

# Tools in Metabonomics: An Integrated Validation Approach for LC-MS Metabolic Profiling of Mercapturic Acids in Human Urine

Silvia Wagner,<sup>†</sup> Karoline Scholz,<sup>†</sup> Maximilian Sieber,<sup>†</sup> Marco Kellert,<sup>†</sup> and Wolfgang Voelkel<sup>†,\*,‡</sup>

Department of Toxicology, University of Wuerzburg, Versbacher Strasse 9, Wuerzburg, Germany, and Environmental Medicine/Biomonitoring, Bavarian Health and Food Safety Authority, Pfarrstrasse 3, Munich, Germany

While for <sup>1</sup>H NMR techniques there already exist common analytical and reporting standards, this does not apply to LC-MS metabolic profiling approaches. These standards are the more recommended when applying metabonomics to human biofluids, particularly urine samples, due to the high degree of biological variation compared to animals. A control study was performed, and urine samples of 30 healthy male and female human subjects were collected at intervals of 8 h twice a day for three consecutive days. Using selective multiple reaction monitoring in combination with a column-switching tool for the analysis of the mercapturate pattern, samples were screened for time and gender differences, the most common confounders. Data preprocessing parameters, alignment, scaling to internal standards, and normalization techniques were optimized by PCA, PLS-DA, and OPLS models. Great care was taken in the validation process of both analytical and chemometric protocols. Additionally, a problem of LC-MS, the combination of “different-batch” data to “one-batch” data could be solved by a batchwise scaling procedure. Based on these results, the use of metabolic profiling via mercapturates will be feasible for the detection of disease or toxicity markers in the future since mercapturates are important biomarkers of reactive metabolites known to be involved in many toxic processes.

Since the ultimate introduction of the concept by Nicholson et al.,<sup>1</sup> metabonomics and metabolomics evolved into an emerging field with an explosion of publications in the past few years.<sup>2</sup> For the “classical” <sup>1</sup>H NMR metabonomics approach, common analytical and reporting standards together with appropriate databases have been progressively established.<sup>3,4</sup> However, a major problem of NMR metabolite screening is the poor sensitivity. To overcome this issue, sensitive LC-MS techniques are increasingly applied,

most notably for the screening of human biofluid samples. In contrast to dealing with NMR spectra, chromatographic “peaks” resulting from LC-MS analyses require a completely different data acquisition, handling, and evaluation protocol. To date, only a few approaches exist concerning integrated analytical and chemometric validation procedures for (large-scale) LC-MS metabonomic studies. This is alarming, as each individual part of a metabonomic study has its own pitfalls, and even minor variations or errors in one of the multiple steps can lead to misinterpretation of the results and collapse of the whole study.

Human metabonomic studies are very difficult to perform because of the wide variability and flexibility of the human organism. Owing to confounders like gender, age, time of day, health state, lifestyle, diet, and varying phenotypes, dealing with terms like “normality”, “control group”, or “abnormality/disease” is not as easy as in animal studies.<sup>5–7</sup> But the aim of human metabonomics, the early detection of risk factors, and thus the prediction of disease or other harmful effects<sup>8</sup> as well as personalized medicine will only succeed if it is possible to gain a good understanding of natural biological variation, i.e., “normal” biofluid metabolite pattern. Otherwise, a decision whether a person is suspected of developing/having a certain disease is not possible. In most of the cases the subjects under investigation cannot serve as their own control group, and thus higher numbers of participants are required for the studies.<sup>9</sup> Recently it was shown that the ratio between the number of variables and the number of samples is a very important factor for validation results and that error rates increased if the number of samples decreased.<sup>10</sup>

In contrast to plasma samples that show little variability under normal physiological conditions,<sup>11</sup> urine samples require a normalization step for comparison due to intersubject dilution factors

\* To whom correspondence should be addressed. Tel: +49 (0)89/2184-248. Fax: +49 (0)89/2184-297. E-mail: wolfgang.voelkel@lgl.bayern.de.

<sup>†</sup> University of Wuerzburg.

<sup>‡</sup> Bavarian Health and Food Safety Authority.

(1) Nicholson, J. K.; Lindon, J. C.; Holmes, E. *Xenobiotica* **1999**, *29*, 1181–1189.

(2) Rochfort, S. J. *Nat. Prod.* **2005**, *68*, 1813–1820.

(3) Lindon, J. C.; Keun, H. C.; Ebbels, T. M.; Pearce, J. M.; Holmes, E.; Nicholson, J. K. *Pharmacogenomics* **2005**, *6*, 691–699.

