analytical chemistry

Article

pubs.acs.org/ac

Nontargeted Modification-Specific Metabolomics Study Based on Liquid Chromatography–High-Resolution Mass Spectrometry

Weidong Dai,[†] Peiyuan Yin,[†] Zhongda Zeng,[†] Hongwei Kong,[†] Hongwei Tong,[‡] Zhiliang Xu,[†] Xin Lu,[†] Rainer Lehmann,*[,§,∥,⊥] and Guowang Xu*[,†]

[†]Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China

[‡]The First Affiliated Hospital of Zhejiang Chinese Medical University, Hangzhou 310006, China

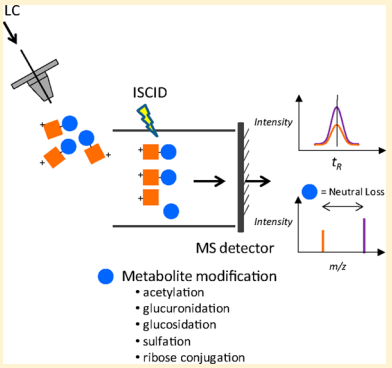
[§]Division of Clinical Chemistry and Pathobiochemistry, Department of Internal Medicine IV, University Hospital Tuebingen, 72076 Tuebingen, Germany

[∥]Institute for Diabetes Research and Metabolic Diseases of the Helmholtz Centre Munich at the University of Tuebingen, 72076 Tuebingen, Germany

[⊥]German Center for Diabetes Research (DZD), 72076 Tuebingen, Germany

Supporting Information

ABSTRACT: Modifications of genes and proteins have been extensively studied in systems biology using comprehensive analytical strategies. Although metabolites are frequently modified, these modifications have not been studied using -omics approaches. Here a general strategy for the nontargeted profiling of modified metabolites, which we call "nontargeted modification-specific metabolomics", is reported. A key aspect of this strategy was the combination of in-source collision-induced dissociation liquid chromatography–mass spectrometry (LC-MS) and global nontargeted LC-MS-based metabolomics. Characteristic neutral loss fragments that are specific for acetylation, sulfation, glucuronidation, glucosidation, or ribose conjugation were reproducibly detected using human urine as a model specimen for method development. The practical application of this method was demonstrated by profiling urine samples from liver cirrhosis patients. Approximately 900 features were identified as modified endogenous metabolites and xenobiotics. Moreover, this strategy supports the identification of compounds not included in traditional metabolomics databases (HMDB, Metlin, and KEGG), which are currently referred to as "unknowns" in metabolomics projects. Nontargeted modification-specific metabolomics opens a new perspective in systems biology.



Modifications of genes and proteins modulate various biological processes and pathogenic events.[1−3] More than 300 types of posttranslational modifications of proteins have been discovered[4] and studied over the past several decades.[5−7] Modifications such as phosphorylation, acetylation, methylation, glucuronidation, and sulfation occur on large molecules as well as on metabolites. These modifications can change the polarity, solubility, structure, bioactivity, and cellular localization of metabolites and can play pivotal roles in the regulation and alteration of metabolic and signaling processes in bacteria, fungi, plants, animals, and humans. Furthermore, the importance of the metabolic modifications of exogenous compounds in animals and humans is now well recognized, and drug metabolism studies have become routine in drug research.[8,9] Endogenous modified metabolites have long been investigated using gas or liquid chromatography[10,11] as well as capillary electrophoresis.[12] However, these analytical methods, since targeted, cover only a small number of known analytes. Large-scale profiling studies of endogenous metabolite modifications have not been reported.

The sensitivity and resolution of liquid chromatography–mass spectrometry (LC-MS) have significantly improved in the past decade.[13,14] An increasing number of mass spectrometric signals can now be detected in nontargeted metabolomic profiles.[15,16] However, processing the signals from a list of several hundreds or even thousands of detected features to identify metabolites remains a major challenge.[17,18] The majority of these features, including numerous modified metabolites, remain unidentified "unknowns", representing a significant loss of valuable information. Therefore, the development of an analytical method to gain insight into these "unknowns" would considerably improve the data obtained from nontargeted metabolomics studies.

In this study, we sought to develop a nontargeted strategy combining novel large-scale detection of modified metabolites and common global metabolomic profiling. An in-source

Received: June 4, 2014
Accepted: August 25, 2014
Published: August 25, 2014

   dx.doi.org/10.1021/ac502045j | *Anal. Chem.* 2014, 86, 9146–9153

collision-induced dissociation (ISCID)-based liquid chromatography–high-resolution mass spectrometry (LC-HRMS) approach was developed to profile modified metabolites in a nontargeted way using mixtures of metabolite standards and a complex model specimen, human urine, as urine contains numerous different unmodified and modified metabolites. This strategy is termed "nontargeted modification-specific metabolomics" (NT-MSM), and its practical application was demonstrated by comparing modified metabolites in urine samples from patients with liver cirrhosis and from healthy individuals.

EXPERIMENTAL SECTION

Chemicals. Taurine was purchased from Acros (Geel, Belgium). Acetic anhydride was purchased from Sinopharm Chemical Reagent (Beijing, China). *N*-Acetylglycine was purchased from Fluka (St. Louis, MO, USA). 3-Methyluridine was purchased from International Laboratory (South San Francisco, CA, USA). 8-Hydroxyguanosine was purchased from J&K Scientific (Beijing, China). Sulfatase (≥2,000 units/mL) and β-glucuronidase (≥85,000 units/mL), both from Type H-2 *Helix pomatia*, pyridine, L-ascorbic acid, and standards of 1-methylhistidine, 1-methyl-L-tryptophan, 1-methylalanine, N_2-methylguanosine, *N*-acetyl-5-hydroxytryptamine, *N*-ω-acetylhistamine, 1-acetylisatin, *N*-acetyl-L-phenylalanine, *N*-acetyl-DL-leucine, estradiol-3-glucuronide-17-sulfate, estrone-glucuronide, estradiol-glucuronide, estriol-glucuronide, 1-cysteine-*S*-sulfate, pregnenolone-sulfate, estradiol-3-sulfate, indoxyl sulfate, cytosine, adenine, hypoxanthine D-riboside, orotic acid, uridine, glycocholic acid, glycodeoxycholic acid, glycochenodeoxycholic acid, glycoursodeoxycholic acid, palmitoyl-DL-carnitine, 1-palmitoyl-*sn*-glycero-3-phosphocholine, palmitoylethanolamide, 5′-UMP, 3′-AMP, 5′-AMP, 5′-ADP, daidzein-7-*O*-glucoside, phenyl-β-D-glucopyranoside, and genistein-7-glucoside were purchased from Sigma (St. Louis, MO, USA). Acetonitrile of HPLC grade was from Merck (Darmstadt, Germany). Deionized water used for all experiments was purified with a Milli-Q system (Millipore, USA).

