



HAL
open science

Chemical Exposure Highlighted without Any A Priori Information in an Epidemiological Study by Metabolomic FT-ICR-MS Fingerprinting at High Throughput and High Resolution

Baninia Habchi, Sandra Alves, Sylvie Streel, Michèle Guillaume, Anne-Françoise Donneau, Brice Appenzeller, Douglas Rutledge, Alain Paris, Estelle Rathahao-Paris

► **To cite this version:**

Baninia Habchi, Sandra Alves, Sylvie Streel, Michèle Guillaume, Anne-Françoise Donneau, et al.. Chemical Exposure Highlighted without Any A Priori Information in an Epidemiological Study by Metabolomic FT-ICR-MS Fingerprinting at High Throughput and High Resolution. *Chemical Research in Toxicology*, 2023, 36 (10), pp.1592-1598. 10.1021/acs.chemrestox.3c00158 . hal-04465837

HAL Id: hal-04465837

<https://hal.science/hal-04465837>

Submitted on 29 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Chemical exposure highlighted without any *a priori*
information in an epidemiological study by
metabolomic FT-ICR-MS fingerprinting at high
throughput and high resolution

Baninia Habchi[†], Sandra Alves^{†}, Sylvie Streef[‡], Michèle Guillaume[‡], Anne-Françoise
Donneau[‡], Brice M.R. Appenzeller[§], Douglas N. Rutledge^{#,1}, Alain Paris^{*#}, Estelle Rathahao-
Paris^{*†,¥}*

[†] Sorbonne Université, Faculté des Sciences et de l'Ingénierie, Institut Parisien de Chimie
Moléculaire (IPCM), 75005 Paris, France;

[‡] Department of Public Health, University of Liège, 4000 Liège, Belgium;

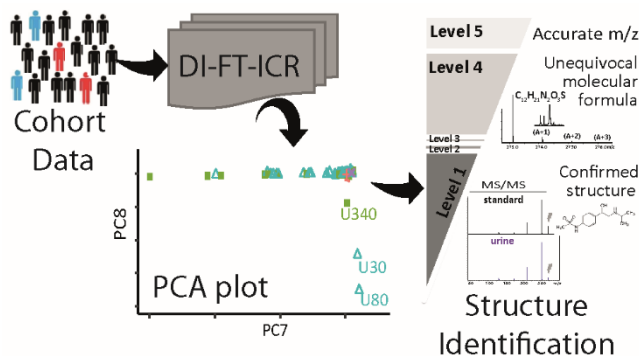
[§] Human Biomonitoring Research Unit, Luxembourg Institute of Health (LIH), 4354, Esch-sur-
Alzette, Luxembourg;

[#] Muséum national d'Histoire naturelle, MCAM, UMR7245 CNRS - MNHN, 75005 Paris,
France;

¹ Université Paris-Saclay, Faculté de Pharmacie, 91400 Orsay, France

[¥] Université Paris-Saclay, CEA, INRAE, Département Médicaments et Technologies pour la
Santé (DMTS), 91191 Gif-sur-Yvette, France

KEYWORDS. Chemical exposome, Epidemiology, Metabolomics, High-Throughput Analysis, Direct Introduction High Resolution Mass Spectrometry (DI-HRMS).



ABSTRACT

Epidemiological studies aim to assess associations between diseases and risk factors. Such investigations involve a large sample size and require powerful analytical methods to measure the effects of risk factors, resulting in a long analysis time. In this study, chemical exposure markers were detected as the main variables strongly affecting two components coming from a Principal Component Analysis (PCA) exploration of the metabolomic data generated from urinary samples collected on a cohort of about 500 individuals using direct introduction coupled with a Fourier Transform Ion Cyclotron Resonance (FT-ICR) instrument. The assignment of their chemical identity was first achieved based on their isotopic fine structures detected at very high resolution ($R_p > 900,000$). Their identification as dimethylbiguanide and sotolol was obtained at level 1 thanks to the available authentic chemical standards, tandem mass spectrometry (MS/MS) experiments and CCS (Collision Cross Section) measurements. Epidemiological data confirmed

that the subjects discriminated by PCA had declared to be prescribed these drugs for either type-II diabetes or cardiac arrhythmia. Concentrations of these drugs in urine samples of interest were also estimated by rapid quantification using an external standard calibration method, direct introduction and MS/MS experiments. Regression analyses showed a good correlation between the estimated drug concentrations and the scores of individuals distributed on these specific PCs. The detection of these chemical exposure markers proved the potential of the proposed high-throughput approach without any prior drug exposure knowledge as a powerful emerging tool for rapid and large-scale phenotyping of subjects enrolled in epidemiological studies to rapidly characterize the chemical exposome and the adherence to medical prescriptions.

INTRODUCTION

Environmental epidemiological studies aim to establish possible relationships between health effects or the occurrence of certain diseases and some external risk factors such as chronic or sub-chronic exposure to chemicals as toxicants or drugs¹. In this way, the collected data defined as metadata, including clinical information and lifestyle details (*e.g.*, nutrient intakes, drug consumption or occupational environment) of participants, are always associated with biobanks of collected biological samples (*e.g.*, urine, blood, etc.). These metadata represent a pivotal information core needed to detect strong canonical links with other information sets (genome, phenome, exposome, microbiome, etc.). Their quality must therefore be rigorously assessed, especially when the association between some environmental factors and a given health risk is strongly suspected².

Recent analytical developments in metabolomics offer attractive opportunities for large-scale population phenotyping and, in particular, for the detection of more and more molecules present

in biological matrices, including xenobiotics and their metabolites³. The most commonly used analytical methods involve a chromatographic system (*e.g.*, liquid or gas chromatography, LC or GC) coupled with mass spectrometry (MS)⁴. This separation step prior to mass analysis offers the advantages of i) a striking reduction in matrix effects within the mass spectrometer, ii) the detection of low concentration compounds, and iii) the separation of certain isomers, enabling comprehensive analysis of complex biological matrices^{4,5}. However, such an approach is less suitable for large-scale analyses such as in epidemiological studies because its throughput is rather limited by the analysis time of a large number of samples. Alternatively, Direct Introduction coupled with Mass Spectrometry (DIMS) allowing data acquisition in a few minutes per sample is a very attractive technique for epidemiological studies⁶. Various injection methods, such as direct sample infusion, nanospray injection or flow injection analysis (FIA) could be used for this purpose. Nanospray injection has the main advantages of small injection sample volume, low flow rate and no carryover problem; a system like Advion nanomate can be easily used to automate injections. Yet, all MS platforms are equipped with a high-performance liquid chromatography (HPLC) system, FIA can also be performed using its autosampler to automate sample injections under isocratic elution conditions. In addition, High Resolution Mass Spectrometry (HRMS) offers the possibility of discriminating isobaric compounds (*i.e.* analytes with very close mass-to-charge (m/z) ratios). Here, the use of a Fourier Transform Ion Cyclotron Resonance (FT-ICR) instrument equipped with a dynamically harmonized cell provides high dynamic range and resolving power (R_p), allowing the detection of a greater number of ions^{7,8}. Therefore, such instrumentation greatly improves the quality of the data generated in terms of both the number of detectable peaks and accurate mass measurements^{6,9}.

