Mass Spectrometry−Based Metabolomics: Accelerating the Characterization of Discriminating Signals by Combining Statistical Correlations and Ultrahigh Resolution

ARTICLE *in* ANALYTICAL CHEMISTRY · AUGUST 2008

Impact Factor: 5.64 · DOI: 10.1021/ac800094p · Source: PubMed

7 AUTHORS, INCLUDING:

Eric Ezan

56 PUBLICATIONS **1,374** CITATIONS

SEE PROFILE

Pierre Chaminade

Université Paris-Sud 11

96 PUBLICATIONS **1,498** CITATIONS

SEE PROFILE

Jean-Claude Tabet

Pierre and Marie Curie University - Paris 6

416 PUBLICATIONS **5,823** CITATIONS

SEE PROFILE

Anal. Chem. **2008,** *80,* 4918–4932

Mass Spectrometry-Based Metabolomics: Accelerating the Characterization of Discriminating Signals by Combining Statistical Correlations and Ultrahigh Resolution

Erwan Werner,[†] Vincent Croixmarie,[‡] Thierry Umbdenstock,[‡] Eric Ezan,[†] Pierre Chaminade,[§] Jean-Claude Tabet,[||] and Christophe Junot*,[†]

CEA, DSV/iBiTec-S/SPI/LEMM, Batiment 136, 91191 Gif-sur-Yvette Cedex, France, Technologie Servier, 25/27 rue Eugène Vignat, 45000 Orléans, France, Groupe de chimie analytique de paris sud−E.A. 4041, Faculté de pharmacie de Châtenay-Malabry, Université Paris XI, 5 rue J.B. Clément, 92296 Châtenay-Malabry cedex, France, Laboratoire de Chimie Structurale Organique et Biologique, UMR 7613-CNRS, Université Paris 6, 4 place Jussieu, 75252 Paris cedex 5, France

A strategy combining autocorrelation matrices and ultrahigh resolution mass spectrometry (MS) was developed to optimize the characterization of discriminating ions highlighted by metabolomics. As an example, urine samples from rats treated with phenobarbital (PB) were analyzed by ultrahigh-pressure chromatography with two different eluting conditions coupled to time-of-flight mass spectrometric detection in both the positive and negative electrospray ionization modes. Multivariate data analyses were performed to highlight discriminating variables from several thousand detected signals: a few hundred signals were found to be affected by PB, whereas a few tenths of them were linked to its metabolism. Autocorrelation matrices were then applied to eliminate adduct and fragment ions. Finally, the characterization of the ions of interest was performed with ultrahigh-resolution mass spectrometry and sequential MS^n experiments, by using a LC-LTQ-Orbitrap system. The use of different eluting conditions was shown to drastically impact on the chromatographic retention and ionization of compounds, thus providing a way to obtain more exhaustive metabolic fingerprints, whereas autocorrelation matrices allowed one to focus the identification work on the most relevant ions. By using such an approach, 14 PB metabolites were characterized in rat urines, some of which have not been reported in the literature.

Over the past few years, "omics tools" (genomics, transcriptomics, proteomics, metabolomics) have been developed to clarify living organisms' responses to a perturbation, of whatever source. The terms metabolomics refer to the untargeted quantitative or semiquantitative analysis of the metabolome (set of all metabolites present in a biological system) by differential analysis of metabolite patterns in different experimental groups.[1–3] Once the metabolic fingerprints have been acquired, they are compared using multivariate statistical analysis[4,5] for discriminating analysis purposes. Multivariate analysis (MVA) visualizes sample distribution and highlights any signals that discriminate between experimental groups.

Almost every analytical technique can be employed for the acquisition of metabolic fingerprints, but nuclear magnetic resonance (NMR) and mass spectrometry (MS)-based techniques have become established as the major tools for metabolomics because of their reliability and selectivity, and also because they identify the metabolites of interest. Initial studies relied on NMR, but despite recent technological developments such as magic angle spinning[6] and cryogenic sample probes,[7,8] the sensitivity and dynamic range of NMR-based techniques are still not as good as those of MS-based methods. As a result, MS-based techniques rapidly emerged as powerful alternatives to NMR for metabolome analysis.

Because of the chemical and physical diversity of the compounds present in a complex biofluid such as urine or plasma, MS detection is classically coupled with a separation technique to maximize the number of metabolites detected. Gas chromatography with electron ionization (GC/EIMS) was first used for metabolic profiling. It provides high chromatographic resolution, good analyte selectivity, and quantification and identification of metabolites thanks to the existing comprehensive EI-mass spectra databases. However, the necessary thermal stability of the investigated compounds, the required derivatization of nonvolatile

* To whom correspondence should be addressed. Phone: +33169084366. Fax: +33169085907. E-mail: christophe.junot@cea.fr.
[†] CEA, DSV/iBiTec-S/SPI/LEMM.
[‡] Technologie Servier.
[§] Université Paris XI.
[||] Université Paris 6.

(1) Fiehn, O. *Plant Mol. Biol.* **2002**, *48*, 155–71.
(2) Villas-Boas, S. G.; Rasmussen, S.; Lane, G. A. *Trends Biotechnol.* **2005**, *23*, 385–6.
(3) Nicholson, J. K.; Lindon, J. C.; Holmes, E. *Xenobiotica* **1999**, *29*, 1181–9.
(4) Holmes, E.; Antti, H. *Analyst* **2002**, *127*, 1549–57.
(5) Trygg, J.; Holmes, E.; Lundstedt, T. *J. Proteome Res.* **2007**, *6*, 469–79.
(6) Wang, Y.; Bollard, M. E.; Keun, H.; Antti, H.; Beckonert, O.; Ebbels, T. M.; Lindon, J. C.; Holmes, E.; Tang, H.; Nicholson, J. K. *Anal. Biochem.* **2003**, *323*, 26–32.
(7) Keun, H. C.; Beckonert, O.; Griffin, J. L.; Richter, C.; Moskau, D.; Lindon, J. C.; Nicholson, J. K. *Anal. Chem.* **2002**, *74*, 4588–93.
(8) Spraul, M.; Freund, A. S.; Nast, R. E.; Withers, R. S.; Maas, W. E.; Corcoran, O. *Anal. Chem.* **2003**, *75*, 1536–41.

10.1021/ac800094p CCC: $40.75
Published on Web 05/31/2008

compounds prior to analysis, and the frequent absence of molecular ions have made liquid chromatography (LC) coupled to atmospheric-pressure ionization mass spectrometry a more and more widely used tool for metabolomics. It classically relies on a separation on reversed-phase (RP) columns, followed by a detection based on electrospray ionization mass spectrometry (ESI-MS). LC/MS is able to separate and detect thousands of compounds with good reproducibility and sensitivity over a wide dynamic range. However, conventional RP-HPLC is often insufficient for a well-resolved separation and comprehensive analysis of complex biological samples. Recently, implementation of ultraperformance liquid chromatography (UPLC) improved chromatographic resolution, peak capacity, and even sensitivity, which is of great value in the context of metabolomics.[9,10]

Another milestone was reached in terms of technological development with the introduction of high-resolution mass spectrometers such as time-of-flight (TOF) and Fourier transform (FT) mass spectrometers. So far, UPLC-TOF/MS has been employed successfully in metabolomics for the detection and characterization of the metabolites of several drugs and natural compounds,[11,12] lipid analysis,[13] and gender or pathology classification.[10] It has spread thanks to good chromatographic separations, full-scan sensitivity, and accurate mass measurement. However, while such a technique is very effective for metabolic profile acquisitions because of a low acquisition cycle duration (from 0.1 to 0.3 s), its averaged resolution of ~5 ppm is a limitation for metabolite characterization.[14] An alternative way of achieving better mass accuracy to facilitate metabolite characterization would be to work with FTMS. Its use has been reported for phenotyping of plant extracts using direct infusion-Fourier transform ion cyclotron resonance mass spectrometry (FTICR)[15] and drug metabolism studies using LC-LTQ-Orbitrap.[16] However, accurate mass measurement is achieved at the expense of longer cycle duration, especially with FTICR, limiting the association of ultrahigh resolution with UPLC, whereas combination with HPLC has already been used for proteomics[17] and metabolite identification.[16]

Despite these technological advances, metabolomic approaches still need to overcome several problems. First, it is impossible to detect all metabolites by a single and universal analytical method. The signal extraction step (i.e., mass spectral background noise suppression and correction of retention time shifts occurring from one chromatogram to the other) must also be improved. It can rely either on manual procedures[18–20] or on dedicated softwares.[21] Although the latter approach seems to be the most attractive, some metabolites are present in trace amounts or ionize poorly and so the resulting low-intensity signals are missed. Some artifactual signals can also be generated due to erratic background noise suppression or to insufficient correction from retention time shifts. Finally, characterization of discriminating signals is a very difficult and time-consuming step.

The present study focuses on this last point and deals with the different reasons that make the identification of discriminating signals a bottleneck of the metabolomic approach. The first one is a direct consequence of the improvements in sensitivity and mass accuracy, which allow the detection of an increased number of signals, leading to hundreds of compounds to be identified after MVA. Moreover, a metabolite does not produce a single m/z peak: adduct and product ions can be produced during the desolvation step following ionization/desorption aggregate formation. Although informative, such signal redundancy slows down the identification procedure. Lastly, profiling experiments are often performed with TOF mass spectrometers that can be coupled to UPLC but that are unsuitable for further CID experiments. Of course, Q-TOF instruments enable MS/MS experiments but do not allow sequential MS^n experiments and, therefore, ion filiation determinations that are very useful for compound identification. As a result, another device may be required for MS/MS or MS^n experiments to get the additional structural information needed for metabolite identification.