Nontargeted Modification-Specific Metabolomics Analysis. All LC-MS analyses were performed with an Acquity UPLC system (Waters, MA, USA) coupled to a LTQ-Orbitrap XL (Thermo Fisher Scientific, Waltham, MA, USA). Chromatographic separation of the metabolite standards and human urine was performed using a BEH T3 column (10 cm × 2.1 mm, 1.7 μm, Waters). Mobile phases A and B were 0.1% formic acid solution and acetonitrile, respectively. The linear gradient program was as follows: from start to 5 min, phase B was linearly increased from 2% to 5%, then increased to 30% over the next 10 min, and then linearly increased to 98% during the next 6 min and held at 98% for 6 min. The flow rate was 0.28 mL/min. The injection volume for the LC-MS analysis with the ISCID voltage (5–45 V) was 10 μL. The MS instrument settings were slightly modified from previous studies.[19,20] The following conditions were applied for mass spectrometry: capillary temperature 325 °C, source voltage 4.5 kV, and capillary voltage 49 V for ESI+ analysis and capillary temperature 325 °C, source voltage −4.0 kV, and capillary voltage −40 V for ESI− analysis. The mass scan range was set to 70–1000. The resolution of the Orbitrap was set to 30,000. The peak picking and alignment were processed using Sieve software (V1.2, Thermo Fisher Scientific, USA) applying a mass width of 0.02 Da and a retention time width of 0.5 min. The number of maximum frames was set at 5,000. Water injection was used to define the background noise. Ions showing an intensity higher than 5000 (ESI+) or 2000 (ESI−) counts in the blank sample were removed as background noise from the peak table.[21] The remaining ions were used for further data processing. The settings for the global nontargeted metabolomics analyses of batch samples, i.e., the LC program, MS instrument parameters, and peak alignment procedure, were the same as the ISCID-based LC-HRMS analysis (without ISCID voltage), with the exception of the injection volume (4 μL) and the mass scan range (m/z 100–1000).

A typical working sequence for the combined ISCID-based LC-HRMS and the global nontargeted metabolomic analysis comprised the following: (i) investigation of a blank sample (H_2O) to define the background noise; (ii) 18 separate nontargeted modification-specific metabolomics analyses of the pooled sample with ISCID voltages of 5, 10, 15, 20, 25, 30, 35, 40, and 45 V respectively applied in positive and negative ionization modes; (iii) individual analysis of a second aliquot of each sample using the conventional global nontargeted metabolomics method in positive and negative ionization modes.

Neutral Loss MSFinder Software and Nontargeted Modification-Specific Metabolomics Data Processing. The data mining procedure for the modified metabolome was performed as follows: the aligned peak tables from the modification-specific metabolomics analysis with different ISCID voltages were analyzed using in-house software for ion pair matching (the neutral loss MSFinder software will be provided at no cost upon request; please contact the corresponding authors). Ions with identical retention times (Δt_R < 0.1 min) and characteristic neutral losses (e.g., NL = 42.01056 (acetylation), 176.03209 (glucuronidation), 79.95681 (sulfation), 132.04226 (ribose conjugation), or 162.05282 (glucosidation), ΔNL < 0.002 Da) were identified as ion pairs of parent ions and fragment ions of modified metabolites. The parent ions were retained to reconstruct a peak table.

$$\left|t_{R(\text{parent ion})} - t_{R(\text{fragment ion})}\right| < 0.1 \text{ min}$$

$$\& \left| m/z_{(\text{parent ion})} - m/z_{(\text{fragment ion})} - \text{NL} \right| < 0.002 \text{ Da}$$

Using the neutral loss MSFinder software, the ions in the reconstructed peak tables were matched with the corresponding ions in the individual peak tables obtained by global metabolomics analysis to obtain the intensities of each peak in the individual samples. The Δt_R and ΔNL parameters were also set to 0.1 min and 0.002 Da. The matched modified metabolite features that were detected by different ISCID voltages were combined to obtain a final modified metabolites peak table (Excel, Microsoft).

Assessment of Reproducibility Using Human Urine. Morning urine samples from 10 healthy volunteers (5 males and 5 females) who fasted overnight were collected under controlled conditions. The samples were pooled and used to develop and establish the method. The urine was immediately frozen in several aliquots at −80 °C and stored until further use. After thawing, the urine aliquots were mixed using a vortex mixer, pooled, and centrifuged at 16,300 g at 4 °C for 10 min (BioFuge Stratos, Thermo Scientific). In total, 500 μL of urine supernatant was diluted with 1000 μL of deionized water for LC-MS analysis. To examine the reproducibility of the ISCID-based LC-HRMS method, three pooled urine samples (diluted 1:3 with water, v/v) were pretreated separately and analyzed by

