Annotation of the Human Adult Urinary Metabolome and Metabolite Identification Using Ultra High Performance Liquid Chromatography Coupled to a Linear Quadrupole Ion Trap-Orbitrap M...

ARTICLE *in* ANALYTICAL CHEMISTRY · JULY 2012

Impact Factor: 5.64 · DOI: 10.1021/ac300829f · Source: PubMed

7 AUTHORS, INCLUDING:

Aurelie Roux
National Institutes of Health
20 PUBLICATIONS **198** CITATIONS
SEE PROFILE

Jean-François Heilier
Université catholique de Louvain
48 PUBLICATIONS **1,593** CITATIONS
SEE PROFILE

Eric Ezan
56 PUBLICATIONS **1,376** CITATIONS
SEE PROFILE

Jean-Claude Tabet
Pierre and Marie Curie University - Paris 6
416 PUBLICATIONS **5,863** CITATIONS
SEE PROFILE

analytical chemistry

Article

pubs.acs.org/ac

Annotation of the Human Adult Urinary Metabolome and Metabolite Identification Using Ultra High Performance Liquid Chromatography Coupled to a Linear Quadrupole Ion Trap-Orbitrap Mass Spectrometer

Aurelie Roux,[†] Ying Xu,[†] Jean-François Heilier,[‡,§] Marie-Françoise Olivier,[†] Eric Ezan,[†] Jean-Claude Tabet,[⊥] and Christophe Junot*[,†]

[†]CEA-Centre d'Etude de Saclay, Gif-sur-Yvette, France

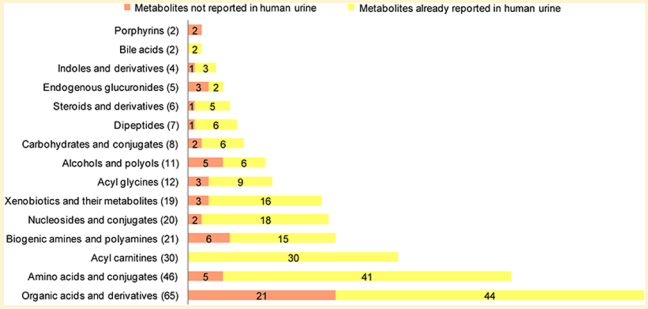
[‡]Institut National de Recherche et de Sécurité, Nancy, France

[§]Université catholique de Louvain, Institut de Recherche Expérimentale et Clinique, Louvain centre for Toxicology and Applied Pharmacology, Brussels, Belgium

[⊥]LCSOB, IPCM, UMR-CNRS 7201, UPMC Paris Universitas, Paris, France

Supporting Information

ABSTRACT: Metabolic profiles of biofluids obtained by atmospheric pressure ionization mass spectrometry-based technologies contain hundreds to thousands of features, most of them remaining unknown or at least not characterized in analytical systems. We report here on the annotation of the human adult urinary metabolome and metabolite identification from electrospray ionization mass spectrometry (ESI-MS)-based metabolomics data sets. Features of biological interest were first of all annotated using the ESI-MS database of the laboratory. They were also grouped, thanks to software tools, and annotated using public databases. Metabolite identification was achieved using two complementary approaches: (i) formal identification by matching chromatographic retention times, mass spectra, and also product ion spectra (if required) of metabolites to be characterized in biological data sets to those of reference compounds and (ii) putative identification from biological data thanks to MS/MS experiments for metabolites not available in our chemical library. By these means, 384 metabolites corresponding to 1484 annotated features (659 in negative ion mode and 825 in positive ion mode) were characterized in human urine samples. Of these metabolites, 192 and 66 were formally and putatively identified, respectively, and 54 are reported in human urine for the first time. These lists of features could be used by other laboratories to annotate their ESI-MS metabolomics data sets.



Metabolomics is a global approach that refers to the large-scale measurement of metabolites (i.e., small molecular mass organic compounds) in cells, tissues, or biofluids. It starts with the acquisition of metabolic fingerprints of a given biological medium. The metabolic fingerprints are preprocessed (i.e., removal of chemical background noise, variable alignment, peak picking) before being pretreated (centering, scaling, ...), and then processed using univariate and/or multivariate statistical analyses,[1] which highlight relevant biological information. Identification of the discriminating signals is undertaken by combining spectra analysis and database queries and by comparing, when available, at least two different physicochemical parameters of the metabolite with those of a reference compound.[2]

Because of the huge chemical diversity of metabolites, there is no universal analytical method to analyze the whole metabolome of a given biological medium. Although all analytical methods can be used to obtain metabolic fingerprints, the most commonly used are nuclear magnetic resonance spectroscopy (NMR)[3,4] and mass spectrometry (MS),[5,6] the latter often being coupled with a separation technique such as liquid or gas chromatography or capillary electrophoresis to maximize the number of detected metabolites.[7,8]

As metabolic systems may comprise hundreds to thousands of metabolites, identification of discriminating signals is one of the most challenging tasks of metabolomics.[9] This complexity of compound identification highlights the necessity to build and use databases reporting analytical information about standards. NMR was one of the first methods used for metabolomics. Nondestructive, rapid, highly robust, and NMR-based databases

Received: January 24, 2012
Accepted: July 3, 2012
Published: July 3, 2012

dx.doi.org/10.1021/ac300829f | *Anal. Chem.* 2012, 84, 6429−6437

are widespread thanks to the good reproducibility of this technique. However, NMR is less sensitive than mass spectrometry and therefore requires larger amounts of samples.

Using gas chromatography coupled with electron impact ionization mass spectrometry (EI-MS), the maximum energy transferred to molecular ions under electron ionization conditions reaches 15 eV and generates odd-electron molecular ions, which significantly and promptly dissociate in the source. Due to the broad internal energy distribution, the ion abundance is not strongly affected by the instrument type, and the pattern of fragmentation is constant. This paved the way to the building of universal spectral libraries such as those of the NIST and Wiley-Interscience[10] (Wiley Registry 9th Edition/NIST 2011 Mass Spectral Library) containing more than 800 000 mass spectra (corresponding to more than 680 000 compounds) and which are of great interest for fast compound identification. However, it is often difficult to detect the molecular species that provides information about the molecular mass of the compound of interest, which makes it difficult to characterize molecules that are not present in spectral libraries.