The objective of this study is to explore the metabolomic data of urine samples collected in the epidemiological study called NESCaV (“Nutrition, Environment and Cardiovascular Health”), which aims to build an epidemiological surveillance tool to establish baseline data on the prevalence of cardiovascular risks related to hypertension, dyslipidemia, obesity, diabetes, smoking and sedentary behaviour¹⁰. The high analytical quality of the data generated by our high throughput DIMS approach using an FT-ICR instrument has been demonstrated previously¹¹. Here, Principal Component Analysis (PCA) was applied to the acquired metabolomic data. The PCA results revealed exposure to certain chemicals or drugs such as dimethylbiguanide and sotalol. Then, epidemiological metadata corroborated the results obtained using our high throughput phenotyping approach without any prior knowledge of drug prescription. Regression analyses were also performed to examine correlations between the PCA scores and the drug concentrations estimated by an external standard quantitation using direct introduction-tandem mass spectrometry (DI-MS/MS).

EXPERIMENTAL SECTION

Chemicals and samples

Ultrapure water with a resistivity of 18.2 MΩ.cm was produced by a Select HP water purification system (Purite France Eau, Lormont, France). Methanol (UHPLC grade) was acquired from VWR Chemicals (Fontenay-sous-Bois, France). 1,1-dimethylbiguanide hydrochloride and sotalol hydrochloride were purchased from Sigma-Aldrich (Saint Quentin Fallavier, France). For external standard quantification, calibration solutions were prepared in methanol/water (1:1, v/v) at different concentrations of 1,1-dimethylbiguanide hydrochloride or sotalol ranging from 0.1 ng.µL⁻¹ to 5 ng.µL⁻¹.

The methodology used for cohort data collection has been published elsewhere¹⁰. NESCaV is a trans-regional project involving Wallonia in Belgium, Lorraine in France, the Grand-Duchy of Luxembourg, and Saarland in Germany. Here, only the human urine samples collected in Wallonia, corresponding to 506 samples collected from subjects aged 20 to 71 years, were phenotyped using the acquisition method described below. All urine samples were first centrifuged and the supernatants were diluted 1000 times in methanol/water (1:1, v/v) for a subsequent spectral fingerprint acquisition. A Quality Control (QC) sample corresponding to a pool of thirty randomly selected human urine samples was used to assess DIMS data quality as previously described¹¹. For ion mobility spectrometry - mass spectrometry (IM-MS) analysis or DI-MS/MS experiments, urine samples were diluted 100 times in methanol/water (1:1, v/v).

Analytical methods

Spectral fingerprint acquisition. The procedures and settings used to generate metabolomic fingerprints have been previously described¹². Before performing mass calibration, the optimization of signal detection is sometime necessary to smooth out some signal imperfections and to reduce second harmonic signals (see **Figure S1** in Supplementary Information – SI). Briefly, spectral data were acquired in the positive electrospray ionization (ESI) mode using a 7T FT-ICR instrument (Solarix-XR, Bruker, Bremen, Germany) equipped with a dynamically harmonized cell^{7,8}. DIMS analyses were conducted using Flow Injection Analysis (FIA) mode through a micro-HPLC System (UltiMate 3000 RSLCnano system, Thermo Fisher Scientific, Villebon-sur-Yvette, France) to inject 5 μL of sample with a flow rate of 15 $\mu\text{L}\cdot\text{min}^{-1}$ of methanol/water (1:1, v/v) plus 0.1% formic acid. All mass spectra were recorded for 4 min in broadband mode with a 100-1000 m/z range using the following MS conditions: time of flight was set at 0.6 s, ion accumulation time at 0.2 s, drying-gas flow 4 $\text{L}\cdot\text{min}^{-1}$ at 250 °C and an electrospray

voltage of -4.5 kV. More than 650 analyses (including 506 urine samples, repeated injections of QC and analytical blank samples) were carried out in a single analytical sequence, lasting less than 74 hours.

Identification and quantification of discriminant m/z variables. Complementary analytical experiments were carried out to identify and quantify ions of interest detected by PCA. All IM-MS analyses and MS/MS experiments were performed on a Q-ToF instrument (timsTOFTM) equipped with a UHPLC Bruker Elute system (Bruker Daltonics, Bremen, Germany) and an electrospray ionization (ESI) source. FIA mode was used to introduce 2 μL of sample in the flow of 50 $\mu\text{L}\cdot\text{min}^{-1}$ of methanol/water (1:1, v/v) into the mass spectrometer. The instrumental parameters were set as follows: electrospray voltage of -4200 V, end plate offset at 500 V, capillary temperature at 250°C. As the nebulizer and drying gas, the nitrogen flow was fixed at 0.3 bar and 3 $\text{L}\cdot\text{min}^{-1}$, respectively. Mass spectra were recorded in a range of 50-500 m/z with a transfer time of 60 μs and a pre-pulse storage of 5 μs . IM-MS analyses were performed to measure the Collision Cross Section (CCS) values of ions of interest using full ion mobility scan with the inverse reduced mobility ratios ($1/K_0$) range of 0.55 - 1.90 $\text{V}\cdot\text{s}\cdot\text{cm}^{-2}$. The following MS/MS conditions were applied to identify and quantify both 1,1-dimethylbiguanide and sotalol: 1 u for the isolation width of precursor ions and an activation energy of 15 eV. Nitrogen was used as the collision gas. From the decomposition of the $[\text{M}+\text{H}]^+$ species, the peak area of the extracted current for the product ions characteristic of each compound was measured to perform specific quantification by considering the following transitions: m/z 130 \rightarrow m/z 71 for 1,1-dimethylbiguanide and m/z 273 \rightarrow m/z 213 for sotalol.