dx.doi.org/10.1021/ac502045j | Anal. Chem. 2014, 86, 9146–9153

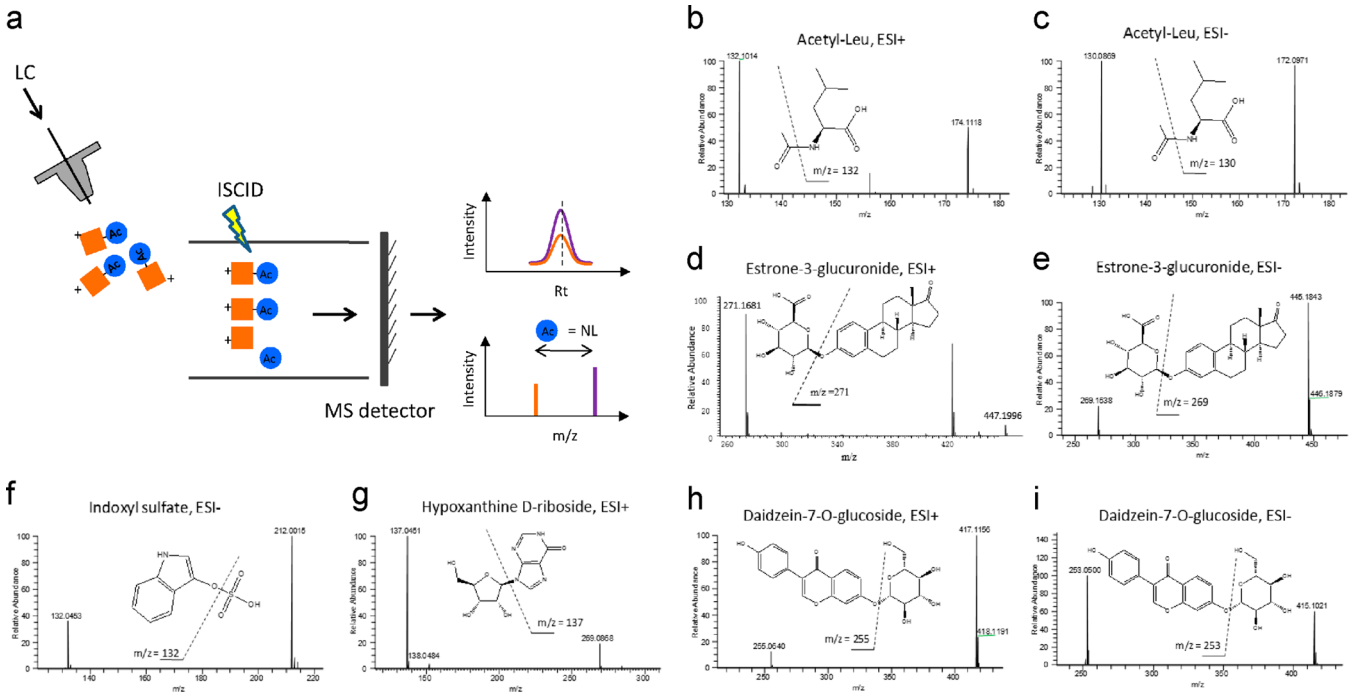


Figure 1. (a) Principle of in-source collision-induced dissociation (ISCID)-based modification-specific profiling exemplarily illustrated for an acetylated metabolite analyzed in the positive ionization mode. Typical ISCID-based fragmentation pattern exemplarily shown for modified metabolites: (b and c) acetylation (theoretical NL m/z = 42.01056) detectable in both ESI+ (ISCID voltage = 20 V for acetylleucine) and ESI− mode (ISCID voltage = −25 V for acetylleucine); (d and e) glucuronidation (theoretical NL m/z = 176.03209) detectable in both ESI+ (ISCID voltage =10 V for estrone-3-glucuronide) and ESI− mode (ISCID voltage = −40 V for estrone-3-glucuronide); (f) sulfation (theoretical NL m/z = 79.95681) detectable in ESI− mode (ISCID voltage = −30 V for indoxyl sulfate); (g) ribose conjugation (theoretical NL m/z = 132.04226) detectable in ESI+ mode (ISCID voltage = 20 V for hypoxanthine D-riboside); (h and i) glucosidation (theoretical NL m/z = 162.05282) detectable in both ESI+ (ISCID voltage = 35 V for daidzein-7-*O*-glucoside) and ESI− mode (ISCID voltage = −35 V for daidzein-7-*O*-glucoside).

nontargeted modification-specific metabolomics analysis and conventional global metabolomics analysis. Following this procedure, the data set was processed to generate a modified metabolite feature list. Between the three different analyses, features with values of Δt_R < 0.1 min and $\Delta m/z$ < 0.002 Da were assigned as originating from identical metabolite features.

Assessment of the Accuracy of Metabolite Modification Identification. To confirm the presence of the detected glucuronic acid and sulfuric acid conjugated metabolites, modification-specific enzymatic cleavage was performed for glucuronidated and sulfated compounds in pooled urine from healthy volunteers using β-glucuronidase and sulfatase solutions. The enzymatic cleavage procedure was as follows: after thawing, the pooled urine of the 10 healthy subjects was centrifuged at 16,300g at 4 °C for 10 min. In total, 500 μL of a freshly prepared enzymatic hydrolysis buffer containing 0.15 M sodium acetate buffer (pH 4.4), 2.0 mg of L-ascorbic acid, and 5 μL of β-glucuronidase solution or sulfatase solution was added to a 500 μL aliquot of urine supernatant. The samples were incubated at 37 °C for 24 h, as described.[22,23] As a control, another aliquot of pooled urine was treated in the same manner without the addition of β-glucuronidase or sulfatase. After the enzymatic cleavage step, the samples were diluted with 500 μL of water and centrifuged at 16,300g at 4 °C for 15 min before LC-HRMS analysis.

Metabolomics Database Searches. The assigned modified metabolite ions were identified by database searches in the HMDB[24] (http://www.hmdb.ca/spectra/ms/search) and Metlin[25] (http://metlin.scripps.edu/) databases. If the HMDB and Metlin database searches resulted in no hits, the metabolite with the assigned modification was searched in the MyCompoundsID[26] (http://www.mycompoundid.org) database. The mass tolerance for the HMDB database search was set at 0.003 Da. The mass tolerance for the Metlin and MyCompoundID database searches was set at 10 ppm. The parameter of "reactions" was set to "1 reaction" for the MyCompoundID database search. The chromatographic retention behavior was also considered to reduce false-positive matches.

Application of the Developed Method for Studying Urine Samples from Patients with Liver Cirrhosis and Healthy Individuals. The second morning, urine samples from age-matched volunteers who fasted overnight were used. These volunteers included 21 male healthy controls (43.4 ± 13.5 years) and 20 male cirrhotic patients (46.9 ± 10.6 years). The urine samples were collected from the Division of Surgery (First Affiliated Hospital of Zhejiang Chinese Medicine University; Hangzhou, China). The experiments were carried out according to the guidelines and approval of the local Ethics Committee of the hospital. Informed consent was obtained from all subjects. After the collection, the urine samples were immediately frozen in several aliquots at −80 °C and stored until further use. The liver cirrhosis patients were diagnosed on the basis of liver function tests, laboratory parameters, ultrasonography, and computed tomography or magnetic resonance imaging results. The ISCID-based LC-HRMS-based profiling of modified metabolites was performed by preparing a pooled urine sample of healthy controls and liver cirrhosis patients by mixing equal volumes of each urine sample, followed by centrifugation at 16,300g at 4 °C for 10

