Anal. Chem. **2003,** *75,* 4784−4792

Screening of Biomarkers in Rat Urine Using LC/Electrospray Ionization-MS and Two-Way Data Analysis

Helena Idborg-Björkman,[†] Per-Olof Edlund,[†] Olav M. Kvalheim,[‡] Ina Schuppe-Koistinen,[§] and Sven P. Jacobsson*,[†]

Department of Analytical Chemistry, Stockholm University, SE-106 91 Stockholm, Sweden, Department of Chemistry, University of Bergen, NO-5007 Bergen, Norway, and Molecular Toxicology, Safety Assessment, AstraZeneca R&D, SE-151 85 Södertälje, Sweden

Biofluids, like urine, form very complex matrixes containing a large number of potential biomarkers, that is, changes of endogenous metabolites in response to xenobiotic exposure. This paper describes a fast and sensitive method of screening biomarkers in rat urine. Biomarkers for phospholipidosis, induced by an antidepressant drug, were studied. Urine samples from rats exposed to citalopram were analyzed using solid-phase extraction (SPE) and liquid chromatography mass spectrometry (LC/MS) analysis detecting negative ions. A fast iterative method, called *Gentle*, was used for the automatic curve resolution, and metabolic fingerprints were obtained. After peak alignment principal component analysis (PCA) was performed for pattern recognition, PCA loadings were studied as a means of discovering potential biomarkers. In this study a number of potential biomarkers of phospholipidosis in rats are discussed. They are reported by their retention time and base peak, as their identification is not within the scope of the study. In addition to the fact that it was possible to differentiate control samples from dosed samples, the data were very easy to interpret, and signals from xenobiotic-related substances were easily removed without affecting the endogenous compounds. The proposed method is a complement or an alternative to NMR for metabolomic applications.

The use of biomarkers in toxicology is becoming increasingly important in assessing the health risks of exposure to potentially toxic drugs and chemicals. Biomarkers are used to measure effects on the catabolism after exposure of a toxic compound and the extent of a toxic response that should be easy to measure and quantify with high sensitivity and specificity, and they should furthermore relate to the biochemical mechanism of a compound. The ultimate biomarker should be species-independent and noninvasive and should predict a toxic effect at realistic doses.[1] There are high expectations that the use of molecular profiling methods such as genomics, proteomics, and metabolomics will improve risk assessment and the identification of biomarkers for drug effects and toxicity.[2] Using these techniques, the expression of thousands of genes, proteins, and endogenous metabolites can be investigated simultaneously following exposure to a toxic compound. The challenge for the toxicologist is to distinguish physiological effects, such as age, gender, diet, etc., and pharmacological effects (induction, diuresis, renal xenobiotic clearance) from toxicity and to offer a valid interpretation of their meaning.

In this study, drug-induced phospholipidosis is used as an example to illustrate the development of a method using LC/MS and two-way analysis, that is, multivariate analysis, for biomarker identification. Phospholipidosis is a phospholipid storage disorder resulting in an excessive accumulation of phospholipids in the tissues and has been observed as a recurrent pathological feature in toxicity studies in both animals and humans.[3,4] To date, the determination of phospholipidosis has relied on the use of histopathology and electron microscopy or the examination of peripheral blood lymphocytes.[5] Accordingly, drug candidates causing phospholipidosis are in general identified at a late stage in drug development, and there is a great need to identify biomarkers for early identification of the insult. In this report, the antidepressant citalopram, a selective serotonin reuptake inhibitor, has been used as a model compound to induce phospholipidosis.

Recently, the use of nuclear magnetic resonance (NMR) as an analytical technique in the screening of biomarkers has gained attention. NMR metabonomics and pattern recognition analysis have been used in several metabolic studies to identify potential biomarkers.[6] NMR has been proven to be a fast and information-rich technique, giving rise to very complex spectra that mainly have to be analyzed by multivariate methods. Owing to sensitivity reasons and the complexity of the data, the proportion of potential biomarkers that can be identified is limited. Thus, there is a need for alternative methods that can fulfill the aim of screening with high capacity and provide a means of identification. The use of

[†] Stockholm University.
[‡] University of Bergen.
[§] AstraZeneca R&D Södertälje.

(1) Timbrell, J. *Toxicology* **1998**, *129*, 1−12.
(2) Hellmold, H.; Nilsson, C. B.; Schuppe-Koistinen, I.; Kenne, K.; Warngard, L. *Toxicol. Lett.* **2002**, *127*, 239−243.
(3) Kodavanti, U. P.; Mehendale, H. M. *Pharmacol. Rev.* **1990**, *42*, 327−54.
(4) Halliwell, H. *Toxicol. Pathol.* **1997**, *25*, 53−60.
(5) Luellmann, H.; Luellmann-Rauch, R.; Wassermann, O. *Crit. Rev. Toxicol.* **1975**, *4*, 185−218.
(6) Nicholls, A. W.; Nicholson, J. K.; Haselden, J. N.; Waterfield, C. J. *Biomarkers* **2000**, *5*, 410−423.

10.1021/ac0341618 CCC: $25.00
Published on Web 08/20/2003

liquid chromatography/mass spectrometry (LC/MS) makes probable the detection of endogenous metabolites at lower concentrations and enables different compounds to be analyzed compared to NMR. In addition, LC/MS is a common technique in most laboratories, easy to automate for large sample series, and suitable for polar substances such as endogenous metabolites. Another advantage is the possibility for analyzing small sample volumes, such as urine samples from mice. Furthermore, the signals are not obtained as one spectrum, giving less complexity, and signals from, for example, xenobiotics can easily be removed. The use of LC/MS combined with pattern recognition in pharmaceutical applications has been described by Gaspari et al.[7] The use of GC/MS in combination with multivariate methods could be found in the field of plant metabolomics[8,9] and in applications such as exposure markers in smokers' breath,[10] molecular markers for sources of combustion,[11] and markers for oil source specific parameters.[12]

Apart from the recently published work of Plumb et al.,[13] the authors are not aware of applications of LC/MS in combination with advanced pattern recognition for capturing optimal information in screening of biomarkers. This is probably due to a number of difficulties. Although one obvious reason for using LC/MS is that it produces abundant data and a lot of information about molecules of interest, it is difficult to extract the information from the data collected. If full scan MS is used, typically the data has a noisy structure and is often dominated by noise at a specific retention time, making it difficult to construct a model for the data. It is not possible to subtract the noise as a specific Thomson number, that is, m/z, since this might contribute to a signal at one retention time but to noise for the remaining time. The major inconvenience is most probably the fact that it is difficult to compare two runs, as the retention time is not constant. In the case of complex matrixes such as urine, the retention time can vary between the injections by almost a minute, with effects on chromatographic separation. At all events complete chromatographic resolution of all peaks is impossible to obtain for such a large number of analytes in a single separation dimension.