Data processing. Data preprocessing procedure have been described previously¹³. Briefly, the data features were extracted from the raw mass spectral data files as '.xlsx' files, then imported

and converted into '.mat' format, using Matlab (MatlabR2008, The MathWorks, Natick, MA, USA). The m/z values of all samples were first rounded to 3 decimal places before creating the data matrix. Data preprocessing was performed, removing variables specific to blank samples with a ratio of mean blank ion intensities to mean QC intensities greater than 1. PCA¹⁴ was then applied to the final data matrix.

RESULTS AND DISCUSSION

Metabolomic fingerprinting of urine samples from approximately 500 subjects enrolled in the NESCaV study was performed in less than 74 hours using a high throughput phenotyping approach without any *a priori* information. The quality of the data generated by direct introduction high resolution mass spectrometry (DI-HRMS) has been assessed in a prior study through 54 repeated injections of a QC sample throughout the analytical process, showing a very good repeatability of peak intensity as well as mass measurement accuracy¹¹.

An unsupervised multivariate analysis, PCA, was applied to mine the metabolomic data without any prior knowledge of the cohort metadata, generating a set of orthogonal principal components (PCs), with the two first PCs containing a significant proportion of the variance (~63%) of the data¹⁴ (**Figure S2** in SI). As expected, a two-dimensional score plot of the first two PCs (*i.e.* PC1 and PC2, see **Figure S3** in SI) shows clusters of samples based on their mass spectral profile similarity. Indeed, the blank injection cluster is outside the cohort subjects, indicating that no spectra signal similarity to the cohort samples was detected, whereas the cluster of the repeatedly injected QC is located in the center of the cohort subjects. This observation reflects the high quality of the data generated in terms of reproducibility in the spectral fingerprints and the absence of memory effects in the blank injections, which was confirmed by the very low total ion currents

(TICs) obtained from the blank injections compared to those from QC injections (see **Figure S4**). Although the first two PCs (with larger variances, see **Figure S2**) contain the most part of the cohort information, no interpretable trend clearly emerged from this first exploration of the score plots. These two PCs are probably related to nutritional and/or physiological factors, which could be further explored by more sophisticated statistical methods such as independent component analysis. Interestingly, other PCs with less variance, namely PC7 and PC8, display potential chemical exposure status. Indeed, PC7 separates sixteen subjects enrolled in the cohort from all the others (**Figure 1a**), and this is mainly explained by the variables at m/z 130.109 and 131.112 (**Figure 1b**). A query of the Human Metabolome Database (HMDB)¹⁵ with a mass error tolerance of 1 ppm was performed based on the accurate m/z value detected in the mass spectra (*i.e.* m/z 130.10878). Only one hit was obtained, suggesting that the m/z 130.10878 ions could be annotated as the $[M+H]^+$ species of 1,1-dimethylbiguanide ($C_4H_{12}N_5$ with an error of 461 ppb), and the m/z 131.112 ions could therefore correspond to its ^{13}C isotope. This compound, marketed as Metformin, Metformax or Glucophage, is used as a drug for the treatment of type II diabetes.¹⁶ On the PC8 component, only three subjects are discriminated from the entire cohort (**Figure 1a**). The loading plot shows that the variables at m/z 273.127, m/z 274.130, m/z 275.123 and m/z 136.563 contribute strongly to this PC (**Figure 1b**). Certainly, the m/z 274.130, m/z 275.123 ions are isotopes of the m/z 273.127 ions while the m/z 136.563 species could correspond to a doubly charged species. According to the HMDB query, the ions at m/z 273.12694 (accurately measured on an FT-ICR instrument) could be annotated as the $[M+H]^+$ ions of sotalol ($C_{12}H_{21}N_2O_3S$ with an error of 732 ppb). Sotalol or methane sulfonanilide is a drug used in the treatment of life-threatening arrhythmias.¹⁷ Concerning the ion at m/z 136.5635 strongly correlated to PC8, only one candidate could be proposed through MS search on HMDB, it could correspond to the

$[M+H+NH_4]^{2+}$ double-charged species of indolo(2,3-b)carbazole, a metabolite of indole-3-carbinol (I3C) found in the Brassica family of vegetables ($C_{18}H_{10}N_2$ with an error of 5 ppm). Unfortunately, no sound structural information could be consistently accumulated to anticipate a possible origin of this analyte. Furthermore, as the PC8 component here concerns a significant discrimination of only three patients treated with sotalol from all the others, the probability that a common analyte is found by chance in these three individuals is high compared to the situation where one PC would relate to 15 or more individuals for whom the dietary or lifestyle may vary and, for whom, consequently, the probability that a common biomarker of these multiple environmental factors may still be detected is very low.

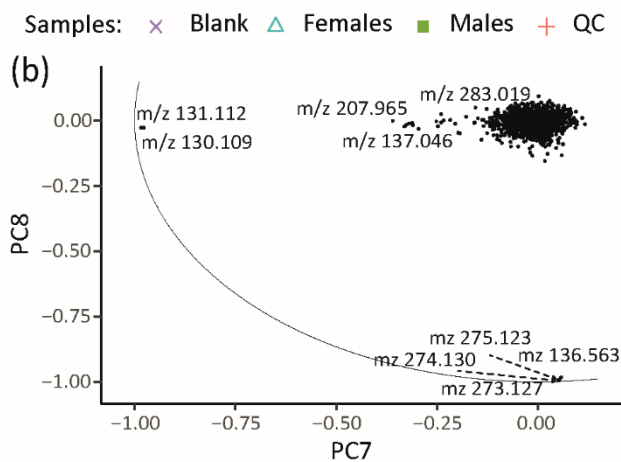
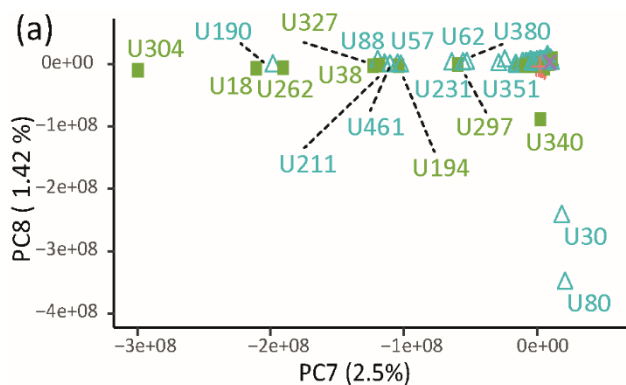


Figure 1. Two-dimensional PCA plots: (a) score plot showing the separation of sixteen subjects on PC7 (2.50% of the total variance) and three subjects on PC8 (1.42% of the total variance), and (b) loading plot showing the most influential variables for PC7 and PC8.

