Practical Approach for the Identification and Isomer Elucidation of Biomarkers Detected in a Metabonomic Study for the Discovery of Individuals at Risk for Diabetes by Integrating...

ARTICLE *in* ANALYTICAL CHEMISTRY · MARCH 2008

Impact Factor: 5.64 · DOI: 10.1021/ac702089h · Source: PubMed

11 AUTHORS, INCLUDING:

Ph. Schmitt-Kopplin
Helmholtz Zentrum München
348 PUBLICATIONS **6,616** CITATIONS
SEE PROFILE

Ning Wu
University of Science and Technology of China
152 PUBLICATIONS **2,819** CITATIONS
SEE PROFILE

Rainer Lehmann
University Hospital of Tuebingen
109 PUBLICATIONS **2,814** CITATIONS
SEE PROFILE

Guowang Xu
Dalian Institute of Chemical Physics, Chinese...
378 PUBLICATIONS **6,335** CITATIONS
SEE PROFILE

Anal. Chem. **2008,** *80,* 1280−1289

Practical Approach for the Identification and Isomer Elucidation of Biomarkers Detected in a Metabonomic Study for the Discovery of Individuals at Risk for Diabetes by Integrating the Chromatographic and Mass Spectrometric Information

Jing Chen,[†] Xinjie Zhao,[†] Jens Fritsche,[‡] Peiyuan Yin,[†] Philippe Schmitt-Kopplin,[§] Wenzhao Wang,[†] Xin Lu,[†] Hans Ulrich Häring,[∥] Erwin D. Schleicher,[‡] Rainer Lehmann,*[,‡] and Guowang Xu*[,†]

National Chromatographic R&A Center, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, 116023 Dalian, China, Division of Clinical Chemistry and Pathobiochemistry, Central Laboratory, University Hospital Tuebingen, D-72076 Tuebingen, Germany, GSF-National Research Center for Environment and Health, Institute for Ecological Chemistry, Ingoldstädter Landstrasse 1, D-85764 Neuherberg, Germany, and Department of Internal Medicine 4, University Hospital Tubingen, D-72076 Tubingen, Germany

Sensitive and high-resolution chromatographic-driven metabonomomics studies experienced major growth with the aid of new analytical technologies and bioinformatics software packages. Hence, data collections by LC−MS and data analyses by multivariate statistical methods are by far the most straightforward steps, and the detection of biomarker candidates can easily be achieved. However, the unequivocal identification of the detected metabolite candidates, including isomer elucidation, is still a crux of current metabonomics studies. Here we present a comprehensive analytical strategy for the elucidation of the molecular structure of metabolite biomarkers detected in a metabonomics study, exemplified analyzing spot urine of a cohort of healthy, insulin sensitive subjects and clinically well characterized prediabetic, insulin resistant individuals. An integrated approach of LC−MS fingerprinting, multivariate statistic analysis, LC−MS^n experiments, micro preparation, FTICR-MS, GC retention index, database search, and generation of an isotope labeled standard was applied. Overall, we could demonstrate the efficiency of our analytical approach by the unambiguous elucidation of the molecular structure of an isomeric biomarker candidate detected in a complex human biofluid. The proposed strategy is a powerful new analytical tool, which will allow the definite identification of physiologically important molecules in metabonomics studies from basic biochemistry to clinical biomarker discovery.

Metabonomics approaches, i.e., top down, nonhypothesis-driven analysis, aim to compare the pattern of thousands of endogenous metabolites under defined temporal conditions as comprehensive as possible. Various goals may be achieved by metabonomics, such as clustering different classes of individuals, identifying biomarkers, understanding biochemical mechanisms, investigating drug metabolism and toxicity, analyzing concentration and fluxes of metabolites, or probing molecular dynamics and pathway interactions.[1−3] A widely used analytical technique in metabonomic projects so far is nuclear magnetic resonance spectroscopy (NMR).[3] Because of the limitations of NMR in terms of sensitivity, mass spectrometry (MS), most commonly in combination with liquid chromatography (LC−MS),[4] and to a lesser extent with gas chromatography (GC/MS)[5] or capillary electrophoresis (CE−MS),[6] is now more and more applied as an alternative analytical tool. The increase in sensitivity results in an amplification of detected metabolites and in the generation of very complex two-dimensional maps (m/z values vs retention time) of thousands of metabolite ions. Following that, multivariate statistical methods are used to achieve separation of different classes and to determine metabolite ions (represented by their m/z values) that exerted the greatest influence on the clustering behavior of certain classes (i.e., biomarker candidates). In the past few years, considerable progress in LC−MS instrumentation and techniques[7−9]

* Corresponding authors. Prof. Dr. Guowang Xu: fax, +86-411-84379559; e-mail, xugw@dicp.ac.cn. Priv. Doz. Dr. R. Lehmann: fax, +49-7071-5348; e-mail, Rainer.Lehmann@med.uni-tuebingen.de.

[†] Dalian Institute of Chemical Physics, Chinese Academy of Sciences.

[‡] Central Laboratory, University Hospital Tuebingen.

[§] GSF-National Research Center for Environment and Health.

[∥] Department of Internal Medicine 4, University Hospital Tubingen.

(1) Lindon, J. C.; Holmes, E.; Nicholson, J. K. *FEBS J.* **2007**, *274*, 1140−1151.
(2) Nicholson, J. K. *Mol. Syst. Biol.* **2006**, *2*, article 52.
(3) Hollywood, K.; Brison, D. R.; Goodacre, R. *Proteomics* **2006**, *6*, 4716−4723.
(4) Want, E. J.; Nordstrom, A.; Morita, H.; Siuzdak, G. *J. Proteome Res.* **2007**, *6*, 459−468.
(5) Yuan, K. L.; Kong, H. W.; Guan, Y. F.; Yang, J.; Xu, G. X. *J. Chromatogr., B* **2007**, *850*, 236−240.
(6) Benavente, F.; van der Heijden, R.; Tjaden, U. R.; van der Greef, J.; Hankemeier, T. *Electrophoresis* **2006**, *27*, 4570−4584.
(7) Mauriala, T.; Chauret, N.; Oballa, R.; Nicoll-Griffith, D. A.; Bateman, K. P. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 1984−1992.
(8) Xia, Y. Q.; Miller, J. D.; Bakhtiar, R.; Franklin, R. B.; Liu, D. Q. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 1137−1145.

10.1021/ac702089h CCC: $40.75
Published on Web 01/15/2008

and in statistical data evaluation[10,11] has been achieved. Notably, the introduction and application of ultraperformance LC (UPLC) greatly improved the chromatographic resolution, sensitivity, and speed of LC-driven biomarker analysis.[12−15] Therefore, current data collection by LC−MS and data analysis by multivariate statistical methods are by far the most straightforward steps in metabonomics studies, and the detection of candidates can easily be achieved.[4] Once potential biomarkers have been selected, precise identification is required. Databases such as NIST, KEGG, METLIN, or Human Metabolome database (HMDB)[16] include data like exact mass, elemental composition, or MS/MS data, etc. for several known metabolites. These metabolite databases are a good starting point to evaluate mass spectrometric data, but their usefulness is limited by incompleteness and they are not suitable for elucidation of isomers. Of note, given that the majority of metabolites consist of two or more isomers, the unequivocal identification of metabolite candidates is still the major analytical challenge. Hence, the crux of all current LC−MS-driven metabonomics studies is to elucidate the identity of the detected biomarkers. Therefore, considerable efforts have gone into the development of new approaches to increase the number of identified biomarkers. This problem has been partially solved in recent reports applying (a) optimization of data evaluation,[11,17−19] e.g., by efficient sorting the principal discriminators from the background of thousands of metabolic features,[11] (b) reduction on a defined class of metabolites,[15,20−22] and (c) technical improvement of LC−MS applications.[9,14] Nevertheless, several signals putatively assigned as biomarker candidates could not be structurally confirmed by MS/MS fragmentation. For most of the metabonomics studies, some unidentified metabolite biomarkers are of no consequence, because the most frequent goal is to understand pathway interactions or investigate metabolic changes in general.[1−3] However, if a metabonomic study is performed to identify diagnostic relevant biomarkers, it is necessary to identify the bioinformatically detected metabolite ions with the greatest influence on the separation of health from disease, because these are the most important diagnostic biomarkers. Ultimately, to reach this goal the combination of many technologies will be required to identify unknown metabolites, including isomers, in biofluids.

