
A Statistical Procedure to Selectively Detect Metabolite Signals in LC-MS Data Based on Using Variable Isotope Ratios

Lung-Cheng Lin,^a Hsin-Yi Wu,^a Vincent Shin-Mu Tseng,^b
Lien-Chin Chen,^b Yu-Chen Chang,^a and Pao-Chi Liao^{a,c}

^a Department of Environmental and Occupational Health, College of Medicine, National Cheng Kung University, Tainan, Taiwan

^b Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan

^c Sustainable Environment Research Center, National Cheng Kung University, University Road, Tainan, Taiwan

The tracing of metabolite signals in LC-MS data using stable isotope-labeled compounds has been described in the literature. However, the filtering efficiency and confidence when mining metabolite signals in complex LC-MS datasets can be improved. Here, we propose an additional statistical procedure to increase the compound-derived signal mining efficiency. This method also provides a highly confident approach to screen out metabolite signals because the correlation of varying concentration ratios of native/stable isotope-labeled compounds and their instrumental response ratio is used. An in-house computational program [signal mining algorithm with isotope tracing (SMAIT)] was developed to perform the statistical procedure. To illustrate the SMAIT concept and its effectiveness for mining metabolite signals in LC-MS data, the plasticizer, di-(2-ethylhexyl) phthalate (DEHP), was used as an example. The statistical procedure effectively filtered 15 probable metabolite signals from 3617 peaks in the LC-MS data. These probable metabolite signals were considered structurally related to DEHP. Results obtained here suggest that the statistical procedure could be used to confidently facilitate the detection of probable metabolites from a compound-derived precursor presented in a complex LC-MS dataset. (J Am Soc Mass Spectrom 2010, 21, 232–241) © 2010 American Society for Mass Spectrometry

Stable isotope-labeled compounds coupled with liquid chromatography mass spectrometry (LC-MS) have been applied to mine metabolite signals. One strategy to mine MS data for metabolite signals uses the isotope cluster that is formed after treating test subjects with an equal mixture of native and isotope-labeled compounds [1–3]. In mineral metabolism research, the stable isotope ratio (1:1) from spiked native and isotope-labeled compounds has been used for metabolite signal detection. The equal response signatures of signal doublets were traced [4]. A strategy using the stable isotope-labeled tracing concept has been developed to detect reactive metabolites, such as glutathione-conjugated metabolites [5].

Another report has demonstrated that equal amounts of native and stable isotope-labeled compounds can be used to detect untargeted metabolite signals with particular isotopic ratios in LC-MS data, where this work also demonstrated the use of a computational program, DoGEX, to facilitate the signal mining process [6]. The program

detects the metabolite signals according to a particular stable isotope ratio and mass difference after data alignment, noise removal, baseline correction, peak detection, and spectral filtering processes. Additionally, the program requires a user-defined isotopic ratio of native to isotope-labeled signal responses to define metabolite signals. However, the isotopic ratios of metabolite signals from equal amounts of spiked native and isotope-labeled compounds are not always closed to 1 because of the influences from analytical variations of sample preparation, instrumental analysis, and matrix interference. This is especially true when signal doublet responses are low [7]. Setting acceptable tolerance for wanted ratios is still empirical, which results in the challenges while tracing one specific ratio [8].

We propose, herein, an additional statistical procedure to assist the conventional isotope-labeling method. We believe that coupling a stable isotope ratio with the statistical procedure could enhance the efficiency and confidence of mining metabolite signals in complex LC-MS data. We used varying isotope concentration ratios and mass measurements to trace the signals of metabolites, which were generated from *in vitro* incubation, in LC-MS data. Varying isotope concentration ratios

Address reprint requests to Dr. P.-C. Liao, Department of Environmental and Occupational Health, College of Medicine, National Cheng Kung University, 138 Sheng-Li Road, Tainan 704, Taiwan. E-mail: liaopc@mail.ncku.edu.tw

of native to stable isotope-labeled compounds in respective samples have been proposed to statistically correlate to their signal responses, which provided higher confidence than the conventional isotope-labeling method. An in-house computational program, comprising a signal mining algorithm with isotope tracing (SMAIT), was developed to perform the statistical procedure.

Di-(2-ethylhexyl) phthalate (DEHP) with a wide range of metabolites, including mono-(2-ethyl-hexyl) phthalate (MEHP), mono-(2-ethyl-hydroxyhexyl) phthalates (OH-MEHPs), mono-(2-ethyl-oxohexyl) phthalates (oxo-MEHPs), and mono-(2-ethyl-carboxypentyl) phthalates (α -MEHPs), has been reported [9–11]. To illustrate the effectiveness of the proposed procedure for detecting metabolite signals within LC-MS data, DEHP was used as a model compound.

Experimental

Materials

DEHP, D₄-DEHP, MEHP, 5OH-MEHP, and 5oxo-MEHP were purchased from Cambridge Isotope Laboratories (Andover, MA, USA). Glucose 6-phosphate, glucose 6-phosphate dehydrogenase, β -nicotinamide adenine dinucleotide phosphate (NADP⁺), formic acid, acetic acid, and buffer salts were purchased from Sigma-Aldrich (St. Louis, MO, USA). Methanol (HPLC grade) was purchased from Merck (Darmstadt, Germany). Deionized water was acquired using a Millipore system (Milford, MA, USA).

Rat Liver Enzyme Preparation and Incubation

For generating metabolites from a parent compound, the simple and commonly used method of *in vitro* liver enzyme incubation was adopted [3, 6]. Liver tissue (0.5 g) from one healthy male Wistar rat, without any drug treatment, was added to 3 mL ice buffer. The liver tissue was homogenized, sonicated, and centrifuged at 9000 \times g for 15 min (4 °C). The supernatant containing proteins was collected and kept at –80 °C before use. The protein concentration was determined using a Bradford assay [12].

In vitro liver enzyme incubation was performed to generate DEHP metabolites. The incubation mixture contained DEHP, D₄-DEHP, 1.3 mM NADP, 3.3 mM MgCl₂, 3.3 mM glucose 6-phosphate, 0.4 U/mL glucose 6-phosphate dehydrogenase, and 1.6 mg/mL enzyme protein in phosphate buffer (100 mM, pH = 7.4). The total concentration of spiked DEHP and D₄-DEHP in a mixture sample was 15 μ M. For illustrating the SMAIT strategy using varying isotope concentration ratios, five samples were incubated. These had concentration ratios of DEHP to D₄-DEHP of 3:7, 4:6, 5:5, 6:4, and 7:3. The metabolic reaction was initiated by adding NADP. Subsequently, the spiked sample was incubated at 37 °C for 1 h. Acetic acid (20%, 100 μ L) was added to terminate the reaction.