The procedure used for the identification of a discriminant variable is illustrated in **Figure 2** for the variable at m/z 273.12694, showing different identification levels according to a five-level classification system.¹⁸ Using HRMS detection ($R_p > 900,000$), the identity assignment of the two drugs could be reinforced by determining their ion elemental compositions unequivocally based on their isotopic fine structures which display isotopic peak distributions, including isotopologues (as shown in **Figure 2** and **Figure S5** in SI). Indeed, the isotopes that contribute to the isotopic clusters can be identified by calculating their relative isotopic mass defects ($R\Delta m_{(A+n)}$), which is defined as the m/z difference between the considered isotopic peak and the monoisotopic peak minus the integer value (i.e. 1, 2, ..., n for (A+1), (A+2)...(A+n), respectively). Each $R\Delta m_{(A+n)}$ value is characteristic of an isotope for a given element^{9,19}. The zoomed HRMS spectrum of a urine sample discriminated on PC8 shows the isotopic pattern of the m/z 273.12694 (level 4 in **Figure 2**). Thanks to the calculated $R\Delta m_{(A+n)}$ values, the identity of the different isotopologues detected could be assigned such as $^{15}\text{N}_1$, $^{33}\text{S}_1$, $^{13}\text{C}_1$ and $^2\text{H}_1$ in the (A+1) cluster at m/z 274 and, $^{34}\text{S}_1$, $^{33}\text{S}_1^{13}\text{C}_1$, $^{18}\text{O}_1$ and $^{13}\text{C}_2$ in the (A+2) cluster at m/z 275 as well as $^{34}\text{S}_1^{13}\text{C}_1$ in the (A+3) at m/z 276 with a mass measurement error of less than 800 ppb. This fine structure of isotopic peak distributions corroborates the presence of the S and O atoms in the proposed elemental composition and the annotation of the discriminant variable at m/z 273.12694 as the $[\text{M}+\text{H}]^+$ ions of sotalol. Similarly, the unequivocal elemental composition supporting the 1,1-dimethylbiguanide identity was also obtained thanks to the isotopic fine structure observed (**Figure S5** in SI).

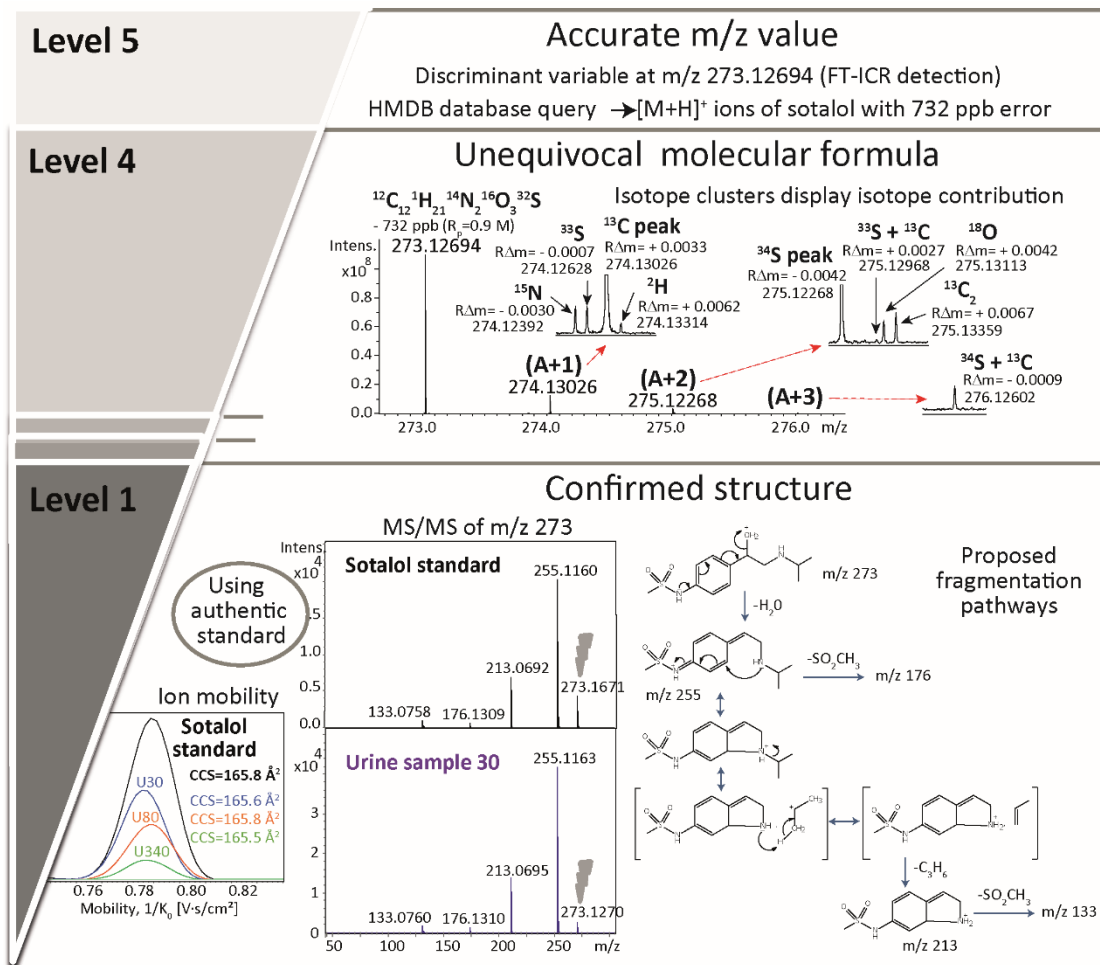


Figure 2. Illustration of the structure identification procedure used for a discriminant variable at m/z 273.12694 according to the five-level classification system of Schymanski *et al.*¹⁸. The level 1 used for confirmed structure requires the acquisition of two experimental data from an authentic standard. Here, two orthogonal methods, the IM-MS analysis and the MS/MS experiment, were performed to obtain CCS values and MS/MS spectra, respectively. The values of $R\Delta m$, relative isotopic mass defects, were obtained using the following equation: $R\Delta m_{(A+n)} = m_{(A+n)} - m_A - n$, with $m_{(A+n)}$ the m/z value of the isotopic peak considered, m_A the m/z value of the monoisotopic peak and n , the difference between the integer values of $m_{(A+n)}$ and m_A . Each $R\Delta m$ value is characteristic of an isotope for a given element.