Accordingly, in the case of incomplete chromatographic separation, it can be improved mathematically using the two-way data obtained. Different approaches to the deconvolution of data from hyphenated chromatography, such as LC-DAD and LC/MS, have been developed.[14] For instance, the approach by Plum et al.[13] relies on a bucketing procedure, summing hundreds of scans into a single scan, resulting in loss of peak resolution and the identification ability of single biomarkers. Alternatively, instead of resolving each sample separately, that is, using two-way data, it might be possible to analyze all sample matrixes at the same time and use tree-way analysis methods such as PARAFAC.[15] It has been demonstrated in the literature that a proper time alignment is essential for LC/MS data prior to using PARAFAC or other three-way methods.[16] Using three-way techniques implicitly assumes, for example, the same noise level and background for all samples. This may provide an unnecessary high noise level causing loss of small peaks.

In this paper an iterative procedure called *Gentle* is used for mathematical resolution of the two-way data.[17,18] This method is similar to alternating regression[19] although the constraints, such as nonnegativity and unimodality, are less harsh. Elementary matrix transformations are used in each step without changing the fit obtained for the data. The samples are resolved independently of each other as, for example, background correction and noise level calculations are based on data from the sample in question. After resolution, the similarity between spectra of peaks between samples is calculated. Peaks within a certain time interval are compared. Principal component analysis (PCA)[20] was used on the matrix obtained, and patterns were studied to discover potential biomarkers.

In this paper we describe a novel method of identifying urinary changes in endogenous metabolites, which might act as specific biomarkers for phospholipidosis. The proposed method uses multivariate analysis to improve the reliability of identification of potential biomarkers. The method, which includes simple solid-phase extraction, is based on reversed-phase LC/MS full scan analysis and, where applicable, on the identification of specific potential biomarkers of phospholipidosis. Enhanced information extraction was obtained by the use of two-way analysis and multivariate visualization approaches.

EXPERIMENTAL SECTION

Urine samples from rats exposed to citalopram were analyzed using SPE and LC/MS. Multivariate curve resolution was used to resolve the data into peaks and spectra, and PCA of resolved peaks was used to study potential biomarkers.

MATERIALS AND METHODS

Animal Experiments. Twelve male Wistar rats (supplier M&B A/S, Denmark, 6−7 weeks old on delivery) were randomly assigned to two groups. They were acclimatized for 10 days prior to dosing, with free access to water and food (RM1.E.SQC, Special Diets Services Ltd., UK), and maintained at a 12-h light/dark cycle. They were dosed for 14 consecutive days by oral gavage with tap water (rats #1−6) or 125 mg/kg/day citalopram (commercially available solution of Seropram, 40 mg/mL, diluted to 15 mg/mL with tap water). For periods of urine collection, animals were housed individually in metabolism cages for 8 h during the daytime, with free access to drinking water. Urine was collected on one occasion prior to dosing (day −5) and on days 1, 3, 7, 10, and 14 of the study. Urine volume and pH were recorded, and

(7) Gaspari, M.; Vogels, J.; Wulfert, F.; Tas, A. C.; Venema, K.; Bijlsma, S.; Vreeken, R.; Van der Greef, J. *Adv. Mass Spectrom.* **2001**, *15*, 283−296.
(8) Fiehn, O.; Kopka, J.; Dörmann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. *Nat. Biotechnol.* **2000**, *18*, 1157−1161.
(9) Sumner, L. W.; Mendes, P.; Dixon, R. A. *Phytochemistry* **2003**, *62*, 817−836.
(10) Gordon, S. M. *J. Chromatogr.* **1990**, *511*, 291−302.
(11) Currie, L. A.; Klouda, G. A.; Benner, B. A., Jr.; Garrity, K.; Eglinton, T. I. *Atmos. Environ.* **1999**, *33*, 2789−2806.
(12) Oberrauch, E.; Salvatori, T. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 137−147.
(13) Plumb, R. S.; Stumpf, C. L.; Gorenstein, M. V.; Castro-Perez, J. M.; Dear, G. J.; Anthony, M.; Sweatman, B. C.; Connor, S. C.; Haselden, J. N. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1991−1996.
(14) Liang, Y.; Kvalheim, O. M. *Fresenius J. Anal. Chem.* **2001**, *370*, 694−704.
(15) Bro, R. *Chemom. Intell. Lab. Syst.* **1997**, *38*, 149−171.
(16) Bylund, D.; Danielsson, R.; Malmquist, G.; Markides, K. E. *J. Chromatogr., A* **2002**, *961*, 237−244.
(17) Manne, R.; Grande, B.-V. *Chemom. Intell. Lab. Syst.* **2000**, *50*, 35−46.
(18) Shen, H.; Grung, B.; Kvalheim, O. M.; Eide, I. *Anal. Chim. Acta* **2001**, *446*, 313−328.
(19) Karjalainen, E. J. *Chemom. Intell. Lab. Syst.* **1989**, *7*, 31−8.
(20) Jackson, J. E. *A users guide to principal components*; Wiley: New York, 1991.

samples were stored at −70 °C until analysis. (No sample from rat 8 on day 10 was available.) For histopathologic examination, liver, lungs, thymus, and spleen tissue were fixed in formalin, paraffin-embedded, cut (4−5 μm), stained with hematoxylin and eosin, and examined by light microscopy.

Standards. Ten milligrams of citalopram (C-7861, Sigma) was dissolved in 1 mL of Milli-Q water. To determine the retention time, 0.5 μg of citalopram was injected. Blank urine was spiked with citalopram and prepared according to the method to confirm that the substance was not discriminated.

Benzoic acid (Sigma B-7521) prepared in an aqueous solution of 10% acetonitrile (10.3 mg/100 mL) was used to study system suitability.

Ten milligrams of phenylacetylglycine, a previously proposed biomarker, was dissolved in 1 mL of water. Blank urine was spiked with phenylacetylglycine and prepared according to the method to establish the retention time and mass spectrum.

