotic metabolites in a metabolomic study.

ACKNOWLEDGEMENTS

Waters Corporation
Amy Davies
Emma Marsden-Edwards
Jeff Goshawk
Richard Gilpin
David Varley
Steve McDonald
Richard Lock
Andy Baker

Umeå University
Jan Trygg
Susanne Wiklund

Umetrics
Erik Johansson

720002252EN

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