Sample Preparation

The incubated sample was passed through a 0.2 μ m PVDF membrane filter (MSF-3; Advantec MFS, Inc., Pleasanton, CA, USA), before off-line solid-phase extraction (SPE). This C₁₈ cartridge (Sigma-Aldrich, St. Louis, MO, USA) was pre-conditioned with MeOH followed by H₂O. After loading the filtered sample, the sample in the cartridge was washed with 2 mL H₂O and eluted with 2 mL MeOH. The eluted sample was dried under nitrogen. Subsequently, 100 μ L of 20% acetic acid and 400 μ L of H₂O were added. The sample was stored at 4 °C until LC-MS analysis was performed.

LC-MS Analysis

Liquid chromatography was carried out on a HPLC system equipped with a PE series 200 autosampler (Perkin Elmer, Norfolk, CT, USA) and a PE series 200 binary pump (Perkin Elmer). MS analyses were performed on a quadrupole-time of flight (Q-TOF) mass spectrometer (QSTAR Pulsar; Applied Biosystems/MDS SCIEX, Foster City, CA, USA). This instrument was equipped with a TurboIonSpray ionization source.

Chromatography was achieved using a 2.0 \times 55 mm C₁₈ column (3- μ m particle size; Merck, Darmstadt, Germany). The chromatographic loading and elution solvents were H₂O (0.1% formic acid) and MeOH (0.1% formic acid), respectively. The flow rate for LC gradient analysis was 300 μ L/min. The incubated sample was loaded onto the C₁₈ column by running solvent (H₂O with 0.1% formic acid) through it for 5 min. At 5 min, the gradient elution was started with the gradient at 3.3%/min from 0 to 100% elution solvent for 30 min. After the gradient elution step, the column was washed with elution solvent for 20 min. Finally, the column was re-equilibrated for 1 min by loading solvent before the next sample injection.

For metabolite ionization via ESI deprotonation, the negative ionization mode was used with an ion spray voltage of –4500 V. Nitrogen was used as the curtain, nebulizer, and turbo gas (400 °C), with settings of 20, 20, and 60 psi, respectively. Full-scan mass spectra were acquired within the mass range of m/z 80–600. Full-scan operating parameters in the TOF-MS mode were as follows: accumulation time = 1 s, mirror voltage = 0.990 kV, plate voltage = 0.339 kV, grid voltage = 0.415 kV, liner voltage = –15 V, and offset voltage = –15 V with resolution at a minimum of 7000. All samples were analyzed using identical instrumental conditions.

Peak Defining in LC-MS Spectra

A commercially available program, Metabolomics Export Script (MES, ver. 1.1.2.0; Applied Biosystems), was used to analyze LC-MS signals and export a raw peak list, which contained m/z , retention time (RT), and signal response information. MES set parameters for each LC-MS spectrum were as follows: subtraction offset = 150

scans; subtraction multiplication factor = 1.0; noise threshold = 10; minimum spectral peak width = 1 ppm; minimum RT peak width = 2 scans; and maximum RT peak width = 100 scans.

MS/MS Analysis

To confirm the presence of metabolites, the Q-TOF product ion scan was used to obtain product ion profiles of metabolite signals via collision-induced dissociation (CID). Metabolite signals were confirmed by comparing their product ion profiles with those of commercial standards.

LC and TurboIonSpray source conditions for MS/MS analysis were set to be identical to those from the TOF-MS mode. Precursor ion selection in the quadrupole (Q_1) was set at unit resolution. The collision energy (CE) was optimized to obtain the highest MEHP signal intensity and was set at -20 eV for the product ion scan of all metabolites. Accumulation time in the TOF analyzer for product ion collection was set at 1 s. Native and isotope-labeled metabolite signals were both analyzed to obtain their product ion profiles.

Results and Discussion

General Summary of the Statistical Procedure

In the proposed statistical procedure, varying isotope concentration ratios of native to stable isotope-labeled

compounds among five mixtures are used as tracers to filter out probable metabolite signals. The procedure uses the assumption that the native to stable labeled-isotope concentration ratios of metabolite signals, the isotopic pairs (IPs), correlate to their spiked varying concentration ratios. As shown in Figure 1, five mixtures with different isotope concentration ratios were incubated with liver enzymes to generate metabolites. The incubation mixtures were desalted and the metabolites concentrated by SPE, which was followed by LC-Q-TOF analysis. LC-MS data from the five mixtures were subjected to the MES program to generate raw peak lists. After the peak stratification step, an in-house computational algorithm (SMAIT) was used to extract probable metabolite signals from the complex LC-MS data. Probable metabolite signals were confirmed by comparing their product ion profiles with those of commercial standards.

SMAIT Computer Program Algorithm

Recently, a computational program was developed to provide an efficient tool for extracting phosphopeptide signals from complex LC-MS data [13]. This program uses the mass difference between phosphopeptide and dephosphopeptide signals to filter out the phosphopeptide signals. Using a similar concept, we developed the SMAIT algorithm. This algorithm mines metabolite signals by not only measuring the specific mass shift