Sample Preparation. The adsorbent was activated and conditioned first with 1 mL of methanol and then with 1 mL of ammonium acetate buffer (10 mM, pH 4). An aliquot of 0.5 mL of rat urine was loaded onto the SPE column (1 cm^3, 30 mg, Waters Oasis HLB). In the washing step 0.5 mL of ammonium acetate buffer (10 mM, pH 4) was used, and 0.5 mL of methanol was used for elution. The eluates were filtered through syringe filters (Gelman GHP Acrodiscs, 13 mm id, 45 μm, Merck) before injection. No chromatographic problems occurred due to the injection of methanol, that is, the eluate, as only 5 μL was injected. Samples awaiting for preparation and newly prepared samples were stored at approximately 8 °C prior to analysis.

The sample preparation procedure described above was developed using a different set of rat urine samples. To verify the adequacy of this method, the washing steps for rats #7−12 after 7 days of citalopram exposure were analyzed. For these samples, a second elution step was performed, with an additional 0.5 mL of methanol.

The samples were not analyzed strictly according to the sample collection, that is, day −5, 1 etc. However, the samples were analyzed in numerical order, that is, rats #1−12, for each particular day. Samples were analyzed as soon as possible after sample preparation, typically within 1 day.

LC/ESI-MS Analysis. For the LC/MS analysis, a Waters 2790 LC-system was coupled to a Quattro Micro mass spectrometer from Micromass.

For the LC-analysis, an Xterra C18-column (3.5 μm, 2.1 mm × 150 mm) from Waters was used together with a precolumn (3.5 μm 2.1 × 10 mm). Mobile phase A consisted of 10 mM ammonium acetate (adjusted to pH 4 with formic acid), and mobile phase B consisted of acetonitrile. The gradient started with 10% B for 2 min and then linearly increased to 90% B within 15 min. This was kept at isocratic conditions for 8 min before the gradient was allowed to reach 10% B in 30 s and then equilibrated for 6.5 min. The total analysis time was 30 min.

An electrospray ionization (ESI) interface was used. The capillary voltage was set to 2.5 kV and the cone voltage to 25 V. The source temperature was 110 °C. Profile data for negative ions from m/z 50 to 600 were recorded at a speed of 1 s/scan. In the same run positive ions were collected, that is, switching between negative and positive ion mode. The data from positive ion mode were not used in this study. Benzoic acid was injected between sets of samples to confirm the performance of the LC/MS analysis.

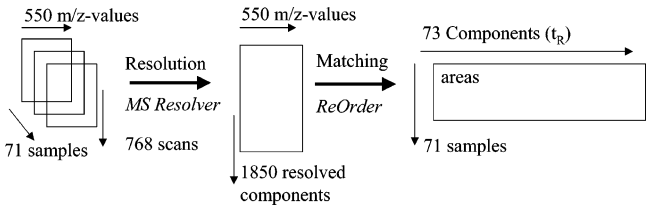

Figure 1. Illustration of the data structure of two-way data analysis using the programs MS Resolver for resolution and ReOrder for peak similarity matching. Seventy-three unique components were found for all samples.

A sample from rat #7 after 10 days of exposure was used to obtain product ion spectra for potential biomarkers (variables 65 and 69) and metabolites of citalopram (variables 71 and 59). In the same run product ions from molecular ion deprotonated molecules 326, 324, 310, and 486 were obtained in negative mode. Two analyses were performed using collision energies of 15 and 20 eV, respectively.

Multivariate Data Analysis. The data obtained from the LC/MS analysis are a two-way matrix for each sample, that is, retention time in one direction and mass spectrum in the other direction, and can therefore be used for automatic curve resolution.[21] Each data file (a matrix of 550 m/z values times 768 scans) converted to cdf format was approximately 30 MB, which is difficult to handle with normal statistical programs. Automatic peak tracking was therefore used to extract information from the data. MS Resolver (Pattern Recognition Systems, Bergen, Norway) was used in this work to mathematically resolve the data into components. The program is based on the *Gentle* iteration method for automatic curve resolution.[17,18] This method starts from a set of key spectra and refines the corresponding concentration profiles by simple elementary transformations, making it a fast method. The resolution obtained is sufficient for classification studies or similar approaches. In this study a total 1850 components were resolved, see Figure 1.

After resolution, the resolved components in each sample were compared between the samples using ReOrder (Pattern Recognition Systems, Bergen, Norway). To decide whether a component in one sample was the same as in another sample (despite a retention time shift), their spectra were matched. The criterion was a similarity index, that is, cross-correlation between two spectra, of 0.8, and the peak tracking was performed within a time interval of ±120 s, which is ∼4 times the peak shift that could be expected. When spectra in a retention time interval were compared, the seven most intense m/z values in the spectrum were used. Using ESI, the molecular ion is expected to dominate the spectrum, although for some compounds in source fragmentation is obtained.

The resulting prediction matrix, consisting of the area for each compound at a determined retention time, was used for principal component analysis (PCA).[20] Before performing PCA, additional preprocessing was needed. The log 10 transformed data were used to reduce the influence from large components and the heteroscedasticity of noise structure, which is suggested by Rietjens.[22]

(21) Hamilton, J. C.; Gemperline, P. J. *J. Chemom.* **1990**, *4*, 1−13.
(22) Rietjens, M. *Anal. Chim. Acta* **1995**, *316*, 205−215.