As an alternative to EI-MS, electrospray ionization mass spectrometry (ESI-MS)-based techniques exhibit good sensitivity and high dynamic range and provide access to the molecular mass of intact molecules from complex mixtures. The internal energy can be controlled during the desolvation step. ESI-MS techniques generate even-electron molecular ions that are relaxed at atmospheric pressure and so are more stable and do not fragment promptly, enabling their detection. The development of high and ultrahigh-resolution mass analyzers (TOF/MS and FT/MS, respectively) has enabled accurate mass measurements useful for elemental composition determination.[11] However, in most cases, ions produced in ESI sources provide only limited structural information. It is therefore necessary to perform either MS/MS or sequential MS^n experiments to go deeper into compound characterization. Unfortunately, ESI-MS/MS techniques exhibit high interinstrument variability in the generation of fragmentation patterns, and this may hamper data sharing and thus constitution of universal databases as done with EI-MS.[12] Despite this drawback, ESI-MS/MS databases began to emerge from the early 2000s (MassBank,[13] Metlin,[14] HMDB,[15] and others[16,17]).

However, before performing MS/MS experiments for metabolite identification, it appears relevant to annotate metabolomics data sets obtained from automatic peak detection software.[18,19] Indeed, ions other than molecular species are produced by ESI methods during the desolvation process, including (i) natural isotopes (^{13}C, ^{15}N, ^{18}O, ^{34}S, ^{37}Cl, ...), (ii) adduct ions with cations (e.g., Na^+, K^+, NH_4^+) and anions (e.g., Cl^-) in the positive and negative ion modes, respectively, (iii) fragment ions formed by spontaneous in-source fragmentations of the precursor ion by release of small size neutrals (loss of H_2O, CO_2, HCOOH, NH_3, ...), and (iv) multimers (e.g., $[2M + H]^+$, $[2M - H]^-$).[20] Of course, such signal redundancy is observed in metabolomics data sets. It leads to the overestimation of the number of metabolites due to irrelevant annotation proposals generated by automatic queries of public chemical and biochemical databases.[20] Conversely, in some cases, signal redundancy enables direct identification of metabolites when the ESI mass spectrum contains in-source fragment ions or specific isotopologue ions.

In this work, we have investigated the human adult urinary metabolome for metabolite identification purposes. We report on the annotation of a large number of features contained in ultrahigh performance coupled to high-resolution mass spectrometry-based metabolomics data sets obtained from automatic ion detection software.

EXPERIMENTAL SECTION

Chemicals. All analytical grade reference compounds were from Sigma (Saint Quentin Fallavier, France), except for lipids, which were from Avanti Polar Lipids (Alabaster, USA) and Steraloids (Newport, USA). The standard mixtures used for the external calibration of the MS instrument (Calmix-positive, for the positive ion mode, consisting of caffeine, L-methionyl-arginyl-phenylalanyl-alanine acetate, and Ultramark 1621, and Calmix-negative, for the negative ion mode, consisting of same mixture plus sodium dodecyl sulfate and sodium taurocholate) were from Thermo Fisher Scientific (Les Ulis, France). Acetonitrile was from SDS (Peypin, France), and formic acid was from Merck (Briare-le-Canal, France). Deionized water was filtered through a Millipore Milli-Q water purification system.

Biological Material. Data sets (i.e., raw data and associated peak tables) were obtained from two studies: the "439020" study, aimed at evaluating the impact of physiological factors such as gender, age, and body mass index on urine concentrations of metabolites from a cohort of 227 subjects, and the "439033" study, designed to evaluate the impact of collection and short-term storage conditions on the metabolite content of human urine samples.

As the results of these studies are not part of the present manuscript and will be published elsewhere, only selected raw data that are representative of investigated metabolites are provided. For the 439020 study, individual and also pooled (referred to as "QC") samples have been selected. For the 439033 study, pooled samples are available. At last, blank samples (i.e., injection of the mobile phase into the LC/MS system) are also available for both studies. Raw data and associated peaktable can be downloaded from the MetaboLights repository (http://www.ebi.ac.uk/metabolights/MTBLS20, accession number: MTBLS20). Peak tables are also provided in the Supporting Information.

Sample Preparation. Urine samples were centrifuged (1430*g* for 5 min) and then diluted in deionized water (dilution 1/5, 20 μL of urine in 80 μL of water) before analysis. Samples were randomly analyzed. A quality control sample consisting of a pool of urine samples analyzed in the course of this study was injected every 10 samples to check for the performance of the analytical system in terms of retention times, accurate mass measurements, and signal intensities.

Liquid Chromatography and Mass Spectrometry. Analyses were performed using an Accela liquid chromatographic system (Thermo Fisher Scientific, Les Ulis, France) coupled to an LTQ-Orbitrap Discovery (Thermo Fisher Scientific, Les Ulis, France) fitted with an electrospray source operated in the positive and negative ion modes. The software interface was Xcalibur (version 2.1.) (Thermo Fisher Scientific, Les Ulis, France). The mass spectrometer was calibrated before each analysis using a calibration solution provided by the manufacturer (external calibration). For singly charged ions, the mass resolution power of the analyzer was set to 30 000 ($m/\Delta m$, fwhm at 400 u) and the mass accuracy was within a 5 ppm range. The ultra high performance liquid chromatographic (UHPLC) separation was performed on a Hypersil GOLD C_{18} 1.9 μm, 2.1 mm × 150 mm column (Thermo Fisher Scientific, Les Ulis, France) equipped with an online prefilter (Interchim,