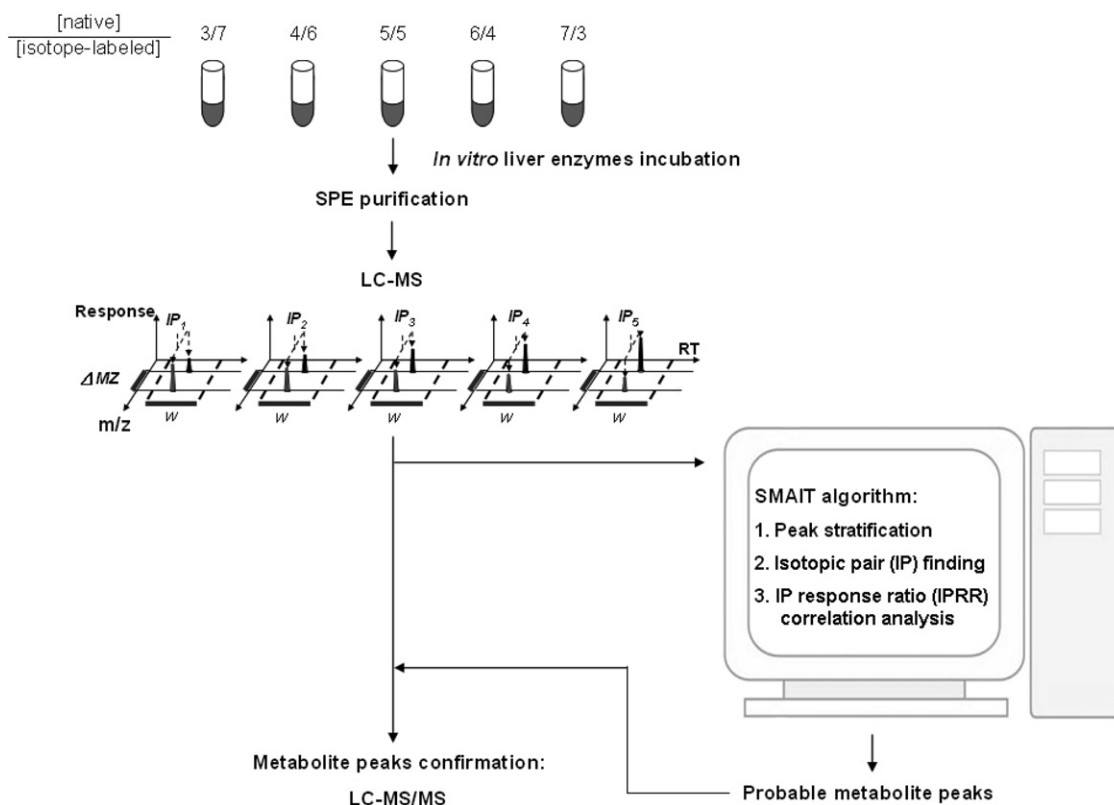


Figure 1. Analytical approach of the SMAIT strategy.

between native and labeled-isotope signals but also calculating the correlation between the varying concentration ratios and the responses. The proposed algorithm contains three steps, including peak stratification, isotopic pair finding, and IP response ratio (IPRR) correlation analysis.

Peak Stratification

LC-MS results were processed by the MES program to generate raw peak lists. The purpose of peak stratification is to consolidate m/z values in the raw peak list to a specific precision, as defined by the user based on instrumental conditions. Initially for LC-MS data, in the m/z dimension, a scan range (M_r) of 80–600 is divided into n parts (M_p) with a user-defined width (K). As shown in eq 1, a peak with M m/z is stratified into its i^{th} part (M_{pi}). If eq 1 has been fulfilled, the median of M_p in the m/z dimension is used to represent the m/z value of a peak, which is in the M_p .

$$M \in M_{pi} \pm \frac{1}{2} K \quad (1)$$

where $i = 1$ to n and $n = (M_r)/K$.

The user-defined width (K) is determined based on the m/z value variation of peaks among analytical runs. According to data from our application experiment, as described below using DEHP incubation samples, the K value was set to 0.2.

After stratification of the m/z dimension, peaks in a particular M_p with similar retention times (RTs) are classified as being in one cluster (C). Thus, the responses of these peaks are consolidated into one peak. To define the RT cluster in a particular M_p , the peaks are sorted by their RTs, and then eq 2 and 3 must be fulfilled. In these equations, R_r is a user-defined range of retention time for different chromatographic conditions. At each given M_p , a newly defined peak was assigned and its response and the RT are from the response sum and RT average of all the peaks in the same cluster. The difference of new peak RT and the first peak of next cluster should be lower than R_r .

$$\{\text{peak}_i, \text{peak}_{i+1}, \dots, \text{peak}_n\} \in C_j$$

$$\left| \overline{RT}_{C_j} - RT_{\text{peak } 1 \text{ of } C_{j+1}} \right| \leq R_r \quad (2)$$

where $i = 1 \sim n$, $j = 1 \sim n$.

i is the peak number, j is the cluster number, and C is a peak cluster defined as a new peak.

LC-MS dataset is put through the peak stratification process separately to generate the stratified peak list. According to the results from our application experiment, as described below using DEHP incubation mixtures, the R_r was set to 0.5 min.

Isotopic Pair Finding

The goal of this step is to discover the IPs of native and labeled-isotope peak doublets in a stratified peak list that have a mass shift (ΔMZ) within a user-defined RT difference (ΔRT). Equation 4 describes the criteria for defining IPs in LC-MS data. If the mass difference between MZ_i and MZ_j of peaks P_i and P_j equals ΔMZ , and the difference of the RTs between two peaks is less than ΔRT , these two peaks are determined as an IP. In this equation, RT_{P_i} and RT_{P_j} are the RTs of peaks with a mass shift, respectively.

$$|RT_{P_i} - RT_{P_j}| \leq \Delta RT \quad (3)$$

where $i = 1 \sim n$, $j = 1 \sim n$, $k = 1 \sim n$.

If eq 4 is fulfilled, the IP response ratio (IPRR) and the average of their retention times are recorded for subsequent IP correlation analysis. In the present study, the ΔMZ setting value was 4.0 and the ΔRT s of IPs in the LC-MS data were consistently smaller than 0.1 min.

IP Response Ratio (IPRR) Correlation Analysis

The goal of this step is to locate the IPs that give response ratios that correlate with expected concentration ratios. It is assumed that the IPs with high correlation coefficients have a higher probability of containing probable metabolite signals. One challenge is that it is common to observe an RT shift between different analytical runs in chromatographic analyses. The RT shift may cause a missing IP for a particular concentration ratio and increase the difficulty of metabolite signal mining using the SMAIT strategy. To solve this problem, we utilized a user-defined sliding window (W) in the RT dimension, instead of a peak alignment process. Peaks of respective samples in the window (W) are grouped to calculate their correlation coefficients. In the RT dimension, sliding window scanning is based about peaks in five LC-MS datasets to find out the IP combinations and their corresponding R values. The SMAIT program finally outputs a table containing m/z values, retention times, response ratio, and IP correlation coefficient. Based on data from the five incubation mixtures of DEHP, RT shifts of the known peaks were always less than 0.5 min. Therefore, the RT sliding window (W) was set as 0.5 min for the current chromatographic condition.