(4) Fiehn, O.; Kristal, B.; Van Ommen, B.; Sumner, L. W.; Sansone, S.-A.; Taylor, C.; Hardy, N.; Kaddurah-Daouk, R. *OMICS* **2006**, *10*, 158–163.

(5) Van der Greef, J.; Smilde, A. K. J. *Chemom.* **2006**, *19*, 376–386.

(6) Kochhar, S.; Jacobs, D. M.; Ramadan, Z.; Berruex, F.; Fuerholz, A.; Fay, L. B. *Anal. Biochem.* **2006**, *352*, 274–281.

(7) 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.

(8) Clayton, T. A.; Lindon, J. C.; Cloarec, O.; Antti, H.; Charuel, C.; Hanton, G.; Provost, J.-P.; Le Net, J.-L.; Baker, D.; Walley, R. J.; Everett, J. R.; Nicholson, J. K. *Nature* **2006**, *440*, 1073–1077.

(9) Hart, B. A.; Vogels, J. T. W. E.; Spijksma, G.; Brok, H. P. M.; Polman, C.; van der Greef, J. J. *Neurol. Sci.* **2003**, *212*, 21–30.

(10) Rubingh, C. M.; Bijlsma, S.; Derks, E. P. P. A.; Bobeldijk, I.; Verheij, E. R.; Kochhar, S.; Smilde, A. K. *Metabolomics* **2006**, *2*, 53–61.

(11) Lenz, E. M.; Bright, J.; Wilson, I. D.; Morgan, S. R.; Nash, A. F. J. *Pharm. Biomed. Anal.* **2003**, *33*, 1103–1115.

up to 2 orders of magnitude. Normalization to the total sum of all signal intensities is routine in NMR metabonomics<sup>12,13</sup> but is also applied in LC-MS approaches.<sup>14</sup> Creatinine normalization as well as normalization to total urine volume are common strategies in conventional quantitative applications but are not widely established in metabonomics.<sup>15,16</sup> The pros and cons of the different normalization approaches are discussed in detail in the literature.<sup>17,18</sup>

The complexity of the urine samples renders analysis difficult, and ion suppression is a drawback of LC-MS approaches, especially when using electrospray ionization technique in combination with survey full scans. Possibilities to overcome this problem are the extraction of the metabolites prior to analysis,<sup>19,20</sup> the use of specific column packing material,<sup>21</sup> or the application of UPLC.<sup>14,22</sup> In the present study, a completely different approach was used, namely on-line sample enrichment and cleanup with a column-switching unit. Through the application of this technique, a decrease in ion suppression effects and thus a significant increase in the signal-to-noise ratio can be gained, resulting in more signals per sample. Since all sample cleanup is done on-line, this system is very time-saving, suitable for high-throughput, and therefore widely used for analytical purposes<sup>23</sup> but has never been applied in the field of metabolic fingerprinting or profiling of biofluids with hundreds or more unknown metabolites. In addition to ion suppression effects, variations in urinary pH values can also cause problems, since the structure of a molecule and thus its ability to be retained by the column is dependent on the milieu. As a consequence, variations in signal intensity and retention time can occur, most notably when using undiluted urine samples. The use of appropriate internal standards can correct these matrix effects, but in contrast to classical quantitative analyses, correction is not yet common practice in fingerprinting approaches.

A crucial if not the all-dominant step in LC-MS data preprocessing is the deconvolution of the chromatograms, i.e., the extraction of the peak variables from the analytical runs. There are several software solutions available, commercial and open source. Sometimes two or even more of them are applied in combination, often using additional "in-house" software.<sup>24,25</sup> Al-

though based on different algorithms, the principal operations for the analyst are common for all software tools. Parameters for peak width, shape, intensity, resolution, etc., have to be set for the peak-picking algorithm to find the variables using predefined tolerances for mass and retention time shifts, baselines, noise structure, etc. To date, only a few publications exist describing these extraction parameters, and even less papers are reporting the optimization procedure. This way, results become uncomparable and the establishment of the long-needed metabolite and LC-MS model databases will fail.