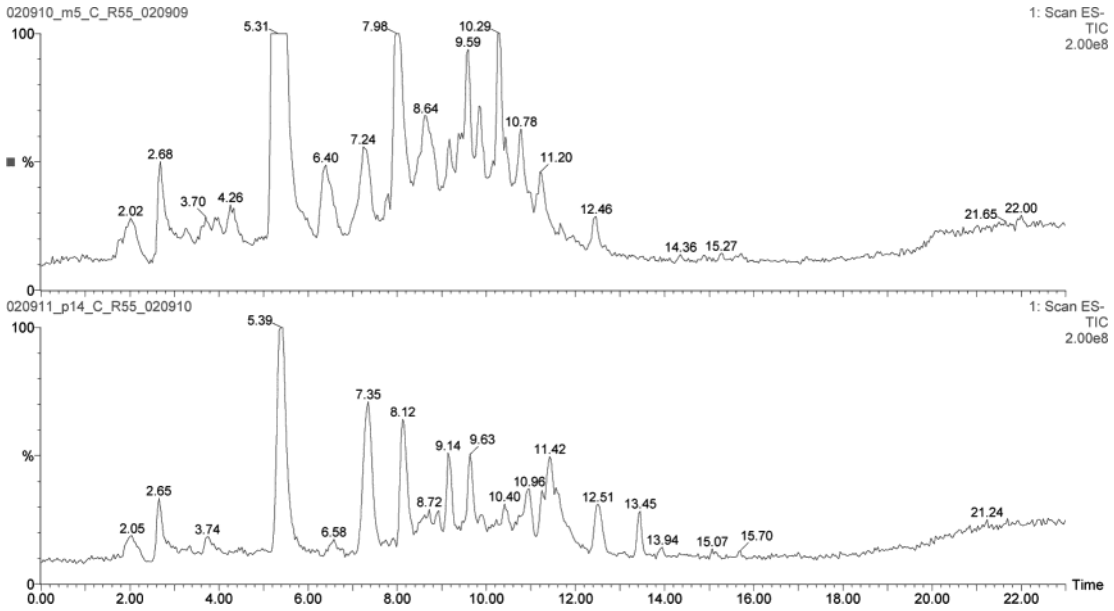

Figure 2. Total ion current (TIC) of urine samples from rat 7 (citalopram-exposed) collected on day −5 (above) and day 14 (below).

As the values for various components in some of the samples were zero, a constant (10^{-9}) was included before the transformation. This is practically similar to setting the logarithm of zero to zero. Another alternative is to use the fourth root transformation suggested by Wold;[23] this was not proceeded with, however, since almost equivalent results were obtained to that of log transform. The data sets were then mean centered before performing PCA. Sole mean centering of data results in variables heavily weighted toward large absolute variance, that is, large peaks. However, since these peaks generally reflect endogenous components in the rat urine not affected by the treatment, less informative visualization plots were obtained. A similar situation was obtained when using scaling based on mean and standard deviations of the variables, that is, autoscaling. In this case the noisy structure, that is, the noninformative part of the data, was too greatly amplified. A number of normalization procedures have been proposed in conjunction with mass spectrometry data, see the papers by Rietjens,[22] Sörensen et al.,[24] and Gottlieb et al.[25] In this study no normalization was carried out to compensate for differences in, for example, urine volume. A straightforward approach for variable normalization was found to be mean centering of logarithmic scaled data. PCA score plots were studied to observe similarities between samples, and loading plots were studied to find important variables, that is, compounds that were found to depend on time and were similar for all rats exposed to citalopram.

In addition to PCA, partial least squares (PLS)[26] was used to confirm that the proposed biomarkers correlated with the day of citalopram exposure. The mean-centered day of exposure was used as a response to determine the variables describing the largest variance and is correlated with the day of exposure.

(23) Wold, S. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109−130.
(24) Sörensen, H. A.; Sperotto, M. M.; Petersen, M.; Kesmir, C.; Radzikowski, L.; Jacobsen, S.; Søndergaard, I. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1232−1237.
(25) Gottlieb, D. M.; Schultz, J.; Petersen, M.; Nesic, L.; Jacobsen, S.; Søndergaard, I. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 2034−2039.
(26) Wold, S.; Ruhe, S.; Wold, H.; Dunn, W. J. *J. Sci. Stat. Comput.* **1984**, *5*, 735−743.

RESULTS AND DISCUSSION

Animal Study. All animals treated with citalopram showed moderate phospholipid accumulation in the lung and minimal to slight accumulation in the thymus and spleen. There was no evidence of phospholipid accumulation in the organs of the control animals. No other significant treatment-related changes in the pathologic and histopathologic examination were observed.

Sample Preparation and LC/MS Analysis. It is difficult to analyze hydrophilic substances in the same run as highly polar substances, that is, unretained compounds such as sugars. When the washing step was analyzed, it was obvious that some of the highly polar compounds had been discarded. However, if they had not been rejected during the sample preparation and had instead been injected onto the column, they would have had no retention and would have been impossible to analyze under these conditions. Some information might have been lost due to the sample preparation. On the other hand, according to the results obtained, the compounds that were analyzed gave a wealth of information. These highly polar compounds may be analyzed in a similar way with a column suitable for more polar compounds, for example, Aquasil or Hypercarbon column. Some of the compounds with $k' < 0$ might also have a very low molecular mass and be detected with lower sensitivity, as the noise is much higher at the low mass end. When the second elution step was analyzed, no major compounds were found. The sample preparation served its purpose of desalting the analytes, leaving a large number of metabolites for analysis. Carryover was observed at a retention time of approximately 5.3 min. This did not, however, conflict with the results.

Although the majority of LC/MS analyses were performed in positive ion mode, it was found in this study that negative ions gave much more information. The complexity of the samples is demonstrated in Figure 2, which shows the total ion current (TIC) chromatogram of urine samples collected on day −5 and day 14 after dosing from rat #7 (citalopram). It is an advantage to be able to study biomarkers without interference from major metabolites