Illustration of the Statistical Procedure Using DEHP and D_4 -DEHP

According to our data from the DEHP/ D_4 -DEHP incubation mixture, when only one concentration ratio (1:1) was used, a total of 787 IPs having a 4-Da shift and within 0.1 min of the RT range were found (data not shown). To ensure that the known DEHP metabolites (MEHP, OH-MEHPs, and oxo-MEHP) were all identified, a 30% response ratio allowance (1.0 ± 0.3) was

used as an additional criterion. With this criterion, the total number of qualified IPs was reduced from 787 to 108. Although the known DEHP metabolites were included in these remaining 108 IPs, there are still many qualified IPs that require further confirmation.

After LC-MS analyses and SMAIT program screening, the IPs, which contain probable metabolite signals, were generated. To demonstrate the different filtering efficiencies of different concentration mixtures, the number of IPs recognized by SMAIT were plotted when three, four, or five concentration mixtures were used. As for three ratios, all combination generated 50 ~ 70 IPs containing 9 ~ 10 IPs ($S/N > 3$). For example, if ratios of 4:6, 5:5, and 6:4 were used, 55 IPs containing 10 IPs can be obtained. Sixty-six IPs containing 9 IPs ($S/N > 3$) were obtained while using ratios of 5:5, 6:4, and 7:3. Another three isotope ratios, 3:7, 5:5, and 7:3, showed 70 IPs containing 9 IPs ($S/N > 3$). As for the combination of using four ratios, 23 ~ 24 IPs containing 9 IPs were selected. A total of 23 IPs containing 9 IPs ($S/N > 3$) were obtained while using ratios of 4:6, 5:5, 6:4, and 7:3. For another case (3:7, 4:6, 5:5, and 6:4), result showed 24 IPs containing 9 IPs ($S/N > 3$). Here, ratios of 4:6, 5:5, and 6:4 were used as an example for three concentration ratios, while 4:6, 5:5, 6:4, and 7:3 were the example for four concentration ratios in Figure 2.

As shown in Figure 2, the correlation coefficient values of qualified IPs greater than 0.9 were plotted against the number of concentration ratios used to generate the IPs. When three native to isotope-labeled concentration ratios (4:6, 5:5, and 6:4) were used, a total of 55 probable metabolite signals were identified. With an increase in the number of concentration ratios to four (4:6, 5:5, 6:4, and 7:3) the number of probable metabolite

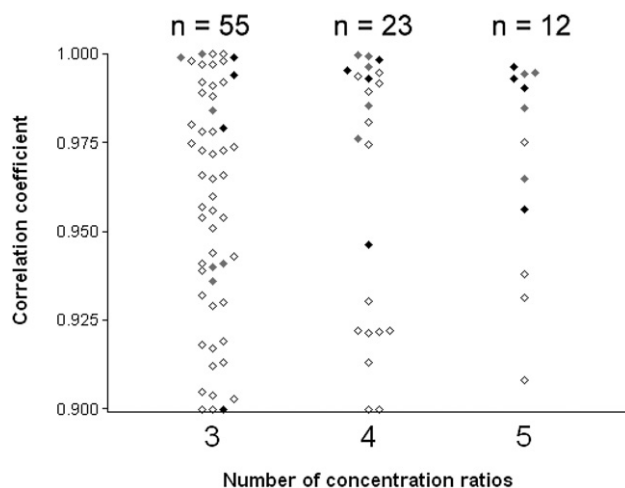


Figure 2. IP results with high R values (>0.9) in the three groups, containing three, four, and five concentration ratios, after SMAIT screening. IPs, containing probable metabolite signals, are further confirmed as to whether they are known DEHP metabolites (black diamonds) or are related but lack standard confirmation (gray diamonds). When IPs at the 5:5 concentration ratio had small signal responses ($S/N < 3$), they were considered as suspicious metabolite signals (open diamonds).

signals was reduced to 23. The number of probable metabolite signals was further reduced to 12 when five concentration ratios were used.

Probable metabolite signals were further confirmed as to whether they are known DEHP metabolites (depicted as black diamonds in Figure 2). When IPs at the 5:5 concentration ratio had small signal responses ($S/N < 3$), they were considered as suspicious metabolite signals (open diamonds). Gray diamonds indicate the S/N s of IPs greater than 3, but without reference material for a structural confirmation. There were 55, 23, and 12 IPs, containing 10, 9, and 8 IPs ($S/N > 3$) related to DEHP metabolites, respectively. These were within the groups with three, four, and five concentration ratios, respectively. Results demonstrated that the use of greater numbers of isotope concentration ratios could efficiently decrease the number of suspicious metabolites that are confirmed.

In summary, more than 3600 signals were found by the MES program when a concentration ratio of 5:5 was used (Figure 3a). Only 8 IPs fulfilled all of the user-defined requirements, including $R > 0.90$ and $S/N > 3$ (Figure 3b), after SMAIT program screening using five concentration ratios. The 8 IPs that contained probable metabolite signals with high confidence were further investigated.

The 8 IPs recognized by SMAIT as containing probable metabolite signals are listed in Table 1. Extract ion chromatograms (XICs) of the 8 IPs at a 5:5 concentration ratio are shown in Figure 4a–g, with Figure 4c containing 2 sets of IPs. In all XICs of the IPs, the D_4 -labeled isotopic peaks were consistently eluted slightly earlier than their native peaks. The RT differences were all less than 5 s. Notice that Figure 4b (m/z 291.1/295.1), c (m/z 293.1/297.1), e (m/z 361.1/365.1), and g (m/z 409.1/413.1) each contain more than one peak. However, SMAIT reported them as having one IP for each chromatographic peak cluster. This was due to the peak defining process in MES. The poorly separated peaks were recognized as one peak in MES and therefore reported as one signal. The resolution of each peak was manually calculated and results are shown in Figure 4. Based on the manual resolution calculation, these four IPs were defined to contain more than one peak, each with resolution greater than 0.75. After manual resolution calculation, 15 probable metabolite signals were defined. The number of peaks within each IP, as redefined with resolution > 0.75 , are listed in Table 1.