dx.doi.org/10.1021/ac300829f | Anal. Chem. 2012, 84, 6429−6437

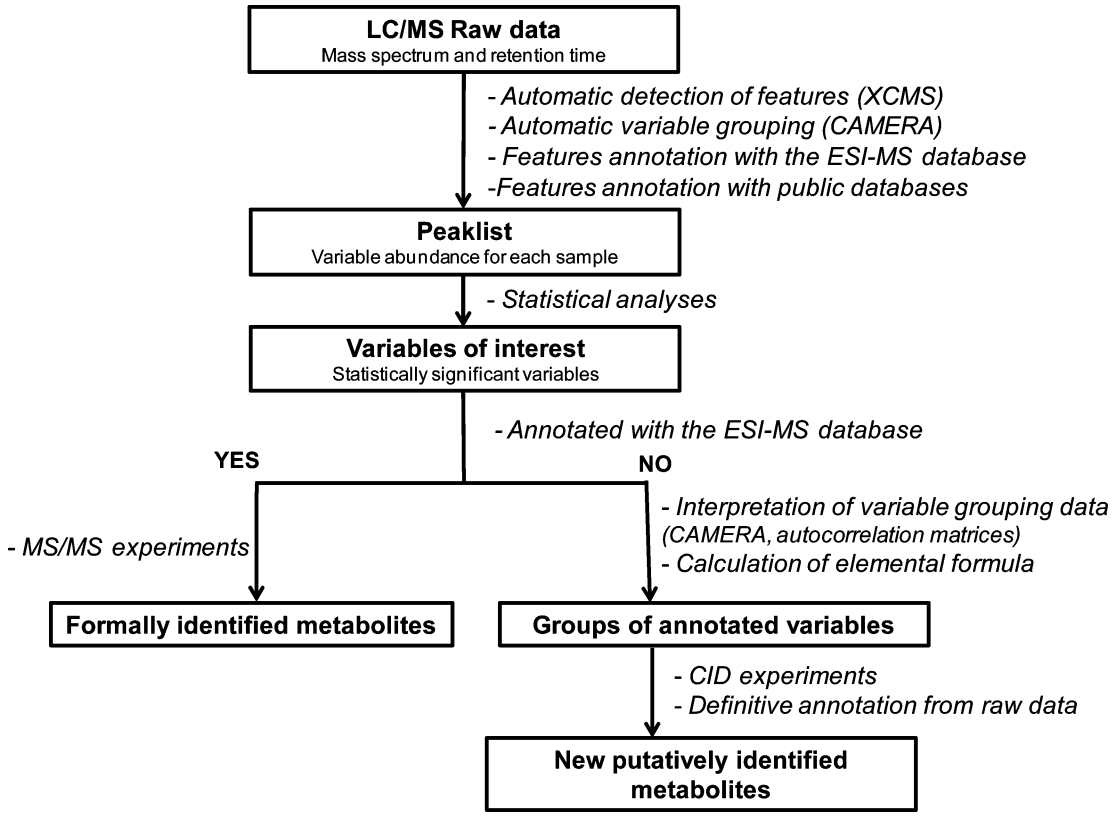


Figure 1. The different steps for the identification of metabolites from a biological data set.

Montlucon, France), and the HPLC chromatographic separation was performed on a Discovery HS F5 PFPP (PentafluorophenylPropyl) 5 μm, 2.1 mm × 250 mm column (Sigma, Saint Quentin Fallavier, France) as described previously.[21,22] Detailed experimental settings, including collision induced dissociation (CID) parameters, are reported in the Supporting Information.

Data Processing. All raw data were manually inspected using the Qualbrowser module of Xcalibur version 2.0.7 (Thermo Fisher Scientific, Les Ulis, France). Automatic peak detection and integration were performed using the XCMS software package.[23] Grouping of features was performed using CAMERA[24] software and autocorrelation matrices.[20] Public database (KEGG,[25] HMDB,[15] and Metlin[14]) annotations were performed on data sets resulting from the XCMS process with a homemade informatics tool developed in R language. Additional details can be found in the Supporting Information.

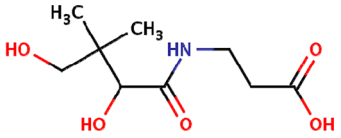
RESULTS AND DISCUSSION

Metabolite Identification Workflow. The metabolite identification in the biological sample workflow is described in Figure 1. After UHPLC/MS analysis of samples and raw data processing by automatic peak detection software, features are automatically grouped and annotated using CAMERA software,[24] the ESI-MS database of the laboratory (see Experimental Section for additional information), and public databases. Statistical tools (see the Supporting Information for experimental settings) such as principal component analysis (PCA) and projection to latent structures (PLS) are then applied to the data matrices in order to highlight variables (i.e., features) of biological interest (i.e., in our study, variables whose levels are affected by gender, age, and/or body mass index).

The identity of metabolites highlighted by feature annotation using the ESI-MS database of the laboratory is then confirmed by additional MS/MS experiments, when required (i.e., metabolites not retained by the chromatographic systems or no specific fragment ion of the metabolite in the mass spectrum). CID spectra are provided in the Supporting Information and can be downloaded from the MetaboLights repository.

Features of interest that have not been annotated using the ESI-MS database are selected for further investigations. Variables with a clear grouping and annotated using public databases are considered a priority. The identification process starts with a careful interpretation of mass spectra, to ensure that the signal of interest is really a monoisotopic ion and not a natural isotopologue or an adduct ion that could be generated during the desolvation process. To this end, results from CAMERA grouping are interpreted, checked, and supplemented by the calculation of autocorrelation coefficients taking into account signal intensities across all samples.[20] Following this ion grouping step, the protonated or deprotonated molecular species is retrieved from a block of ions by ESI mass spectrum interpretation. Its elemental formula is then calculated from accurate mass measurement using Qualbrowser (Thermo-Fisher) and its chemical formula generator. For single-charged ions of mass to charge ratio above 300 u, restrictive criteria such as nitrogen rule, valence consideration, and isotopic patterns are required in order to help to remove irrelevant proposals. The molecular mass of the putative metabolite is obtained by adding or subtracting (depending on the ionization mode) the mass of one proton (=1.00728 u) to

dx.doi.org/10.1021/ac300829f | Anal. Chem. 2012, 84, 6429−6437



m/z (experimental)	Abundance	Relative Abundance	m/z (Theoretical)	δ (ppm)	Molecule	Elemental Formula	Generated Ions	RT (min) (Theoretical)	RT (min) (experimental)
90.0549	8.1E+06	1.65	90.0550	-0.50	pantothenic acid	C3 H8 O2 N	[(M+H)-(C6H10O3)]+	4.77	4.74
202.1076	1.8E+06	0.36	202.1074	1.21	pantothenic acid	C9 H16 O4 N	[(M+H)-(H2O)]+	4.77	4.74
220.1175	**4.9E+08**	**100**	**220.1180**	**-1.85**	**pantothenic acid**	**C9 H18 N O5**	**[(M+H)]+**	**4.77**	**4.74**
221.1217	2.4E+07	4.92	221.1213	1.70	pantothenic acid	C8 13C H18 N O5	[(M+H)]+ (13C)	4.77	4.74
222.1221	1.5E+06	0.31	222.1222	-0.48	pantothenic acid	C9 H18 N O4 18O	[(M+H)]+ (18O)	4.77	4.74
242.1006	2.1E+06	0.43	242.0999	3.06	pantothenic acid	C9 H17 N Na O5	[(M+Na)]+	4.77	4.74
258.0753	4.9E+05	0.10	258.0738	5.74	pantothenic acid	C9 H17 K N O5	[(M+K)]+	4.77	4.74
259.0753	3.6E+04	0.01	259.0772	-7.47	pantothenic acid	C8 13C H17 K N O5	[(M+K)]+ (13C)	4.77	4.74

Figure 2. Annotation and formal identification of pantothenic acid in a urinary metabolomics data set. The annotated list shows molecular species and its most abundant isotope but also two in-source fragment ions including a specific neutral loss, adduct ions, and a related isotope. Of note, the [^{13}C] natural isotope of the sodium adduct has not been detected by XCMS.

that of the molecular species. Queries in chemical or metabolomic public databases can provide one or more structures that are confirmed or not by interpretation of MS/MS experiments performed in the laboratory.

Formal Metabolite Identification from Biological Data Sets Using the ESI-MS Database and Reference Compounds. Formal identification regarding the metabolomic standard initiative (MSI) criteria[2] is achieved when at least two physicochemical parameters are similar to those a reference compound. In the present case, these parameters can be at least either an accurate mass and chromatographic retention time or chromatographic retention time and a product ion spectrum.