An additional problem occurring when performing large-scale LC-MS metabonomic studies is the so-called "batch-to-batch" variation. It is very challenging to compare data that were analyzed in different batches (different laboratories, time points, mass spectrometers, columns, sample containers, storage procedure, etc.) because of the susceptibility of the multivariate approach to every single variable in the sample.<sup>26,27</sup> As a consequence, great care must be taken in the fusion process of "different-batch" data to "one-batch" data. One way to evaluate differences and shifts between batches is the use of quality control samples over the whole time scale of the analyses, but this has not yet been investigated for the combination of multivariate datasets.<sup>28</sup>

Here an integrated validation approach for human urine LC-MS metabonomic data based on a nonrestricted control study is presented. The aim was to carefully check the analytical and chemometric pitfalls mentioned before and to optimize parameters and settings wherever possible. On the basis of a mercapturic acid-specific multiple reaction monitoring scan<sup>29</sup> in combination with a column-switching tool, a screening for gender and diurnal pattern was carried out. Thus, an overview of normal biological variation in urine patterns of healthy unexposed human subjects was obtained. In the context of oxidative stress and the onset of certain diseases, mercapturic acids are of special interest since reactive compounds from various endogenous and exogenous sources are excreted this way. This class of metabolites, known effect markers, is therefore considered appropriate for the evaluation of the electrophilic burden of an organism as previously described.<sup>30</sup>

For all deconvolution and preprocessing steps of the chromatographic data, only MarkerView software was used. Peak picking, filtering, and alignment parameters within the software were optimized as far as possible. Appropriate scaling and normalization techniques were investigated and monitored by multivariate data analysis tools like PCA, PLS-DA, and OPLS. Quality control samples, together with internal standards, were used as an evaluation tool to check the quality of both raw and processed data.

- (12) Solanky, K. S.; Bailey, N. J.; Beckwith-Hall, B. M.; Bingham, S.; Davis, A.; Holmes, E.; Nicholson, J. K.; Cassidy, A. J. *Nutr. Biochem.* **2005**, *16*, 236–244.
- (13) Brindle, J. T.; Antti, H.; Holmes, E.; Tranter, G.; Nicholson, J. K.; Bethell, H. W.; Clarke, S.; Schofield, P. M.; McKilligin, E.; Mosedale, D. E.; Grainger, D. J. *Nat. Med.* **2002**, *8*, 1439–1444.
- (14) 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.
- (15) Lu, G.; Wang, J.; Zhao, X.; Kong, H.; Xu, G. *Chin. J. Chromatogr.* **2006**, *24*, 109–113.
- (16) Lutz, U.; Lutz, R. W.; Lutz, W. K. *Anal. Chem.* **2006**, *78*, 4564–4571.
- (17) Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. *Anal. Chem.* **2006**, *78*, 4281–4290.
- (18) Craig, A.; Cloarec, O.; Holmes, E.; Nicholson, J. K.; Lindon, J. C. *Anal. Chem.* **2006**, *78*, 2262–2267.
- (19) Yang, J.; Xu, G.; Zheng, Y.; Kong, H.; Pang, T.; Lv, S.; Yang, Q. *J. Chromatogr. B* **2004**, *813*, 59–65.
- (20) Wang, C.; Kong, H.; Guan, Y.; Yang, J.; Gu, J.; Yang, S.; Xu, G. *Anal. Chem.* **2005**, *77*, 4108–4116.
- (21) Idborg, H.; Zamani, L.; Edlund, P. O.; Schuppe-Koistinen, I.; Jacobsson, S. *P. J. Chromatogr. B* **2005**, *828*, 9–13.
- (22) Nordstroem, A.; O'Maille, G.; Qin, C.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 3289–3295.
- (23) Zell, M.; Husser, C.; Erdin, R.; Hopfgartner, G. *J. Chromatogr. B* **1997**, *694*, 135–143.

- (24) 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.
- (25) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779–787.
- (26) Teahan, O.; Gamble, S.; Holmes, E.; Waxman, J.; Nicholson, J. K.; Bevan, C.; Keun, H. C. *Anal. Chem.* **2006**, *78*, 4307–4318.
- (27) Tate, A. R.; Damment, S. J.; Lindon, J. C. *Anal. Biochem.* **2001**, *291*, 17–26.
- (28) Sangster, T.; Major, H.; Plumb, R.; Wilson, A. J.; Wilson, I. D. *Analyst* **2006**, *131*, 1075–1078.
- (29) Scholz, K.; Dekant, W.; Völkel, W.; Pöhler, A. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1976–1984.
- (30) Wagner, S.; Scholz, K.; Donegan, M.; Burton, L.; Wingate, J.; Voelkel, W. *Anal. Chem.* **2006**, *78*, 1296–1305.