Recent epidemiological data impressively demonstrated that the increasing incidence of diabetes mellitus is not only a disease of the "western world" but a serious and rapidly increasing worldwide problem.[23,24] Diabetes is a condition with a long "silent" or asymptomatic period where the patient is not aware of the disease but complications gradually develop. At the time of clinical diagnosis, as many as 20−30% have microvascular complications and diabetes is often diagnosed after retinal or renal problems are clinically overt.[25] Furthermore, diabetes, particularly type 2 diabetes, is frequently diagnosed in intensive care units in patients with acute myocardial infarction[26] indicating that also macrovascular complications develop clinically silent during the undiagnosed period. Important to note, insulin resistant (IR), i.e., a prediabetic state of impaired insulin sensitivity precedes the manifestation of overt type 2 diabetes for decades.[27] Since, early therapeutic and/or lifestyle interventions can retard and even prevent the development of type 2 diabetes,[28,29] the need of early diagnosis of this "prediabetic" state with an easy-to-perform screening assay from easy-to-obtain samples is obvious. However, up to now no laboratory blood test evaluated for the diagnosis of IR, like triglyceride to high-density lipoprotein−cholesterol ratio,[30] fasting plasma insulin levels,[31] or impaired fasting glucose,[32] allowed a reliable detection of IR individuals. At present, time-consuming and laborious provocation tests of the glucose metabolism by an oral or intravenous glucose tolerance test or an insulin clamp are the only diagnostic options.[33,34] Therefore, we set out to search for biomarkers in easy-to-obtain samples possibly suitable for screening tests, i.e., spot urines. Since it cannot be readily expected to find biomarkers in urine from apparently healthy individuals differing only in their future potential to develop diabetes the study groups have to be vigorously defined.

Here we describe the complete procedure of our metabonomic analysis. Starting out with urine samples from a well-characterized insulin resistant, prediabetic, and insulin sensitive cohort, a distinct clustering of the groups was achieved by applying ultraperformance LC−Q-TOF MS (UPLC−Q-TOF MS) for data collection

(9) Plumb, R. S.; Johnson, K. A.; Rainville, P.; Smith, B. W.; Wilson, I. D.; Castro-Perez, J. M.; Nicholson, J. K. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 1989−1994.
(10) Eriksson, L.; Antti, H.; Gottfries, J.; Holmes, E.; Johansson, E.; Lindgren, F.; Long, I.; Lundstedt, T.; Trygg, J.; Wold, S. *Anal. Bioanal. Chem.* **2004**, *380*, 419−429.
(11) Crockford, D. J.; Lindon, J. C.; Cloarec, O.; Plumb, R. S.; Bruce, S. J.; Zirah, S.; Rainville, P.; Stumpf, C. L.; Johnson, K.; Holmes, E.; Nicholson, J. K. *Anal. Chem.* **2006**, *78*, 4398−4408.
(12) Plumb, R.; Castro-Perez, J.; Granger, J.; Beattie, I.; Joncour, K.; Wright, A. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2331−2337.
(13) Wilson, I. D.; Nicholson, J. K.; Castro-Perez, J.; Granger, J. H.; Johnson, K. A.; Smith, B. W.; Plumb, R. S. *J. Proteome Res.* **2005**, *4*, 591−598.
(14) Plumb, R. S.; Rainville, P.; Smith, B. W.; Johnson, K. A.; Castro-Perez, J.; Wilson, I. D.; Nicholson, J. K. *Anal. Chem.* **2006**, *78*, 7278−7283.
(15) Rainville, P. D.; Stumpf, C. L.; Shockcor, J. P.; Plumb, R. S.; Nicholson, J. K. *J. Proteome Res.* **2007**, *6*, 552−558.
(16) Wishart, D. S. *Pharmacogenomics* **2007**, *8*, 683−686.
(17) Bijlsma, S.; Bobeldijk, L.; Verheij, E. R.; Ramaker, R.; Kochhar, S.; Macdonald, I. A.; van Ommen, B.; Smilde, A. K. *Anal. Chem.* **2006**, *78*, 567−574.
(18) Nordstrom, A.; O'Maille, G.; Qin, C.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 3289−3295.
(19) Ebbels, T. M. D.; Keun, H. C.; Beckonert, O. P.; Bollard, M. E.; Lindon, J. C.; Holmes, E.; Nicholson, J. K. *J. Proteome Res.* **2007**, *6*, 4407−4422.
(20) Wang, C.; Kong, H. W.; Guan, Y. F.; Yang, J.; Gu, J. R.; Yang, S. L.; Xu, G. W. *Anal. Chem.* **2005**, *77*, 4108−4116.
(21) Wagner, S.; Scholz, K.; Donegan, M.; Burton, L.; Wingate, J.; Volkel, W. *Anal. Chem.* **2006**, *78*, 1296−1305.
(22) Lutz, U.; Lutz, R. W.; Lutz, W. K. *Anal. Chem.* **2006**, *78*, 4564−4571.
(23) Zimmet, P.; Alberti, K. G. M. M.; Shaw, J. *Nature* **2001**, *414*, 782−787.
(24) Yoon, K. H.; Lee, J. H.; Kim, J. W.; Cho, J. H.; Choi, Y. H.; Ko, S. H.; Zimmet, P.; Son, H. Y. *Lancet* **2006**, *368*, 1681−1688.
(25) Rajala, U.; Laakso, M.; Qiao, Q.; Keinanen-Kiukaanniemi, S. *Diabetes Care* **1998**, *21*, 1664−1669.
(26) Maki, K. C.; Abraira, C.; Cooper, R. S. *J. Cardiopulm. Rehabil.* **1995**, *15*, 97−102.
(27) DeFronzo, R. A. *Med. Clin. North Am.* **2004**, *88*, 787−835.
(28) Knowler, W. C.; Barrett-Connor, E.; Fowler, S. E.; Hamman, R. F.; Lachin, J. M.; Walker, E. A.; Nathan, D. M. *N. Engl. J. Med.* **2002**, *346*, 393−403.
(29) Tuomilehto, J.; Lindstrom, J.; Eriksson, J. G.; Valle, T. T.; Hamalainen, H.; Ilanne-Parikka, P.; Keinanen-Kiukaanniemi, S.; Laakso, M.; Louheranta, A.; Rastas, M.; Salminen, V.; Uusitupa, M.; Aunola, S.; Cepaitis, Z.; Moltchanov, V.; Hakumaki, M.; Mannelin, M.; Martikkala, V.; Sundvall, J. *N. Engl. J. Med.* **2001**, *344*, 1343−1350.
(30) McLaughlin, T.; Reaven, G.; Abbasi, F.; Lamendola, C.; Saad, M.; Waters, D.; Simon, J.; Krauss, R. M. *Am. J. Cardiol.* **2005**, *96*, 399−404.
(31) McLaughlin, T.; Abbasi, F.; Cheal, K.; Chu, J.; Lamendola, C.; Reaven, G. *Ann. Intern. Med.* **2003**, *139*, 802−809.
(32) Ford, E. S.; Abbasi, F.; Reaven, G. M. *Atherosclerosis* **2005**, *181*, 143−148.
(33) Stumvoll, M.; Wahl, H. G.; Loblein, K.; Becker, R.; Volk, A.; Renn, W.; Jacob, S.; Haring, H. *Horm. Metab. Res.* **2001**, *33*, 89−95.
(34) Bartoli, E.; Castello, L.; Sainaghi, P. P.; Schianca, G. P. C. *Clin. Lab.* **2005**, *51*, 613−624.

and the most common multivariate statistical methods principal components analysis (PCA) and partial least-squares-discriminant analysis (PLS-DA) for data evaluation. The strategy for the identification of unknown biomarkers is exemplarily shown by the analysis of a urinary metabolite ion. This biomarker candidate showed a great outlier position in the PLS-DA loading plot, which indicates a major influence on the clustering, but the initial identification based on UPLC Q-TOF MS and MS/MS analysis and database search failed. We could successfully adopt a combination of micropreparative UPLC, FTICR-MS, database search, re-evaluation of UPLC Q-TOF MS/MS spectra, GC/MS and GC retention indices, and stable isotope labeled metabolite synthesis for the identification and validation of this isomeric metabolite.