In order to reach the level 1 of structure identification¹⁸, ion mobility measurements and MS/MS experiments were performed using a timsTOF instrument to analyze 1,1-dimethylbiguanide and sotalol, as authentic commercial standards, and the sixteen urine samples selected by PC7 and the three samples detected on PC8 as exemplified in the **Figure 2** for the identification of sotalol. CCS values of 242.8 Å², and 165.8 Å² were obtained for 1,1-dimethylbiguanide and sotalol, respectively, which are consistent with the CCS values measured in the analyzed urine samples, with bias of 0.8% for 1,1-dimethylbiguanide and 0.1% for sotalol (**Figure 2** and **Figure S6** in SI). In addition, the fragmentation patterns generated from the authentic chemical standards were very similar to those obtained from urine samples (**Figure 2**, **Figures S7** and **S8** in SI). All these data obtained by two orthogonal analytical methods support the identity of 1,1-dimethylbiguanide and sotalol (**Figure 3**), and therefore their presence in the sixteen samples separated by PC7 and the three samples separated by PC8, respectively.

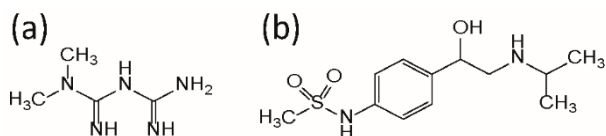


Figure 3. Chemical structures of (a) 1,1-dimethylbiguanide, and (b) sotalol.

These results obtained using our high throughput phenotyping approach without any prior knowledge of metadata information were reinforced by the complementary epidemiological data, providing further information on the prescription of these two drugs. Indeed, such a direct access to the chemical exposome characterization through our processing workflow (e.g., multivariate statistical mining, structural characterization) is helpful to further anticipate in more sophisticated

modeling studies the existence of possible metabolic disruptions related to this exposomic information.

Among the sixteen subjects discriminated by PC7, thirteen declared non-insulin-dependent diabetes, while the three others did not mention this disease, but on the other hand, all of them affirmed they were taking a drug such as Metformin, Metformax or Glucophage. Interestingly, our approach allows to check the consistency of the information reported in the metadata as demonstrated for the three subjects selected on PC7 who declared to be non-diabetic despite taking 1,1-dimethylbiguanide. Probably, in this specific latter case, an atypical medical prescription may have been given, since this drug may also be used for other diseases associated with insulin resistance such as gestational diabetes²⁰ or polycystic ovary syndrome^{21,22} or, eventually, to prevent type-II diabetes in so-called predisposed subjects²³, even if this disease was not yet diagnosed, and, as stated more recently, to reinforce insulin sensitivity in young type 1 diabetes²⁴. It should be emphasized that 1,1-dimethylbiguanide is neither metabolized nor excreted in bile, the clearance of this drug depending exclusively on the renal excretion²⁵. Therefore, it should be found in its unchanged form, which is consistent with its detection here in the urine.

However, some other subjects (n = 13) also declared this disease and taking the drug for its treatment, but none were selected by PC7. Careful examination of the FT-ICR raw data for these subjects shows a very low abundance (or absence) of the *m/z* 130 peak. In addition, MS/MS experiments performed on urinary samples coming from these thirteen subjects yielded no results due to insufficient signal detection.

Metadata mining also displays other subjects with cardiac arrhythmia in addition to the three detected on PC8. But, among all the subjects, only the three ones selected from PC8 mentioned being under sotalol treatment. This compound is a β -adrenergic blocking agent, which is

considered as an important class III antiarrhythmic drug, and is known to be almost completely absorbed after oral dosing and excreted mainly in urine as the intact compound¹⁷.

The results obtained from the PCA processing are in good agreement with the metadata. However, some questions remain unanswered regarding the thirteen subjects not selected by PC7 who had nevertheless declared to be prescribed 1,1-dimethylbiguanide. This could be explained by a low concentration of this drug in the urinary samples of these subjects, which could be below the analytical detection threshold. In this case, the distribution of individuals projected on the PCs could be related to the drug concentration.

Therefore, to fully assess the amount of these drugs found in urine samples separated thanks to the PC7 or PC8 components, external standard quantitation was performed using the DI-MS/MS approach to analyze standard solutions at different concentrations. Calibration curves were constructed from the responses of the diagnostic product ions resulting from the decomposition of the $[M+H]^+$ parent ions, considering the transitions m/z 130 \rightarrow m/z 71 for 1,1-dimethylbiguanide and m/z 273 \rightarrow m/z 213 for sotalol, thereby improving the specificity of the quantification. The detector response was linear in the range of 0.1 -2.5 ng. μ L⁻¹ for 1,1-dimethylbiguanide with $R^2 = 0.94$ and in the range of 0.1 -5 ng. μ L⁻¹ for sotalol with $R^2 = 1.00$ (**Figure S9** in SI). The obtained regression equations were then used to calculate the 1,1-dimethylbiguanide and sotalol concentrations found in the urine samples, *i.e.* the sixteen samples selected by PC7 and the three samples selected by PC8, analyzed under the same DI-MS/MS conditions (**Table S1**). The objective of these complementary assays of 1,1-dimethylbiguanide and sotalol in urine samples was not to obtain the absolute concentrations of these two drugs in the samples of interest but only to compare drug concentrations between subjects, in order to relate these values to the distribution of the scores found on the PC7 and PC8 components. This is why no extensive validation

parameters (*e.g.*, limit of detection (LoD), limit of quantification (LoQ), repeatability, etc.) were investigated as a priority here, only the regression parameters over the range of concentrations studied for 1,1-dimethylbiguanide and sotalol (**Figure S9**).