dx.doi.org/10.1021/ac502045j | Anal. Chem. 2014, 86, 9146−9153

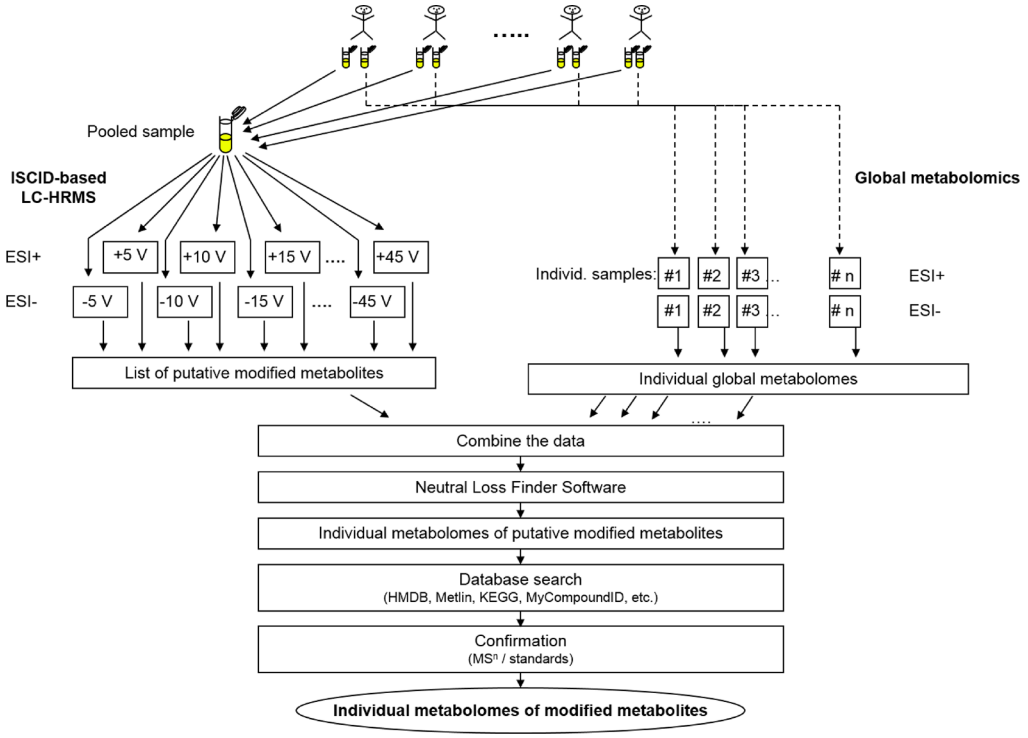


Figure 2. Overview of the multistep procedure of nontargeted modification-specific metabolomics, exemplified by a biomedical study.

min. The supernatant of the pooled urine sample was then diluted with two volumes of deionized water, and the analysis was performed as described above. In addition, a separate aliquot of each urine sample from healthy controls and liver cirrhosis patients was centrifuged at 16,300*g* at 4 °C for 10 min, and 100 μL of the supernatant was subsequently diluted with 200 μL of deionized water for conventional global nontargeted metabolomics LC-MS analysis in ESI+ and ESI− mode as described above.

Data Pretreatment for the Liver Cirrhosis Study. Before the univariate analysis and multivariate analysis, the intensity of each detected ion was normalized to the total ion intensity in positive or negative mode to minimize the effects associated with the use of spot urine. The intensities were then multiplied by a constant (10,000) and subjected to the "80% rule".[27,28] Univariate analysis of Student's *t*-test and nonparametric test were performed to select significantly changed modified metabolites using PASWstat software (Version 18.0, USA). Multivariate analysis was performed using SIMCA-P 11.0 software (Umetrics AB, Umeå, Sweden).

RESULTS AND DISCUSSION

Overview of the Strategy for Combined Nontargeted Modification-Specific and Global Metabolomics. We assumed that the successful identification of metabolite modification includes the following preconditions: modified metabolite molecules are first ionized in the ion source of the mass spectrometer, some of which dissociate at the bond between the substrate and the modifying group when a fragmentation voltage is applied. The precursor ions of the modified metabolite and its fragment ions are then simultaneously detected by a high-resolution mass spectrometer analyzer to produce a specific neutral loss (NL) value uniquely representing the modifying group (Figure 1a). The key to nontargeted profiling of the modified metabolites was the introduction of ISCID, a nonselective dissociation that occurs during the ionization or ion-sampling process and is generally used to remove adducts of H_2O and other solvent molecules for high-resolution metabolomics analysis. In this study, we used ISCID to remove modifying groups from the modified metabolites and then used NL matches of high-resolution mass spectral features to acquire the structural information for the metabolite modifications (Figure 1a).

This NT-MSM strategy was designed as a multistep procedure for high-resolution nontargeted profiling of modified metabolites (Figure 2). Two aliquots of each sample were utilized: one for common global metabolic profiling and one for NT-MSM analysis. In step 1, one aliquot of the pooled sample set was used for ISCID-based LC-HRMS. Samples were pooled for nontargeted profiling of modified metabolites based on the assumption that this pooled sample was representative of the modified metabolites present in the sample set. In step 2, ISCID-based LC-HRMS was performed at different ISCID voltages in positive and negative ionization modes to generate modification-specific neutral loss fragments. For each applied ISCID voltage, a separate run was performed. In step 3, a separate aliquot of each sample was analyzed using a common global nontargeted metabolomics method. In step 4, raw data files from step 2 and step 3 were subjected to peak detection and alignment to generate peak tables. Step 5 consisted of applying the neutral loss MSFinder software (self-developed) for ion pair matches to the ISCID-based LC-HRMS data to identify modified metabolite features. In step 6, the putative modified metabolite list from the ISCID-based LC-HRMS analysis was combined with the global metabolomics results using the neutral loss MSFinder software to acquire individual responses for putative modified metabolites. Finally, a table containing the final peaks for the individual metabolomes of modified metabolites was generated following a database search and confirmation of modified metabolites. Details of the

dx.doi.org/10.1021/ac502045j | *Anal. Chem.* 2014, 86, 9146−9153

method development, LC-MS analysis, and data processing procedure are provided below and in the Experimental Section.