To confirm the chemical structures of probable metabolites, the LC-MS/MS product ion profiles of all 15 probable metabolites were obtained and compared with their commercially available standards. Tables 2 and 3 list the top four most intense product ions. Among the previously reported DEHP metabolites [9–11], only three metabolite standards were commercially available (MEHP, 5OH-MEHP, and 5oxo-MEHP). These three standards were used to confirm the chemical structures of these three metabolites. As shown in Table 2, the precursor ion of MEHP (m/z 277.1; RT 34.11 min) is a

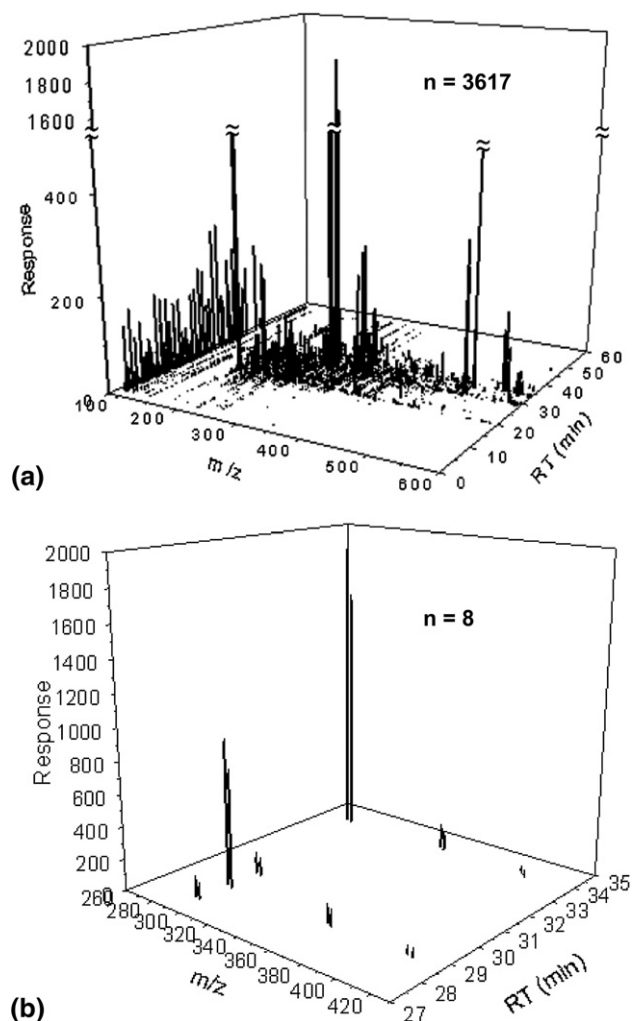


Figure 3. LC-MS data of an incubated sample with DEHP and D_4 -DEHP (5:5) plotted from the (a) raw peak list of MES and (b) IPs, containing probable metabolite signals, mined by SMAIT analysis in the group with five concentration ratios.

deprotonated ion with a single charge $[M - H]^-$ in the negative ionization mode. This generated product ions at m/z 121.0509, 127.1334, 134.0591, and 233.1851. The RT value, precursor ion mass, and product ion profile of the SMAIT recognized signal (m/z 277.1) were similar to those characteristics for the MEHP standard. This result suggests that metabolite I is MEHP. Further evidence for this was given by the product ion profile of metabolite Id (precursor ion m/z 281.1). The m/z 281.1 precursor ion generated the following product ions: m/z 125.0757, 127.1331, 138.0855, and 237.1204. Compared with the product ion profile of metabolite I, m/z 125.0757, 138.0855, and 237.1204 contained the D_4 -labeled aromatic group and yielded a 4-Da shift. Therefore, the probable metabolite signals of m/z 277.1 and 281.1 were from MEHP and D_4 -labeled MEHP, respectively.

Product ion profiles of the probable metabolite II (m/z 291.1, 28.1 min) and IV (m/z 293.1, 29.0 min) were compared with those of their corresponding commer-

cial standards (Table 2). The two probable metabolites (II and IV) were identified as 5OH-MEHP and 5oxo-MEHP, respectively. The precursor ions and product ion profiles of metabolite III were very similar to those of the 5oxo-MEHP standard. Furthermore, the small RT difference between the 5oxo-MEHP standard and metabolite III shows that metabolite III was an oxo-MEHP isomer. Precursor ions and product ion profiles of metabolites V, VI, and VII were very similar to those of the 5OH-MEHP standard. The small RT difference led us to believe that these are part of the five OH-MEHP isomers. Additionally, for metabolites V, VI, and VII, we also observed 4-Da shifts of precursor ions and product ions between native and stable isotope-labeled compounds. Therefore, metabolites V, VI, and VII were identified as the three isomers of OH-MEHPs.

Product ion profiles of the four probable metabolite signals of m/z 345.1 (VIII) and 361.1 (IX, X, XI) contained product ions at m/z 277.1 and 293.1. Furthermore, the product ions of their isotopic probable metabolites, at m/z 345.1 and 361.1, contained 4-Da shift signals, m/z 277.1 to 281.1, and m/z 293.1 to 297.1, respectively. This 4-Da shift of product ions demonstrated that the probable metabolites contained benzene and D_4 -labeled benzene groups. Additionally, the 148 Da mass difference between precursor (m/z 345) and product (m/z 197) ions indicated primarily the loss of phthalic anhydride. A similar mass difference of 148 Da was observed with ions from m/z 293 to 145. Data indicated that the metabolite VIII could be generated and modified at the side-chain of DEHP via in vitro liver enzyme incubation. The metabolites XII, XIII, XIV, and XV also showed a mass difference of 148 Da. This corresponds to a loss of the phthalic anhydride from DEHP, between the precursor and product ions. Data demonstrated that these metabolites should be modified at the side-chain of DEHP via in vitro liver enzyme incubation. We also considered that these metabolites were DEHP-derived.

The confirmed metabolites were compared with reported metabolites in the literature [9–11]. According to the DEHP metabolic pathway, DEHP is oxidized to yield MEHP, a mono-ester, in the first stage. Subsequently, MEHP is further metabolized to generate five

Table 1. DEHP metabolite signals recognized by SMAIT strategy in LC-MS data

IP (isotope-labeled)	Average RTs of IPs among five samples (min)	Correlation coefficient	Number of peaks ^a
m/z 277.1 (281.1)	34.1	0.956	1
m/z 291.1 (295.1)	28.0	0.993	2
m/z 293.1 (297.1)	28.9	0.997	3
m/z 293.1 (297.1)	29.9	0.991	1
m/z 345.1 (349.1)	34.1	0.985	1
m/z 361.1 (365.1)	28.9	0.995	3
m/z 397.1 (401.1)	34.1	0.965	1
m/z 409.1 (413.1)	28.9	0.995	3

^aTotal number of peaks present in both XICs.