Nevertheless, regression analyses were further used to assess the relationship between estimated 1,1-dimethylbiguanide and sotalol contents in urine samples and the respective PCA scores. The resulting curves showed good linearity for both drugs with $R^2 \geq 0.89$, indicating a good correlation between the scores of individuals distributed on the specific PCs from the DI-HRMS data (using a FT-ICR instrument) on the one hand, and the drug concentrations estimated by external calibration curves from DI-MS/MS (using a QToF instrument) on the other hand (**Figure 4**). Interestingly, these results demonstrate that our direct approach for high-throughput metabolomic fingerprinting could provide confident quantitative data. The estimated concentrations of sotalol in (1:100) diluted urine samples were from 0.04 to 0.47 mg.L⁻¹ and those of 1,1-dimethylbiguanide ranged from 0.38 to 1.91 mg.L⁻¹ (**Table S1**). Variations in drug concentrations between individuals could be due to inter-individual differences in metabolism, drug disposition or posology, which nevertheless did not impact the detection of these chemicals.

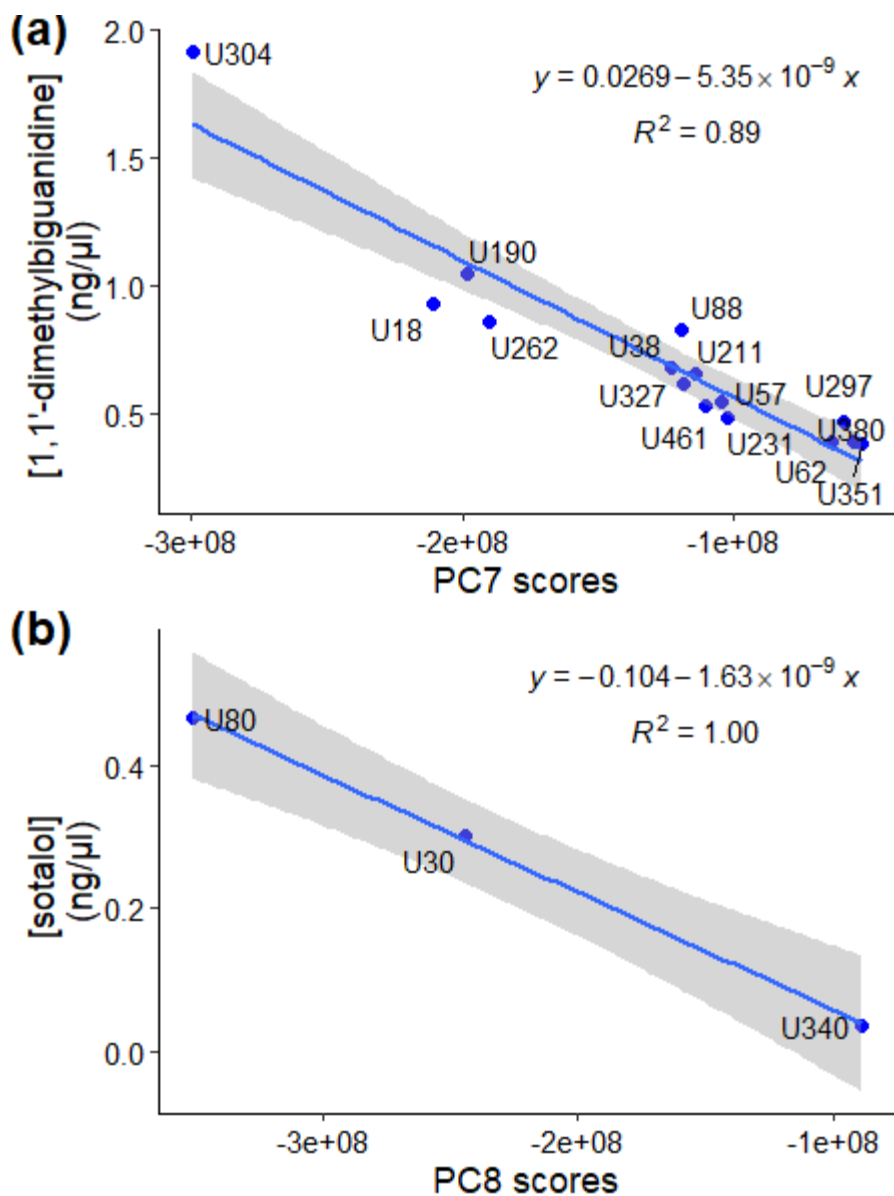


Figure 4. Regression analysis showing the relationship between estimated (a) 1,1-dimethylbiguanidine and (b) sotalol content and their respective PCA scores displayed on PC7 and PC8 (see **Figure 1**).

CONCLUSIONS

In conclusion, a cohort phenotyping of approximately 500 individuals was performed in less than 74 hours thanks to direct analysis using an FT-ICR instrument equipped with a dynamically harmonized cell. A clear separation of subjects who were prescribed drugs such as 1,1-dimethylbiguanide or sotalol was simply obtained from the PCA analysis. This unsupervised detection demonstrated the efficacy of our direct approach to highlight exposure to xenobiotics without any prior knowledge of metadata. Hence, such an untargeted DI-HRMS methodology appears to be a very promising approach for high throughput phenotyping of cohorts in epidemiological studies, particularly for characterizing the human chemical exposome. The knowledge of the chemical exposures is very useful for evaluating *a posteriori* their possible associations with the occurrence of certain diseases or clinical symptoms.

FIGURE AND SCHEME CAPTIONS

Figure 1. Two-dimensional PCA plots: (a) score plot showing the separation of sixteen subjects on PC7 (2.50% of the total variance) and three subjects on PC8 (1.42% of the total variance), and (b) loading plot showing the most influential variables for PC7 and PC8.

Figure 2. Illustration of the structure identification procedure used for a discriminant variable at m/z 273.12694 according to the five-level classification system of Schymanski et al.¹⁸. The level 1 used for confirmed structure requires the acquisition of two experimental data from an authentic standard. Here, two orthogonal methods, the IM-MS analysis and the MS/MS experiment, were performed to obtain CCS values and MS/MS spectra, respectively. The values of $R\Delta m$, relative isotopic mass defects, were obtained using the following equation: $R\Delta m_{(A+n)} = m_{(A+n)} - m_A - n$, with $m_{(A+n)}$ the m/z value of the isotopic peak considered, m_A the m/z value of the monoisotopic

peak and n the difference between the integer values of $m_{(A+n)}$ and m_A . Each $R\Delta m$ value is characteristic of an isotope for a given element.