Through application of our approach to “different-batch” data, the possibility of combining three separate data matrices to a new “one-batch” data set was tested and validated. The focus of the present study was on the optimization and validation procedure rather than on the identification of markers. The biological interpretation of the data is under investigation and will be published elsewhere.

## EXPERIMENTAL SECTION

**Chemicals.** All solvents were HPLC grade and purchased from Roth (Karlsruhe, Germany). Formic acid (98–100%) and hydrochloric acid (25%) were of analytical grade and obtained from Merck (Darmstadt, Germany). The internal standard *S*-phenylmercapturic acid was purchased from Toronto Research Chemicals (TRC, North York, Canada), and 4-*tert*-butylbenzyl bromide and *N*-acetylcysteine are from Sigma-Aldrich (Taufkirchen, Germany).

**Synthesis of Acrylamide-*d*<sub>3</sub> Mercapturic Acid.** Acrylamide-*d*<sub>3</sub> mercapturic acid was synthesized and characterized following the protocol recently described.<sup>31</sup>

**Synthesis of *S*-(4-*tert*-Butylbenzyl)mercapturic Acid.** *S*-(4-*tert*-Butylbenzyl)mercapturic acid was synthesized by the reaction of 4-*tert*-butylbenzyl bromide with *N*-acetylcysteine as previously described.<sup>32</sup>

**Human Study.** Urine samples of 10 healthy male and 20 healthy female human subjects (age 21–62) were collected at intervals of 8 h twice a day (“overnight” 11 pm to 7 am and “during the day” 7 am to 3 pm) for three consecutive days, resulting in a set of 180 urine samples. Total urinary volume was determined for each collecting period, and aliquots were stored at –20 °C until analysis. A questionnaire was used to query the participants for lifestyle activities and dietary habits (including drug intake) during the study. All subjects enlisted in the study did not abuse alcohol and were either nonsmokers or only occasional smokers. The study was carried out according to the Declaration of Helsinki, after approval by the Regional Ethical Committee of the University of Wuerzburg, Germany, and after written informed consent by the subjects.

**Internal Standard Mix.** An aqueous internal standard mix was prepared containing acrylamide-*d*<sub>3</sub> mercapturic acid, *S*-phenylmercapturic acid, and *S*-(4-*tert*-butylbenzyl)mercapturic acid at concentrations of 100 mg/L, 10 mg/L, and 1 mg/L, respectively.

**Sample Preparation of Human Urine.** Urine aliquots were allowed to thaw at room temperature. After centrifugation at 4 °C and 14 000*g* for 10 min, 1.0 mL of urine was directly transferred into an autosampler vial. A 5 µL amount of hydrochloric acid (25%) and 10 µL of the internal standard mix were added, vortex-mixed, and directly used for LC-MS/MS analysis. For validation purposes, a randomly chosen urine sample was used. This sample was independently prepared 22 times and used as a quality control (QC) sample.

Two 400 µL injections from the same vial were made: the first one for the “one-batch” measurement, the second one for the “different-batches” analysis. All samples were analyzed in random order. Between the two measurements the specimens were stored

at –20 °C. Prior to the second analysis, all samples were thawed at room temperature and centrifuged at 4 °C and 14 000*g* for 10 min.

**Creatinine Analysis.** Creatinine analysis was carried out by the laboratory of the university hospital of Wuerzburg using an enzymatic procedure<sup>33</sup> on a Roche Diagnostics device (COBAS INTEGRA).

**Chromatography.** For LC-MS/MS analysis, a column-switching unit was used, consisting of an autosampler and two solvent pumps, binary and quaternary (all Agilent Series 1100, Waldbronn, Germany), controlled by an electrical valve. In the loading position the binary pump carried 400 µL of the sample at a flow rate of 0.750 mL/min 0.1% formic acid from the autosampler vial onto the trap column (ReproSil-Pur C18-AQ 5 µm, 33 × 3 mm, Dr. Maisch, Ammerbuch, Germany). After 2.0 min, sample loading and washing steps were completed and the valve switched to the elution position. Then the quaternary pump back-flushed the trapped analytes from the trap column onto the analytical column (ReproSil-Pur C18-AQ 3 µm, 150 × 2 mm, Dr. Maisch) using a linear gradient at a flow rate of 0.200 mL/min. Elution of the urinary metabolites was carried out at 95% solvent A (0.1% formic acid)/5% solvent B (acetonitrile) and held isocratically for 2 min followed by a linear gradient of 5–50% B in 23 min and further to 90% B in 2 min. These conditions were held for 2 min prior to switching to the starting conditions in 2 min and held for 9 min to reequilibrate the column. During the elution period, the trap column was washed with 90% acetonitrile and equilibrated within 13 min to the initial conditions of 100% of 0.1% formic acid.