EXPERIMENTAL SECTION

Chemicals. Acetonitrile (HPLC grade) was purchased from Fisher (Fairlawn, NJ), formic acid (HPLC/SPECTRO grade) was purchased from Tedia. Distilled water was filtered through a Milli-Q system (Millipore, MA). Leucine-enkephalin, 2-hydroxyhippuric acid, and BSTFA were from Sigma-Aldrich (St. Louis, MO). ^{15}N-3-hydroxyhippuric acid was synthesized in commission by SOHENA (Tuebingen, Germany).

Human Sample Collection and Oral Glucose Tolerance Test. Spot urine samples were collected from 51 individuals after 12 h of overnight fasting under standardized conditions in the metabolic ward at the University Clinic in Tuebingen (Germany). The samples were stored immediately at −80 °C. The protocol of the study was approved by the Ethics Committee of the University Tuebingen conformed to the 1975 Declaration of Helsinki before commencement, and all subjects gave written informed consent. The investigation was conducted in accordance with the ethical principles of Good Clinical Practice. All individuals underwent a 75 g oral glucose tolerance test (oGTT) according to the WHO to determine the insulin sensitivity index (ISI). Venous blood samples were obtained at 0, 30, 60, 90, and 120 min for determination of plasma glucose and insulin. Insulin sensitivity was calculated from glucose and insulin values during the oGTT as proposed by Matsuda and DeFronzo[35] using the formula

$$\mathrm{ISI} = \frac{10\,000}{\sqrt{(\mathrm{FGP})\,(\mathrm{FPI})\,(\bar{x}\ \mathrm{oGTT\ glu})\,(\bar{x}\ \mathrm{oGTT\ ins})}} \qquad (1)$$

where FGP = fasting plasma glucose; FPI = fasting plasma insulin concentration; $\bar{x}$ oGTT glu = average of glucose concentration during the oGTT; $\bar{x}$ oGTT ins = average insulin concentration during the oGTT. Low levels of the ISI (<8.5) indicate that the body is more resistant to insulin action. On the basis of the ISI results, 23 subjects of the cohort were found to be insulin resistant (IR) and 28 individuals were detected to be insulin sensitive (IS).

Analysis of Glucose and Insulin. Blood was collected in kaolin and sodium fluoride-vacuum containers; and serum, respectively, plasma was separated by low-speed centrifugation at 2000*g* for 7 min. Glucose was measured with the ADVIA 1650 clinical chemical analyzer ,and insulin was analyzed with the ADVIA Centaur immunoassay system according to the instructions of the manufacturers. Both analyzers were from Bayer HealthCare (Fernwald, Germany).

Sample Preparation for the Metabonomic Analysis. Urine samples were thawed, deproteinized with 2 volume parts of acetonitrile (final concentration 66%), centrifuged (13 000*g* for 20 min), divided in several aliquots, run to dryness in a vacuum centrifuge, and stored at −20 °C. For the UPLC−MS analysis, the samples were reconstituted in 200 μL of acetonitrile and water (2 + 8).

UPLC−MS^n Analysis. UPLC chromatographic separation was performed on a 2.1 mm × 100 mm ACQUITY 1.7 μm C18 column (Waters, Milford, USA) using an ACQUITY ultraperformance liquid chromatography system (Waters, Milford MA). The column was maintained at 35 °C. The gradient duration was 35 min at a flow rate of 0.35 mL/min. The gradient program was 98% A for 0.5 min, then changed to 70% B linearly within 30 min, then changed to 100% B and held for 3.5 min, then back to 98% A (A = 0.1% formic acid in water and B = acetonitrile). A 5 μL aliquot of each sample was injected onto the column.

Mass spectrometry was performed on a Waters Q-TOF micro (Waters MS Technologies, Manchester, UK) operating in ESI$^+$- and ESI$^-$-ion modes. The scan range was from 100 to 700 *m/z*. The nebulization gas was set to 500 L/h at a temperature of 300 °C, the cone gas was set to 50 L/h, and the source temperature was set to 120 °C; capillary voltage and cone voltage were set to 3100 and 35 V, respectively. The MCP detector voltage was set to 2600 V. The Q-TOF micro MS acquisition rate was set to 0.4 s with a 0.1 s interscan delay. In the MS/MS experiments, argon was employed as the collision gas and collision energy was set to 25 eV. All other parameters were the same as mentioned above.

All analyses were acquired using the lock spray to ensure accuracy and reproducibility. Leucine enkephalin was used as the lock mass in the positive ion mode ($[M + H]^+ = 556.2771$) and in the negative ion mode ($[M - H]^- = 554.2615$). The lock spray frequency was set at 20 s.

Micropreparation. Micropreparation was conducted on an UPLC system (Waters, Milford MA). A Waters Acquity TUV was employed as the detector (Waters, Milford MA); the wavelength was set to 254 nm. Injection volume was 10 μL, and the other chromatogram parameters were the same as described for UPLC−MS analysis. The peak at retention time 6.0−6.25 min was collected for further analysis.

Exact Mass Analysis. High-resolution mass spectra were acquired on a Bruker (Bremen, Germany) APEX Qe Fourier transform ion cyclotron resonance MS equipped with a 12 T superconducting magnet and an Apollo II ESI source in the negative ion mode. Spectra were externally calibrated on clusters of arginine (10 mg/L in methanol); calibration errors in the relevant mass range were always below 100 ppb.

Gas Chromatography/Mass Spectrometry. The collected fraction was silylated using *N*,*O*-bis-(trimethylsilyl)trifluoroacetamide (BSTFA) at 60 °C in a water bath for 30 min and analyzed on a 50 m × 0.25 mm × 0.25 μm DB-5 column by GC/MS with a QP-2010 instrument (Shimadzu, Japan). The temperature of the GC column was maintained at 200 °C, and the linear velocity was set to 39.6 cm/s. Injector temperature was 280 °C, and detector

(35) Matsuda, M.; DeFronzo, R. A. *Diabetes Care* **1999**, *22*, 1462−1470.