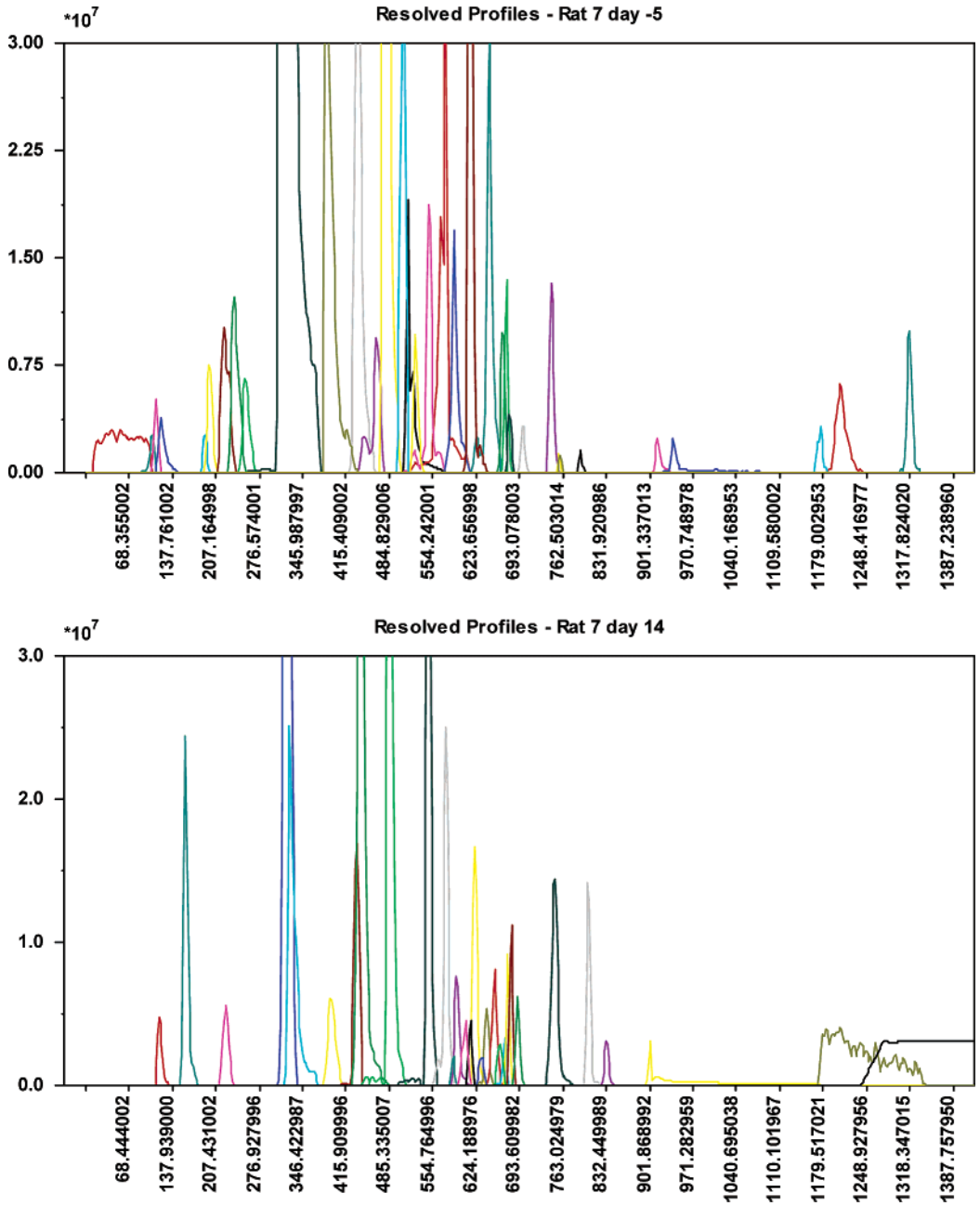

Figure 3. Mathematically resolved profiles of urine samples collected on day −5 (above) and on day 14 (below) for rat 7 (citalopram-exposed). Retention time in seconds can be seen on the *x*-axis and intensity (10^7) is the response.

of citalopram[27] as they are not detected in negative mode. However, two of the major metabolites of citalopram, citalopram acid and the glucuronide of citalopram acid, were detected in negative mode and will therefore be identified as false biomarkers.

To demonstrate that citalopram is not detected in negative ion mode a spiked urine sample was prepared and analyzed in both positive and negative ion mode. The retention time was 11.8 min for $[M + H]^+$. In negative ion mode no peak was detected at that retention time corresponding to $[M - H]^-$ m/z 323. This also shows that citalopram was not discriminated during the sample cleanup.

Multivariate Analysis. From each sample approximately 30 components are resolved. In Figure 3 the resolved chromatograms of the profile of rat #7 collected on day −5 and day 14 are shown. When creating the predictor matrix, it is possible to remove components that are artifacts. The variable selection was not based on visual judgment but on retention time. Peaks eluting without any significant retention, that is, eluting before 2.9 min, and components eluting after 16 min were removed as no compounds with typical peak shapes were observed after 16 min. According to this study, it is possible to shorten the gradient from 23 min to at least 16 min, not including the time to equilibrate the column. It should be noted that MS Resolver resolves the late components correctly, comparing reconstructed chromatograms, although they

(27) Dalgaard, L.; Larsen, C. *Xenobiotica* **1999**, *29*, 1033−1041.

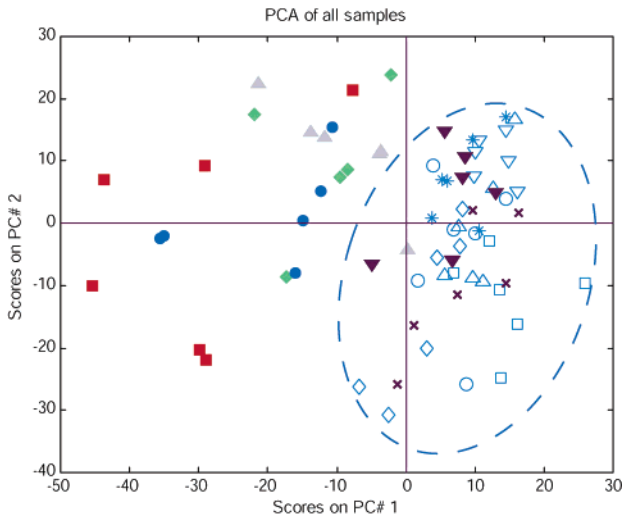

Figure 4. Scores of PC1 and PC2 obtained from PCA of all samples are plotted. Control samples (rats #1−6) are separated from the dosed animal samples collected after day 3 and are shown as nonsolid markers on the right. Nonsolid markers of different kinds were used to separate each rat, e.g., data for rat #4 on different collection days are shown as nonsolid diamonds (◇, blue). Predosed samples (rats #7−12) are labeled ×. Post-dose samples (filled markers) collected after 1 (▼, purple), 3 (▲, gray), 7 (●, blue), 10 (♦, green) and 14 (■, red) days are distributed to the left of the plot.

are not considered to be real compounds because of their odd peak shape (elution profiles), see Figure 3.

Not all components can be correctly resolved using MS Resolver, for example, partially separated isomers with very similar mass spectra. This is not necessary as long as the data are sufficient for classification and identification of potential biomarkers, which seemed to be the case. It is not practicable, for example, to resolve components without an iterative approach, as the number of samples and components is very large and manual analysis becomes prohibitive. A score plot of all samples demonstrates the success of this method. It can also be noted that the peak alignment using spectral matching by ReOrder was successful. The number of components (1850 variables) was reduced to 73 unique variables (see Figure 1).