For a relevant metabolite annotation in a biological data set, its experimental accurate mass and retention times have to match at least that of the molecular species related to the reference compound or the most abundant ion observed in the mass spectrum of the ESI-MS database. The latter can be an adduct ion or an ion generated by collision induced dissociation. A typical example of metabolite annotation in a metabolomics data set is shown in Figure 2; 8 ions were retrieved from a biological data set at the same retention time as that of the reference compound: protonated molecules and their ^{13}C and ^{18}O natural isotopes, Na and K adduct ions, loss of H_2O, and one other $C_6H_{10}O_3$ specific loss. Thus, pantothenic acid was not only annotated but also identified by the correspondence of two physicochemical parameters with the standard (accurate mass and retention time) and also by the presence of a specific fragment of pantothenic acid in the mass spectrum.

Such an approach applied to our data set of human urine led to the annotation of 825 and 659 features present in urinary metabolomics data sets recorded in positive ion (Table 1s, Supporting Information) and negative ion (Table 2s, Supporting Information) modes, respectively (see the Experimental Section for the description of the cohort, LC/MS conditions, data processing, and statistical analyses). These features correspond to 191 metabolites, 34 of which are, to our knowledge, reported in human urine for the first time (Table 3s, Supporting Information).

Of note, huge signal redundancy can be observed from ESI mass spectra. For example, 12 and 16 ions related to citric acid were detected in urine samples from ESI mass spectra recorded in the positive (Table 1s, Supporting Information) and the negative (Table 2s, Supporting Information) modes, respectively. In this case, such a high number of annotated features may be attributed to the ESI process but also to the chromatographic peak shape, which is broad and poorly resolved for citric acid.

Isomers Occurring in Biological Fluids Complicate Metabolite Identification. Urine samples and reference compounds were first of all analyzed using reverse phase chromatography (C_{18}). As many isomers from our chemical library had similar retention times, another kind of chromatographic system consisting of a pentafluorophenylpropyl (PFPP)-based stationary phase was used in a complementary way to discriminate between them. PFPP-based stationary phases have been reported to be efficient in retaining both nonpolar metabolites and basic organic molecules[26,27] and could thus help to discriminate between some isomers that may coelute in C_{18} conditions. We thus observed that 38 compounds detected in urine using the C_{18} system actually correspond to 83 metabolites that are separated in PFPP conditions. Of these 38 metabolites, 37 were eluted near the dead volume of the C_{18} chromatographic system.

For example, a single chromatographic peak related to methylnicotinamide was detected in urine at a retention time of 0.88 min in C_{18} conditions, thus suggesting the presence of a single metabolite in human urine. However, two different chromatographic peaks with similar MS and MS/MS spectra were observed with the PFPP chromatographic system. According to our ESI-MS database, the metabolite eluted at 8.10 min corresponds to 2-methylnicotinamide, whereas the other one, which exhibits a retention time of 9.17 min, corresponds to 1-methylnicotinamide (Figure 3).

This result underlines the complexity of the metabolome and demonstrates how difficult metabolite identification can be with a single chromatographic system which lacks resolution and thus leads to an underestimation of the number of detected metabolites. Furthermore, the existence of several isomeric molecules remaining chromatographically unresolved hampers further statistical and biological interpretations. Indeed, significant concentration variations cannot be attributed to one of the isomers but rather to the sum of all contributions of each molecule. Conversely, biological significance can be masked if concentration trends of each isomer are opposite.

Taken together, these results highlight that at least two complementary chromatographic methods have to be implemented in order to achieve relevant metabolite identification. They also indicate that retention factors below the value of one (i.e., the compound is not retained on the chromatographic

dx.doi.org/10.1021/ac300829f | Anal. Chem. 2012, 84, 6429−6437

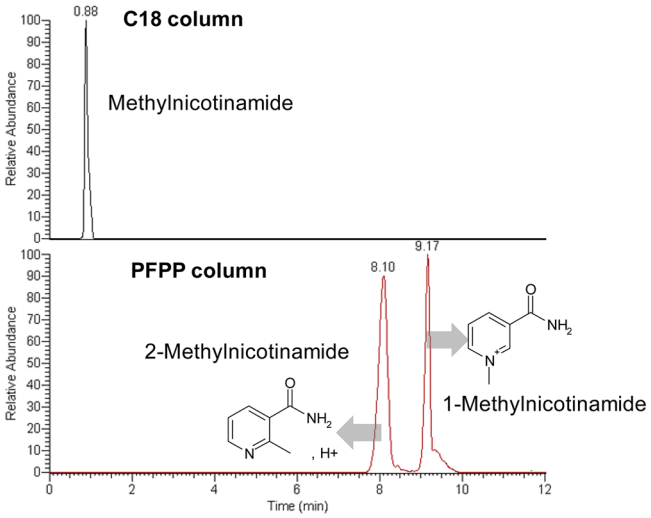


Figure 3. Analysis of methylnicotinamide isomers in human urine using two different chromatographic systems. Methylnicotinamide $[(M + H)^+]$ at m/z 137.071 was detected in urine at a retention time of 0.88 min with a C_{18} retention system. However, two chromatographic peaks were found at 8.10 and 9.17 min using the PFPP system. According to reference compounds, the 8.10 min peak corresponds to 2-methylnicotinamide, and the other one at 9.17 min corresponds to 1-methylnicotinamide.

column) cannot be considered as valuable identification criteria because of the lack of discrimination between many coeluting isomers.

Metabolite Identification from Biological Data Sets by Interpretation of CID Spectra. Many metabolites are not commercially available and cannot be formally identified. However, they can be putatively identified from biological data sets (i.e., human urine for the present study) using mathematical tools and investigation of mass spectra and product ion spectra. Using such an approach, 173 and 96 ions were annotated from our urinary metabolomics data set in the positive ion (Table 1s, Supporting Information) and the negative ion (Table 2s, Supporting Information) modes, respectively. All these ions correspond to 66 putatively identified metabolites for which reference compounds are not commercially available. Interestingly, 54 of these 66 metabolites have not yet been described in human urine according to HMDB (Table 4s, Supporting Information).