Figure 3. Chemical structures of (a) 1,1-dimethylbiguanide, and (b) sotalol.

Figure 4. Regression analysis showing the relationship between estimated (a) 1,1-dimethylbiguanide and (b) sotalol content and the respective PCA scores displayed on PC7 and PC8 (see **Figure 1**).

SUPPORTING INFORMATION

The supporting information can be found online. The following file is available free of charge. It contains figures of some second harmonic signals $2\nu_+$ detected in our data, PCA analysis figures, figures of the IM-MS analysis and the MS/MS spectra, calibration curves and a table of estimated concentrations of sotalol and 1,1-dimethylbiguanide in (1:100) diluted urine samples (PDF).

AUTHOR INFORMATION

Corresponding Authors

Sandra Alves - Sorbonne Université, Faculté des Sciences et de l'Ingénierie, Tour 42-43, 4^{ème} étage, BP 45, 4 place Jussieu F-75005, Paris, France; Email: sandra.alves@sorbonne-universite.fr

or

Alain Paris - Muséum national d'Histoire naturelle, MCAM, UMR7245 CNRS - MNHN, 75005 Paris, France; E-mail: alain.paris@mnhn.fr

or

Estelle Rathahao-Paris - UMR MTS, Laboratoire d'Immuno-Allergie Alimentaire, CEA de Saclay- Bat 136, 91191, Gif-sur-Yvette cedex, France; E-mail: estelle.paris@inrae.fr

Author Contributions

SS, MG, A-FD and BA designed the epidemiological study and built the database containing all metadata. BH, SA, AP and ER-P performed the analytical experiments, interpreted the results and wrote the manuscript. BH, AP and DNR supported the data mining. All authors reviewed the manuscript and have given approval to the final version of the manuscript.

Notes

The authors declare no conflict of interest.

Research involving Human Participants: All procedures performed in this study were in accordance with the ethical standards of the national research committee and with the 1964 Helsinki declaration and its later amendments. The protocol of the NESCaV study was approved by the Ethics Committee of the Faculty of Medicine of the University of Liège (Wallonia) as institutional review board representing the Wallonia part of the study (B70720097541).

ACKNOWLEDGMENT

Author acknowledgements: Dr J. Fioramonti (Toxalim, INRAE, Toulouse, France) for his support in this project, Pr J.P. Chapelle (Service de Chimie médicale, University Hospital Centre of Liège, Belgium) for providing human urine samples, Region Ile-de-France and DIM Analytics for the financial support of B. Habchi's thesis, the DIM ASTREA program for financing the upgrade of the FT-ICR instrument, the French FR-2054 Infranalytics research infrastructure for access to the

FT-ICR instrument and the French MetaboHUB infrastructure (ANR-11-INBS-0010 grant) for the funding.

ABBREVIATIONS

CCS, Collision Cross Section; DIMS, Direct Introduction coupled with Mass Spectrometry; FT-ICR, Fourier Transform Ion Cyclotron Resonance; IM-MS, ion mobility spectrometry - mass spectrometry; HRMS, High Resolution Mass Spectrometry; PCA, Principal Component Analysis; PCs, principal components; QC, Quality Control; Rp, resolving power; MS/MS, tandem mass spectrometry.

REFERENCES

- (1) National Research Council (US) Committee on Environmental Epidemiology. *Environmental Epidemiology* Vol 1: Public Health and Hazardous Wastes; Washington (DC): National Academies Press (US), **1991**; Vol. 1.
- (2) Prado-Aguilar, C. A.; Martínez, Y. V.; Segovia-Bernal, Y.; Reyes-Martínez, R.; Arias-Ulloa, R. Performance of Two Questionnaires to Measure Treatment Adherence in Patients with Type-2 Diabetes. *BMC Public Health* **2009**, 9 (1), 38. <https://doi.org/10.1186/1471-2458-9-38>.
- (3) Haring, R.; Wallaschofski, H. Diving Through the “-Omics”: The Case for Deep Phenotyping and Systems Epidemiology. *OMICS A J. Integr. Biol.* **2012**, 16 (5), 231–234. <https://doi.org/10.1089/omi.2011.0108>.

- (4) Ren, J. L.; Zhang, A. H.; Kong, L.; Wang, X. J. Advances in Mass Spectrometry-Based Metabolomics for Investigation of Metabolites. *RSC Adv.* **2018**, *8*, 22335–22350. <https://doi.org/10.1039/c8ra01574k>.
- (5) Gika, H.; Virgiliou, C.; Theodoridis, G.; Plumb, R. S.; Wilson, I. D. Untargeted LC/MS-Based Metabolic Phenotyping (Metabonomics/Metabolomics): The State of the Art. *J. Chromatogr. B* **2019**, *1117*, 136–147. <https://doi.org/10.1016/j.jchromb.2019.04.009>.
- (6) Habchi, B.; Alves, S.; Paris, A.; Rutledge, D. N.; Rathahao-Paris, E. How to Really Perform High Throughput Metabolomic Analyses Efficiently? *Trends Anal. Chem.* **2016**, *85*, 128–139. <https://doi.org/10.1016/j.trac.2016.09.005>.
- (7) Boldin, I. A.; Nikolaev, E. N. Fourier Transform Ion Cyclotron Resonance Cell with Dynamic Harmonization of the Electric Field in the Whole Volume by Shaping of the Excitation and Detection Electrode Assembly. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 122–126. <https://doi.org/10.1002/rcm.4838>.
- (8) Nikolaev, E. N.; Boldin, I. a.; Jertz, R.; Baykut, G. Initial Experimental Characterization of a New Ultra-High Resolution FTICR Cell with Dynamic Harmonization. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1125–1133. <https://doi.org/10.1007/s13361-011-0125-9>.
- (9) Alves, S.; Rathahao-Paris, E.; Tabet, J. C. Potential of Fourier Transform Mass Spectrometry for High-Throughput Metabolomics Analysis. In *Metabolomics coming of age with its technological diversity*; Rolin, D., Ed.; *Adv. Bot. Res.* Elsevier Amst. Ltd, **2013**, *67*, 219–303.