A score plot based on PCA of all samples is rather complex (Figure 4). However, the difference between exposed (rats #7−12) and nonexposed rats is obvious. The difference between individuals (when studying male rats) is small compared to differences caused by exposure. After only 1 day of exposure no significant change has occurred. PC1 and PC2 explain 19% of the variance, and it is possible to explain the change related to the exposure of citalopram using only one rotated component. A clearer time dependence was observed when performing PCA on the average of rats #1−6 (nonexposed) and rats #7−12 (exposed) (the corresponding scores plot is shown in Figure 5a).

A time dependence can be studied for samples, that is, scores, obtained from rats exposed to citalopram (Figure 5a), whereas for rats only drinking tap water, a more random distribution was obtained. Additionally, PCA was performed on each rat separately (not shown), and the loadings were studied carefully. If a variable has a high absolute value in all PCA models performed on citalopram-exposed rats but low absolute loading values for PCA performed on tap water rats, this variable is a potential biomarker.

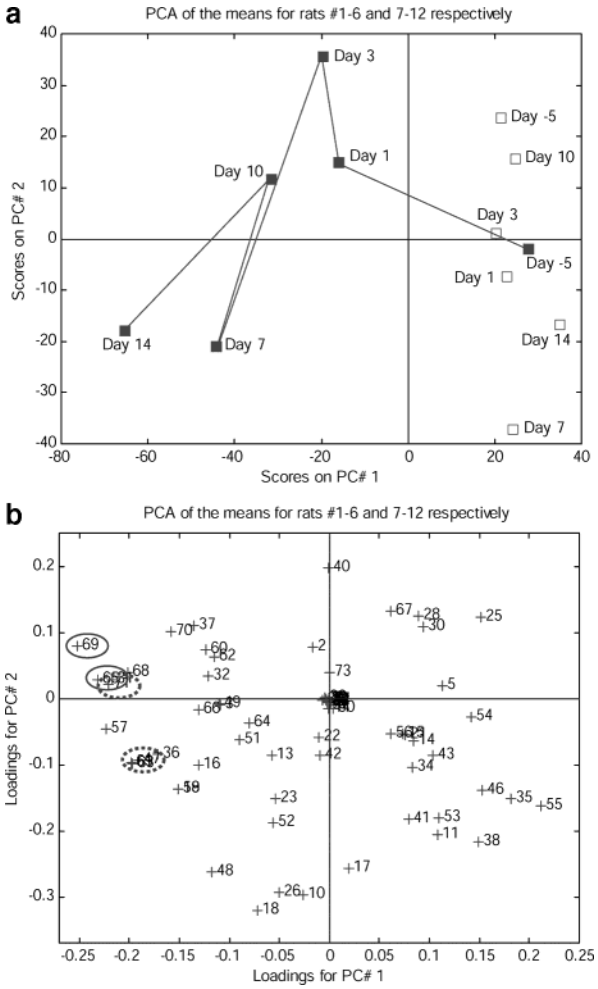

Figure 5. a. Scores of PCA performed on the average of samples from rats #1−6 (tap water) and rats #7−12 (citalopram) collected on different days. Nonexposed rats (□) are not separated in PC1 but citalopram-exposed rats (■) are. The average for day 10 of the citalopram-exposed rats is only based on 5 values instead of 6 values as no sample for rat #8 on day 10 was available. b. Loadings of PCA performed on the average of samples from rats #1−6 (tap water) and rats #7−12 (citalopram) collected on different days. Variables distributed to the left are important for samples after dosage. Two of the most obvious potential biomarkers (variables 69 and 65) are encircled with a solid line, and two variables that correspond to metabolites of citalopram, i.e., variables 71 and 59, are encircled with a dashed line.

In Figure 5b the loadings for PCA of the average of rats #1−6 (nonexposed) and rats #7−12 (exposed), respectively, are plotted. In Figure 5b variables to the left are important for rats exposed to citalopram and may be used as biomarkers. For example, variables 69 and 65 are potential biomarkers. Citalopram acid (m/z 310, variable 71) and the glucuronide of citalopram acid (m/z 486, variable 59) were falsely detected as potential biomarkers. On the other hand, this indicates that the method described actually captured changes in the metabolic pattern and time dependence. It is also possible to remove signals related to, for example, xenobiotics without losing signals from endogenous compounds. The variable most likely corresponding to citalopram acid and glucuronide citalopram acid was discarded before PCA, and the same results, that is, the same potential biomarkers, were obtained.

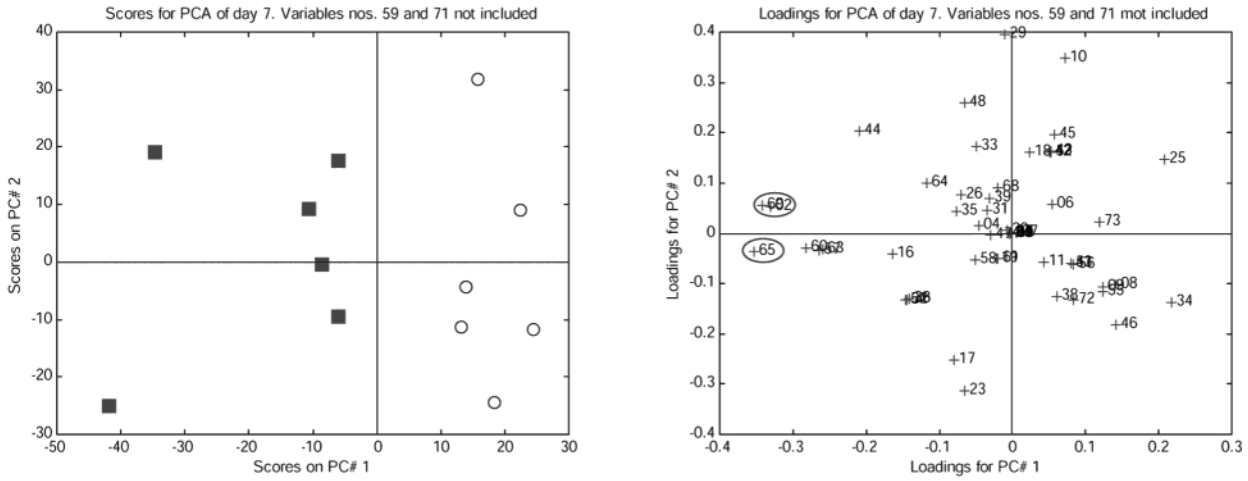

Figure 6. Scores obtained from PCA of all rats on day 7 (left) are plotted for PC1 and PC2. Samples from citalopram-exposed rats (■) are distributed to the left and control samples (○) to the right. Loadings obtained from PCA of all rats at 7 days for PC1 and PC2 are plotted in the right picture. Variables 65 and 69 are encircled with a solid line and are regarded as potential biomarkers. The same variables were important when comparing control and dosed samples for days 3, 10, and 14, respectively. In the above pictures variables 71 and 59 were excluded before performing PCA as they correspond to metabolites of citalopram.