Identification of L-Urobilinogen in Human Urine. Data matrices obtained from the XCMS software and corresponding to data recorded in the negative ion mode are displayed in Figure 4A. The variable n°4679 (m/z 595.3463, retention time of 9.40 min) of the negative ion mode peaklist, which has been annotated as a molecular species by CAMERA, belongs to the group n°394 with two other ions (Figure 4A). Differential annotation indicates the presence of a molecular species, its ^{13}C natural isotope, and chloride adduct anion. Abundances observed in the mass spectrum are consistent with CAMERA annotations (Figure 4C). Correlation coefficients obtained from autocorrelation matrices are above 0.8, confirming the grouping relevance[20] (Figure 4A). For this anion, two compounds were returned by databases: C-curarine and L-urobilinogen (Figure 4B). Because of the structure of C-curarine (positively doubly charged compound that cannot be observed in negative ion mode) and also its biological relevance (it is a xenobiotic), we putatively identified L-urobilinogen (or another unreported molecule). CID experiments at a collision energy of 20% in the negative ion mode indicated losses of CO_2, $C_3H_4O_2$, $C_{17}H_{24}N_2O_3$, and $C_{17}H_{24}N_2O_5$ (Figure 4D), which are consistent with the L-urobilinogen structure. Interestingly, this metabolite, which is an uribiniloid, a product of bilirubin, has not hitherto been reported as present in human urine.

It is important to point out that some ions relevant for metabolite annotation and identification, such as the ^{18}O natural isotope of the molecular species, ^{13}C and ^{37}Cl natural isotopes of chlorine adduct ion, and $[(M - 2H) + Na]^-$, are not present in the XCMS peaklist because they were not detected by this software (Figure 4A). It is thus necessary to perform definitive metabolite annotation on raw data rather than on peaklists generated by automatic peak detection software. As a consequence, once the new metabolite is characterized and its molecular formula is known, theoretical masses of ions can then be calculated to filter the experimental mass spectrum, as performed for reference compounds (see Figure 4E). Finally, the resulting list was imported into the spectral database and included both ions detected by XCMS (highlighted in red), and those of lower abundance missed by XCMS (colored in blue).

Overview of Identified Metabolites. This work led to the characterization of 384 metabolites, including 258 that are identified (Table 3s, Supporting Information). The remaining 126 were only characterized thanks to accurate mass measurements (i.e., no MS/MS available). These 384 metabolites represent 659 and 825 annotated ions in the positive and the negative ion mode, respectively.

Of the 258 identified metabolites, 192 are formally identified thanks to reference compounds (by matching of at least two physicochemical parameters with those of reference compounds contained in the ESI-MS database), and 66 are putatively identified (based on information from databases and the interpretation of MS and CID spectra; Figure 5A, right side). 170 of them are reported in human urine according to HMDB, whereas 88 are not (Figure 5A, left side). A simultaneous search on PubMed demonstrates that 34 of these 88 metabolites have actually been reported in human urine in previous studies[28−32] (bringing the number of unreported metabolites to 54). Most of them (29 out of 34 metabolites) are acylcarnitines characterized in the urine of 6 healthy volunteers.[32] Otherwise, 100 metabolites which are part of our spectral database and supposed to be present in human urine according to HMDB were not detected in our urine samples, perhaps because of insufficient sensitivity of the method or also because of dynamic range and selectivity issues. Lastly, the ESI-MS analysis of 388 metabolites of our chemical library that are supposed to occur in urine according to HMDB is under way.

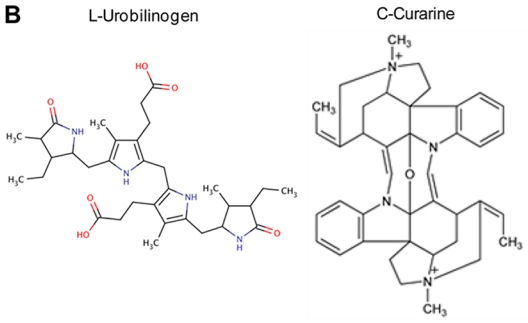
To sum up, identified metabolites are distributed over 15 chemical families (Figure 5B), including organic acids, amino acids, acylcarnitines, biogenic amines and polyamines, nucleosides, and conjugates and, to a lesser extent, xenobiotics, such as caffeine, or xenometabolites from tobacco or drugs widely used such as acetaminophen. High resolution mass spectrometers fitted with an electrospray source enable the detection of a wide range of involatile compounds with simple sample preparation protocols. Although the coupling of high resolution mass spectrometers with reverse phase chromatography is a reference method for metabolomics,[33] highly polar metabolites such as sugars and small organic acids are not efficiently retained on

dx.doi.org/10.1021/ac300829f | Anal. Chem. 2012, 84, 6429−6437

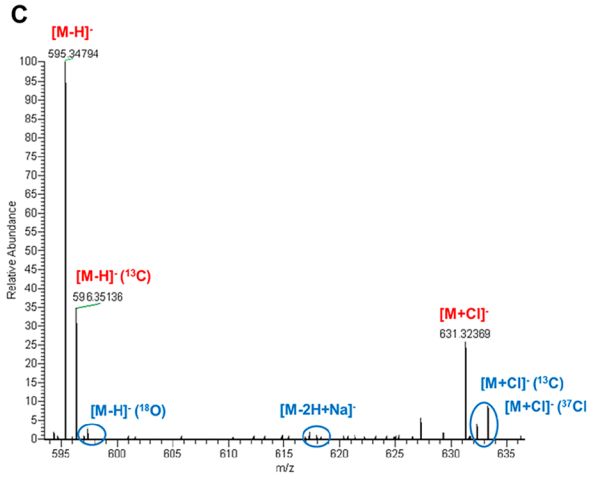
A

XCMS output			CAMERA output			Inter-sample correlation	Public database annotation
Variable number	m/z	Retention time (min)	isotopes	adduct	pcgroup		
1806	303.1443	9.33	**	**	531	NA	**
4663	593.3312	9.34	[681][M]+	**	512	NA	L-Urobilin
4668	594.3368	9.34	[681][M+1]+	**	512	NA	**
4679	595.3463	9.40	**[650][M]+**	**[M-H]-**	**394**	**1.00**	C-Curarine / L-Urobilinogen
4682	596.3514	9.40	**[650][M+1]+**	**	**394**	**0.98**	**
4878	631.3256	9.40	**	**[M+Cl]-**	**394**	**0.96**	**
3797	481.2789	9.46	**	**	552	NA	GPCho(10:0/4:0) / GPCho(12:0/2:0)
2763	381.1910	9.53	**	**	627	NA	**
3834	485.1792	9.61	**	**	544	NA	Rutaevin / Nafenopinglucuronide
1255	253.1440	9.67	**	**	556	NA	**