Nontargeted Profiling Analysis of Modified Metabolites by ISCID-Based LC-HRMS. In the initial step of our method development process, we aimed to establish a procedure that allowed profiling of different metabolite modifications simultaneously. This goal could not be reached by applying only a single ISCID voltage. Accordingly, a range of standards covering nine different metabolite modifications was investigated in positive and negative ionization modes at different ISCID voltages ranging from 5 to 55 V to study the fragmentation of methylated, acetylated, glucuronidated, sulfated, phosphorylated, ribose conjugated, glycine conjugated, glucosidated, and palmitoylated metabolites. Acetylation (Figure 1b,c), glucuronidation (Figure 1d,e), sulfation (Figure 1f), ribose conjugation (Figure 1g), and glucosidation modifications (Figure 1h,i) resulted in characteristic NL fragments, facilitating unequivocal detection based on specific theoretical NL mass-to-charge (m/z) values of 42.01056 ($-C_2H_2O$), 176.03209 ($-C_6H_8O_6$), 79.95681 ($-SO_3$), 132.04226 ($-C_5H_8O_4$), and 162.05282 ($-C_6H_{10}O_5$), respectively. However, the applied ISCID voltages did not lead to NL fragments suitable for profiling methylated, palmitoylated, phosphorylated, and glycine conjugated metabolites.

We proceeded to evaluate the developed conditions to profile acetylation, glucuronidation, sulfation, ribose conjugation, and glucosidation modifications in a more complex sample, pooled human urine, which contains numerous unmodified and modified metabolites. As shown in Figure 3, the application of different ISCID voltages was required for the analysis of molecules with various distinct chemical bonding characteristics between the substrate and the modification group to generate specific neutral loss fragments. First, we needed to define a suitable number of ISCID voltages because each applied ISCID condition required a separate LC-MS run. The pooled human urine was investigated in the ESI+ and ESI− modes with applied ISCID voltages of 5 to 55 V. Figure 3 shows the sum of the detected modified metabolite features in pooled urine per modification per applied voltage from an ISCID voltage of 5 V. Finally, we selected 18 different ISCID voltages with intervals of 5 V from 5 to 45 V and from −5 V to −45 V for subsequent investigations.

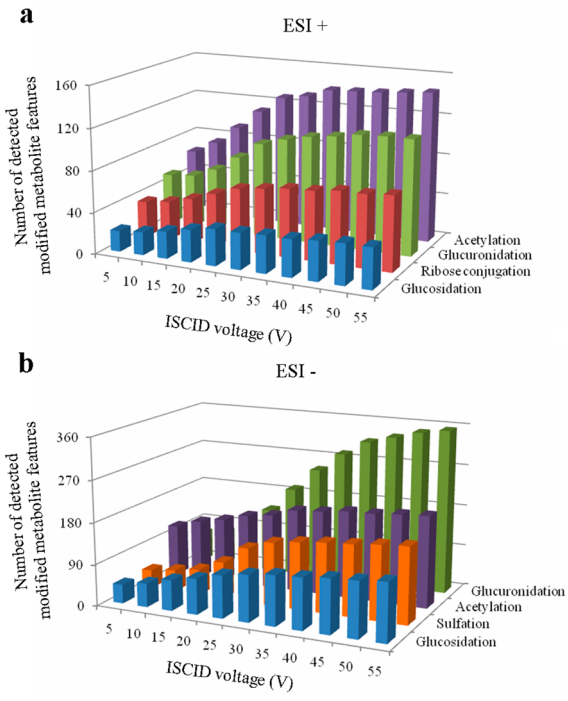


Figure 3. Effect of different in-source collision-induced dissociation (ISCID) offset voltages on the number of detected modified metabolite features in pooled human urine (a) in positive ion mode (voltage ramped from 5 to 55 V; step width: 5 V) and (b) in negative ion mode (voltage ramped from −5 V to −55 V; step width: 5 V).

Next, we studied the reproducibility of the results by analyzing the pooled urine in the ESI+ and ESI− modes in triplicate. Each replicate was processed separately, i.e., separate sample pretreatment, LC-MS analysis, and data processing. These three independent analyses of pooled urine resulted in the detection of 1075 ± 36 putative modified metabolite features (Figure S1). Among them, 896 features were reproducibly detected in all replicates. Several other modified metabolite features in the complex pattern were detected once or twice. The primary reason for the lower reproducibility of some features was the low signal intensity of the parent ion or the production of a fragment ion that was close to background noise. These demonstrated that the application of a ramp of ISCID voltages was an effective and robust way to large scale profile acetylated, glucuronidated, sulfated, ribose conjugated, and glucosidated compounds simultaneously, regardless of the modification type and bond strength.

A prerequisite for the correct annotation of these modifications was the close setting of the window for shifts in retention time (Δ 0.1 min) and mass (Δ 0.002 Da) for the NL matches. To evaluate the accuracy of ISCID-based-LC-HRMS, the modified metabolites in pooled human urine were subjected to enzymatic hydrolysis. Following enzymatic sample pretreatment, a full-scan MS analysis was conducted and compared to the global nontargeted feature profile of the controls. The controls were identically pretreated, with the exception that no enzymes were added. Two enzymes, β-glucuronidase and sulfatase, were selected for these experiments, which allowed us to study the changes in distinct masses of metabolites identified as either glucuronidated or sulfated by our novel approach. Significant decrease ($p < 0.05$) of putative modified metabolites in enzyme treated sample indicates correct identification of the metabolite modification. Compared to the nonenzymatically treated urine, 96% of the putative glucuronidated metabolites in ESI+ mode and 98% in ESI− mode exhibited a significant decrease in the enzymatically treated samples (Figure S2). For putative sulfated metabolite features identified by ISCID-based LC-HRMS, the coverage of the enzyme-reduced ion signals was 93% (Figure S2). The reduction caused by sulfatase treatment was not as pronounced as that observed for the β-glucuronidase-treated samples, owing to the lower activity of sulfatase under the applied conditions. These proof-of-principle experiments showed an accuracy of 93−98% using ISCID-based LC-HRMS for the detection of glucuronidated and sulfated metabolites, even though part of the glucuronidated and sulfated metabolites might not have responded to β-glucuronidase and sulfatase treatment. Furthermore, that over 30 putative modified metabolite features in pooled human urine were exemplarily confirmed by a comparative analysis with standard compounds demonstrated correct annotation of metabolite modification by the novel method (Table S1).