**Mass Spectrometry.** A linear ion trap (QTRAP 2000, Applied Biosystems/MDS Sciex, Concord, Canada) with a TurboIonSpray (ESI) source operating in the negative ion mode was used for MS/MS experiments. Survey scans were carried out in a “constant neutral loss (CNL) like” multiple reaction monitoring (MRM) mode and recorded by Analyst 1.4.1 software. For this purpose, a method with all transitions *m/z* → *m/z*–129 amu in the range of 200–450 amu were generated, i.e., 200 → 71, 201 → 72, 202 → 73... 449 → 320, 450 → 321, resulting in 251 mercapturic acid specific mass transitions. For the acrylamide-*d*<sub>3</sub> mercapturic acid standard the transition 236 → 104 was monitored because of the CNL of 132 amu specific for mercapturic acids labeled at the *N*-acetyl moiety of the molecule. Dwell time was set at 5 ms for each transition with a mass range pause of 5 ms.

Source voltage was set at –4.2 kV and vaporizer temperature at 400 °C. The mass spectrometer operated with gas settings of 50 psi for turbo gas, 45 psi for nebulizer gas, 30 psi for curtain gas, and 10 psi for collision gas. Entrance potential (EP), cell exit potential (CXP), declustering potential (DP), and collision energy (CE) were set at –10, –2, –50, and –20 V, respectively.

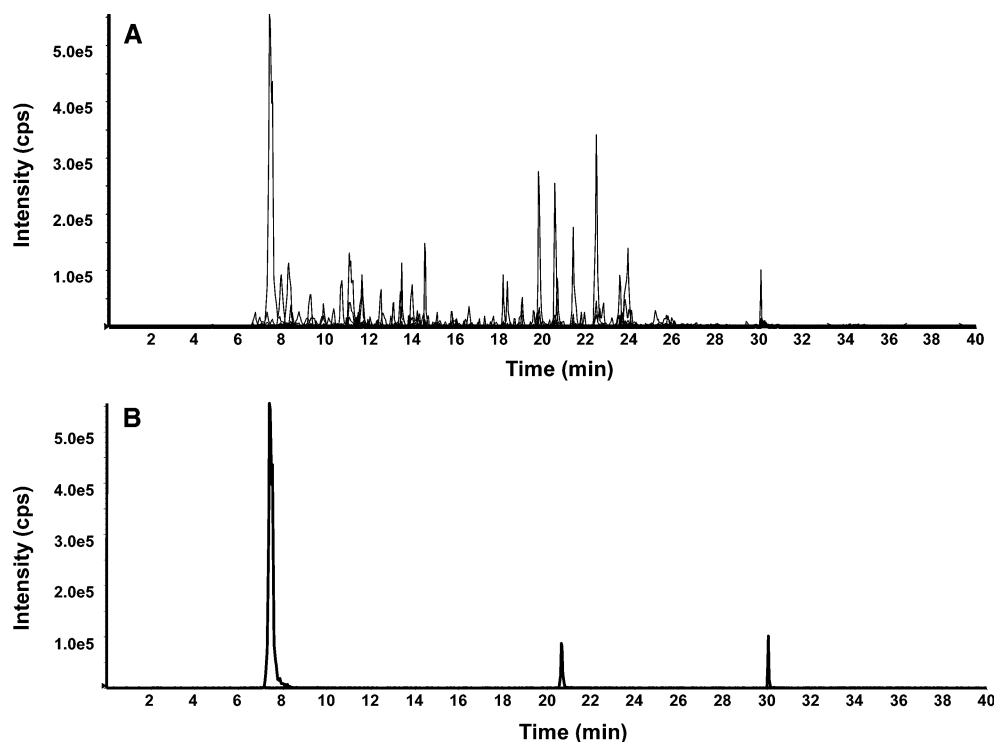
**Data Preprocessing and Multivariate Data Analysis.** Peak finding, filtering, and alignment as well as scaling (to internal standards) and normalization (to creatinine or total urine volume) were carried out using MarkerView software 1.2.0.0 (Applied Biosystems/MDS Sciex) capable of processing MRM data. Unless otherwise stated, default preprocessing parameters were as following: smoothing half-width 1 point, baseline subtraction window 1.0 min, noise percentage 50%, peak-splitting factor 4,

(31) Kellert, M.; Scholz, K.; Wagner, S.; Dekant, W.; Volkel, W. *J. Chromatogr. A* **2006**, *1131*, 58–66.