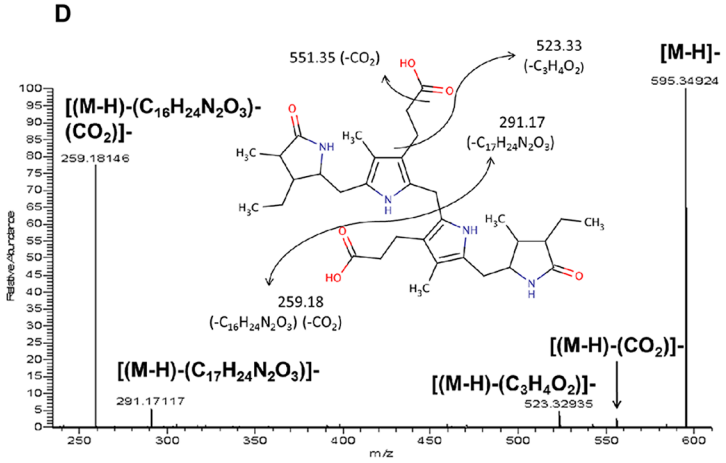
B



C



D



E

M/Z	Intensity	Relative	M/Z (theo)	Delta* (ppm)	Composition	Attribution
259.1885	1.74E+03	0.02	259.1816	3.45	C16 H23 O N2	[(M-H)-(C16H24O3N2)-(CO2)]-
595.3480	**1.21E+05**	**1.32**	**595.3501**	**0**	**C33 H47 N4 O6**	**[(M-H)]-**
596.3514	4.40E+04	0.48	596.3535	0.08	C32 13C H47 N4 O6	[(M-H)]- (13C)
597.3536	3.24E+03	0.04	597.3544	2.28	C33 H47 N4 O5 18O	[(M-H)]- (18O)
617.3341	2.19E+03	0.02	617.3321	6.91	C33 H46 N4 Na O6	[(M-2H+Na)]-
618.3318	9.14E+02	0.01	618.3354	-2.34	C32 13C H46 N4 Na O6	[(M-2H+Na)]- (13C)
631.3237	3.01E+04	0.33	631.3268	-1.33	C33 H48 Cl N4 O6	[(M+Cl)]-
632.3273	4.80E+03	0.05	632.3301	-0.98	C32 13C H48 Cl N4 O6	[(M+Cl)]- (13C)
633.3241	1.80E+03	0.02	633.3238	4.03	C33 H48 Cl0 37Cl N4 O6	[(M+Cl)]- (37Cl)

* Relatively to molecular species

Figure 4. Identification of L-urobilinogen in human urine. (A) Annotated biological peaklist in negative ion mode. The XCMS output is reported in the three first columns, including the variable number, m/z of ions, and retention time. Columns four to six represent the output of the CAMERA software; the correlation coefficient from autocorrelation matrices is reported in the column seven, and the results obtained from queries of public biochemical and metabolomics databases are in the ninth. Peaklists show cluster of ions related to the same variable corresponding to m/z 595.3463 at retention time 9.40 min. (B) Public database annotations. Variable corresponding to m/z 595.3463 at retention time of 9.40 min is annotated by public databases as C-curarine or L-urobilinogen. (C) ESI mass spectrum in negative ion mode. In addition to the most abundant ions detected by XCMS (in red, molecular species and its natural ^{13}C isotope, chloride adduct ion), other less abundant ions (in blue) are observed in the ESI mass spectrum: molecular species ^{18}O isotope, chloride adduct ion ^{13}C and ^{37}Cl isotope, and the sodium adduct ion. (D) MS/MS (CID 20%) for m/z 595.3463 in negative ion mode. MS/MS pattern shows losses of CO_2, $C_3H_4O_2$, $C_{17}H_{24}N_2O_3$, and $C_{17}H_{24}N_2O_5$ and can correspond to L-urobilinogen. (E) Peaklist of all generated ions in negative ion mode. Ions in red are those detected by XCMS, and those in blue were not.

these columns. As a consequence, it is often difficult to discriminate between isomers and the occurrence of matrix effect limits the sensitivity of detection. Despite recent improvements brought by the implementation of HILIC (hydrophilic interaction liquid chromatography)[34] and derivatization strategies,[35,36] GC/MS[37] and also now two-dimensional CG/MS[38] are the most relevant for the analysis of such metabolites in biofluids. Furthermore, apolar metabolites such as steroid and sterol derivatives are not easily analyzed using ESI-MS, and again, GC/MS is still a relevant method for these compounds,[39] although LC-MS/MS methods are emerging.[40]

Otherwise, sample preparation and insufficient dynamic concentration range are other shortcomings of our UHPLC/MS profiling method in terms of metabolic diversity coverage. This is particularly the case with phospholipids. Only few of them were detected in our experimental conditions, although they ionize well under electrospray conditions.[41] As a matter of fact, Kim et al. reported on the detection of a few tenths of phosphatidylcholine and phosphatidylethanolamine in human urine. However, they used a Folch extraction protocol followed by a sample concentration step, which are more appropriate for lipid detection than the simple dilution protocol used in the present study.[42]

Regarding metabolite identification, even though the ESI-MS process generates many ions from a single molecule, these ions often provide limited information about chemical structures, with the exception of in-source fragment ions and [M + 2] isotopes such as ^{34}S. Actually, these latter ions were only detected in about 20% of metabolites putatively identified in the course of the present study. In most situations, information