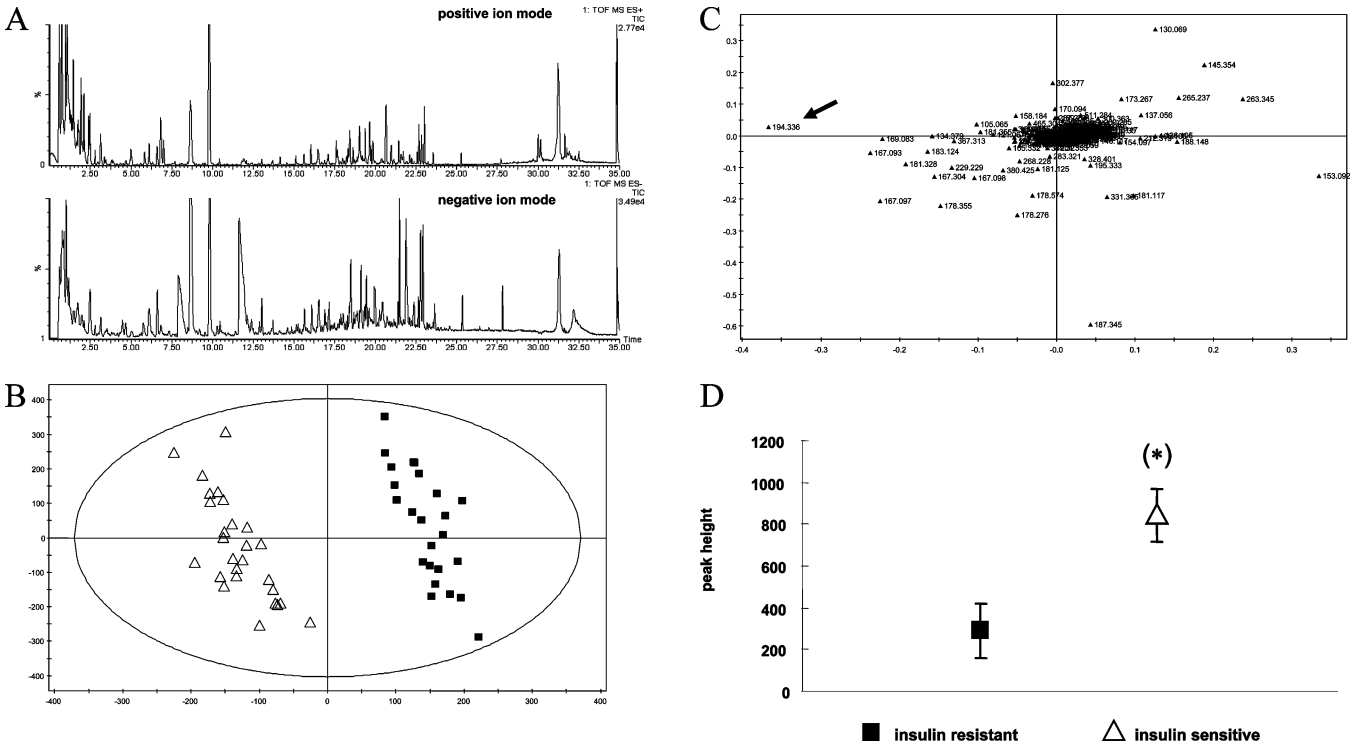

Figure 1. **(**A**)** Representative reversed phase UPLC−Q-TOF MS total ion chromatograms (TIC) of a urine sample analyzed in positive and negative ion modes; **(**B**)** PLS-DA scores plot (OSC filtered) of healthy, insulin sensitive subjects (△) and prediabetic, insulin resistant individuals (■); **(**C**)** corresponding PLS-DA loading plot. The variables are labeled with *m*/*z*, and *m*/*z* 194.34 is labeled by an arrow; and **(**D**)** differences in the TIC peak height of *m*/*z* 194 between insulin sensitive subjects (△) and insulin resistant individuals (■). ∗ $p < 0.05$ vs insulin resistant individuals analyzed by the Wilcoxon Rank Sum Test.

temperature was 200 °C. An amount of 1 μL of the derivatized sample was injected in split mode (split ratio 10:1).

Retention index analysis was performed on a GC 6890 (Agilent, USA) equipped with an OV-1701 column (30 m × 0.32 mm × 0.25 μm). The temperature of the GC column was programmed from 70 to 270 °C at a rate of 5 °C /min, and kept at 270 °C for 20 min. Linear velocity was set to 25 cm/s. Injector temperature was 270 °C, and the detector temperature was 300 °C. An amount of 1 μL of the derivatized sample was injected in the split mode (split ratio 30:1). Additionally, 1 μL volume of light diesel oil was analyzed under the same parameters for the retention index calculation.

Data Collection. Mass spectra were digitally analyzed using the Micromass MarkerLynx Applications Manager version 4.0 (Waters Ltd, USA). The statistical calculation was performed using the maximum intensity of the metabolite ion peaks. The data were combined into a single matrix by aligning peaks with the same mass-retention time pair together from each data file in the data set. The intensities for each peak were normalized to the sum of the peak intensity for each data set.

Data Analysis. The pre-processed UPLC−MS data were exported into Soft Independent Modeling of Class Analogy (SIMCA)-P (version 10.0, Umetrics AB, Umea, Sweden) for analysis and visualization by multivariate statistical methods. After Pareto scaling and OSC according to Wold et al.,[36] PLS-DA was applied. The PLS loading plot was used to identify metabolites with major influence on the group membership. To support the metabolite identification, the following databases have been used: KEGG (http://www.kegg.com/), PubChem compound database (http://ncbi.nim.nih.gov/), METLIN (http://metlin-.scripps.edu/), and the human metabolite data base (http://www.hmdb.ca/).

(36) Wold, S.; Antti, H.; Lindgren, F.; Ohman, J. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 175−185.

RESULTS AND DISCUSSION

Detection of Biomarker Candidates. In this study we performed the metabonomics analysis of the spot urine of 51 overnight fasted subjects using UPLC−Q-TOF MS in the positive and negative ESI mode. A typical total ion chromatogram (TIC) of spot urine for both ionization modes obtained from the same individual is shown in Figure 1A. The pre-processed UPLC−MS data were investigated further using principal components analysis (PCA) resulting in a separation of healthy, insulin sensitive subjects and prediabetic, insulin resistant individuals in the PCA scores plot of the first and second PCA components (data not shown). Following that, partial least-squares-discriminant analysis (PLS-DA) with the orthogonal signal correction (OSC) data filter using randomized sample order was applied. The OSC-PLS-DA scores plot for the two component model clearly divided the 28 IS- and the 23 IR-individuals in two distinct clusters (Figure 1B) with a goodness of fit of $R^2Y = 0.966$ and a goodness of prediction of $Q^2Y = 0.844$ based on 7-fold cross validation. Figure 1C shows the corresponding OSC-PLS-DA loading plot. The distance of an ion from the origin in this plot, its loading, represents the influence of that ion on the PLS components and therefore on the clustering in the OSC-PLS-DA scores plot shown in Figure 1B. The ion with *m*/*z* 194.34 (retention time 6.1 min) shows one of the greatest loadings in component 1 which is the PLS component mainly