In this context, the present work was designed to provide a suitable strategy for highlighting and characterizing discriminating metabolites generated by LC/MS-based metabolomics. The proof of concept was performed on urine from phenobarbital (PB)-treated rats. PB is of interest for metabolomics because it exerts pleiotropic pharmacological and toxicological effects. It is also metabolized[22,23] and is known to be a cytochrome P450 inducer in rat.[24,25] UPLC/TOFMS experiments were performed using two different eluting conditions and the detection was performed in both the positive and negative ESI modes, with the aim of improving the recovery of metabolites. MVAs were then performed to highlight discriminating variables. Before identification of the ions of biological relevance, the signal redundancy due to the presence of adduct and product ions was removed by using a mathematical tool based on autocorrelation matrices. The characterization of the PB-related ions was finally achieved by performing MS^2 and MS^3 experiments on a LC-LTQ-Orbitrap mass spectrometer. Behind the linear ion trap (LTQ, Thermo-Fisher), the Orbitrap cell is based on a new type of mass analyzer designed

(9) Wilson, I. D.; Plumb, R.; Granger, J.; Major, H.; Williams, R.; Lenz, E. M. *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* **2005**, *817*, 67–76.
(10) Plumb, R. S.; Granger, J. H.; Stumpf, C. L.; Johnson, K. A.; Smith, B. W.; Gaulitz, S.; Wilson, I. D.; Castro-Perez, J. *Analyst* **2005**, *130*, 844–9.
(11) Giri, S.; Idle, J. R.; Chen, C.; Zabriskie, T. M.; Krausz, K. W.; Gonzalez, F. J. *Chem. Res. Toxicol.* **2006**, *19*, 818–27.
(12) Chen, C.; Ma, X.; Malfatti, M. A.; Krausz, K. W.; Kimura, S.; Felton, J. S.; Idle, J. R.; Gonzalez, F. J. *Chem. Res. Toxicol.* **2007**, *20*, 531–42.
(13) Rainville, P. D.; Stumpf, C. L.; Shockcor, J. P.; Plumb, R. S.; Nicholson, J. K. *J. Proteome Res.* **2007**, *6*, 552–8.
(14) Sleno, L.; Volmer, D. A.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 183–98.
(15) Aharoni, A.; Ricd, V.; Verhoeven, H. A.; Maliepaard, C. A.; Kruppa, G.; Bino, R.; Goodenowe, D. B. *OMICS* **2002**, *6*, 217–34.
(16) Peterman, S. M.; Duczak, N.; Kalgutkar, A. S.; Lame, M. E.; Soglia, J. R. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 363–75.
(17) Meng, F.; Wiener, M. C.; Sachs, J. R.; Burns, C.; Verma, P.; Paweletz, C. P.; Mazur, M. T.; Deyanova, E. G.; Yates, N. A.; Hendrickson, R. C. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 226–33.
(18) Plumb, R. S.; Stumpf, C. L.; Gorenstein, M. V.; Castro-Perez, J. M.; Dear, G. J.; Anthony, M.; Sweatman, B. C.; Connor, S. C.; Haselden, J. N. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1991–6.
(19) Idborg, H.; Zamani, L.; Edlund, P. O.; Schuppe-Koistinen, I.; Jacobsson, S. P. *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* **2005**, *828*, 14–20.
(20) Ducruix, C.; Vailhen, D.; Werner, E.; Fievet, J. B.; Bourguignon, J.; Tabet, J. C.; Ezan, E.; Junot, C. Chemom. Intell. Lab. Syst. In press.
(21) Katajamaa, M.; Oresic, M. *J. Chromatogr., A* **2007**, *1158*, 318–28.
(22) Harvey, D. L.; Glazener, L.; Stratton, L. *Res. Commum. Chem. Pathol. Pharmocol.* **1972**, *3*, 557–65.
(23) Tang, B. K.; Kalow, W.; Grey, A. A. *Drug Metab. Dispos.* **1979**, *7*, 315–8.
(24) Handschin, C.; Meyer, U. A. *Pharmacol. Rev.* **2003**, *55*, 649–73.
(25) Pustylnyak, V. O.; Gulyaeva, L. F.; Lyakhovich, V. V. *Toxicology* **2005**, *216*, 147–53.

by Makarov and co-workers.[26] As an ingeniously modified Kingdon ion trap, it operates by radially trapping ions about a central spindle electrode and generates a transient ion signal related to the axial ion motion rather than to the radial detection used for FTICR. It achieves high mass resolution (up to 100 000, as expressed as full width at half-maximum peak height, i.e. fwhm), high mass accuracy (below 5 ppm with external calibration), a mass/charge range of 6000, and a dynamic range greater than 3 orders of magnitude. Moreover, it allows MS^n experiments to be performed with a large space charge capacity due to the linear ion trap, which is the front part of this device.[27–29]

EXPERIMENTAL SECTION

Chemicals. Acetonitrile, ammonium acetate, phenobarbital, and standards, used for tuning the Orbitrap mass spectrometer (i.e., testosterone glucuronide, aldosterone, DHEA sulfate, and adipic acid), were from Sigma (Saint Quentin Fallavier, France). Methanol and formic acid were from Merck (Briare-le-canal, France). Water was deionized and filtered through a Millipore Milli-Q water purification system. The standard mixtures used for the external calibration of the LTQ-Orbitrap instrument (Calmix-positive and Calmix-negative, for both the positive and negative ionization modes, respectively) were from ThermoFisher Scientific (Les Ulis, France). Metabonomics MS system test mix used as reference during UPLC-TOF fingerprint acquisition was purchased from Waters.

Animals and Treatments. Forty Wistar male rats (SPF Glx/BRL/Han) from Charles River Laboratories France (Domaine des Oncins, 69210 Saint-Germain-sur-l'Arbresle, France) were acclimated to housing conditions for a period of one week. The animals were 10 weeks old at the start of the dosing period. Their bodyweight ranged approximately from 300 to 325 g based on supplier data.

PB was formulated as a 40 mg/ml suspension in 0.9% w/v aqueous sodium chloride and was administered during 4 days (from day 1 to day 4) at 2 mL/kg to male Wistar rats ($n = 20$), via intraperitoneal injection. Urine samples were collected twice (day −3 and day 5) over a 16-h period on a refrigerated surface using metabolism cages. They were then stored at −80 °C until analysis. All studies on animals complied with the *Décret sur l'Expérimentation Animale* (French law on rules for animal experimentation, Decree 87−848, October 19, 1987).

Ultraperformance Liquid Chromatography Coupled to Electrospray Time-of-Flight Mass Spectrometry. Urine samples were centrifuged at 10 000 rpm for 5 min to remove particles and proteins and then diluted with two volumes of water prior to sample injection. Predose and postdose urine samples were randomly analyzed together by UPLC/TOFMS. Urine samples (10 μL/injection) were separated on a 100 × 2.1 mm Acquity BEH 1.7-μm C_{18} column (Waters, Saint Quentin en Yvelines, France) using an Acquity UPLC system (Waters) with a gradient mobile phase. Two eluting conditions were used: a first gradient mobile phase (hereafter referred to as "acetonitrile-acidic eluting conditions") comprising 0.1% formic acid (A) and acetonitrile containing 0.1% formic acid (B) and a second one (hereafter referred to as "methanol-neutral eluting conditions") comprising 10 mM ammonium acetate (A) and methanol containing 10 mM ammonium acetate (B). In both cases, a 0.5 mL/min flow rate was maintained in a 35-min run with relative proportions of A and B as mentioned in Table 1.

Table 1. UPLC Gradient Conditions

time (min)	A (%)	B (%)
0	100	0
5	100	0
15	80	20
25	20	80
30	5	95
32	5	95
32.1	100	0
35	100	0

Mass spectrometry experiments were performed on a Waters LCT Premier operating in negative and positive ion modes with W-optics enabled. For the negative ion mode, the nebulizer gas flow was 100 L/h, the desolvation gas flow 700 L/h at 400 °C with the cone gas flow 50 L/h, and the source temperature 150 °C. The capillary voltage and the cone voltage were 2700 and 50 V, respectively. Deprotonated leucine-enkephalin was used as the lock mass (m/z 554.2614) for accurate mass calibration and introduced at a concentration of 1.5 μg/mL in 50% aqueous methanol using the LockSpray interface at 50 μL/min. For the positive ion mode, parameters were similar, except for the capillary voltage and the lock mass, which were 3000 V and m/z 556.2771 (corresponding to protonated leucine-enkephalin), respectively. Data were acquired in centroid mode from m/z 100 to 1000 with dynamic range enhancement. Following data acquisition, UPLC-MS chromatograms and spectra were further processed by MassLynx 4.1 and MarkerLynx application softwares (Waters) on an IBM intellistation Z Pro. The metabonomics MS system test mix (Waters) was injected at regular intervals (every 10 injections) to check for consistency of analytical results.

Data Preprocessing and Multivariate Statistical Analysis. Raw data were analyzed using the Micromass MarkerLynx Application Manager (version 1.0, Waters). The UPLC/TOFMS data were peak-detected and noise-reduced in both the LC and MS domains to exclude peaks related to noise. A list of the intensities of the peaks detected was then generated for each sample, using a combination of the retention time (RT) and the m/z ratio as an identifier. An arbitrary number was assigned to each of these RT−m/z pairs in order of elution. The data were then combined into a single matrix by aligning peaks with the same mass−retention time pair together from each data file in the data set. In the case of either nondetected signals or detected signal intensities under the set cutoff value, MarkerLynx assigned a zero value to the variable in the matrix for the samples concerned. Once this was completed, the ion intensities for each detected peak were normalized, within each sample, to the sum of the peak intensities in that sample. Difference between urinary volumes before and after treatment was not statistically significant (paired t test, $\alpha = 0.05$). As a consequence, there was no need for a normalization step against creatinine. The resulting data were

(26) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham, C. R. *J. Mass Spectrom.* **2005**, *40*, 430–43.
(27) Makarov, A.; Denisov, E.; Lange, O.; Horning, S. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 977–82.
(28) Hu, Q.; Makarov, A. A.; Cooks, R. G.; Noll, R. J. *J. Phys. Chem. A* **2006**, *110*, 2682–9.
(29) Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S. *Anal. Chem.* **2006**, *78*, 2113–20.

introduced into SIMCA P (version11.0; Umetrics) for multivariate analyses (MVAs). Prior to MVAs by unsupervised and supervised methods, principal component analysis (PCA), partial least-squares or projection on latent structure-discriminant analysis (PLS-DA), and orthogonal partial least-squares or projection on latent structure (OPLS) in discriminant analysis mode, respectively, data exported from MarkerLynx were first mean-centered and scaled either to unit variance or to Pareto variance. Models were validated by using the cross-validation function of SIMCA P. PLS-DA models were tested with permutation tests ($k = 20$).

Autocorrelation Matrices. Such an approach was originally developed for the processing of NMR data. As described by Cloarec et al.,[30] it is based upon the calculation of a correlation matrix C according to the following equation:

$$C = (1/(n-1))X_1 \times {}^tX_2 \quad (1)$$

where, for MS experiments, X_1 and X_2 denote the autoscaled experimental matrices of $n \times v_1$ and $n \times v_2$, respectively; n is the number of samples and v_1 and v_2 are the number of variables (i.e., couple RT−m/z) in the chromatograms for each matrix. ${}^tX_2$ is the transpose of the matrix X_2, and C is therefore a matrix of $v_1 \times v_2$, where each value is a correlation coefficient between two variables of the matrices X_1 and X_2. X_1 is equal to X_2 for autocorrelation analysis. In this latter case, the matrix displays the correlation among the intensities of the variables across the whole sample. By assuming that the different ions from a single metabolite should exhibit the same retention time and the same change in concentration in response to a perturbation, and that instrumental conditions were kept identical from one sample to another, the relative intensities of such ions should theoretically be fully correlated (correlation coefficient $r = 1$). In complex samples such as biofluids, r was always less than 1 because spectral noise or peak overlaps occur from other molecules, thus leading to a matrix effect or ion suppression. However, in practice, the correlation matrix from a set of mass spectra containing different amounts of the same molecules showed very high correlations between the variables corresponding to the various ions related to the same molecule (adduct ions, in-source CID product ions). Autocorrelation matrices were calculated from mean-centered and scaled to unit variance signal intensities, by using Mathlab 6.