(32) Dekant, W.; Metzler, M.; Henschler, D. *J. Biochem. Toxicol.* **1986**, *1*, 57–72.

(33) Junge, W.; Wilke, B.; Halabi, A.; Klein, G. *Clin. Chim. Acta* **2004**, *344*, 137–148.





**Figure 1.** (a) Total ion chromatogram (TIC) of 252 mercapturic acid-specific MRM transitions of an “am” urine sample (female) of study day 3. (b) Extracted ion chromatogram (XIC) of the three internal standards: acrylamide-*d*<sub>3</sub>-MA at 7.4 min, *S*-phenyl-MA at 20.7 min, and 4-*tert*-butylbenzyl-MA at 30.1 min.

minimum required intensity 500, minimum peak width 3 points, minimum signal-to-noise 5.0, maximum number of peaks 500, retention time tolerance 1.0 min. Linear retention time (RT) correction and scaling of the samples was performed using the internal mercapturic acid (MA) standard variables ( $m/z$   $m/z$ -132\_RT and  $m/z$   $m/z$ -129\_RT, respectively).

Chemometric analyses including PCA, PLS-DA, and OPLS were performed with SIMCA-P software 11.0 (Umetrics, Umeå, Sweden). All data were autoscaled (mean-centered and scaled to unit variance). Significant components were calculated using 7-fold cross-validation, the default settings within the SIMCA-P software. The  $Q^2$  value denotes the quality of the resulting models. Preprocessing that could not be done in MarkerView software (multiplication of scale factors with normalization factors) was performed with Excel 2002 (Microsoft Germany, Unterschleisheim, Germany).

## RESULTS AND DISCUSSION

### Theoretical Multiple Reaction Monitoring Survey Scan.

Recently, a new approach for metabolic profiling of human urine samples was introduced using a mercapturic acid-specific constant neutral loss (CNL) mode.<sup>30</sup> The problem with CNL data together with the even more complicated full scan data is that it is continuous and therefore hard to interpret. Selection and optimization of the peak extraction parameters was very difficult to perform, and often correct peak alignment failed completely with continuous data. Therefore the continuous CNL survey scan was converted into a discrete theoretical multiple reaction monitoring scan (thMRM),<sup>29</sup> i.e. every CNL of 129 amu was replaced with a MRM transition in the form of  $m/z \rightarrow m/z - 129$ . A characteristic LC-ESI-MS/MS total ion chromatogram (TIC of 252 thMRMs) is

shown in Figure 1A. In Figure 1B the transitions of the three internal standards are extracted (acrylamide-*d*<sub>3</sub>-MA 236  $\rightarrow$  104 at 7.5 min, *S*-phenyl-MA 238  $\rightarrow$  109 at 20.7 min and 4-*tert*-butylbenzyl-MA 308  $\rightarrow$  179 at 30.1 min). The concentrations of the internal standards were carefully adjusted to the mean intensities of the urinary mercapturic acids in the same region of the chromatograms.

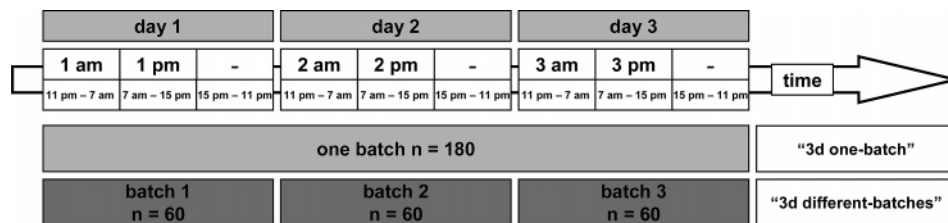
Acrylamide-MA is a well-known metabolite of acrylamide that is formed in the baking and roasting process of food stuff like French fries, crispbread, or coffee beans.<sup>34</sup> The labeled acrylamide-MA was taken as internal standard for two reasons: the ubiquitous occurrence of acrylamide and its high polarity marking the beginning of the chromatograms. *S*-Phenyl-MA is formed by bioactivation of benzene, a toxic organic solvent and component of tobacco smoke.<sup>35</sup> In this study only non-smokers and occasional smokers participated. Even in the urines of the occasional smokers, no *S*-phenyl-MA could be detected, and thus it was taken as internal standard for the center region of the chromatograms. 4-*tert*-Butylbenzyl-MA has neither an endogenous nor an exogenous origin. It was synthesized to have an appropriate internal standard for the nonpolar region of the chromatograms.