(10) Alkerwi, A.; Guillaume, M.; Zannad, F.; Laufs, U.; Lair, M.-L. Nutrition, Environment and Cardiovascular Health (NESCAV): Protocol of an Inter-Regional Cross-Sectional Study. *BMC Public Health* **2010**, 10, 698. <https://doi.org/10.1186/1471-2458-10-698>.

(11) Habchi, B.; Alves, S.; Jouan-Rimbaud Bouveresse, D.; Appenzeller, B.; Paris, A.; Rutledge, D. N.; Rathahao-Paris, E. Potential of Dynamically Harmonized Fourier Transform Ion Cyclotron Resonance Cell for High-Throughput Metabolomics Fingerprinting: Control of Data Quality. *Anal. Bioanal. Chem.* **2018**, 410 (2), 483–490. <https://doi.org/10.1007/s00216-017-0738-3>.

(12) Rathahao-Paris, E.; Alves, S.; Paris, A. High-Throughput Metabolomics Using Flow Injection Analysis and Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. In *Metabolomics*; Wood, P. L., Ed.; *Neuromethods*; Springer US: New York, NY, **2021**; Vol. 159, pp 9–23. https://doi.org/10.1007/978-1-0716-0864-7_2.

(13) Habchi, B.; Alves, S.; Jouan-Rimbaud Bouveresse, D.; Moslah, B.; Paris, A.; Lécluse, Y.; Gauduchon, P.; Lebailly, P.; Rutledge, D. N.; Rathahao-Paris, E. An Innovative Chemometric Method for Processing Direct Introduction High Resolution Mass Spectrometry Metabolomic Data: Independent Component–Discriminant Analysis (IC–DA). *Metabolomics* **2017**, 13 (4). <https://doi.org/10.1007/s11306-017-1179-x>.

(14) Wold, S.; Esbensen, K.; Geladi, P. Principal Component Analysis. *Chemom. Intell. Lab. Syst.* **1987**, 2, 37–52. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9).

(15) Wishart, D. S.; Feunang, Y. D.; Marcu, A.; Guo, A. C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; Sayeeda, Z.; Lo, E.; Assempour, N.; Berjanskii, M.; Singhal, S.; Arndt, D.; Liang, Y.; Badran, H.; Grant, J.; Serra-Cayuela, A.; Liu, Y.; Mandal, R.;

Neveu, V.; Pon, A.; Knox, C.; Wilson, M.; Manach, C.; Scalbert, A. HMDB 4.0: The Human Metabolome Database for 2018. *Nucleic Acids Res.* **2018**, 46 (D1), D608–D617. <https://doi.org/10.1093/nar/gkx1089>.

(16) Bailey, C. J. Metformin: Historical Overview. *Diabetologia* **2017**, 60, 1566–1576. <https://doi.org/10.1007/s00125-017-4318-z>.

(17) Singh, B. N.; Deedwania, P.; Nademane, K.; Ward, A.; Sorkin, E. M. Sotalol. A Review of Its Pharmacodynamic and Pharmacokinetic Properties, and Therapeutic Use. *Drugs* **1987**, 34, 311–349.

(18) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Matthias, R.; Singer, H. P.; Hollender, J. Viewpoint Pubs.Acs.Org/Est Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidenc. *Env. Sci Technol* **2014**, 48 (4), 2097–2098.

(19) Thurman, E. M.; Ferrer, I. The Isotopic Mass Defect: A Tool for Limiting Molecular Formulas by Accurate Mass. *Anal. Bioanal. Chem.* **2010**, 397 (7), 2807–2816. <https://doi.org/10.1007/s00216-010-3562-6>.

(20) Moore, L. E.; Clokey, D.; Rappaport, V. J.; Curet, L. B. Metformin Compared With Glyburide in Gestational Diabetes. *Obstet. Gynecol.* **2010**, 115, 55–59. <https://doi.org/10.1097/aog.0b013e3181c52132>.

(21) Velazquez, E. M.; Glueck, C. J.; Mendoza, S.; Hamer, T.; Sosa, F. Metformin Therapy in Polycystic Ovary Syndrome Reduces Hyperinsulinemia, Insulin Resistance, Hyperandrogenemia, and Systolic Blood Pressure, While Facilitating Normal Menses and Pregnancy. *Metabolism* **1994**, 43, 647–654. [https://doi.org/10.1016/0026-0495\(94\)90209-7](https://doi.org/10.1016/0026-0495(94)90209-7).

(22) Tang, T.; Norman, R. J.; Balen, A. H.; Lord, J. M. Insulin-Sensitising Drugs (Metformin, Troglitazone, Rosiglitazone, Pioglitazone, D-Chiro-Inositol) for Polycystic Ovary Syndrome. In *Cochrane Database of Systematic Reviews*; The Cochrane Collaboration, Ed.; John Wiley & Sons, Ltd: Chichester, UK, **2003**; p CD003053. <https://doi.org/10.1002/14651858.CD003053>.

(23) Knowler, W.; Barrett-Connor, E.; Fowler, S.; Hamman, R.; Lachin, J.; Walker, E. Reduction in the Incidence of Type 2 Diabetes Mellitus with Lifestyle Intervention or Metformin. *N Engl J Med* **2002**, 346, 393–403. <https://doi.org/10.1007/bf03344412>.

(24) Cree-Green, M.; Bergman, B. C.; Cengiz, E.; Fox, L. A.; Hannon, T. S.; Miller, K.; Nathan, B.; Pyle, L.; Kahn, D.; Tansey, M.; Tichy, E.; Tsalikian, E.; Libman, I.; Nadeau, K. J. Metformin Improves Peripheral Insulin Sensitivity in Youth With Type 1 Diabetes. *J. Clin. Endocrinol. Metab.* **2019**, 104 (8), 3265–3278. <https://doi.org/10.1210/jc.2019-00129>.

(25) Sirtori, C. R.; Franceschini, G.; Galli-Kienle, M.; Cighetti, G.; Galli, G.; Bondioli, A.; Conti, F. Disposition of Metformin (N,N-Dimethylbiguanide) in Man. *Clin. Pharmacol. Ther.* **1978**, 24 (6), 683–693. <https://doi.org/10.1002/cpt1978246683>.