Liquid Chromatography Coupled to Electrospray-LTQ-Orbitrap Mass Spectrometry. An HP1100 series LC system (Agilent Technologies, Massy, France) was used for sample introduction and separation. The chromatographic separation was performed on a XTerra MS C_{18} 5-μm, 2.1 × 150 mm column (Waters) equipped with an online prefilter (Interchim, Montluçon, France). The mobile phases were either (A) 100% water and (B) 100% methanol, both containing 10 mM ammonium acetate buffer pH 7.0 or (A) 100% water and (B) 100% acetonitrile, both containing 0.1% formic acid. After an isocratic step of 5 min at 100% phase A, a linear gradient from 0 to 100% B was run over the next 40 min with a mobile phase flow of 0.2 mL/min. Returning to 100% A over 1 min, the column was then allowed to equilibrate for 10 min leading to a total run time of 60 min. These will later be referred to as "methanol-neutral eluting conditions" and "acetonitrile-acidic eluting conditions", respectively.

ESI-MS metabolic profiles were acquired within the range m/z 100−1000, successively in both positive and negative ion modes with a resolution set to 60 000 ($M/\Delta M$ at fwhm) in centroid mode. For the negative ion mode, the sheath gas was 50 (arbitrary units), the auxiliary gas 10 (arbitrary units), and the capillary temperature 275 °C. The source voltage and the tube lens voltage were 3500 and −100 V, respectively. Lens parameters were optimized using a tune solution containing testosterone glucuronide, aldosterone, DHEA sulfate, and adipic acid at 1 μg/mL in methanol/water (50:50), infused at 5 μL/min. Gas flows were then optimized with the same compounds but with a 0.2 mL/min HPLC flow. For the positive ion detection mode, the sheath gas was 50 (arbitrary units), the auxiliary gas 5 (arbitrary units), and the capillary temperature 275 °C. The source voltage and the tube lens voltage were 5000 and +100 V, respectively. Orbitrap calibration was performed in the positive and negative ion modes by using calibration mixture from Thermo-Fisher.

CID spectra were acquired using data-dependent scanning function. The scan event cycle comprised a full scan mass spectrum at a resolution power of 15 000 and two data-dependent (MS^2 and MS^3) events acquired with a resolution set to 7500. The most intense ion detected during full scan underwent data-dependent scanning (MS^2) whose most intense ion (MS^2 parent ion excluded) triggered MS^3. Microscan count was set to unity, and a repeat count for dynamic exclusion was set to 3. MS^n acquisition parameters were an isolation width of 1 u, normalized collision energy of 35%, and activation time of 30 ms.

All data were processed using Qualbrowser (Thermo-Fisher), and its chemical formula generator was used to provide elemental compositions.

RESULTS AND DISCUSSION

This study deals with an approach to analyzing UPLC/MS-based metabolomics data sets, from the analytical method to the characterization of discriminating variables, highlighted by multivariate analyses. The proof of concept was performed on urine samples from PB-treated rats.

Data Acquisition. The analysis of urine samples was carried out using a 35-min UPLC reversed-phase gradient coupled with ESI-TOFMS detection in both the positive and the negative electrospray modes. The mass accuracy of the mass spectrometer was ensured by the use of the lockspray interface. Moreover, the metabonomics MS system test mix (Waters) was injected every 10 samples, to check for consistency of analytical results. This mix was compatible with the different chromatographic elution conditions in terms of distribution of reference compounds along the gradient. It allowed to assess the performances of both the chromatographic column (consistency of retention times and peak widths of reference compounds) and the mass spectrometer during the experiments (consistency of mass accuracy and signal intensity of reference compounds). Experiments were judged suitable for further biological interpretation if the mass accuracy of the test mix standards was below 5 ppm and if the coefficients of variation obtained from the retention times and signal intensities of the chromatographic peaks of the standards were below 5 and 25%, respectively.

(30) Cloarec, O.; Dumas, M. E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J. *Anal. Chem.* **2005**, *77*, 1282–9.

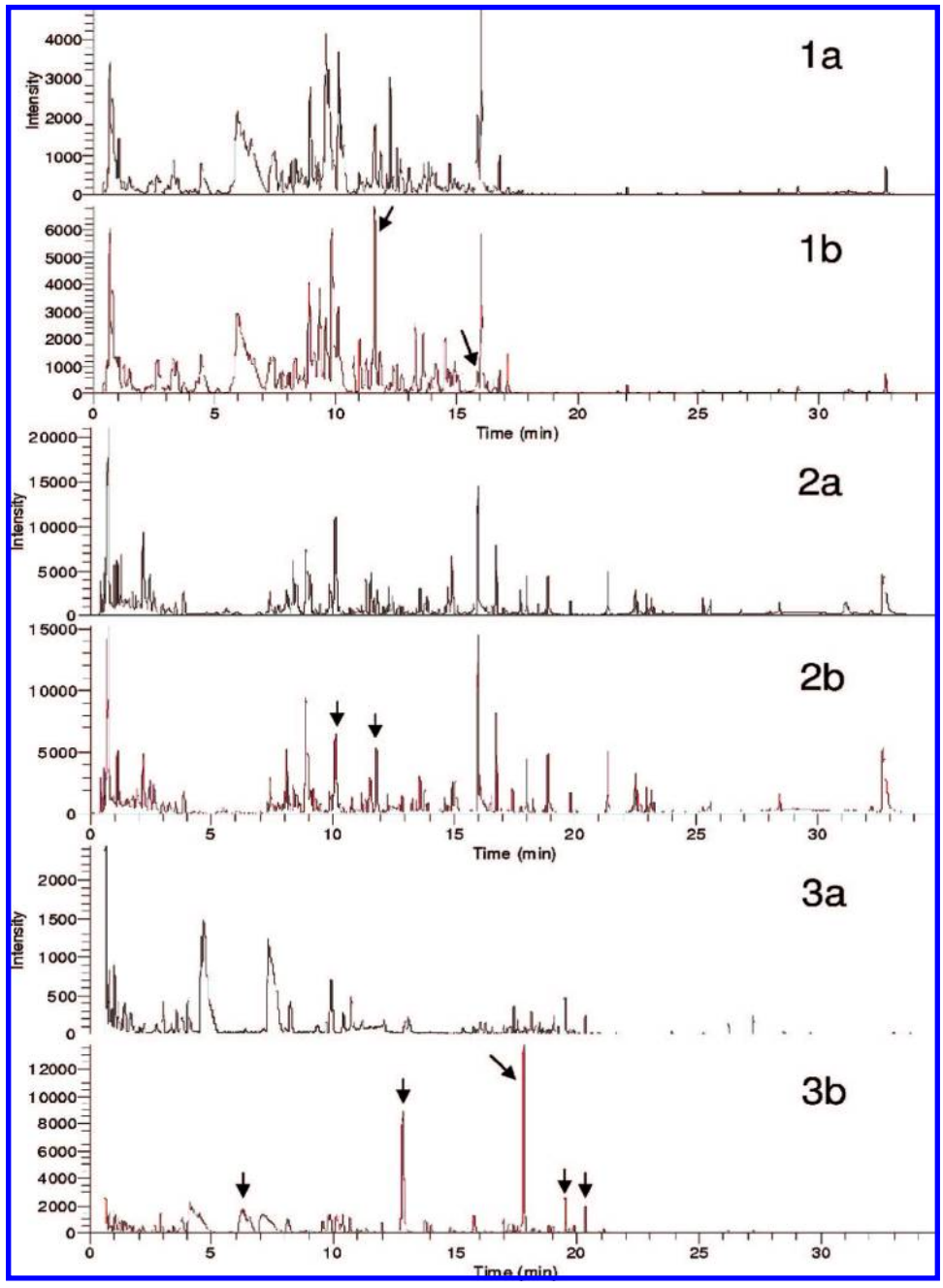


Figure 1. Representative base peak ion reconstructed chromatogram resulting from the injection of urine samples from male Wistar rats obtained before (a) and after (b) a four-day treatment with PB. Acquisition was performed in negative (1) and positive (2) ion electrospray with a UPLC-TOFMS separation on a 2.1 mm × 150 mm, 1.7μm C_{18} bonded column using a water/acetonitrile (both with 0.1% formic acid) gradient over 30 min at 500 μL/min or using a water/methanol (both containing ammonium acetate 10 mM) gradient (same chromatographic conditions) in the negative ion detection mode (3). Arrows highlight peaks that visually underwent intensity modifications.

Two sets of eluting conditions were used. Typical UPLC/TOF base peak ion chromatograms of predose (a) and treated (b) rat urine are displayed in Figure 1 for each experimental condition. No major modifications of the metabolic fingerprints were observed from visual inspection of the reconstructed chromatograms obtained in acetonitrile-acidic eluting conditions for both positive and negative electrospray modes. This was, however, not the case for the chromatograms obtained with the methanol-neutral eluting conditions, for which dramatic modifications were observed in metabolic fingerprints between predose and treated animal urine samples. This suggests that different mobile-phase compositions lead to significant variations of ionization efficiency. Furthermore, the use of different solvents allowed us to exploit complementary retention mechanisms on reversed-phase columns. Indeed, two different organic modifiers affecting preferentially hydrogen bonds (methanol) or dipole−dipole interactions (acetonitrile) may discriminate coeluting compounds in one or the other condition. Moreover, changing pH-eluting conditions modifies the ionization state of ionizable compounds during the chromatographic migration and, therefore, modifies their chromatographic retention behavior. In this way, compounds that were eluted in the dead volume under one set of conditions due to an ionized state may be retained on the column under the other set of conditions. Finally, regarding mass spectrometry, the use of

two organic modifiers is likely to modify the relative ionization efficiency of some compounds in the ESI source. Although such an experimental design did not allow us to discriminate between the effects related to either pH or solvent changes, different signals were highlighted from one condition to the other, showing that the use of different chromatographic conditions has a strong impact on the peak-picking step and is mandatory for maximizing information exhaustiveness when using a single chromatographic system.

Mass spectrometric detection was performed by using a TOF analyzer operated in the W-mode, which provided accurate mass measurements within 5 ppm, using internal calibration. This is of particular interest for the analysis of the chromatographic dead volume, which contains highly polar metabolites that are not retained on the reversed-phase column. Such metabolites represented 27, 7, and 18% of the signals that were detected in methanol-neutral eluting conditions with negative electrospray, and in acetonitrile-acidic eluting conditions with both the negative and positive modes, respectively. As a consequence, a significant part of the information that is lost with low-resolution measurements can be taken into account. However, matrix effects and ion suppressions are still more likely to occur in such circumstances and isomers remain indiscernible.