**LC-MS Analysis with a Column-Switching Unit.** Only a short overview of the mode of operation of the column-switching system is given here (for details see ref 36). The first step is the sample loading, i.e., analytes are trapped and accumulated on the trap column, followed by washing the trap column in order to

(34) Fennell, T. R.; Sumner, S. C.; Snyder, R. W.; Burgess, J.; Friedman, M. A. *Toxicol. Sci.* **2006**, *93*, 256–267.

(35) Liao, P. C.; Li, C. M.; Lin, L. C.; Hung, C. W.; Shih, T. S. *J. Anal. Toxicol.* **2002**, *26*, 205–210.

(36) Brink, A.; Lutz, U.; Voelkel, W.; Lutz, W. K. *J. Chromatogr. B* **2006**, *830*, 255–261.



**Figure 2.** Experimental flowchart: scheme of study and analytical design.

remove matrix compounds (salts). With the switching of the valve, flow direction changes and the enriched analytes are back-flushed onto the analytical column and eluted by a linear gradient. Switching the valve from the loading to the eluting position after 2.0 min resulted in highest intensities and peak numbers. The same was true for an injection volume of 400  $\mu\text{L}$  of human urine. Compared to conventional screening methods using injection volumes of 5–20  $\mu\text{L}$ , this was a 20–80-fold increase resulting in significantly higher signal-to-noise ratios, total intensities, and thus more peaks per sample. As monitored by the internal standards, a wide range from polar to nonpolar mercapturic acids could be trapped this way, showing the performance of the new approach. Column-switching units are commonly used for the measurement of single analytes, e.g., drug metabolites. However, to our knowledge, this was the first time that a column-switching tool was successfully used for an untargeted screening method.

Two principles could be united using this technique: an on-line extraction of the mercapturic acid metabolites together with a significant reduction of ion suppression effects. As a consequence, the noise level was reduced and the detection of the metabolites improved, altogether important prerequisites for the subsequent deconvolution steps and multivariate data analysis.

**Design of Experiments.** The human study was designed as a control study with no interventions, i.e., all participating subjects were healthy, and no restrictions were made concerning food intake or life style activities. To check for gender and diurnal patterns, urine samples of 10 male and 20 female human subjects were collected twice a day on three consecutive days (3d). The overnight collection period was denoted as “am”, the period during the day as “pm” time point. The resulting 180 samples of the human study together with 22 quality control samples (QC) were analyzed following two different protocols.

The first one comprised the measurement of the 202 specimens in one single batch (“3d one-batch” data set) whereas the second protocol consisted of three different batches, one for each of the 3 days (“3d different-batches” data set). The three different batches were analyzed independently within a time scale of 14 days. The experimental flowchart is shown in Figure 2. Other samples were run on the same mass spectrometer in between. Prior to the measurements, the instrument was routinely cleaned. The aim of our study and the different analysis protocols was the following: (i) to optimize the parameters for peak picking and filtering using a specific thMRM screening method; (ii) to check appropriate alignment, scaling and normalization techniques for human biofluid data sets; (iii) to evaluate biological variation in urine patterns of healthy unexposed human subjects such as diurnal and gender differences; (iv) to verify the possibility of combining data sets that were analyzed in different batches and to compare them to one-batch data.