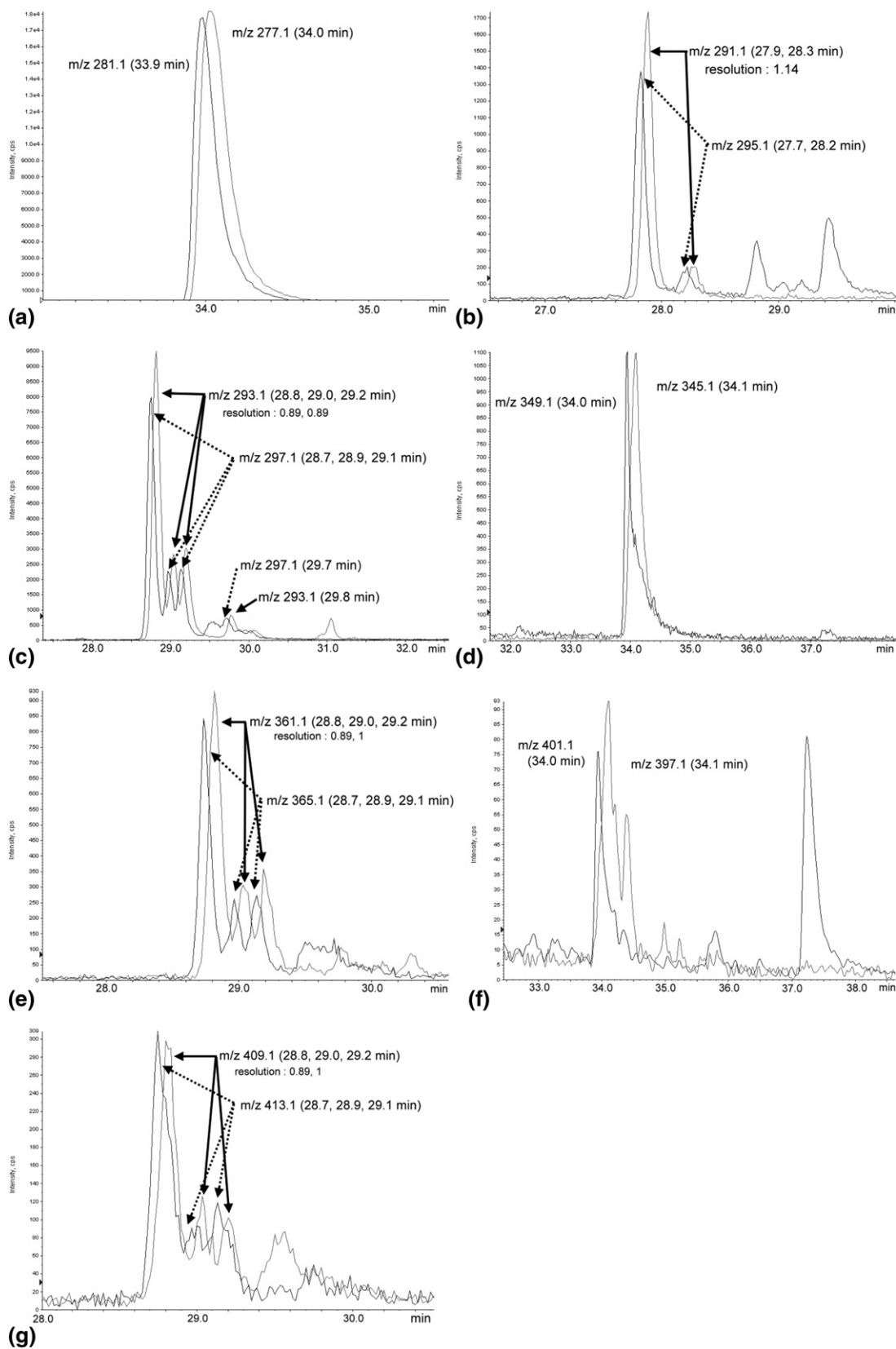


Figure 4. Extract ion chromatograms of the probable metabolite signals mined via the SMAIT strategy in a sample with DEHP and D₄-DEHP (5:5). These IPs contained (a) m/z 277.1 and 288.1, (b) m/z 291.1, and 295.1, (c) m/z 293.1 and 297.1, (d) m/z 345.1 and 349.1, (e) m/z 361.1 and 365.1, (f) m/z 397.1 and 401.1, and (g) m/z 409.1 and 413.1.

Table 2. Comparison of LC-MS/MS characteristics of seven probable metabolite signals with commercially available metabolite standards

Probable metabolites				Metabolite standards			
Number	RT (min)	Precursor ion (<i>m/z</i>)	Product ion <i>m/z</i> (%) [*]	Name	RT (min)	Precursor ion (<i>m/z</i>)	Product ion <i>m/z</i> (%) [*]
I	34.1	277.1	121.0444 (30) 127.1337 (35) 134.0612 (100) 233.1862 (20)	MEHP	34.1	277.1	121.0509 (10) 127.1334 (25) 134.0591 (100) 233.1851 (5)
Id	34.1	281.2	125.0757 (20) 127.1331 (47) 138.0855 (100) 237.1204 (14)				
II	28.1	291.1	120.9248 (66) 143.1285 (100)	5oxo-MEHP	28.0	291.1	121.0310 (70) 143.1062 (100)
IIId	28.0	295.1	125.0553 (100) 143.1046 (97)				
III	28.3	291.1	120.9358 (50) 143.0033 (100)				
IIIId	28.2	295.1	124.9691 (30) 143.0034 (100)				
IV	29.0	293.1	121.0497 (83) 145.1455 (100)	5OH-MEHP	28.6	293.1	121.0263 (66) 145.1173 (100)
IVd	29.0	297.1	125.0742 (67) 145.1455 (100)				
V	29.2	293.1	121.0335 (20) 145.1345 (100)				
Vd	29.1	297.1	125.0732 (47) 145.1450 (100)				
VI	29.3	293.1	121.0450 (21) 145.1457 (100)				
VIId	29.2	297.1	125.0715 (11) 145.1462 (100)				
VII	30.1	293.1	121.0401 (13) 145.1435 (100)				
VIIId	30.0	297.1	124.9714 (17) 145.1447 (100)				