dx.doi.org/10.1021/ac300829f | Anal. Chem. 2012, 84, 6429−6437

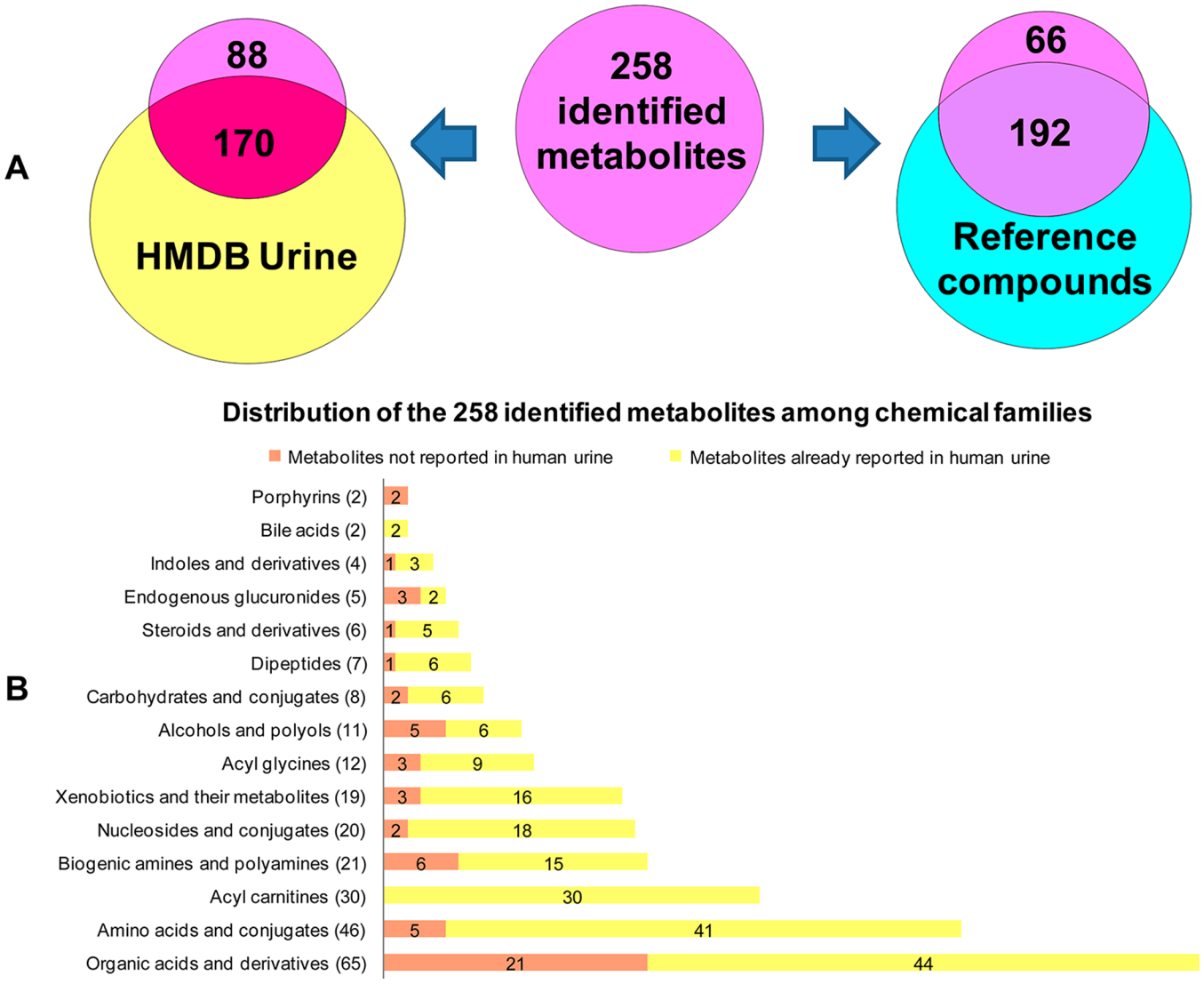


Figure 5. Overview of identified metabolites. (A) Venn diagrams of the metabolites identified in this study and metabolites reported in HMDB as present in human urine. This study reports 258 identified metabolites, including 192 identified thanks to reference compounds and 66 directly from biological data. Of these 258 metabolites, 170 are reported in urine according to HMDB, while 88 are not. (B) Distribution of the 258 identified metabolites according to their chemical families and their occurrence in human urine (according to HMDB and PubMed).

provided by LC/MS (polarity, accurate mass, natural isotopes, and in-source fragment ions when available) and the CID spectra are sufficient for putative identification of a metabolite. However, in some cases, it is difficult to discriminate between isomers with the same pattern of fragmentation and a very close polarity. The only alternative is their chemical synthesis if they are not commercially available.

CONCLUSION

The aim of this work was to annotate the human adult urinary metabolome and to identify metabolites using UHPLC/MS combined with mathematical and informatics tools. Metabolite identification in urine samples was achieved using two complementary approaches: (i) formal metabolite identification using reference compounds from a chemical library and (ii) identification from biological data sets of metabolites that are not commercially available. By these means, 384 metabolites were characterized in urine. Of these metabolites, 192 were formally identified by matching at least two of their physicochemical parameters to those of reference compounds; 66 were putatively identified, on the basis of the public database annotation and the interpretation of CID spectra, and 126 were characterized on the basis of the public database annotation and only accurate mass measurements. Of the identified metabolites, 54 are reported in human urine for the first time.

The identified metabolites are distributed over 15 chemical families, including organic acids, amino acids and derivatives, and acylcarnitines, which are among the most represented and also the most frequently observed in this cohort. Of course, this picture of the metabolic diversity observed in urine is clearly dependent on our data treatment procedure, which focused on the most frequently observed metabolites. Other data mining will have to be implemented to complete this picture and to improve the detection and identification of xenometabolites which could be of interest to document host−environment interactions.

This study has also raised the issue of criteria for identifying metabolites in biofluids. It is conventionally accepted in the research community that a metabolite is positively identified when it has at least two physicochemical parameters identical to

dx.doi.org/10.1021/ac300829f | Anal. Chem. 2012, 84, 6429−6437

those of a synthesized reference compound. Such a criterion is unfortunately not suitable for most situations, as many compounds are not commercially available and also because of the biological presence of many isomers. The presence of isomers in biological fluids raises the issue of chromatographic resolution. Indeed, many isomers are coeluted and/or not retained by a single chromatographic system, which leads to an underestimation of the number of metabolites occurring in biological media. At least two different separation methods should therefore be used to achieve the largest metabolome coverage and to refine identification proposals with regards to isomers.

Finally, our study led to the annotation of 659 and 825 features related to formally and putatively identified metabolites in human urine, in the positive and negative mode, respectively. These lists of annotated features could be used by other laboratories to annotate data sets obtained from any kind of high-resolution mass spectrometers fitted with an ESI source, before performing metabolite identification using MS/MS databases. Due to interinstrument variability, some annotated ions observed in our experimental conditions will not be retrieved by other laboratories. This is the reason why such lists of annotated features have to be publicly available in order to be shared and enriched.

■ ASSOCIATED CONTENT

Supporting Information

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

■ AUTHOR INFORMATION

Corresponding Author

*Phone: 33-1-69-08-43-66. Fax: 33-1-1-69-08-59-07. E-mail: christophe.junot@cea.fr.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

A.R. is supported by a grant from the Commissariat à l'Energie Atomique (CEA) and the Délégation Générale à l'Armement (DGA). The authors would like to thank Dr. Michel Tournier (Occupational Medicine Department of CEA-Saclay), and Drs. Xavier Millot and Catherine Bourcier and their staff (Clinical Chemistry laboratory of CEA-Saclay) for providing us with urine samples. The authors would also like to thank Arnaud Martel and Cyrille Petat for informatics development and Drs. François Fenaille and Benoit Colsch for helpful discussion. At last, the authors wish to thank the reviewers for their comments and Dr. Reza Salek for help regarding the Metabolights repository.