dx.doi.org/10.1021/ac502045j | Anal. Chem. 2014, 86, 9146−9153

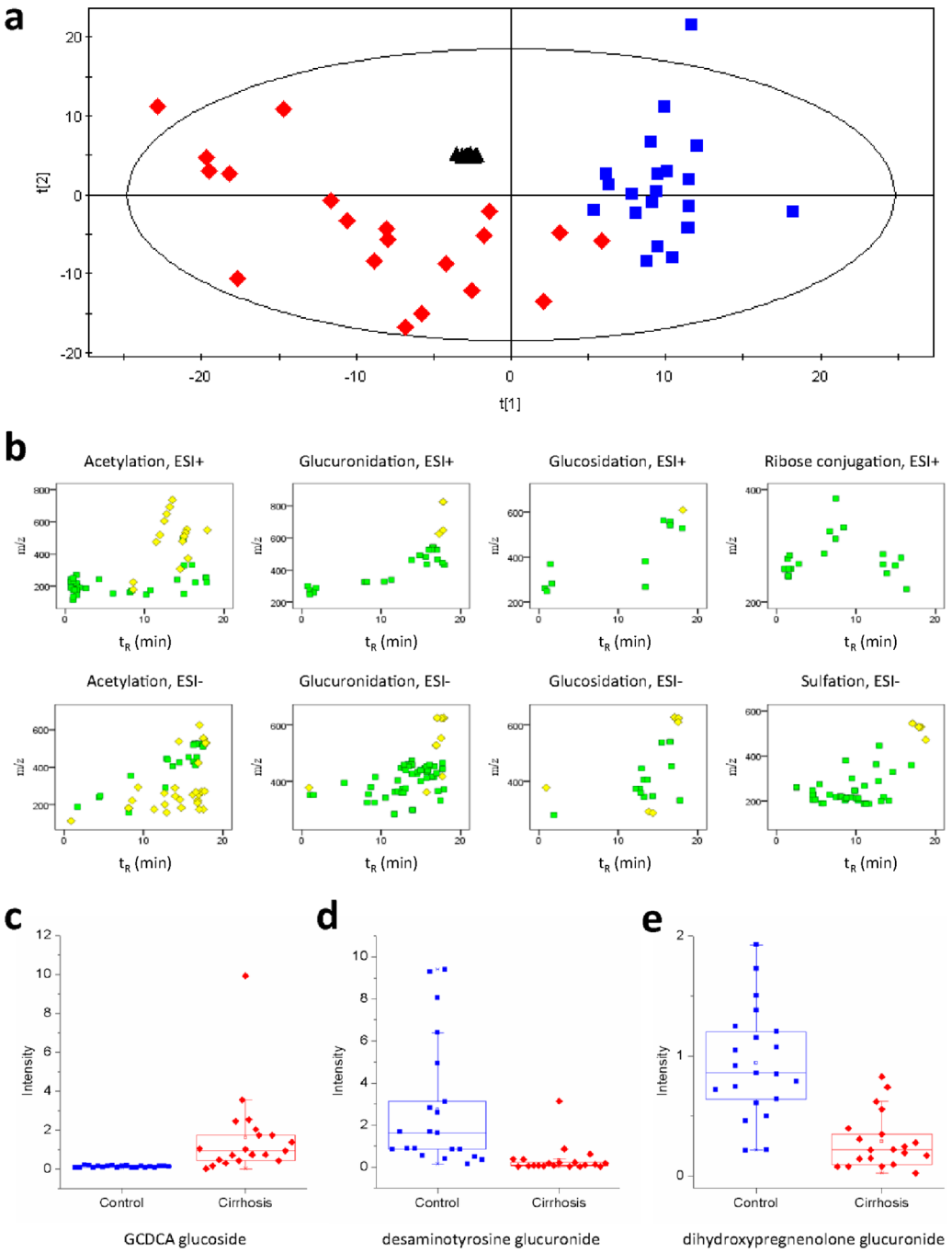


Figure 4. Urinary modification-specific metabolomics analysis of samples from patients with liver cirrhosis and healthy individuals. (a) The PCA score plot shows distinct differences between healthy subjects (marked with blue squares) and liver cirrhosis patients (marked with red diamonds). Quality control (QC) samples were shown as black triangles. Features detected in ESI+ and ESI− modes are included in the evaluation. (b) Overview of altered metabolite modifications in liver cirrhosis patients. Yellow diamonds represent significantly increased modified metabolite features ($p < 0.01$), and green squares represent significantly decreased modified metabolite features ($p < 0.01$). Examples of box plots showing differences in the signal intensities of (c) glycochenodeoxycholic acid (GCDCA) glucoside, (d) desaminotyrosine glucuronide, and (e) dihydroxypregnenolone glucuronide in healthy subjects (marked with blue squares) and liver cirrhosis patients (marked with red diamonds).

One major bottleneck in current nontargeted metabolomics studies is the considerable number of signals that remain unidentified.[16] Since the addition of the metabolite modification information to the mass feature, database searches for the identification of unknown compounds were facilitated by the NT-MSM approach in two ways. First, the addition of the detected modifying group to the mass feature greatly reduced the number of database-suggested candidates when searching metabolomics databases. For example, each metabolite feature in Table S1 resulted in an average hit number of 32.1 (HMDB) and 49.3 (Metlin) metabolite identities within a mass error of 10 ppm (Table S2). Specifying the modification type as supplementary information greatly reduced the average number of putative metabolite identities to 2.1 (HMDB) and 3.2 (Metlin) (Table S2), which could greatly reduce both the difficulty and the time cost for metabolite identification. More importantly, modification-specific metabolomics combined with the recently constructed database MyCompoundID (http://

dx.doi.org/10.1021/ac502045j | Anal. Chem. 2014, 86, 9146−9153

www.mycompoundid.org), which is based on the in silico biotransformation of 76 metabolic reactions for 8,021 metabolites entries in HMDB,[26] is a promising approach to support the identification of compounds not included in traditional metabolomics databases (HMDB, Metlin, and KEGG, among others), which are currently referred to as "unknowns" in metabolomic studies. For instance, an ion feature (m/z = 166.0183, ESI−; Table S1) was assigned as an acetylated metabolite, but the m/z value did not match any acetylated metabolite in the HMDB, Metlin, or KEGG databases owing to the underrepresentation of modified metabolites. However, specifying an acetyl group caused the MyCompoundID database to assign this molecule as acetyltaurine, which was confirmed following the synthesis of an acetyltaurine standard (Figure S3).