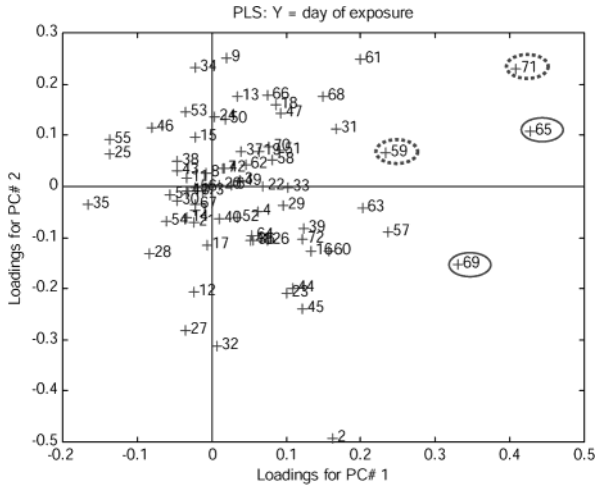

Figure 7. Weights for PLS-component 1 and 2 are plotted above. Variables 65 and 69, encircled with a solid line, were found to be correlated with the day of exposure, which confirms the earlier suggestion of potential biomarkers. Variables 71 and 59, encircled with a dashed line, correspond to metabolites of citalopram and were, of course, correlated with the day of exposure.

This can also be confirmed when PCA is performed on control and dosed samples from different exposure days, that is, samples after 1, 3, 7, 10, and 14 days, respectively. Before performing PCA, variables nos. 59 and 71 were removed, as they are most likely metabolites of citalopram. When scores and loadings from day 14 are studied in Figure 6, variables such as 65 and 69 have high loadings in PCs important for rats #7−12 (citalopram). The same variables were found for all rats after some days of exposure and can therefore, in principle, be used as biomarkers. To verify that the high loadings are correlated with the day of exposure, PLS was performed. The two first latent variables described 16% of the variance in x and 91% of the variance in y. The x scores (t) and weights (w) are plotted, and weights are shown in Figure 7. The same variables as above were still found to correlate with the day of citalopram exposure. Variable no. 65 describes 18% of the variance described by PC1 (which accounts for 16% of the total variance in data), and variables 71, 69, and 57 describe 16, 11, and 5% respectively.

Potential Biomarkers. Several potential biomarkers were tentatively identified in this study. The most obvious were components corresponding to variables nos. 69 and 65. Variable 69 corresponded to a compound with a retention time of approximately 13.1 min and a base peak of m/z 324 and variable 65 with a retention time of 12.6 min at m/z 326. The spectrum of variable no. 69 and product ion spectrum of variable no. 65 are shown in Figure 8. For variable 69, in-source fragmentation was obtained. Variable 71 (13.8 min at m/z 310) and variable 59 (11.2 min at m/z 486) might correspond to two of citalopram's major metabolites, citalopram acid and the glucuronide of the acid metabolite, respectively. Product ion spectra verify this assumption.

An examination of the two potential biomarkers revealed an increase in the concentration profile of rats exposed to citalopram and the kinetics being shown in Figure 9. None of these compounds could be detected in the control samples. Previous investigations by NMR suggest phenylacetylglycine (PAG) as a urinary biomarker for phospholipidosis.[6] When examining the reconstructed ion chromatograms of m/z 192 at 7.9 min corresponding to PAG, the changes are plotted by day of dosage in Figure 10. PAG (variable no. 20) was also detected in control samples and according to the present study is at best only a weak biomarker for phospholipidosis. However, if a different preprocessing is used, for example, normalization, PAG might be used as a biomarker, although there are better candidates. It could also be noted, comparing Figures 9 and 10, that the proposed biomarkers, that is, variable nos. 65 and 69, are present at a much lower intensity compared to PAG, although distinguished as markers using LC/MS. These markers are the most obvious difference between control and dosed samples as they are not present at detectable levels in the control samples. When a different preprocessing was used, such as peak area normalization, other less pronounced markers could be detected, for example,