^{*}The top four intensive product ions are shown. The proportion of relative abundance to the most intense product ion is shown in parentheses.

isomers of OH-MEHPs, called secondary metabolites, in the second stage. The hydroxyl metabolites can be oxidized to generate oxo-metabolites. For the DEHP metabolism pathway, secondary metabolites coming from the primary metabolite (MEHP) were reported for five OH-MEHP isomers having different hydroxyl group binding positions. In results, seven metabolite signals, MEHP, four OH-MEHP isomers, and two oxo-MEHP isomers, were confirmed by commercial standards. The four OH-MEHP isomers were recognized, although only one metabolite (5OH-MEHP) could be confirmed by a commercial standard. In the third stage of DEHP metabolism, two oxo-MEHP isomers were observed. The 5oxo-MEHP showed a higher response compared with the other oxo-MEHP isomer. Similar results have been reported [9]. Other reported metabolites that were not recognized by the SMAIT program may be the result of metabolite signal intensities that were too low to be detected by LC-MS.

Recently, LC-MS instrumentation with high sensitivity, selectivity, and productivity has shown its usefulness in the development of metabolite identification

strategies [14–17]. Several LC-MS-coupled strategies, such as the radio isotope-labeling technology [18–23], metabolite prediction strategy [15, 15], exact mass filter strategy [24–26], and signal tracing by particular isotopic ratio strategy [6], have been reported for the assistance of metabolite identification. It is considered that radio isotope-labeling technology is a golden method for metabolite ID, but it has several advantages, including the associated health hazards and waste disposal issues. The metabolite prediction strategy has been utilized to predict metabolic routes that a parent compound will likely undergo. This method then specifically looks for mass changes between the parent compound and its metabolites to confirm the prediction. However, this strategy relies on previous knowledge of the metabolic pathways. Meanwhile, the exact mass filtering strategy uses mass defects as a specific filter. This is based on the exact mass and mass deficiencies specific to each parent drug of interest. The strategy can help to remove unwanted peaks in LC-MS data and dramatically reduce the number of potential metabolite signals.

Table 3. Product ion profiles of DEHP structurally-related metabolites identified by SMAIT

Number	IP (<i>m/z</i>)	RT (min)	Precursor ion (<i>m/z</i>)	Product ion (%)*
VIII	345.1	34.1	345.1	277.1358 (100), 197.1162 (25)
VIII _d	349.1	34.1	349.1	281.1605 (100), 197.1094 (20)
IX	361.1	29.0	361.1	293.1294 (100), 145.1232 (22)
IX _d	365.1	28.9	365.1	297.1578 (100), 145.1260 (17)
X	361.1	29.2	361.1	293.1270 (100), 145.1211 (22)
X _d	365.1	29.2	365.1	297.1564 (100), 145.1265 (16)
XI	361.1	29.4	361.1	293.1291 (100), 145.1261 (63)
XI _d	365.1	29.3	365.1	297.1599 (100), 145.1263 (50)
XII	397.1	34.1	397.1	249.0506 (100)
XII _d	401.1	34.0	401.1	249.0484 (100)
XIII	409.1	28.8	409.0	261.0939 (100)
XIII _d	413.1	28.9	413.1	261.0912 (100)
XIV	409.1	29.0	409.0	261.1086 (100)
XIV _d	413.1	29.1	413.1	261.0937 (100)
XV	409.1	29.2	409.0	261.0909 (100)
XV _d	413.1	29.3	413.1	261.0922 (100)

*Intense product ions are shown. The proportion of relative abundance to the most intense product ion is shown in parentheses.

Isotope-labeling method coupled with LC-MS analysis provides a route to trace compound-derived metabolite signals. As for SMAIT, the calculation of the correlation coefficient among five different ratios and their responses, rather than only one ratio, are carried out to enhance the compound-derived ion detection efficiency. SMAIT is currently available through the website located at <http://binfo.csie.ncku.edu.tw:8080/SMAIT/>. Noise removal, baseline correction, and peak detection procedures should be accomplished by other peak-picking software (e.g., MES) before SMAIT program. The input data structure of SMAIT is a *.txt file with three columns, peak *m/z*, retention time and response. (As illustrated in Supporting Information A, which can be found in the electronic version of this article).

Additionally, another approach was suggested that uses the direct point out method (e.g., radio-isotope tracer and stable isotope tracer) combined with the exact mass filter strategy. This could also increase the accuracy of metabolite detection [17].

Conclusions

In the present study, the proposed statistical procedure could enhance the efficiency and confidence of detecting compound-derived signals in LC-MS data. A computational program (SMAIT) was developed to execute the data processing and statistical analysis. However, the procedure has not yet been applied to an in vivo experiment. Signal levels within in vivo matrices are often low and masked by endogenous signals, which may result in peak loss. Besides, in vivo experiments with this strategy may be costly, laborious, and time-prohibitive for five different ratios used. Thus, further work should be carried out to investigate the applicability and the flexibility of this strategy to in vivo experiments. Additionally, the procedure was limited by the artificial isotopes, which prohibit its utility in

mining fixed ratio natural isotopes, such as Cl, Br, etc. [6].

Acknowledgments

The authors acknowledge support for this study by grants NSC97-2113-M-006-005-MY3, NSC94-2113-M006-06, and NSC97-2113-M-006-005-MY3 from the National Science Council, Landmark Project of National Cheng Kung University, and the National Cheng Kung University Project of Promoting Academic Excellence and Developing World Class Research Centers from the Ministry of Education of Taiwan.

Appendix A Supplementary Material

Supplementary material associated with this article may be found in the online version at [doi:10.1016/j.jasms.2009.10.002](https://doi.org/10.1016/j.jasms.2009.10.002).