Statistical Data Reduction and Analysis. A multivariate analysis was conducted in order to visualize the data and to highlight signals allowing the discrimination of samples from treated and predose animals. Before performing statistical analysis, raw data were preprocessed in order to extract the analytically relevant variables after having removed background noise and isotopes. MarkerLynx extracted 2458 variables in the negative ion mode for methanol-neutral eluting conditions and 8524 or 3627 variables for acetonitrile-acidic eluting conditions for both the negative and positive ion modes, respectively. In the latter case, the high number of variables that were extracted in the negative ion mode could be attributed to lower signal and chemical noise intensities than those observed in the positive ion mode, thus allowing extraction of variables with a lower intensity cutoff limit. Although some of these extracted signals may be erratic, all of the discriminating variables that were investigated in the course of this study were found to be analytically relevant.

The main issues encountered during software-assisted signal treatment are first to ensure selection of all the relevant peaks without extracting peak spikes or background noise and then to avoid peak splitting. In this context, UPLC offered great advantages compared with HPLC. First, it reduced analysis time for the same or even better peak resolution, thus improving signal recoveries by limiting ion suppression in the latter case. It also improved signal-to-noise ratios due to narrower peak widths. These latter were of 9.5, 9.5, and 10 s for acidic positive, acidic negative, and neutral negative chromatographic conditions, respectively. As a consequence, UPLC allowed the detection of compounds occurring at lower levels and improved software-assisted peak detection and variable extraction compared with conventional LC/MS-based systems (data not shown). Moreover, UPLC provided good reproducibility and robustness of data generated. There were no, or at least minor, retention time shifts, and so peak realignment step was unnecessary (e.g., standard deviations on retention time were 1.63 (40 samples) and 0.55 s (20 samples) for hippuric acid and phenobarbital, respectively).

Processed data were then submitted to multivariate analysis using the software SIMCA P11 (Umetrics, Umea, Sweden). Before and after normalization (mean-centering and scaling to "unit variance" or to "Pareto variance"), the preprocessed data were first submitted to PCA, an unsupervised projection method used to visualize the structure of the data set. Strong clustering in the PCA scores between predose and treated samples was observed with the first two components, suggesting that phenobarbital administration was the main source of variance in the samples (data not shown).

Supervised methods such as projection on latent structure-discriminant analysis (PLS-DA) and orthogonal-PLS (O-PLS) were then used to facilitate the isolation of the variables responsible for the discrimination between postdose and predose samples. Two-component OPLS models were constructed for each experimental condition and led in any case to R2Y and Q2 above 0.9 and 0.85, respectively (detailed results are provided in Table S1 as Supporting Information). As an example, Figure 2 displays the results obtained with the acetonitrile-acidic eluting conditions coupled to negative ion mode detection. SIMCA returned a variable importance on projection (VIP) score, which reflects the contribution of the variables to the model. A variable is considered as important for the model when its VIP is above 1.0. To focus on highly significant variables, only those with a VIP above 1.1 were considered. Removing variables with a VIP within the range 1.0−1.1 excluded variables with, for instance, 20 zero values for a group and 16 zero values for the other. Such variables are not very meaningful because of their poor robustness, even if they exhibit a statistical significance in the discrimination between the two groups. Moreover, as the identification step is already the most time- and work-consuming, it is worth limiting the number of variables to be characterized.

The numbers of variables identified by multivariate statistical analyses, in each set of conditions, are displayed in Table 2. It is possible to classify these variables according to their presence in both the predose and treated groups or just in the treated group. Those occurring in both groups could be related to susceptibility biomarkers, i.e., endogenous metabolites whose concentrations were altered due to PB treatment (pharmacological or toxicological effects, CYP induction). In contrast, the signals that are only present in the treated group could be attributed to either PB metabolites (i.e., biomarkers of exposure) or to endogenous signals that are not detected in the predose group and whose levels are increased following PB administration. In the latter case, further MS^n experiments are required in order to assess the exogenous or endogenous origin of these signals. This is the aim of the following parts of this work, by using LC-ESI-LTQ-Orbitrap. However, a high number of variables had to be investigated, i.e., 42 signals for methanol-neutral eluting conditions, with the negative ion mode detection, and 71 signals for acetonitrile-acid eluting conditions, with both positive and negative ion mode detection. Furthermore, among the 113 (42 + 71) signals highlighted by MVA in the negative mode, only 19 of them were common to the two sets of conditions. In the same way, only 10 common variables (i.e., same retention times and back calculated molecular weights) were found among the 142 (71 + 71) signals

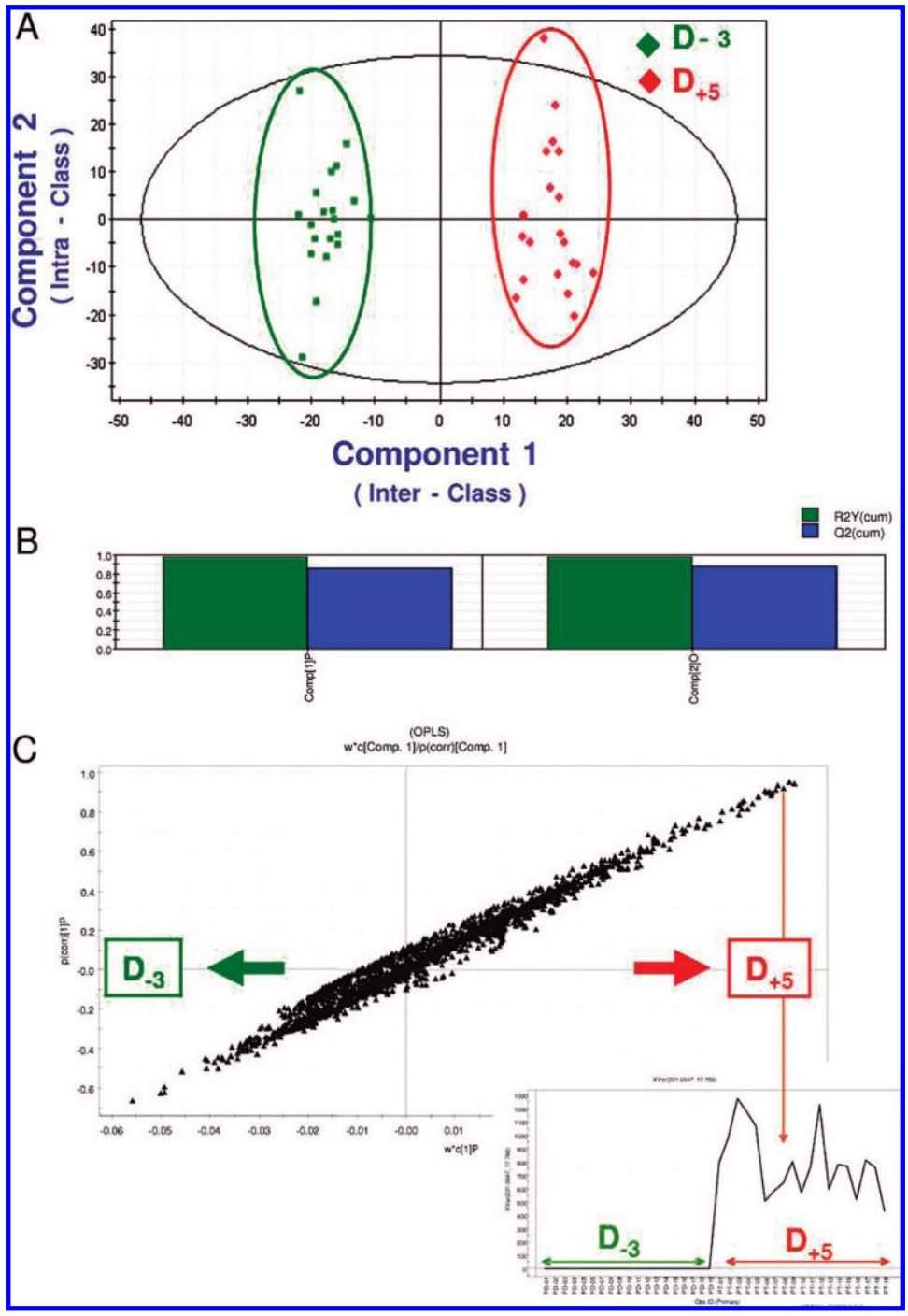


Figure 2. OPLS model based on metabolite profiles of urine samples from 20 male Wistar rats three days before (D_{-3}) and five days after (D_{+5}) a four-day treatment with daily ip injection of phenobarbital. Profiles were acquired by using UPLC-TOFMS in the negative ion mode with acetonitrile-acidic eluting conditions. (A) Scores plot. Data were mean-centered and scaled to unit variance. (B) Model overview: R2Y (cumulative fraction of the variation of the *Y* variable explained after the selected component) and Q2 (cumulative predicted fraction (cross-validation) of the variation of *Y*). (C) OPLS S-plot. Data points in the top right corner indicate increased ions responsible for variation between predose and treated samples. The inset displays the intensity of one of them in each sample. This ion is only detected in treated sample, as a result its high discriminating power.

highlighted by MVA for both the positive and the negative mode in acetonitrile-acidic conditions. A statistical tool based on autocorrelation matrices was therefore implemented to reduce the number of signals to be identified via the suppression of redundant signals.

Autocorrelation Matrices. This mathematical tool was originally developed for the processing of NMR data, for which signal redundancy is frequently observed. In this context, the aim is to group the different chemical shifts originating from a single molecule. It is based on the fact that different signals from a single metabolite should exhibit the same change in concentration in response to a perturbation. By analogy, autocorrelation matrices were here used to structure the data set by trying to group signals related to the same "parent molecule". With ESI-MS, in the course of the desolvation process, a single compound may produce several signals corresponding to the formation of adduct ions with

Table 2. Number of Discriminating Signals Identified by Multivariate Statistical Analyses

UPLC/TOF conditions[a]	extracted variables (MarkerLynx)	discriminating variables (Vip > 1.1) total	present at D_{-3} and D_5	absent at D_{-3}
I	2458	670	628	42
II	8254	407	336	71
III	3627	1433	1362	71

[a] I, neutral-water−methanol eluting conditions, negative electrospray detection. II, acid−water−acetonitrile eluting conditions, negative electrospray detection. III, acid−water−acetonitrile eluting conditions, positive electrospray detection.

cations such as sodium, potassium, or ammonium or to product ions formed by the spontaneous in-source fragmentation of the parent molecule.