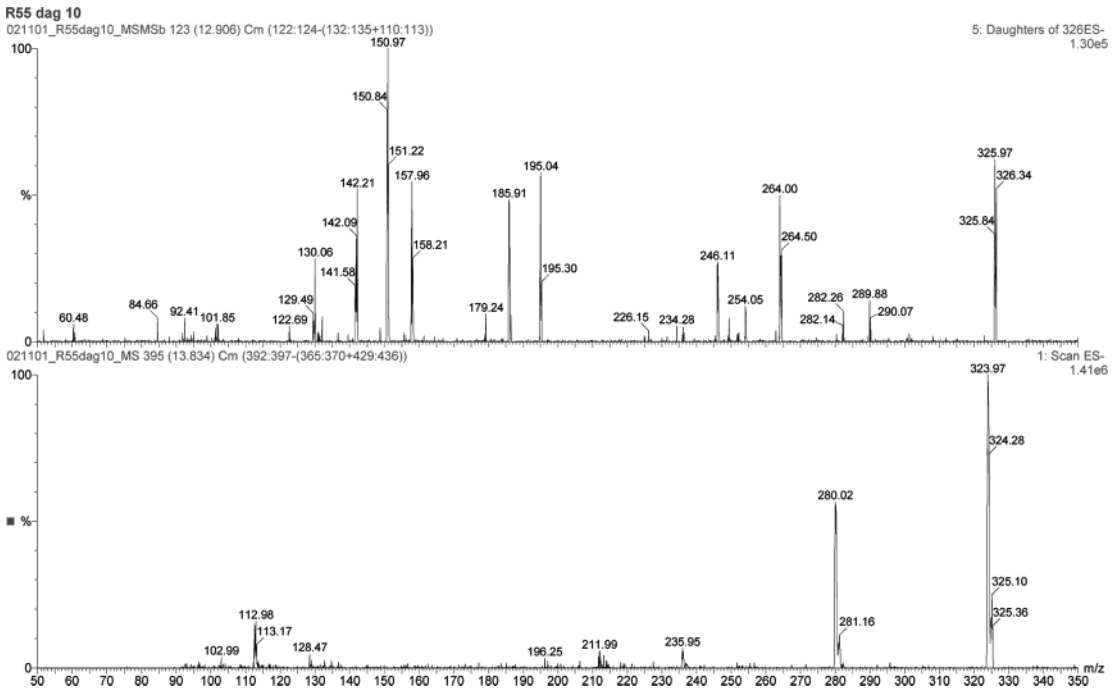

Figure 8. Product ion spectrum of *m*/*z* 326 (variable no. 65) was obtained (upper picture) using collision-induced dissociation. Product ion spectrum was obtained for *m*/*z* 324 (variable no. 69) but not shown as the collision energy used was too high. Instead the MS spectrum for variable no. 69 is shown (lower picture), demonstrating in-source fragmentation.

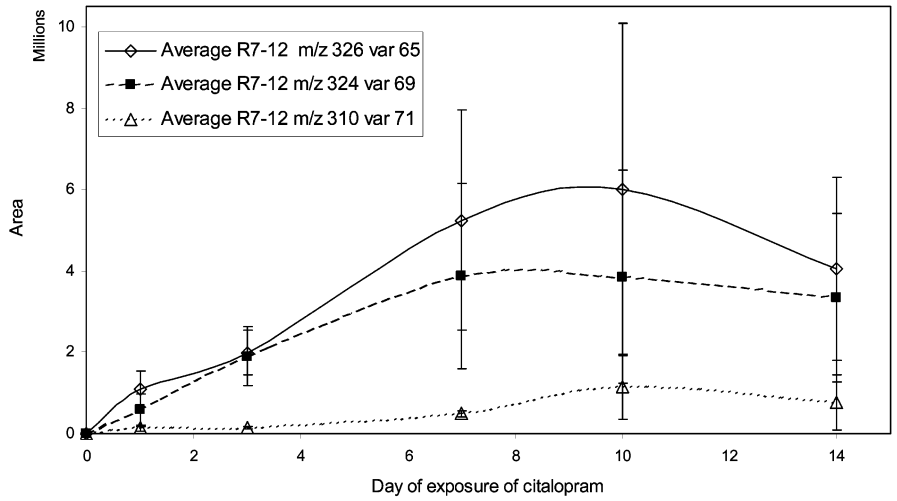

Figure 9. Reconstructed ion chromatograms (RIC) for *m*/*z* 326, corresponding to variable no. 65, and *m*/*z* 324, corresponding to variable no. 69, were obtained. The average of the area of RIC for rats #7−12 (citalopram) on different days after exposure are plotted above to illustrate the kinetics of the potential biomarkers. In rats nos. 1−6 (tap water) no peak corresponding to variables 65 and 69 were detected.

m/*z* 297 at tR 9 min. These were present both in the controlled and dosed samples, but a significant increase or decrease was found. It should be pointed out that the discussion above concerning potential biomarkers is of methodological nature and not unambiguous identification of biomarkers of phospholipidos. If further, more extensive data preprocessing approaches as various normalization procedures and specific studies of variables present in both control samples and dosed samples other potential biomarkers could be singled out.

It might be possible to identify these important biomarkers by LC/MS/MS and NMR. However, this was beyond the scope of this work. After determination of the biomarkers it is possible to optimize a method for the substances, by selective ion monitoring (SIM) or multiple reaction monitoring (MRM), to increase the sensitivity. The analysis time might also be shortened. A more reliable result, in such studies, is probably obtained if several biomarkers are used. It is proposed that SIM of, say, five Thomson numbers could be used in the same run, thus substantially increasing the sensitivity of the method.

CONCLUSIONS

The method developed using LC/MS and two-way analysis was found to be suitable for metabolomic applications. In comparison

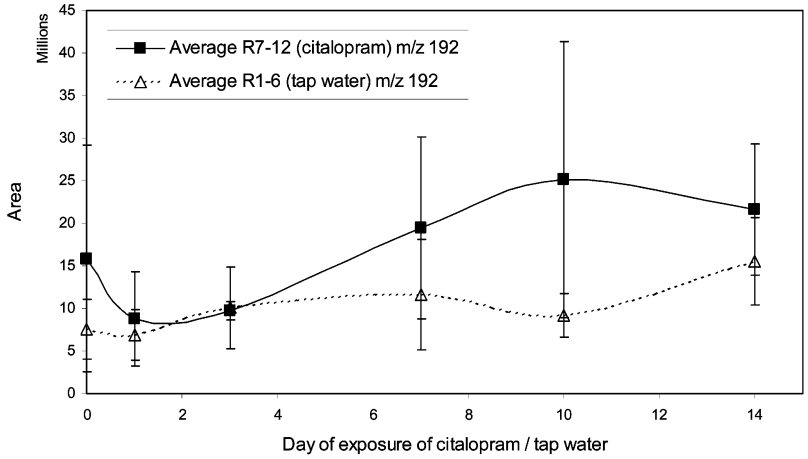

Figure 10. Reconstructed ion chromatograms (RIC) of a previously proposed biomarker, phenylacetylglycine (variable no. 20, *m*/*z* 192, tR 7.3 min), were obtained for rats #1−6 (tap water) and rats #7−12 (citalopram) on different days. No significant differences between control and dosed samples in the concentration of PAG were found. According to this study, PAG is therefore of limited value as a biomarker for phospholipidos.

to metabonomic NMR the data generated by this approach are very easy to interpret, and the greater part of the data are resolved into uniquely defined components. By means of two-way analysis the identity of the obtained peaks are ensured, and the peak shifts are compensated for.

It was possible to differentiate between control and dosed samples using obtained LC/MS fingerprints. Using this method, it is possible to detect potential biomarkers that are present at low concentrations in a complex matrix, for example, urine. It is also possible to study a *set* of substances that can be used as a marker for the observed phenomenon.

ACKNOWLEDGMENT

The authors would like to thank Pattern Recognition systems in Bergen, Norway for providing the software of MS-Resolver and for good support and assistance. The authors are also grateful for the assistance of Micromass Sweden for providing the LC/MS instrumentation.

Received for review February 18, 2003. Accepted July 10, 2003.

AC0341618