■ REFERENCES

(1) Sumner, L. W.; Urbanczyk-Wochniak, E.; Broeckling, C. D. *Methods Mol. Biol.* **2007**, *406*, 409−436.
(2) Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W. M.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane, A. N.; Lindon, J. C.; Marriott, P.; Nicholls, A. W.; Reily, M. D.; Thaden, J. J.; Viant, M. R. *Metabolomics* **2007**, *3*, 211−221.
(3) Holmes, E.; Foxall, P. J.; Nicholson, J. K.; Neild, G. H.; Brown, S. M.; Beddell, C. R.; Sweatman, B. C.; Rahr, E.; Lindon, J. C.; Spraul, M. *Anal. Biochem.* **1994**, *220*, 284−296.
(4) Nicholson, J. K.; Wilson, I. D. *Prog. Nucl. Magn. Reson. Spectrosc.* **1989**, *21*, 449−501.
(5) Dumas, M. E.; Debrauwer, L.; Beyet, L.; Lesage, D.; Andre, F.; Paris, A.; Tabet, J. C. *Anal. Chem.* **2002**, *74*, 5393−5404.
(6) Lafaye, A.; Junot, C.; Ramounet-Le Gall, B.; Fritsch, P.; Ezan, E.; Tabet, J. C. *J Mass Spectrom.* **2004**, *39*, 655−664.
(7) Bowers, L. D.; Sanaullah. *J. Chromatogr., B* **1996**, *687*, 61−68.
(8) Monton, M. R. N.; Soga, T. *J. Chromatogr., A* **2007**, *1168*, 237−246.
(9) Werner, E.; Heilier, J. F.; Ducruix, C.; Ezan, E.; Junot, C.; Tabet, J. C. *J.Chromatogr., B* **2008**, *871*, 143−163.
(10) McLafferty, F. W.; Stauffer, D. A.; Loh, S. Y.; Wesdemiotis, C. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 1229−1240.
(11) Kind, T.; Fiehn, O. *BMC Bioinf.* **2007**, *8*, 105.
(12) Jansen, R.; Lachatre, G.; Marquet, P. *Clin. Biochem.* **2005**, *38*, 362−372.
(13) Horai, H.; Suwa, K.; Arita, M.; Nihei, Y.; Nishioka, T. Mass Spectral Database for Metabolome Analysis. *55th ASMS Conference on Mass Spectrometry and Allied Topics*, Indianapolis, Indiana, 2007.
(14) Smith, C. A.; O'maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. *Ther. Drug Monit.* **2005**, *27*, 747−751.
(15) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M. A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; Macinnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. *Nucleic Acids Res.* **2007**, *35*, D521−D526.
(16) Marquet, P.; Venisse, N.; Lacassie, E.; Lachatre, G. *Analusis* **2000**, *28*, 925−934A.
(17) Oberacher, H.; Pavlic, M.; Libiseller, K.; Schubert, B.; Sulyok, M.; Schuhmacher, R.; Csaszar, E.; Kofeler, H. C. *J. Mass Spectrom.* **2009**, *44*, 494−502.
(18) Brown, M.; Dunn, W. B.; Dobson, P.; Patel, Y.; Winder, C. L.; Francis-McIntyre, S.; Begley, P.; Carroll, K.; Broadhurst, D.; Tseng, A.; Swainston, N.; Spasic, I.; Goodacre, R.; Kell, D. B. *Analyst* **2009**, *134*, 1322−1332.
(19) Mohamed, R.; Varesio, E.; Ivosev, G.; Burton, L.; Bonner, R.; Hopfgartner, G. *Anal. Chem.* **2009**, *81*, 7677−7694.
(20) Werner, E.; Croixmarie, V.; Umbdenstock, T.; Ezan, E.; Chaminade, P.; Tabet, J. C.; Junot, C. *Anal. Chem.* **2008**, *80*, 4918−4932.
(21) Darghouth, D.; Koehl, B.; Madalinski, G.; Heilier, J. F.; Bovee, P.; Xu, Y.; Olivier, M. F.; Bartolucci, P.; Benkerrou, M.; Pissard, S.; Colin, Y.; Galacteros, F.; Bosman, G.; Junot, C.; Romeo, P. H. *Blood* **2011**, *117*, e57−e66.
(22) Forquin, M. P.; Hebert, A.; Roux, A.; Aubert, J.; Proux, C.; Heilier, J. F.; Landaud, S.; Junot, C.; Bonnarme, P.; Martin-Verstraete, I. *Appl. Environ. Microbiol.* **2011**, *77*, 1449−1459.
(23) Smith, C. A.; Want, E. J.; O'maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779−787.
(24) Kuhl, C.; Tautenhahn, R.; Böttcher, C.; Larson, T. R.; Neumann, S. *Anal. Chem.* **2012**, *84* (1), 283−289.
(25) Ogata, H.; Goto, S.; Sato, K.; Fujibuchi, W.; Bono, H.; Kanehisa, M. *Nucleic Acids Res.* **1999**, *27*, 29−34.
(26) Nichthauser, J.; Stepnowski, P. *J. Chromatogr. Sci.* **2009**, *47*, 247−253.
(27) Havlikova, L.; Matysova, L.; Hajkova, R.; Satinsky, D.; Solich, P. *Talanta* **2008**, *76*, 597−601.
(28) Duran, M.; Ketting, D.; Vanvossen, R.; Beckeringh, T. E.; Dorland, L.; Bruinvis, L.; Wadman, S. K. *Clin. Chim. Acta* **1985**, *152*, 253−260.
(29) Jellum, E.; Horn, L.; Thoresen, O.; Kvittingen, E. A.; Stokke, O. *Scand. J. Clin. Lab. Invest.* **1986**, *46*, 21−26.
(30) Ogata, M.; Yamasaki, Y.; Kawai, T. *Int. Arch. Occup. Environ. Health* **1986**, *58*, 197−202.

(31) VanBocxlaer, J. F.; Clauwaert, K. M.; Lambert, W. E.; DeLeenheer, A. P. *Clin. Chem.* **1997**, *43*, 627−634.
(32) Zuniga, A.; Li, L. A. *Anal. Chim. Acta* **2011**, *689*, 77−84.
(33) Dettmer, K.; Aronov, P. A.; Hammock, B. D. *Mass Spectrom. Rev.* **2007**, *26*, 51−78.
(34) Zhang, T.; Creek, D. J.; Barrett, M. P.; Blackburn, G.; Watson, D. G. *Anal. Chem.* **2012**, *84*, 1994−2001.
(35) Guo, K.; Li, L. *Anal. Chem.* **2009**, *81*, 3919−3932.
(36) Guo, K.; Li, L. *Anal. Chem.* **2010**, *82* (21), 8789−8793.
(37) Pasikanti, K. K.; Ho, P. C.; Chan, E. C. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 2984−2992.
(38) Almstetter, M. F.; Oefner, P. J.; Dettmer, K. *Anal. Bioanal. Chem.* **2012**, *402*, 1993−2013.
(39) Krone, N.; Hughes, B. A.; Lavery, G. G.; Stewart, P. M.; Arlt, W.; Shackleton, C. H. *J. Steroid Biochem. Mol. Biol.* **2010**, *121*, 496−504.
(40) Griffiths, W. J.; Wang, Y. *Biochim. Biophys. Acta* **2011**, *1811*, 784−799.
(41) Pulfer, M.; Murphy, R. C. *Mass Spectrom. Rev.* **2003**, *22*, 332−364.
(42) Kim, H.; Min, H. K.; Kong, G.; Moon, M. H. *Anal. Bioanal. Chem.* **2009**, *393*, 1649−1656.