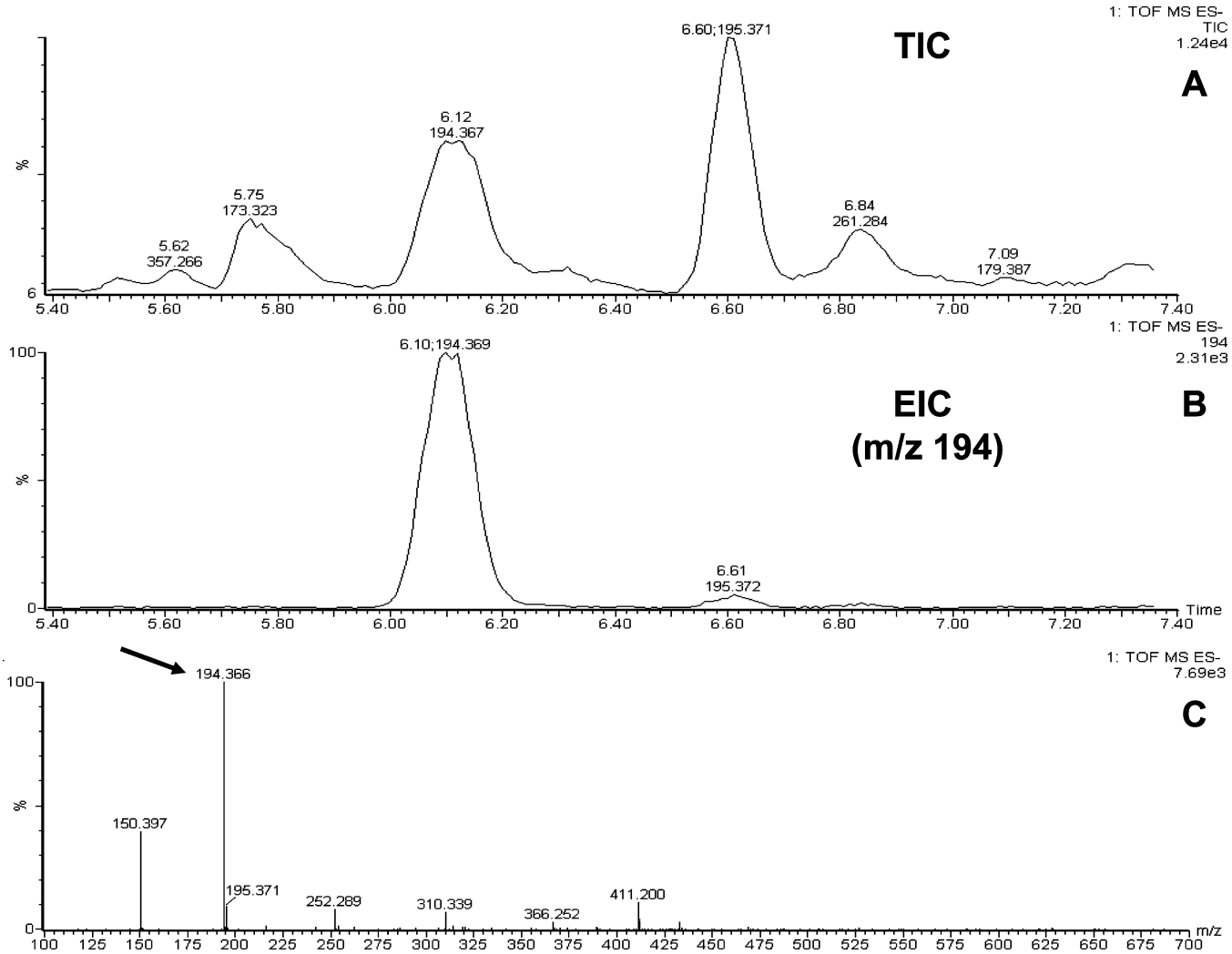

Figure 2. **(**A**)** Detail of the reversed phase UPLC−Q-TOF MS TIC containing *m*/*z* 194 scanned by ESI$^-$; **(**B**)** the extracted ion chromatogram (EIC) of the ion at *m*/*z* 194; and **(**C**)** corresponding mass spectrum of the TIC peak eluting at a retention time of 6.1 min.

responsible for the class separation. Therefore, it may be an interesting biomarker candidate to reflect metabolic traits between healthy and insulin resistant individuals. Additionally, to analyze the possible difference between IS and IR individuals, we performed a comparison of the peak height of m/z 194.34 in the spectra of IS and IR subjects and found a significant lower peak area in insulin resistant individuals ($p < 0.05$) according to the Wilcoxon Rank Sum Test (Figure 1D). This result clearly indicates that the elucidation of the identity of m/z 194.34 would be quite worthwhile. Therefore, we exemplarily show in the following sections step by step the analysis of the urinary metabolite ion with m/z 194.34 as a practical approach for the unambiguous identification of unknown biomarker candidates in metabonomics studies.

Elucidation of the Molecular Structure of the Biomarker Candidate. *Step 1: Detection of the Quasi-Molecular Ion.* First we excluded that m/z is a product or adduct ion. Parts A and B of Figure 2 represent a detail of the negative ion mode TIC and the corresponding extracted ion chromatogram (EIC) of m/z 194 of this section. The mass spectrum of the peak at 6.1 min demonstrates that several additional ions are present (m/z 150, 195, 252, 310, 366, 411) (Figure 2C). Ions from the same compound, i.e., having a tight relationship, are represented by correlation coefficients > 0.8.[37] No correlation between m/z 194 and m/z values 252, 310, 366, and 411 was detected. The linear correlation coefficient of m/z 195 vs m/z 194 was 0.9082 and of m/z 150 vs m/z 194 was 0.8954. The ion of m/z 195 could be an isotope peak of m/z 194 and m/z 150 a potential fragment. Inspecting the TIC obtained in the positive ion mode, we detected also at a retention time of 6.1 min the corresponding $[M + H]^+$ ion (m/z 196.12), but the average response intensity was 10 times smaller than the signal of m/z 194.34, the $[M - H]^-$ ion (data not shown). Hence, the molecular weight of this biomarker candidate was considered to be 195 Da.

Step 2: Fragmentation Pattern of the Biomarker Candidate. Structural elucidation of metabolites is routinely performed by the acquisition of additional MS data. Therefore in the second step we applied MS/MS experiments in the negative ion mode to get structural information via the interpretation of the fragmentation pattern of the biomarker candidate. The transition from the molecular anion m/z 194 to the specific fragments resulted in the formation of dominating diagnostic ions at m/z 93, m/z 150, and

(37) Clish, C. B.; Davidov, E.; Oresic, M.; Plasterer, T. N.; Lavine, G.; Londo, T.; Meys, M.; Snell, P.; Stochaj, W.; Adourian, A.; Zhang, X.; Morel, N.; Neumann, E.; Verheij, E.; Vogels, J. T. W. E.; Havekes, L. M.; Afeyan, N.; Regnier, F.; van der Greef, J.; Naylor, S. *OMICS* **2004**, *8*, 3−13.

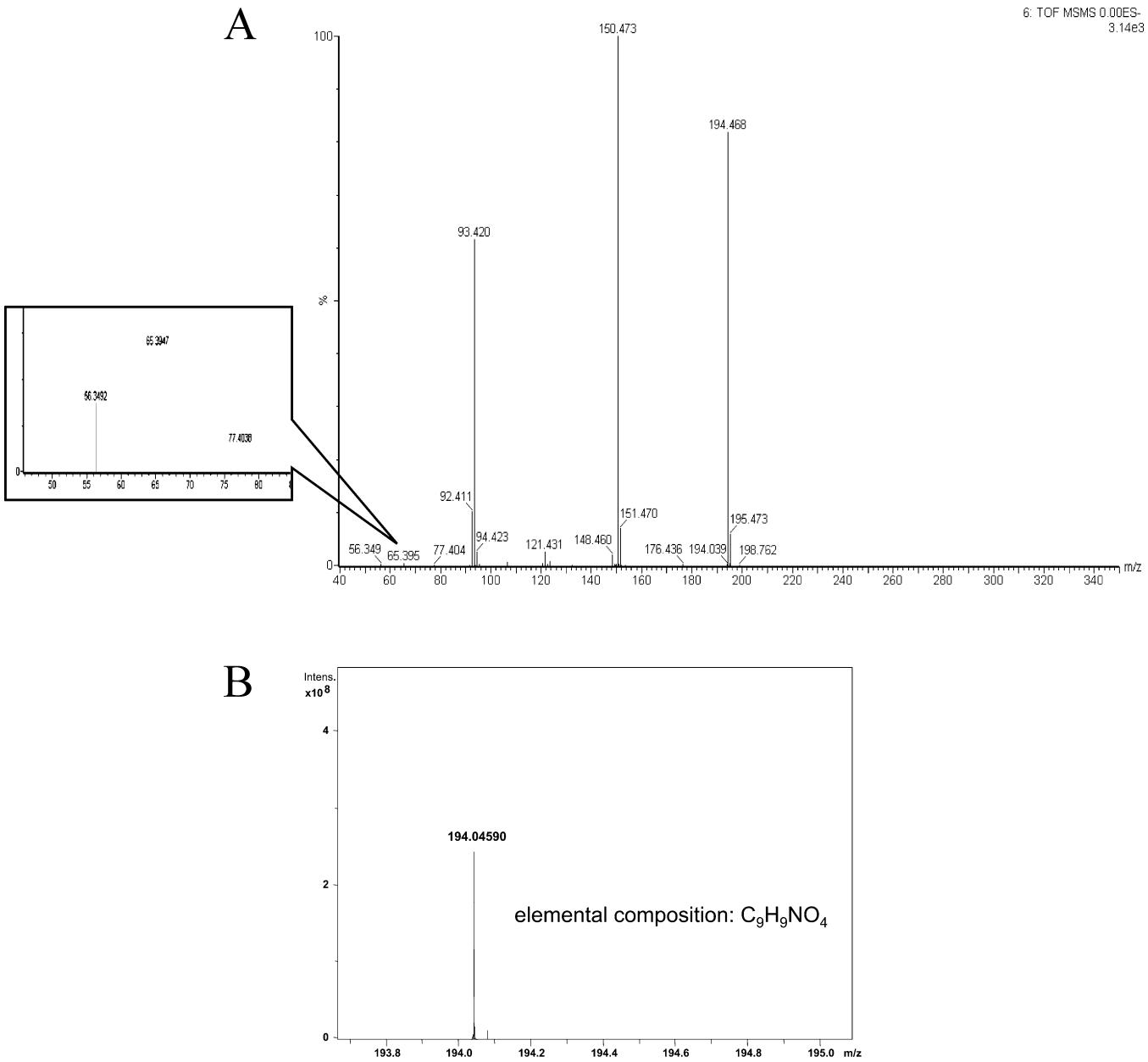

Figure 3. UPLC−Q-TOF MS/MS data. **(A)** Negative product ion spectrum of *m*/*z* 194, the potential biomarker to be identified; **(B)** exact mass determination of the metabolite ion *m*/*z* 194.04589 of the loading plot shown in Figure 1C measured with a 12 T electrospray ionization Fourier transform ion cyclotron resonance mass spectrometer in negative ion mode (attribution of the formula [$C_9H_9NO_4$] with an accuracy of 120 ppb).