Although adduct ions can easily be evidenced by accurate mass measurement and use of scripts, it is far more difficult to link each molecular species with its own in-source fragment ions in the case of coeluting compounds. Figure 3A shows the color-scaled autocorrelation matrix calculated from the 71 signals sorted as discriminant by MVA in the acetonitrile-acidic eluting conditions and positive ion mode. On the *X* and *Y* axis, signals (i.e., variables = couples RT−*m/z*) are numbered from 1 to 71 according to their retention time. As all variables are present on both axes in the same order, the matrix exhibits a symmetry line corresponding to its diagonal. Each point of this diagonal corresponds to the correlation of one variable with itself and therefore returns a correlation coefficient equal to 1. The use of autocorrelation matrices visualized the global redundancy of data (Figure 3A). This plot reveals the presence of aggregates of colored spots around the diagonal (e.g., between variables 10−20, 40−50, and 50−60). These spots correspond to high correlation coefficients between two ions because of similar behavior during PB treatment, and as they are grouped into aggregates around the diagonal, they exhibit similar retention times. They are therefore likely to represent a parent molecule with its related signals. As an example, Figure 3B displays six ions (M_1−M_6) coeluted at 13.27 min on the UPLC/TOF system for which we extracted the related part of the autocorrelation matrix (Figure 3C). The resulting correlation coefficients between the intensity of each pair of ions across all samples were used to determine whether or not there is a link between these coeluting signals. Plots of intensity of M_1, M_3, and M_4 as a function of M_2 intensity for each sample (Figure 3D) confirmed that these four ions are related to the same parent molecule, as previously attested by correlation coefficients (Figure 3C). M_2 (*m/z* 409) can be attributed to the protonated molecule (MH^+), whereas M_3 (*m/z* 391) corresponds to the loss of a molecule of water from M_2, M_4 (*m/z* 431) to the sodium cationized molecule ($[M + Na]^+$), and M_1 (*m/z* 839) to the sodium cationized dimer ($[2M + Na]^+$). For M_5 and M_6, it is difficult to establish a chemical link with either M_2 or between these two ions. However, plotting their intensities as a function of the intensity of M_2 (*m/z* 409) (Figure 3E) indicates that there is no correlation between M_2 and these two ions. Finally, plotting the intensity of M_5 as a function of M_6 intensity (Figure 3F) highlighted a link between these two ions. In this case, autocorrelation matrix calculation indicated that these six signals were due to only two parent molecules, the first one generating four signals (M_1 to M_4) and the second one generating two signals (M_5 and M_6). Furthermore, neither M_5 nor M_6 was detected using the HPLC-Orbitrap system for signal identification (Figure 3B), indicating that part of the information provided by UPLC/TOFMS has been lost, but that only one compound was undetected despite the loss of two signals when switching from one instrument to the other. However, further structural information regarding M_5 and M_6 could not be acquired to confirm or validate the link that was highlighted by autocorrelation matrices.

Thus, autocorrelation matrices allowed the grouping of ions, adduct ions, and product ions generated by in-source CID. It was possible in this way to exclude some redundant information generated by MS detection so as to focus the further identification process only on protonated or deprotonated molecules. However, this redundant information, such as cationized molecules or dimers, may be helpful to confirm the relevance of elemental compositions as it allowed several accurate mass measurements on ions related to the same molecule. Finally, when applied to the experimental conditions, this statistical tool pointed out 88 signals to be investigated from the 184 signals that were obtained for all the detection conditions.

The identification process is summarized in Table 3. Autocorrelation matrices allowed us to remove 55% of the initial variables that were highlighted in the negative electrospray mode from methanol-neutral elution conditions. Most of the variables (i.e., 16 signals or group of signals) were detected by using the LC-LTQ-Orbitrap, leading to the characterization of PB and 10 of its metabolites by further MS^n experiments. For acetonitrile-acidic conditions, ∼51% of the initial variables were removed by autocorrelation matrices. Of note, depending on the chromatographic conditions and ionization modes, 20−30% of the UPLC/TOF signals were not detected by using the LC-LTQ-Orbitrap system. Finally, around 20 and 30 variables or groups of variable were investigated, leading to the identification of 9 and 10 PB metabolites plus PB excreted unchanged by further sequential MS^n experiments, in the positive and negative ion modes, repectively. Taken together, these results indicate that although autocorrelation matrices allow removal of more than a half of the signals to be characterized, some of them, corresponding to product and adduct ions exhibiting poor correlation with protonated or deprotonated ions, still require investigation. Finally, 31 signals of interest for all the detection conditions, corresponding to either protonated or deprotonated PB-related ions, had to be investigated by further sequential MS^n experiments.

Identification of PB Metabolites Using LC-ESI-LTQ-Orbitrap. Signals highlighted by multivariate statistical analyses and by autocorrelation matrices were characterized using LC-ESI-LTQ-Orbitrap. To this end, the 20 predose samples and the 20 treated samples were pooled, thus saving time and averaging interindividual variations.

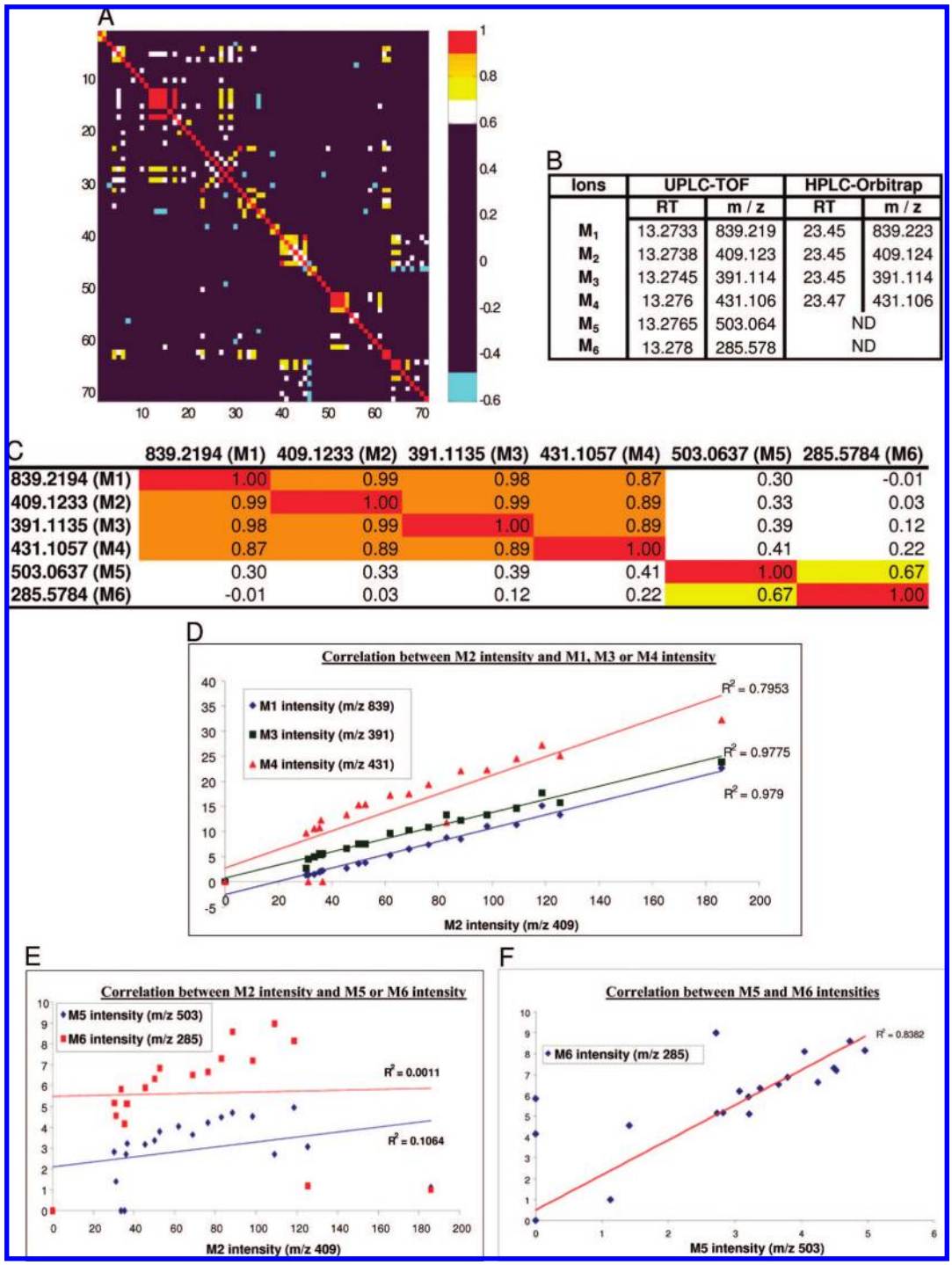


Figure 3. Overview of autocorrelation matrices. (A) Autocorrelation plot from discriminating PB-related signals detected using UPLC/TOF in the positive ion mode. (B) Extract of discriminating variables: RT−*m*/*z* pairs for UPLC/TOF and correspondence with HPLC/Orbitrap. (C) Autocorrelation matrix. (D) Plot of M_1, M_3, and M_4 ion intensities as a function of M_2 intensity across the 20 treated samples. (E) Plot of M_5 and M_6 ion intensities as a function of M_2 intensity across the 20 treated samples. (F) Plot of M_6 ion intensities as a function of M_6 intensity across the 20 treated samples.

The identification process was as follows. First, HPLC-Orbitrap and UPLC-TOF signals were matched. As the chromatographic systems were different, the matching was based on mass accuracy and also on the absence of PB-related signals in predose samples. Ions to be characterized were extracted from HPLC-Orbitrap chromatograms and compared with those occurring in UPLC-TOF within a mass range of 20 ppm.

Despite the presence of many variations in chromatographic conditions (gradient, column, UPLC vs HPLC), elution order was almost always maintained. Accurate mass filtering of the full scan mass spectra often led to the extraction across the entire chromatogram of several ions with an identical *m*/*z* ratio, but comparison of predose and treated samples and coeluted ions ensured matching between the signals of the two methods.

Formulas were then generated using the Qualbrowser elemental composition calculator (ThermoFisher) for each UPLC-TOF variable detected with the HPLC-Orbitrap system with a mass accuracy of 5 ppm. Whatever the eluting conditions, none of the

Table 3. Impact of Autocorrelation Matrices on the Number of Discriminating Variables To Be Identified

UPLC/TOF conditions[a]	discriminating variables [b]	variables or groups of variables [c]	variables or groups of variables to identify [d]	metabolites identified
I	42	19	16	11, including PB
II	71	34	28	11, including PB
III	71	35	23	9

[a] I. neutral-water−methanol eluting conditions, negative electrospray detection. II, acid−water−acetonitrile eluting conditions, negative electrospray detection. III, acid−water−acetonitrile eluting conditions, positive electrospray detection. [b] Vip > 1.1, absent at D_{-3}. [c] Highlighted by autocorrelation matrices. [d] After suppression of nondetected signals on the HPLC/LTQ-Orbitrap system.

measured masses differed from theoretical masses by more than 1.5 ppm.