References

- Griffin, I. J. Using Stable Isotopes and Isotope Ratio Mass Spectrometry to Study Mineral Metabolism in Humans. *J. Anal. At. Spectrom.* **2002**, *17*, 1186–1193.
- Yan, Z.; Caldwell, G. W. Stable-Isotope Trapping and High-Throughput Screenings of Reactive Metabolites Using the Isotope MS Signature. *Anal. Chem.* **2004**, *76*, 6835–6847.
- Sanchez-Ponce, R.; Guengerich, F. P. Untargeted Analysis of Mass Spectrometry Data for Elucidation of Metabolites and Function of Enzymes. *Anal. Chem.* **2007**, *79*, 3355–3362.
- Abramson, F. P. The Use of Stable Isotopes in Drug Metabolism Studies. *Semin. Perinatol.* **2001**, *25*, 133–138.
- Chen, C.; Gonzalez, F. J.; Idle, J. R. LC-MS-based Metabolomics in Drug Metabolism. *Drug Metab. Rev.* **2007**, *39*, 581–597.
- Zhang, H. A Software Filter to Remove Interference Ions from Drug Metabolites in Accurate Mass Liquid Chromatography/Mass Spectrometric Analyses. *J. Mass Spectrom.* **2003**, *38*, 1110–1112.
- Zhu, M. Detection and Characterization of Metabolites in Biological Matrices Using Mass Defect Filtering of Liquid Chromatography/High Resolution Mass Spectrometry Data. *Drug Metab. Dispos.* **2006**, *34*, 1722–1733.
- Balogh, M. P. Spectral Interpretation, Part II: Tools of the Trade. *LC-GC North Am.* **2006**, *24*, 762–769.
- Bradford, M. M. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.* **1976**, *72*, 248–254.

10. Wu, H. Y.; Tseng, V. S.; Liao, P. C. Mining Phosphopeptide Signals in Liquid Chromatography-Mass Spectrometry Data for Protein Phosphorylation Analysis. *J. Proteome Res.* **2007**, *6*, 1812–1821.
11. Lee, M. S.; Kerns, E. H. LC-MS Applications in Drug Development. *Mass Spectrom. Rev.* **1999**, *18*, 187–279.
12. Lafaye, A.; Junot, C.; Ramounet-Le Gall, B.; Fritsch, P.; Ezan, E.; Tabet, J. C. Profiling of Sulfoconjugates in Urine Using Precursor Ion and Neutral Loss Scans in Tandem Mass Spectrometry. Application to the Investigation of Heavy Metal Toxicity in Rats. *J. Mass Spectrom.* **2004**, *39*, 655–664.
13. Castro-Perez, J.; Plumb, R.; Liang, L.; Yang, E. A High-Throughput Liquid Chromatography/Tandem Mass Spectrometry Method for Screening Glutathione Conjugates Using Exact Mass Neutral Loss Acquisition. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 798–804.
14. Liu, D. Q.; Karanam, B. V.; Doss, G. A.; Sidler, R. R.; Vincent, S. H.; Hop, C. E. C. A. In Vitro Metabolism of MK-0767 [(±)-5-[(2,4-Dioxothiazolidin-5-yl) Methyl]-2-Methoxy-N-[(4-Trifluoromethyl)-Phenyl] Methyl] Benzamide], a Prexisome Proliferator-Activated Receptor Alpha/Gamma Agonist. II. Identification of Metabolites by Liquid Chromatography-Tandem Mass Spectrometry. *Drug Metab. Dispos.* **2004**, *32*, 1023–1031.
15. Testa, A. L.; Balmat, A. L.; Long, A. Predicting Drug Metabolism: Concepts and Challenges. *Pure Appl. Chem.* **2004**, *76*, 907–914.
16. Testa, A. L.; Balmat, A. L.; Long, A.; Judson, P. Predicting Drug Metabolism—An Evaluation of the Expert System METEOR. *Chem. Biodiv.* **2005**, *2*, 872–885.
17. Castro-Perez, J. M. Current and Future Trends in the Application of HPLC-MS to Metabolite-Identification Studies. *Drug Discov. Today.* **2007**, *12*, 249–256.
18. Albro, P. W.; Hass, J. R.; Peck, C. C.; Odam, D. G.; Gorbett, J. T.; Bailey, F. J.; Blatt, H. E.; Barrett, B. B. Identification of the Metabolites of Di-(2-Ethylhexyl) Phthalate in Urine from the African Green Monkey. *Drug Metab. Dispos.* **1981**, *9*, 223–225.
19. Ostin, A.; Kowalyczak, M.; Bhalerao, R. P.; Sandberg, G. Metabolism of Indole-3-Acetic Acid in Arabidopsis. *Plant Physiol.* **1998**, *118*, 285–296.
20. Grova, N.; Feidt, C.; Laurent, C.; Rychen, G. [14C] Milk, Urine, and Feces Excretion Kinetics in Lactating Goats After an Oral Administration of [14C]Polycyclic Aromatic Hydrocarbons. *Int. Dairy J.* **2002**, *12*, 1025–1031.
21. Branfman, A. R.; McComish, M. F.; Bruni, R. J.; Callahan, M. M.; Robertson, R.; Yesair, D. W. Characterization of Diaminouracil Metabolites of Caffeine in Human Urine. *Drug Metab. Dispos.* **1983**, *11*, 206–210.
22. VandenHeuvel, W. J. A. Drug Metabolite Identification: Stable Isotope Methods. *J. Clin. Pharmacol.* **1986**, *26*, 427–434.
23. Baranczewski, P.; Stanczak, A.; Kautiainen, A.; Sandin, P.; Edlund, P. Introduction to Early In Vitro Identification of Metabolites of New Chemical Entities in Drug Discovery and Development. *Pharmacol. Rep.* **2006**, *58*, 341–352.
24. Albro, P. W.; Tondeur, I.; Marbury, D.; Jordan, S.; Schroeder, J.; Corbett, J. T. Polar Metabolites of Di-(2-Ethylhexyl)Phthalate in the Rat. *Biochim. Biophys. Acta* **1983**, *760*, 283–292.
25. ATSDR. *Toxicological Profile for Di(2-Ethylhexyl) Phthalate*; Atlanta; 2002.
26. Koch, H. M.; Bolt, H. M.; Preuss, R.; Angerer, J. New Metabolites of Di(2-Ethylhexyl)Phthalate (DEHP) in Human Urine and Serum After Single Oral Doses of Deuterium-Labeled DEHP. *Arch. Toxicol.* **2005**, *79*, 367–376.