several ions with very low intensities (Figure 3A). This result confirmed that the ion at *m*/*z* 150 found in the mass spectrum of the EIC in Figure 2C was a fragment of *m*/*z* 194. On trhe basis of the lack of comprehensive MS/MS databases it was still difficult to deduce the structure of the molecule. The mass difference of *m*/*z* 44 between *m*/*z* 194 and *m*/*z* 150 may be related to several candidate structures including a carboxyl, a propyl, an amide, an ethylamino, and an acetaldehyde group. The product ion at *m*/*z* 93 could be caused by the fragmentation of a phenoxy group or some heterocyclic structures. Although the molecular structure may be elucidated by MS/MS,[38−40] in our case the product ion data could not reveal the exact molecular structure of *m*/*z* 194.

Step 3: Micropreparative UPLC, Accurate Mass Measurement, and Determination of Elemental Composition. Detailed information about the elemental composition of a biomarker candidate is a valuable tool to proceed. This analysis represents an initial, however, crucial step in metabolite annotation, e.g., by use of various targeted databases (KEGG, METLIN, etc.). The recent advances in ultrahigh resolution mass spectrometry enable the assignment of the elementary compositions of metabolites (CHON-SP, 100 ppb mass accuracy) in a mass range from 150 to 800 Da.[41,42] Therefore our analyte of interest (*m*/*z* 194) was isolated

(38) Ma, S. G.; Chowdhury, S. K.; Alton, K. B. *Curr. Drug Metab.* **2006**, *7*, 503−523.
(39) Prakash, C.; Shaffer, C. L.; Nedderman, A. *Mass Spectrom. Rev.* **2007**, *26*, 340−369.
(40) Triolo, A.; Altamura, M.; Dimoulas, T.; Guidi, A.; Lecci, A.; Tramontana, M. *J. Mass Spectrom.* **2005**, *40*, 1572−1582.
(41) Aharoni, A.; Ric de Vos, C. H.; Verhoeven, H. A.; Maliepaard, C. A.; Kruppa, G.; Bino, R.; Goodenowe, D. B. *OMICS* **2002**, *6*, 217−234.
(42) Breitling, R.; Pitt, A. R.; Barrett, M. P. *Trends Biotechnol.* **2006**, *24*, 543−548.

by micropreparative UPLC and subjected to negative electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS). Figure 3B shows the accurate anion mass (m/z 194.04590; 120 ppb accuracy) resulting in an elementary composition of $C_9H_9NO_4$ of our biomarker candidate.

Step 4: Database Search, Evaluation of the Metabolite Candidate List, and Reinterpretation of LC−MS/MS Fragmentation Pattern. The measured exact mass and the elemental composition of m/z 194 were used in the next step for a comparative analysis by online database searching using KEGG, PubChem compound, METLIN, and HMDB database revealing 18 candidate compounds. To reduce this candidate list we inspected all compounds under chemico-physical and physiological aspects. In the negative ion mode a molecule with a nitryl group cannot be detected. Therefore, we removed molecules containing nitryl groups in the structure, thereby reducing the metabolite candidates down to the number of 10. In accordance with IUPAC nomenclature, the remaining biomarker candidates were 2-amino-3-(4,5-dioxocyclohex-1-enyl)propanoic acid; 5,6-dihydroxyindoline-2-carboxylic acid; 3-amino-4-methoxycarbonylbenzoic acid; 4-acetamido-2-hydeoxybenzoic acid; 5-acetamido-2-hydeoxybenzoic acid; 4-(amino(carboxyl)methyl)benzoic acid; 2-benzamido-2-hydroxyacetic acid; 2-(2-hydroxybenzamido)acetic acid; 2-(3-hydroxybenzamido)acetic acid; and 2-(4-hydroxybenzamido)acetic acid.

Next, we re-evaluated the product ion spectra shown in Figure 3A to study the structural information of m/z 194 in comparison to the remaining molecules of the candidate list. The product ions at m/z 77 and 65 represent most likely the cleavage of a phenyl group. The fragment anion at m/z 93 may be a diagnostic ion of a phenoxy group. In addition, the mass difference of 44 Da (m/z 194 to m/z 150) was rationalized by the possible loss of a carboxyl group in the structure. Following that, we excluded molecules from the list which do not contain a phenoxy group and a carboxyl group. Finally, the three isomers of 2-(hydroxybenzamido)acetic acid (= hydroxyhippuric acid) with the hydroxyl group at position 2 (ortho), 3 (meta), or 4 (para) of the benzyl-ring remained as the potential biomarker candidate. All isomers were described as metabolites detected in human urine.[43−46] Consequently, the next step was to elucidate the isomeric structure of hydroxyhippuric acid to finally identify our biomarker candidate.

Step 5: Isomer Elucidation and Confirmation of the Biomarker Candidate. The differentiation of distinct isomers to elucidate the molecular structure of our metabolite of interest is the next analytical challenge. Initially we chose the most pragmatic way and searched for standard substances to compare the retention time and spiked a urine sample with this standard. But unfortunately, only 2-hydroxyhippuric acid (= salicyluric acid) was commercially available. In the following reversed phase UPLC analysis, the detected retention time for salicyluric acid was 11.6 min (data not shown) and our biomarker candidate eluted at 6.1 min (Figure 2). Considering the molecular structure of the two remaining isomers, the retention time order should be 4-hydroxyhippuric acid < 3-hydroxyhippuric acid on a reversed phase (RP)-column. This consideration could be confirmed by a literature search where we found a representative chromatogram of a RP-HPLC-analysis of urine showing a retention time of 13.7 min for 4-hydroxyhippuric acid and of 15.3 min for 3-hydroxyhippuric acid.[47] Hence, we re-inspected our UPLC−MS raw data and detected two peaks in the EIC of m/z 194. Figure 4A shows a detail of the EIC including one peak at 5.1 min and the other peak at 6.1 min containing our biomarker candidate. Therefore, based on the retention time our candidate is most likely 3-hydroxyhippuric acid. This speculation was supported by the recently reported major diagnostic product ions for 3-hydroxyhippuric acid (m/z 150) and for 4-hydroxyhippuric acid (m/z 100),[46] which we also found in the product ion spectra of each peak as expected (Figure 4A). To confirm this result by a valid analytical procedure, i.e., to unambiguously identify the isomer of hydroxyhippuric acid we applied GC/MS and GC retention index. Compared to other structure analytical technologies such as IR, UV, and NMR, the retention index of GC/MS is the most sensitive technology, robust to impurities and furthermore the results from GC/MS are easy to interpret. A lyophilized fraction of the peak at 6.1 min was derivatized with BSTFA to obtain better volatility resulting in the detection of a molecular ion at m/z 339 (bis(trimethylsilyl) ester (di TMS)) in the GC/MS spectrum (data not shown). A comparison with NIST standard MS spectra confirmed that the target compound is hydroxyhippuric acid di TMS. The NIST score for 3-hydroxyhippuric acid di TMS and 4-hydroxyhippuric acid di TMS are 88 and 83, respectively. Because of this close similarity, we applied the GC retention index method. The equation for our experimental condition to calculate a retention index was[48]

$$I(x) = 100n_1 + 100\,\frac{T_R(x) - T_R(n_1)}{T_R(n_2) - T_R(n_1)} \qquad (2)$$

where $T_R(x)$, $T_R(n_1)$, and $T_R(n_2)$ are the retention temperatures of the target compound, the *n*-alkanes with n_1 and n_2 carbon atoms, respectively. T_R can be replaced by corresponding retention time at the linear temperature program. The GC chromatograms to calculate the retention index for our metabolite are shown in Figure 4B. Retention index $I(x) = 2434$ of the target compound was calculated according to eq 2. The literature value[49] for 3-hydroxyhippuric acid di TMS retention index is 2423 and for 4-hydroxyhippuric acid di TMS $I(x) = 2528$. On the basis of this finding, we finally conclude that our biomarker candidate is unambiguously identified to be 3-hydroxyhippuric acid. A scheme of the MS/MS fragmentation of 3-hydroxyhippuric acid is demonstrated in Figure 4C, confirming the product ions found in the spectra of Figures 3A and 4A.