Experiments performed with LTQ-Orbitrap not only confirmed results obtained from UPLC/TOFMS experiments but also dispelled ambiguities about formulas previously generated with lower mass accuracy measurements. However, even a mass accuracy of 1 ppm does not always enable a unique assignment to a molecular formula.[31] But, analysis of isotopic distribution helped to determine the nature of elements to be selected for calculating elemental compositions.[32] As an example, Figure 4 displays two ions of nominal *m/z* 263. These two ions exhibit a *m/z* difference of 0.018 and similar retention times (24.02 and 24.52 min on HPLC-Orbitrap system). One contains one sulfur atom, as attested by a 4.5% relative isotopic abundance at monoisotopic peak plus 1.9958 (theoretically 1.9959), whereas the other does not. It is therefore possible to limit the number of elements in the elemental composition of the compound to be identified and thus to decrease the number of formulas generated by the software. In fact, these two signals corresponded to fragment ions generated by in-source CID of the sulfoconjugate and the mercapturic acid conjugate of PB.

MS^n experiments are also useful to confirm or to discriminate between relevant formulas. CID spectra provide a second accurate mass measurement of the parent ion and accurate mass determination of its product ions. Product ion formulas can be generated, thus reinforcing or invalidating initial hypotheses. MS^n fragment ion or neutral loss formula determination may also highlight the presence of a particular functional group or signal a metabolite when a product ion or a neutral loss is common to the parent drug fragmentation pattern. These points are illustrated in Figures 5 and 6.

The negative ion CID spectra of the PB−mercapturic acid conjugate are shown in Figure 5. The isotope pattern analysis in full scan MS revealed the presence of a sulfur atom. MS^2 experiments removed biotransformation add-on (i.e., the *N*-acetylcysteine moiety). Finally, the MS^3 product spectrum confirmed the presence of PB with two ions *m/z* 234 and 177 corresponding to typical neutral losses for para-substituted PB (29u C_2H_5 and 43u NCHO). Of note, autocorrelation matrix showed a link between ions at *m/z* 821, 410, and 392, corresponding to $(2M + H)^+$, $(M + H)^+$, and $(M + H - H_2O)^+$, respectively. The ion at *m/z* 234 that was observed on mass spectrum and that may be a product ion from the ion at *m/z* 263, was not highlighted by MVA (i.e., not sorted as a discriminating signal). As a consequence, it was not included in the autocorrelation matrix which was built from MVA results.

Another example is given in Figure 6. In this case, the isotope pattern excluded the occurrence of elements with natural isotopes at the monoisotopic peak plus 2u, such as K, Cl, S, and P. MS^2 experiments removed the biotransformation add-on (*N*-acetylhexopyranosyl, 221u). Sequential MS^3 CID spectra demonstrated linkage of the compound with PB through a fragment ion corresponding to *p*-OH-PB (*m/z* 247). A proposed mechanism of fragmentation is presented in Figure 6D. This metabolite could correspond to an *O*-glucuronide-*N*-acetylhexopyranosylamine conjugate of PB. However, this structure must be confirmed by further investigations relying on D_2O exchange and NMR experiments. The latter example clearly highlights the limitations of conventional CID experiments as a stand-alone tool for the identification of discriminating variables identified by metabolomic approaches.

Finally, such an approach detected PB, excreted unchanged in urine, as well as 14 of its metabolites, as summarized in Table 4, and could be used to propose a metabolic pathway of PB in the rat (Scheme 1). The metabolism of phenobarbital has been investigated in vitro, in rat hepatocytes,[34] and also in vivo in rodent biofluids,[36] thus leading to the characterization of *p*- and *m*-hydroxyphenobarbital, together with the phenobarbital-*N*-glucoside (in mice),[37] the phenobarbital-*N*- (in mice),[38] and the hydroxyphenobarbital-*O*-glucuronides. A 2-phenyl-γ-butyrolactone derivative of phenobarbital was also reported from hepatocyte incubations.[34] This metabolite was detected in human biofluids following PB intoxication[39] but was not found in our experimental conditions. According to our knowledge, all the other metabolites that were highlighted by our metabolomic approach have not been identified in rats, but rather from experiments with human materials,[22,23,40–42] except for the PB-*O*-glucuronide-*N*-acetylhexopyranosylamine metabolite, that is reported here for the first time.

Even if complete characterization can rarely be achieved using exclusively MS data, ultrahigh resolution combined with sequential MS^n experiments provides great structural information that can be used to predict metabolic pathways. Of note, the fragment ions obtained from some CID experiments on PB and its related metabolites are displayed in the Table S2 (Supporting Information). Furthermore, this study showed the value of performing several acquisitions in different eluting and detection conditions. Some metabolites such as deethyl-PB, dihydroxy-PB, dihydrodiol-PB,

(31) Kind, T.; Fiehn, O. *BMC Bioinformatics* **2006**, *7*, 234.
(32) Stoll, N.; Schmidt, E.; Thurow, K. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1692–9.
(33) Levin, S. S.; Vars, H. M.; Schleyer, H.; Cooper, D. Y. *Xenobiotica* **1986**, *16*, 213–24.
(34) Verite, P.; Cave, C.; Menager, S.; Davy, J.; Andre, D.; Farnoux, C. C.; Lafont, O. *Drug Metabol. Drug Interact.* **1996**, *13*, 41–55.
(35) Caldwell, J.; Croft, J. E.; Smith, R. L.; Sneddent, W. *Br. J. Pharmacol.* **1977**, *60*, 295P–6P
(36) Anderson, G. D.; Levy, R. H. *Pharm. Res.* **1992**, *9*, 1622–8.
(37) Soine, W. H.; Soine, P. J.; England, T. M.; Ferkany, J. W.; Agriesti, B. E. *J. Pharm. Sci.* **1991**, *80*, 99–103.
(38) Neighbors, S. M.; Soine, W. H. *Drug Metab. Dispos.* **1995**, *23*, 548–52.
(39) Andersen, B. D.; Davis, F. T.; Templeton, J. L.; Hammer, R. H.; Panzik, H. L. *Res. Commun. Chem. Pathol. Pharmacol.* **1976**, *15*, 21–9.

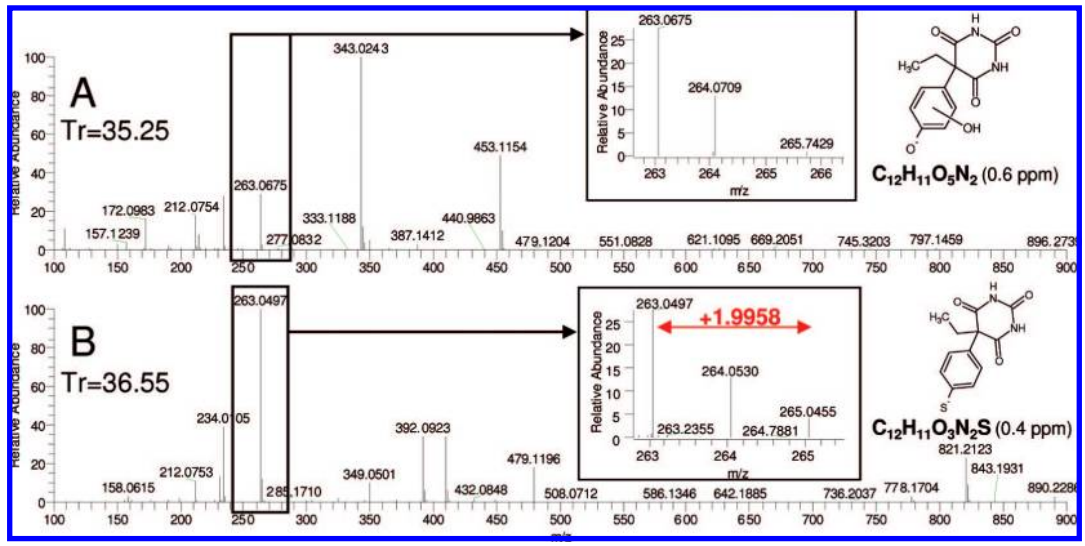


Figure 4. Orbitrap electrospray mass spectra of two LC peaks containing *m*/*z* 263: RT = 35.25 min (A) and RT = 36.55 min (B). The resolution was set to 60 000, and the mass spectra were recorded in the negative ion mode. Detailed isotopic distribution for each ion is provided as inset.

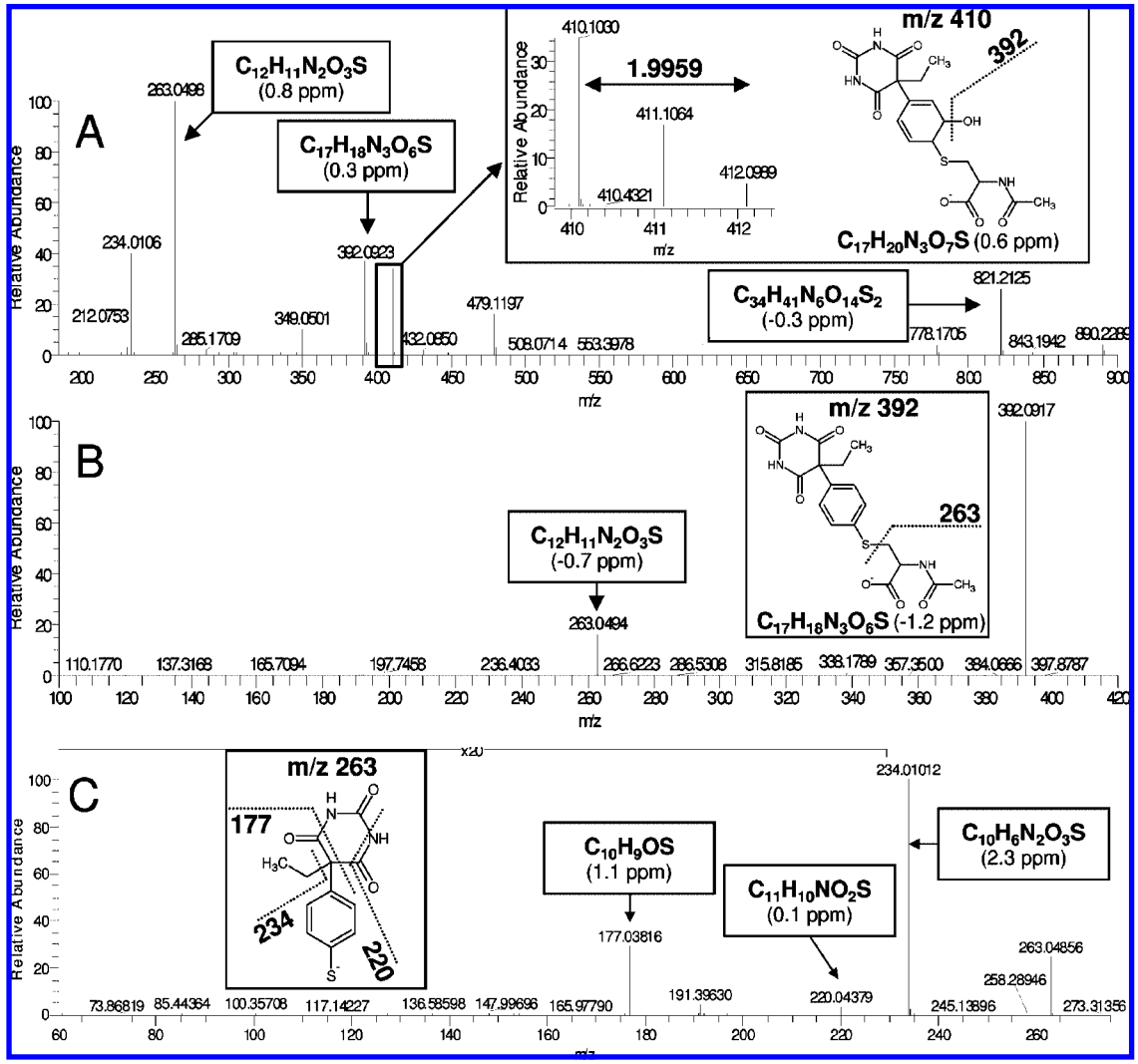


Figure 5. (A) Orbitrap electrospray mass spectrum of the LC peak containing *m*/*z* 410 recorded in the negative ion mode. The resolution was set to 60 000, and detailed natural isotopic distribution is provided as inset. (B) CID product mass spectrum of the molecular ion *m*/*z* 410 with a resolution set to 30 000. (C) Sequential MS^3 spectrum of the most intense ion produced in MS^2 spectrum of *m*/*z* 410 (*m*/*z* 392).