Application of NT-MSM for the Investigation of Urine Samples Obtained from Patients with Liver Cirrhosis. As a proof of concept, the second morning urine of 21 healthy male subjects and 20 male patients with liver cirrhosis matched by age (for clinical characteristics, see Table S3) was investigated, since the liver is a central organ of metabolism in mammals, particularly for the biotransformation of endogenous metabolites and xenobiotics. A total of 910 metabolite features with acetylation, glucuronidation, sulfation, ribose conjugation, or glucosidation were detected in the urine of healthy subjects and patients with liver cirrhosis (263 in ESI+ and 647 in ESI− mode). These modified metabolites represented 11.9% of the total ion features detected in the global metabolomic analysis of the urine. A clear clustering of the two groups reflecting distinct differences was detected by applying principal component analysis (PCA) to study the fingerprints of urinary modified metabolites of patients with liver cirrhosis compared with healthy control subjects (Figure 4a). Figure 4b shows an overview of altered metabolite modifications caused by cirrhosis. Modifications of glucuronidation, glucosidation, ribose conjugation, and sulfation were severely down-regulated. The most pronounced differences in the modified metabolites between the two groups (selection criteria: >3-fold difference in the metabolite signal compared with healthy controls and $p < 0.01$) are given in Table S4. Only 30% of these 89 most substantially different urinary modified compounds could be identified by HMDB and Metlin database searches, even though the modifications were known (Table S4). Another 34% of these compounds could be interpreted by combining modifying groups information with the MyCompoundID database searches. Based on the considerable number of modified metabolites in the metabolome (in the present study, 11.9% of the human urinary metabolome was acetylated, glucuronidated, sulfated, ribose conjugated, or glucosidated), future projects will probably reveal a largely unexplored facet of metabolomes toward the comprehensive coverage.

Among the 89 most substantially different urinary modified compounds, 26 modified metabolites exhibited higher signal intensities (Table S4), dominated, as expected, by bile acids.[29,30] Metabolites from steroid, tryptophan, and tyrosine metabolism dominated the pattern of the 63 signals that were lower in the urine of liver cirrhosis patients. Parts c−e of Figure 4 show three examples of promising novel urine markers of liver cirrhosis, with the area under the receiver operating characteristic (ROC) curve of 0.933, 0.912, and 0.907, namely, glycochenodeoxycholic acid glucoside, desaminotyrosine glucuronide, and dihydroxypregnenolone glucuronide, respectively. These modified metabolites were not included in the HMDB, Metlin, or KEGG databases but were structurally interpreted by the modification-specific metabolomics approach in combination with the MyCompoundID database searches and confirmation by MS^3 (Figure S4). The results of this proof of principle study demonstrated the potential to gain further insights into complex (patho)biochemical processes, possibly leading to the discovery of novel therapeutic targets. This study also suggests the potential of this technique to identify new distinct modified metabolites, which could lead to the development of novel distinct biomarkers or patterns as potential diagnostic tools (Figure 4c−e).

CONCLUSIONS

Modifications of metabolites are common and important for a multitude of biological functions in all types of organisms. In this study, we report the development and application of a novel strategy for the reproducible, large-scale, nontargeted discovery and annotation of acetylated, glucuronidated, sulfated, ribose conjugated, and glucosidated metabolites in complex samples. Our study of human urine demonstrated that NT-MSM is complementary to conventional nontargeted metabolomic analysis. In general, this strategy was capable of detecting unexpected metabolites and offers a new perspective of endogenous metabolilte modification for systems biology projects investigating (patho)physiological mechanisms of cells, plants, organs, microbiomes, and higher living organisms. In addition, it facilitated the identification of features, reducing the number of "unknowns" in current global metabolomics studies. NT-MSM is a tool that will facilitate the comprehensive coverage of the metabolome. Further work is still needed for the large-scale profiling of other types of metabolite modification, such as methylation, phosphorylation, and many others, for a more comprehensive coverage of modification-specific metabolites.

ASSOCIATED CONTENT

Supporting Information

Assessment of the reproducibility of NT-MSM of pooled human urine. Investigation of differences in the signal intensities of modified metabolite features in pooled human urine samples with and without enzymatic pretreatment to verify the accuracy of ISCID-based-HRMS. Confirmation of *N*-acetyltaurine in the urine sample using a synthesized standard. Representative MS^3 identification of three urinary modified metabolites in patients with liver cirrhosis compared with healthy controls. Representative modified metabolite features detected in the pooled urine of healthy individuals and identified by modification-specific metabolomics, database searches, and comparative analysis with standard compounds. Comparison of the number of HMDB and Metlin database hits for the modified metabolite features in human urine given in Table S1 with/without specifying the modifying group. Clinical characteristics of the age-matched 21 healthy controls and 20 liver cirrhosis patients. Urinary modified endogenous metabolites and xenobiotics in liver cirrhosis patients displaying >3-fold differences compared with healthy controls and a p-value < 0.01. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*Tel: +49 7071 29 83193. Fax: +49 7071 29 5348. E-mail: rainer.lehmann@med.uni-tuebingen.de.

*Tel./Fax: +86-411-84379530. E-mail: xugw@dicp.ac.cn.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study was supported in part by the State Key Science & Technology Project for Infectious Diseases (2012ZX10002011, 2012ZX10002009); the foundation (No. 21175132) and the creative research group project (No. 21321064) from National Natural Science Foundation of China; the Sino-German Center for Research Promotion (GZ 753 by NSFC and DFG); and the German Federal Ministry of Education and Research (BMBF) to the German Centre for Diabetes Research (DZD e.V.; Grant 01GI0925).