Step 6 (Final Step): Confirmation of the Identified Metabolite and Potential Physiological Relevance. To confirm the identity of the metabolite and to generate a standard for future quantitative validation of the biomarker candidate, a stable isotope labeled

(43) Armstrong, M. D.; Wall, P. E.; Parker, V. J. *J. Biol. Chem.* **1956**, *218*, 921−927.
(44) Baxter, G. J.; Lawrence, J. R.; Graham, A. B.; Wiles, D.; Paterson, J. R. *Ann. Clin. Biochem.* **2002**, *39*, 50−55.
(45) Finnie, M. D. A.; Ersser, R. S.; Seakins, J. W. T.; Snedden, W. *Clin. Chim. Acta* **1976**, *70*, 171−178.
(46) Gonthier, M. P.; Cheynier, V.; Donovan, J. L.; Manach, C.; Morand, C.; Mila, I.; Lapierre, C.; Remesy, C.; Scalbert, A. *J. Nutr.* **2003**, *133*, 461−467.
(47) Rechner, A. R.; Kuhnle, G.; Bremner, P.; Hubbard, G. P.; Moore, K. P.; Rice-Evans, C. A. *Free Radical Biol. Med.* **2002**, *33*, 220−235.
(48) Garcia-Dominguez, J. A.; Lebron-Aguilar, R.; Quintanilla-Lopez, J. E. *J. Sep. Sci.* **2006**, *29*, 2785−2792.
(49) Lefevere, M. F.; Verhaeghe, B. J.; Declerck, D. H.; Van Bocxlaer, J. F.; De Leenheer, A. P.; De Sagher, R. M. *J. Chromatogr. Sci.* **1989**, *27*, 23−29.

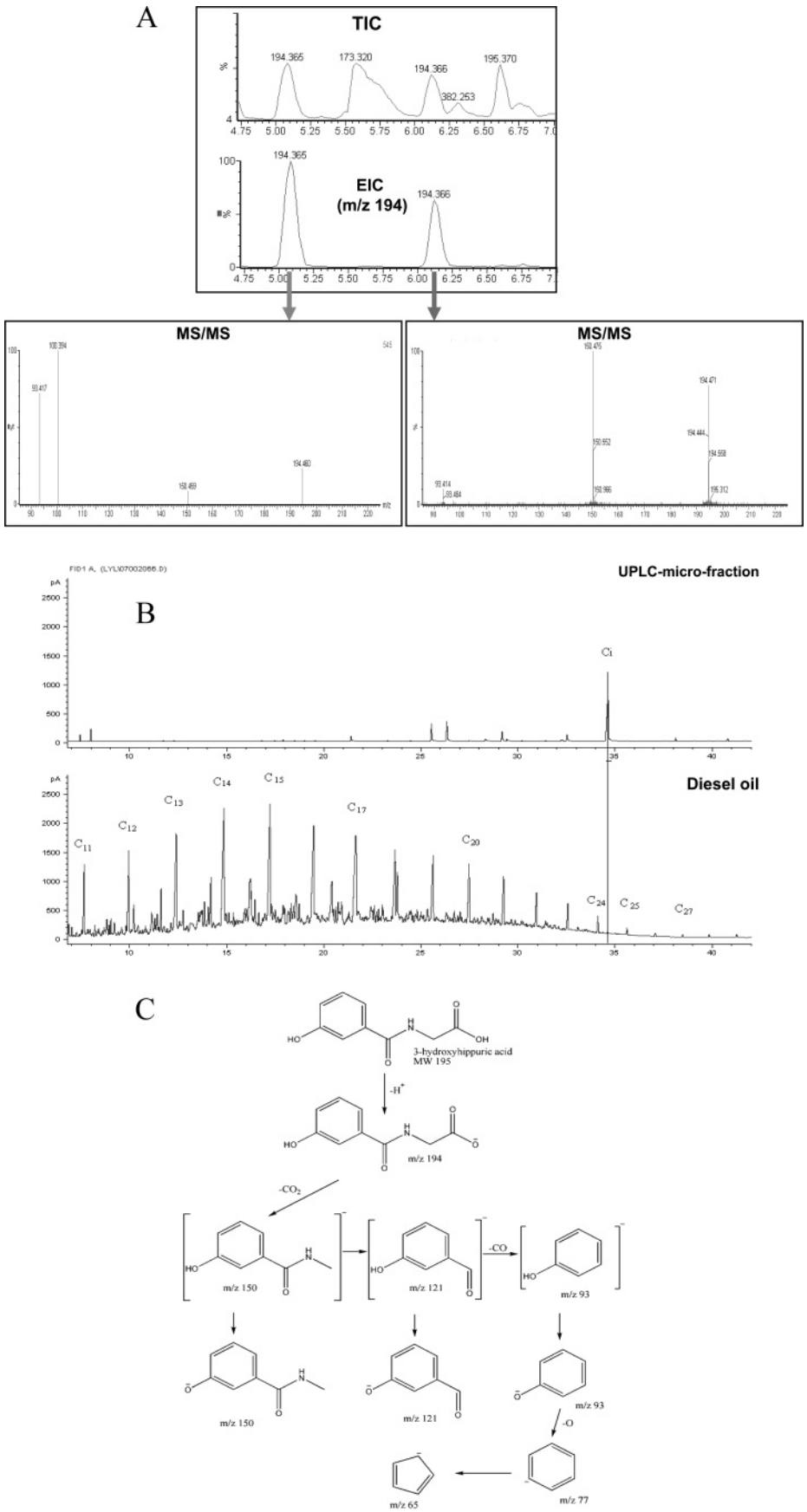

Figure 4. **(A)** Section of the UPLC−Q-TOF MS TIC and EIC of $[M - H]^-$ anions of *m*/*z* 194 and corresponding negative product ion spectra of both *m*/*z* 194 containing peaks. **(B)** GC/MS spectrum of BSTFA derivatived target compound and diesel oil to calculate retention indices (details see eq 2 in the text); **(C)** proposed fragmentation scheme of the theoretical fragmentation mechanism of 3-hydroxyhippuric acid in the negative ESI mode.