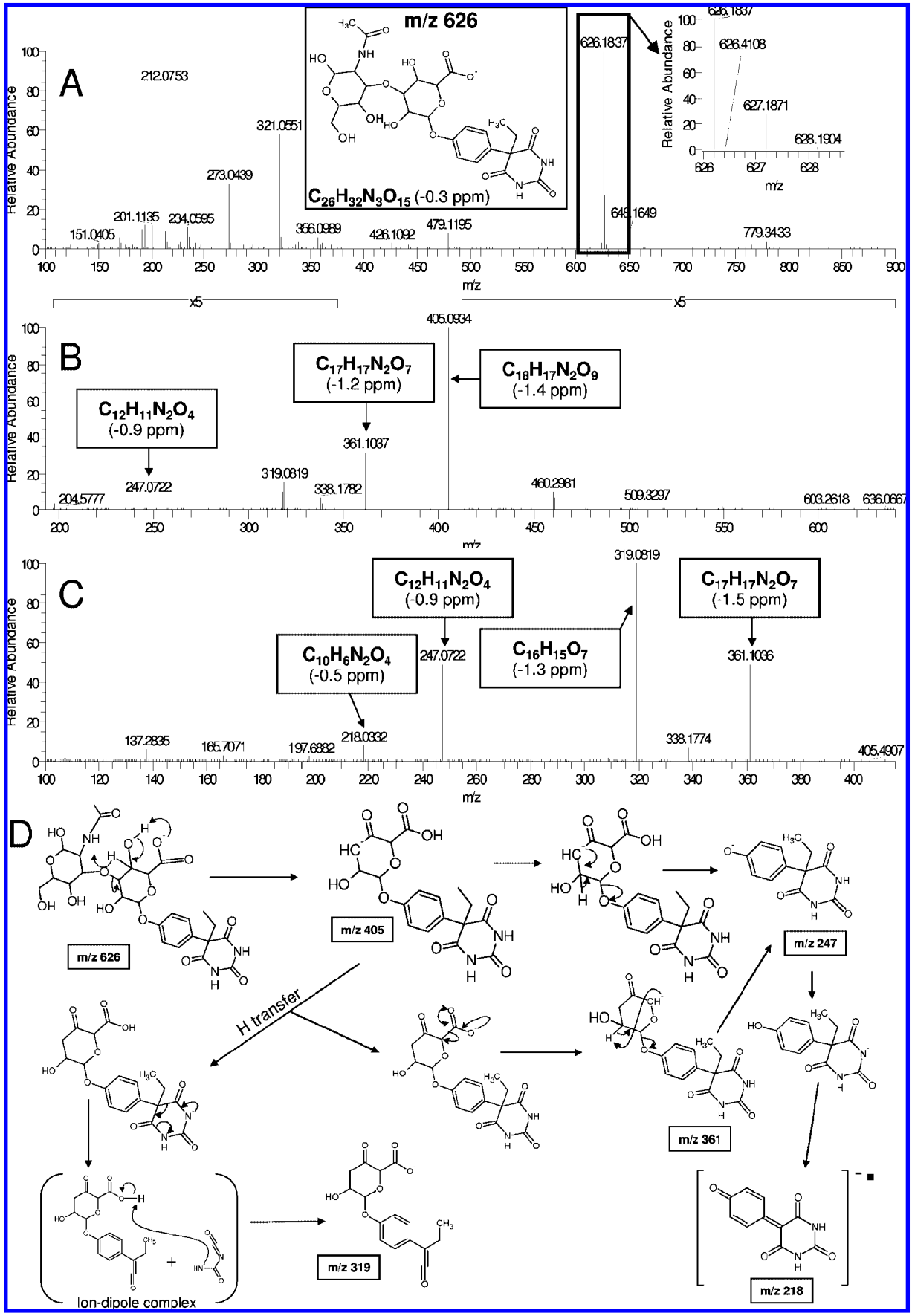


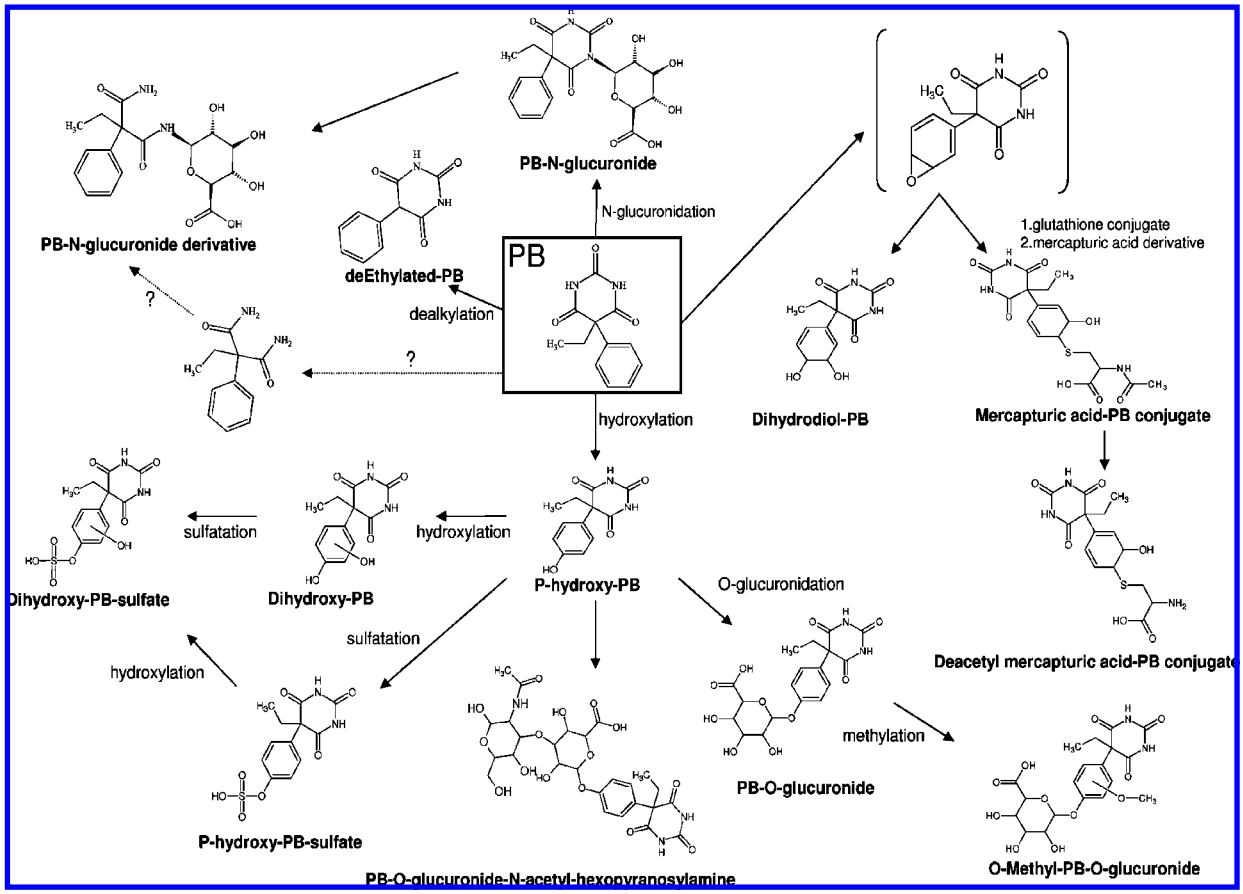
Figure 6. (A) Orbitrap electrospray mass spectrum of the peak containing *m*/*z* 626 recorded in the negative ion mode. The resolution was set to 60 000. (B) The inset shows the detailed isotopic distribution. CID product spectrum of the molecular ion (*m*/*z* 626) with resolution set to 30 000. (C) Sequential MS^3 spectrum of the most intense ion produced in MS^2 for *m*/*z* 626 (*m*/*z* 405). (D) Proposed mechanism of fragmentation.

Table 4. Overview of Detected Metabolites According to Detection Conditions

composition	putative structure	detection conditions[a]
$C_{10}H_8O_3N_2$	PB without ethyl lateral chain	I
$C_{12}H_{12}O_3N_2$	PB	I, II
$C_{12}H_{12}O_4N_2$	*p*-hydroxy-PB	I, II, III
$C_{12}H_{12}O_5N_2$	dihydroxy-PB	I
$C_{12}H_{14}O_5N_2$	dihydrodiol-PB	I[b], II
$C_{12}H_{12}O_7N_2S$	*p*-hydroxy-PB-sulfate	I, II
$C_{12}H_{12}O_8N_2S$	dihydroxy-PB-sulfate	I[b], II
$C_{17}H_{22}O_8N_2$	PB-*N*-glucuronide with an opening of the barbital moiety (loss of CO)	I, II, III
$C_{17}H_{20}O_9N_2$	hydroxy-PB-O-glucuronide with an opening of the barbital moiety (loss of CO)	II[b], III
$C_{17}H_{21}O_7N_3S$	PB-mercapturic acid conjugate	I, II, III
$C_{18}H_{19}O_9N_2$	PB-*N*-glucuronide enantiomer	I, II, III
$C_{18}H_{19}O_9N_2$	PB-*N*-glucuronide	I, II, III
$C_{18}H_{20}O_{10}N_2$	*p*-hydroxy-PB-*O*-glucuronide	I, II, III
$C_{19}H_{22}O_{11}N_2$	*O*-methyl-PB-*O*-glucuronide	I, II, III
$C_{26}H_{33}O_{15}N_3$	PB-*O*-glucuronide-*N*-acetylhexopyranosylamine	I[b], III

[a] I, neutral-water−methanol eluting conditions, negative electrospray detection. II, acid−water−acetonitrile eluting conditions, negative electrospray detection. III, acid−water−acetonitrile eluting conditions, positive electrospray detection. [b] Detected under the mentioned condition but not extracted by Markerlynx.