■ REFERENCES

(1) Kouzarides, T. *Cell* **2007**, *128*, 693−705.
(2) Zhao, S.; Xu, W.; Jiang, W.; Yu, W.; Lin, Y.; Zhang, T.; Yao, J.; Zhou, L.; Zeng, Y.; Li, H.; Li, Y.; Shi, J.; An, W.; Hancock, S. M.; He, F.; Qin, L.; Chin, J.; Yang, P.; Chen, X.; Lei, Q.; Xiong, Y.; Guan, K. L. *Science* **2010**, *327*, 1000−1004.
(3) Wang, Q.; Zhang, Y.; Yang, C.; Xiong, H.; Lin, Y.; Yao, J.; Li, H.; Xie, L.; Zhao, W.; Yao, Y.; Ning, Z. B.; Zeng, R.; Xiong, Y.; Guan, K. L.; Zhao, S.; Zhao, G. P. *Science* **2010**, *327*, 1004−1007.
(4) Zhao, Y.; Jensen, O. N. *Proteomics* **2009**, *9*, 4632−4641.
(5) Nørregaard Jensen, O. *Curr. Opin. Chem. Biol.* **2004**, *8*, 33−41.
(6) Swaney, D. L.; Beltrao, P.; Starita, L.; Guo, A.; Rush, J.; Fields, S.; Krogan, N. J.; Villén, J. *Nat. Methods* **2013**, *10*, 676−682.
(7) Mertins, P.; Qiao, J. W.; Patel, J.; Udeshi, N. D.; Clauser, K. R.; Mani, D. R.; Burgess, M. W.; Gillette, M. A.; Jaffe, J. D.; Carr, S. A. *Nat. Methods* **2013**, *10*, 634−637.
(8) Jennen, D. G. J.; Gaj, S.; Giesbertz, P. J.; van Delft, J. H. M.; Evelo, C. T.; Kleinjans, J. C. S. *Drug Discovery Today* **2010**, *15*, 851−858.
(9) Hoffmann, M. F.; Preissner, S. C.; Nickel, J.; Dunkel, M.; Preissner, R.; Preissner, S. *Nucleic Acids Res.* **2013**, *42*, D1113−D1117.
(10) Hofmann, U.; Heinkele, G.; Angelberger, S.; Schaeffeler, E.; Lichtenberger, C.; Jaeger, S.; Reinisch, W.; Schwab, M. *Anal. Chem.* **2012**, *84*, 1294−1301.
(11) Yang, J.; Xu, G.; Zheng, Y.; Kong, H.; Pang, T.; Lv, S.; Yang, Q. *J. Chromatogr., B* **2004**, *813*, 59−65.
(12) Zhao, R.; Xu, G.; Yue, B.; Liebich, H.; Zhang, Y. *J. Chromatogr., A* **1998**, *828*, 489−496.
(13) Herrero, M.; Simó, C.; García-Cañas, V.; Ibáñez, E.; Cifuentes, A. *Mass Spectrom. Rev.* **2012**, *31*, 49−69.
(14) Ni, Q.; Reid, K. R.; Burant, C. F.; Kennedy, R. T. *Anal. Chem.* **2008**, *80*, 3539−3546.
(15) Nicholson, J. K.; Holmes, E.; Kinross, J. M.; Darzi, A. W.; Takats, Z.; Lindon, J. C. *Nature* **2012**, *491*, 384−392.
(16) Baker, M. *Nat. Methods* **2011**, *8*, 117−121.
(17) Chen, J.; Zhao, X.; Fritsche, J.; Yin, P.; Schmitt-Kopplin, P.; Wang, W.; Lu, X.; Häring, H. U.; Schleicher, E. D.; Lehmann, R. *Anal. Chem.* **2008**, *80*, 1280−1289.
(18) Leichtle, A. B.; Dufour, J.-F.; Fiedler, G. M. *Swiss Med. Wkly* **2013**, *143*, w13801.
(19) Dai, W.; Huang, Q.; Yin, P.; Li, J.; Zhou, J.; Kong, H.; Zhao, C.; Lu, X.; Xu, G. *Anal. Chem.* **2012**, *84*, 10245−10251.
(20) van Wietmarschen, H. A.; Dai, W.; van der Kooij, A. J.; Reijmers, T. H.; Schroën, Y.; Wang, M.; Xu, Z.; Wang, X.; Kong, H.; Xu, G.; Hankemeier, T.; Meulman, J. J.; van der Greef, J. *PLoS One* **2012**, *7*, e44331.
(21) Kind, T.; Tolstikov, V.; Fiehn, O.; Weiss, R. H. *Anal. Biochem.* **2007**, *363*, 185−195.
(22) Xu, X.; Veenstra, T. D.; Fox, S. D.; Roman, J. M.; Issaq, H. J.; Falk, R.; Saavedra, J. E.; Keefer, L. K.; Ziegler, R. G. *Anal. Chem.* **2005**, *77*, 6646−6654.
(23) Huang, J.; Sun, J.; Chen, Y.; Song, Y.; Dong, L.; Zhan, Q.; Zhang, R.; Abliz, Z. *Anal. Chim. Acta* **2012**, *711*, 60−68.
(24) Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S.; Mandal, R.; Sinelnikov, I.; Xia, J.; Jia, L.; Cruz, J. A.; Lim, E.; Sobsey, C. A.; Shrivastava, S.; Huang, P.; Liu, P.; Fang, L.; Peng, J.; Fradette, R.; Cheng, D.; Tzur, D.; Clements, M.; Lewis, A.; De Souza, A.; Zuniga, A.; Dawe, M.; Xiong, Y.; Clive, D.; Greiner, R.; Nazyrova, A.; Shaykhutdinov, R.; Li, L.; Vogel, H. J.; Forsythe, I. *Nucleic Acids Res.* **2009**, *37*, D603−D610.
(25) Tautenhahn, R.; Cho, K.; Uritboonthai, W.; Zhu, Z. J.; Patti, G. J.; Siuzdak, G. *Nat. Biotechnol.* **2012**, *30*, 826−828.
(26) Li, L.; Li, R.; Zhou, J.; Zuniga, A.; Stanislaus, A. E.; Wu, Y.; Huan, T.; Zheng, J.; Shi, Y.; Wishart, D. S.; Lin, G. *Anal. Chem.* **2013**, *85*, 3401−3408.
(27) Dai, W. D.; Wei, C.; Kong, H. W.; Jia, Z. H.; Han, J. K.; Zhang, F. X.; Wu, Z. M.; Gu, Y.; Chen, S. L.; Gu, Q.; Lu, X.; Wu, Y. L.; Xu, G. W. *J. Pharmaceut. Biomed.* **2011**, *56*, 86−92.
(28) Bijlsma, S.; Bobeldijk, I.; Verheij, E. R.; Ramaker, R.; Kochhar, S.; Macdonald, I. A.; van Ommen, B.; Smilde, A. K. *Anal. Chem.* **2006**, *78*, 567−574.
(29) Danielsson, H.; Sjovall, J. *Annu. Rev. Biochem.* **1975**, *44*, 233−253.
(30) Batta, A. K.; Arora, R.; Salen, G.; Tint, G. S.; Eskreis, D.; Katz, S. *J. Lipid Res.* **1989**, *30*, 1953−1962.