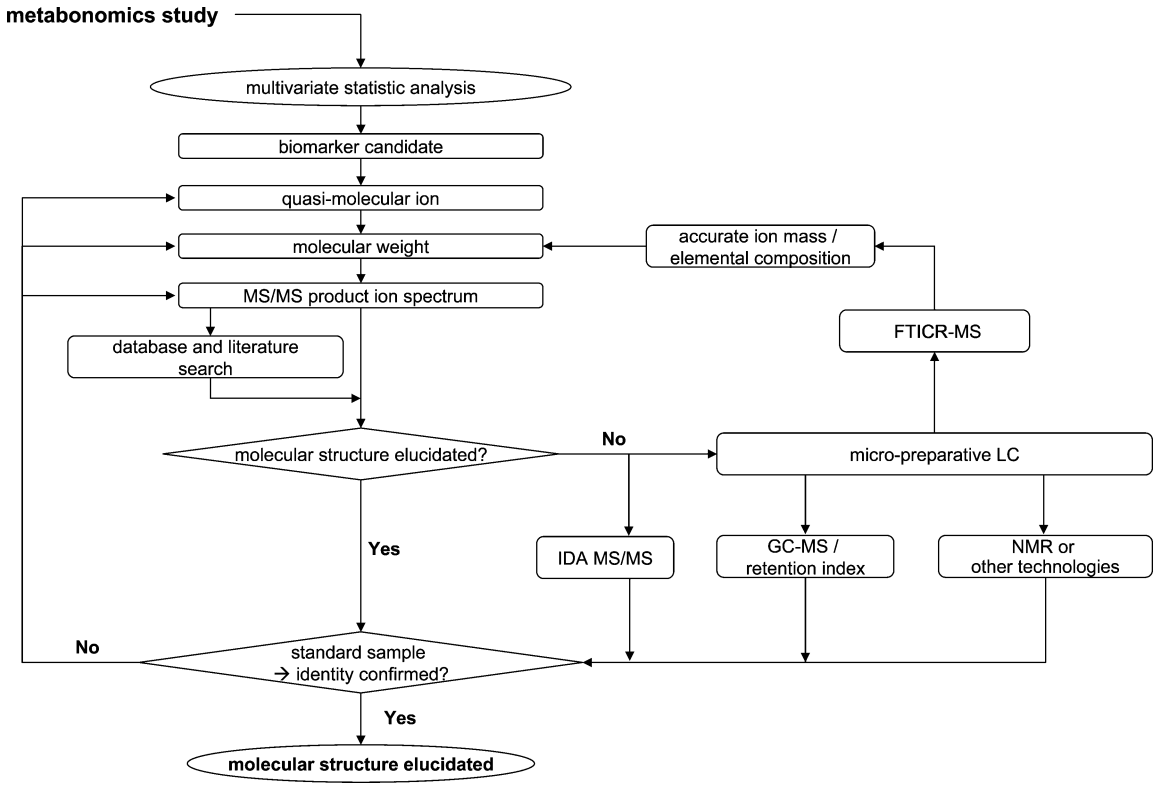

Figure 5. Flow chart of a proposed strategy for the identification and isomer elucidation of biomarkers detected in metabonomic studies applying complementary chromatographic, mass spectrometric, and spectroscopic techniques. (IDA = information-dependent acquisition)

metabolite is needed as a final step (if the metabolite of interest is not commercially available). ^{15}N-labeled 3-hydroxyhippuric acid was synthesized in commission according to a modified procedure of Armstrong et al.[43] Using this standard, we performed again RP-UPLC−MS/MS analysis confirming 3-hydroxyhippuric acid as our target biomarker candidate based on the retention time and the product ion spectrum (data not shown). In Figure 5 the overall strategy for the unambiguous identification of metabolite biomarkers detected in metabonomics studies is summarized in a general flow chart. With application of this strategy, the other biomarker candidates detected in the loading plot shown in Figure 1C can also be identified. Finally, we performed a database and literature search to detect the metabolic pathway of 3-hydroxyhippuric acid. The formation of 3-hydroxyhippuric acid seems to be a central metabolic pathway for dietary flavonoids, in which the colon microflora and the liver are active metabolic sites.[50] Differences in the population of microflora among individuals may explain differences in the production of this metabolite.[51] Relationships between gut microbial disruption and metabolic phenotype change have been intensively investigated in several recent studies by the group of J. K. Nicholson.[52−54] Of note, there is growing evidence from studies in animals and humans that variation in gut microbial composition is an important factor in obesity and insulin resistance.[55,56] Furthermore, recent metabonomic studies by ^{1}H NMR in plasma and urine of dietary-induced insulin resistant mice revealed subtle changes in metabolic regulation underlying disease-promoting and compensatory mechanisms, collectively contributing to diet-induced insulin resistance and nonalcoholic fatty liver disease.[57,58] Although the gut flora come more and more in the focus, we cannot exclude at the moment that the urine marker 3-hydroxyhippuric acid could also reflect a difference in the composition of dietary intake. Future studies may reveal the potential role of this metabolic pathway in the pathogenesis of insulin resistance in prediabetic individuals and clarify the diagnostic relevance of 3-hydroxyhippuric acid to discriminate insulin resistant and insulin sensitive individuals.

CONCLUSIONS

The results presented here indicate the potential of complementary chromatographic and mass spectrometric techniques to unambiguous identify metabolite biomarkers detected in meta-

(50) Spencer, J. P. E. *J. Nutr.* **2003**, *133*, 3255S−3261S.
(51) Manach, C.; Donovan, J. L. *Free Radical Res.* **2004**, *38*, 771−785.
(52) Nicholson, J. K.; Holmes, E.; Wilson, I. D. *Nat. Rev. Microbiol.* **2005**, *3*, 431−438.
(53) Stella, C.; Beckwith-Hall, B.; Cloarec, O.; Holmes, E.; Lindon, J. C.; Powell, J.; van der Ouderaa, F.; Bingham, S.; Cross, A. J.; Nicholson, J. K. *J. Proteome Res.* **2006**, *5*, 2780−2788.
(54) Martin, F. P. J.; Dumas, M. E.; Wang, Y. L.; Legido-Quigley, C.; Yap, I. K. S.; Tang, H. R.; Zirah, S.; Murphy, G. M.; Cloarec, O.; Lindon, J. C.; Sprenger, N.; Fay, L. B.; Kochhar, S.; van Bladeren, P.; Holmes, E.; Nicholson, J. K. *Mol. Syst. Biol.* **2007**, 3, article 112.
(55) Turnbaugh, P. J.; Ley, R. E.; Mahowald, M. A.; Magrini, V.; Mardis, E. R.; Gordon, J. I. *Nature* **2006**, *444*, 1027−1031.
(56) Ley, R. E.; Turnbaugh, P. J.; Klein, S.; Gordon, J. I. *Nature* **2006**, *444*, 1022−1023.
(57) Dumas, M. E.; Barton, R. H.; Toye, A.; Cloarec, O.; Blancher, C.; Rothwell, A.; Fearnside, J.; Tatoud, R.; Blanc, V.; Lindon, J. C.; Mitchell, S. C.; Holmes, E.; McCarthy, M. I.; Scott, J.; Gauguier, D.; Nicholson, J. K. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 12511−12516.
(58) Toye, A. A.; Dumas, M. E.; Blancher, C.; Rothwell, A. R.; Fearnside, J. F.; Wilder, S. P.; Bihoreau, M. T.; Cloarec, O.; Azzouzi, I.; Young, S.; Barton, R. H.; Holmes, E.; McCarthy, M. I.; Tatoud, R.; Nicholson, J. K.; Scott, J.; Gauguier, D. *Diabetologia* **2007**, *50*, 1867−1879.

bonomics studies. Besides the generally applied metabonomics tools, i.e., LC−MS fingerprinting, multivariate statistic analysis, LC−MS^n experiments, the important contribution of the determination of the elemental composition by FTICR-MS as well as the elucidation of the isomeric structure of the metabolite in our approach investigated by GC retention index was clearly demonstrated. The overall strategy reflects the necessity but also the effectiveness of close interactions of various high-end analytical devices supplemented by database and literature searchs to elucidate the identity of the detected biomarkers thus providing a powerful tool for chromatographic-driven metabonomics studies.

ACKNOWLEDGMENT

The study has been supported by the foundation (Grant No. 20425516) for Distinguished Young Scholars, the foundation (Grant No. 20675082) from National Natural Science Foundation of China, the Sino-German Center for Research Promotion (DFG and NSFC, Grant No. GZ 364), the National Basic Research Program of China (Grant No. 2006CB503902), the High Technical Research and Development Program of China (Grant No. 2006AA02Z342), and National Key Project of Scientific and Technical Supporting Programs (Grant No. 2006BAK02A12) from State Ministry of Science & Technology of China.

Received for review October 10, 2007. Accepted November 15, 2007.

AC702089H