Scheme 1. Proposed Metabolic Pathway of Phenobarbital from Metabolites Identified in Rat Urine Samples



dihydroxy-PB-sulfate, a hydroxy-PB-O-glucuronide derivative and PB-O-glucuronide-N-acetyl-hexopyranosylamine were only observed in one of the three experimental conditions. Although these metabolites were detected in the other conditions, they were not extracted by MarkerLynx, despite careful optimizations of the parameters of this software, due to poor ionization efficiency in these conditions.

Other approaches such as neutral loss monitoring and the use of mass defect filters are also currently applied in metabolism studies.[43] However, they have some drawbacks and are not as exhaustive as the metabolomic approach. The fragmentation

(40) Kallberg, N.; Agurell, S.; Ericsson, O.; Bucht, E.; Jalling, B.; Boreus, L. O. *Eur. J. Clin. Pharmacol.* **1975**, *9*, 161–8.
(41) Soine, W. H.; Soine, P. J.; Mongrain, S. E.; England, T. M. *Pharm. Res.* **1990**, *7*, 402–6.
(42) Bernus, I.; Dickinson, R. G.; Hooper, W. D.; Eadie, M. J. *Eur. J. Clin. Pharmacol.* **1994**, *46*, 473–5.
(43) Kostiainen, R.; Kotiaho, T.; Kuuranne, T.; Auriola, S. *J. Mass Spectrom.* **2003**, *38*, 357–72.

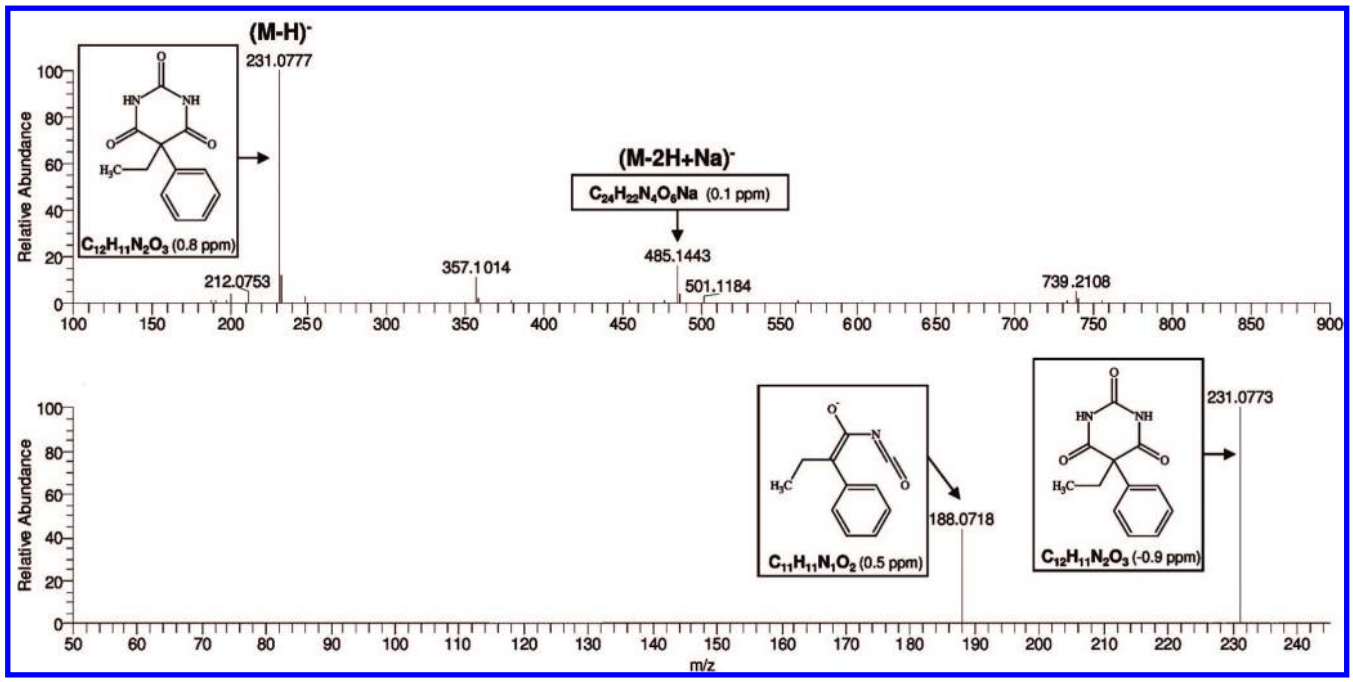


Figure 7. (A) Negative Orbitrap electrospray mass spectrum of phenobarbital. The resolution was set to 60 000. (B) CID product spectrum of PB molecular ion (resolution 30 000) showing the particular fragmentation pattern of the drug compared with its metabolites: a neutral loss of 43u corresponding to the opening of the barbital moiety was observed, whereas no neutral loss of 29u corresponding to the release of the ethyl lateral chain was found.

pattern of PB, which is shown in Figure 7, differs from that of its metabolites. This can be explained by the fact that most PB metabolites are derived from *p*-OH-PB, which is the main phase I metabolite of PB. However, the hydroxyl group, which is located in the para position of the phenyl ring, was responsible for a drastic change in the fragmentation pattern since it led to a further fragmentation of the barbital moiety and to the loss of the ethyl lateral chain, which were the common losses retrieved for other metabolites (two neutral losses of 43u instead of one and a loss of 29u in negative ion mode, respectively). It is also interesting to note that the PB mercapturic derivative produced a first fragment with sulfur in the para position, which resulted in the same fragmentation pattern. As a consequence, most of the PB metabolites that have been identified by the metabolomic approach would not have been detected by a neutral loss procedure based on PB CID fragmentation. A neutral loss procedure characteristic of phase 2 biotransformations should be more efficient. However, even in this latter case, the putative PB-*O*-glucuronide-*N*-acetylhexopyranosylamine metabolite, which does not exhibit any neutral loss related to classical phase 2 biotransformations in its MS^2 spectrum, would not have been detected.

The same kinds of drawbacks may apply to the mass defect filter strategy. Such filters are generally applied within the range ±20−50 mDa.[44] They are therefore designed to identify metabolites undergoing only minor metabolic transformations. These transformations correspond to phase I metabolites obtained via oxidation, reduction, and loss or gain of small aliphatic groups. If applied to the present results, *m/z* 626 (neg) or 628 (pos), *m/z* 381 (neg) or 383 (pos), and *m/z* 343 (neg), corresponding to PB-*O*-glucuronide-*N*-acetylhexopyranosylamine, PB-*N*-glucuronide with an opening of the barbital moiety (loss of CO) and dihydroxy-PB-sulfate, respectively, would have been discarded with a mass defect tolerance of 50 mDa. However, computational prediction of metabolite masses undergoing common biotransformations such as sulfation or glucuronidation may have helped to select these ions, but would also have led to the selection of many other endogenous compounds.

Nevertheless, metabolomics-based approaches to metabolism studies also have some drawbacks. They are complex and time-consuming since they require the development of several statistical procedures such as multivariate statistical analysis and autocorrelation matrices, followed by the acquisition and interpretation of CID spectra. Another limitation is the access to only relative quantitation. This is acceptable as far as metabolite identification is concerned. However, additional information about the abundances of the metabolites will require further experiments in order to achieve absolute quantitation. Moreover, low-intensity variables often result in the absence of CID spectra and reliable isotopic distribution, which are of value to constrain the elements entering molecular formulas. Further sample pretreatment, such as preconcentration may therefore be required to obtain a signal intensity suitable for sequential MS^n experiments or for well-defined isotopic pattern so as to identify compounds over 300 Da.

CONCLUSION

Our data illustrate a new strategy for the analysis of MS metabolomic data, especially in the field of metabolite identification. First, the use of different chromatographic conditions can drastically impact on the chromatographic migration and ionization efficiency of compounds, thus providing a way to obtain more exhaustive information during metabolic fingerprint acquisition. Second, autocorrelation matrices combined with pattern recogni-

(44) Zhu, M.; Ma, L.; Zhang, D.; Ray, K.; Zhao, W.; Humphreys, W. G.; Skiles, G.; Sanders, M.; Zhang, H. *Drug Metab. Dispos.* **2006**, *34*, 1722–33.

tion deciphered more than 50% of the MS data redundancy to focus the identification process on the most relevant discriminating ions. Third, the LTQ-Orbitrap rapidly identifies ions of interest. These results also confirm that the metabolomic approach provides effective tools for in vivo drug metabolite screening.[45] However, two main limitations remain: the absence of absolute quantification with the MS-based untargeted metabolomic approach and the impossibility of a fully comprehensive analysis of the metabolome with one single analytical tool.

In our view, a promising strategy may be to combine UPLC with Orbitrap detection. In this way, UPLC would provide robustness, better sensitivity and better software-assisted peak extraction, whereas the Orbitrap would allow accurate mass measurement and multistep fragmentation for identification purposes. However, accurate mass measurements at ultrahigh resolution (over 30 000) with the Orbitrap require scan times of ~1 s, and typical peak widths achieved with UPLC in our experimental conditions are ~10 s. This could limit the performances of the UPLC/LTQ-Orbitrap coupling for data-dependent MS^n acquisitions at ultrahigh resolution, which are very useful for ion filiation determination.

Compound identification remains a challenging task and thus highlights the crucial need for any reliable mathematical tool or complementary technique that simplifies it. Further NMR experiments are necessary to achieve comprehensive identification of compounds and to obtain quantitative results. However, NMR is less sensitive than LC/MS and often requires isolation and concentration of compounds prior to its implementation as a complementary technique to formally characterize signals initially detected by an LC/MS-based approach. Databases of compounds identified by LC/MS have to be enriched to take full advantage of the tremendous potential applications of metabolomics, as a tool for elucidation of interrelated and interregulated biological processes. In the latter application, autocorrelation may highlight correlation or anticorrelation between ions eluted at different retention times and thus could help piece together metabolic networks. Finally, once more comprehensive databases become available; the use of isotopic dilution should provide a way to achieve absolute quantitation as soon as fingerprints are acquired.

(45) Plumb, R. S.; Stumpf, C. L.; Granger, J. H.; Castro-Perez, J.; Haselden, J. N.; Dear, G. J. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2632–8.

ACKNOWLEDGMENT

This work was supported by a research contract between CEA and Technologie Servier. We thank Arnaud Martel (GIPSI, CEA) for providing a very useful visual basic program run on Excel able to search for multiple databases and to sort and link results, and François Fenaille and Elisabeth Métreau-Grognet for their help with data handling and useful comments.

SUPPORTING INFORMATION AVAILABLE

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review January 14, 2008. Accepted April 16, 2008.

AC800094P