**Optimization of Peak Picking and Filtering.** In order to facilitate the optimization process only QC samples of the “one-batch” analysis were taken into consideration. The first QC sample, the start of the batch, was discarded because of untypical chromatographic behavior. The optimization was then carried out on the remaining 21 QC samples. Most of the MarkerView peak picking parameters were chosen intuitively by simple inspection of a typical MRM chromatogram. The average peak width at the baseline was in the range of 0.3–0.8 min; thus, the baseline subtraction window was set to 1.0 min. Within this total window a baseline was calculated by connecting the minimum intensity value on the left and right side of the peak and subtracting this noise level from the data. The peak-splitting factor defines whether a peak cluster is split into different peaks or not. A factor of, for example, 4 means that there have to be at least 4 consecutive data points from a local minimum in the cluster to the next maximum. By inspection of different peak clusters, this value was considered as appropriate and taken as default. The minimum required intensity (after baseline subtraction) was set to 500 cps, the minimum signal-to-noise level to 5.0, respectively, due to the fact that these are the minimum standard values for quantitative analyses. Since the 4-*tert*-butylbenzyl-MA standard showed a very low peak width, minimum baseline peak width was set to 3 points. For smoothing half-width and noise percentage, the software default values of 1 point (slight smoothing) and 50% were taken. Altogether, changing these parameters (only one or different combinations of them) resulted in very similar data matrices. In contrast, the setting of the retention time ( $t_R$ ) tolerance for peak alignment and the maximum number of peaks was crucial for the quality of the output. Although the  $t_R$  of both standards and endogenous metabolites generally varied less than 0.5 min over the 201 measurements, setting this parameter to 0.5 min resulted in incorrect peak alignment. By manual inspection of the matrix, it was noticed that especially the  $d_3$ -acrylamide-MA was spread over several rows with very similar retention times, i.e., the same internal standard peak was registered as different peaks (variables) in some samples. Such misalignments have to be eliminated or at least minimized as far as possible; otherwise the biomarker discovery process will fail. With a  $t_R$  tolerance of 1.0 min, this problem could be overcome. Also using a  $t_R$  tolerance of 1.0 min instead of 0.5 min reduced the number of zeros from 29% to 24%. This effect was also noticed when reducing the maximum number of peaks stepwise from 10 000 to 250. The number of zeros decreased from 63% to 11%. Though multivariate data analysis tools can deal with up to 50% zeros or missing values in the data matrix,<sup>37</sup> it is recommended to limit zeros as far as possible, since

(37) Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Wold, S. *Multi- and Megavariate Data Analysis: Principles and Applications*; Umetrics Academy: Umeå, Sweden, 2001.

a high number of zeros in the data matrix can contribute to chance variation and thus false clustering results. The maximum peak number was therefore set to 500 peaks (24% zeros), allowing two peaks per mass transition (252 MRMs) to be put into the peak list. Following this approach, of course, only the major signals were registered and minor signals were lost. However, manual inspection of the matrix, and especially checking the correct alignment of the internal standard variables, revealed that this was the best way to map the raw data. With an increasing number of variables, and thus zeros, the quality of the processed data decreased dramatically, i.e., variables were increasingly misaligned into the data matrix. Correct peak alignment is the prerequisite for all subsequent multivariate data analysis tools and has to be examined with care. Alternative strategies for reducing both the number of variables and the occurrence of missing values for improvement of data analysis are known.<sup>24,38</sup> All parameter optimization procedures were rechecked with the “3d one-batch” data set to ensure that the observed trends are not only limited to the QC samples.

**Alignment, Scaling, and Normalization Techniques.** In contrast to <sup>1</sup>H NMR, LC-MS data acquisition shows higher sensitivity but less stability and reproducibility. It is therefore recommended to ascertain the quality of the analyses by using internal standards. Every sample was spiked with three internal standards (3ISTs) for evaluation of  $t_R$  shifts and variations in detector sensitivity. These variations are often observed in urine samples of different salt content and pH values, leading to more or less pronounced ion suppression. To overcome these effects, all samples were acidified (pH 1–2) and analysis was performed using a column-switching unit (see above). The following alignment and scaling procedures were tested: no  $t_R$  correction ( $t_{R\text{-corr}}$ )/no scaling to internal standards, only  $t_{R\text{-corr}}$ , only scaling to the three internal standards,  $t_{R\text{-corr}}$  together with scaling to 3ISTs. Using 3ISTs, linear regression  $t_{R\text{-corr}}$  was carried out by calculating a vector for the correction of intersample  $t_R$  shifts. Although there may be more sophisticated correction algorithms available in other software solutions, the use of linear regression retention time correction was adequate since the drifts were rather small in our analytical runs.

Scaling to 3ISTs was done by calculating an average scale factor from the response of all standards in the sample that was then applied to every single variable. A time window-wise scaling procedure may lead to even better results but is not integrated in the MarkerView software used in our study. Since the internal standards were evenly distributed over the complete time scale of the chromatographic run, averaging scale factors was supposed to be a good compromise. A drawback of urine samples is that volume can vary in a wide range, especially when sampling human spot urine. Normalization is then required in order to be able to compare the metabolite levels. In this study, no normalization, normalization to total urine volume, and normalization to creatinine content were checked. Normalization to a constant sum was not performed because of the fact that the screening procedure only comprised one class of metabolites, namely mercapturic acids, whose concentrations depend rather on the electrophilic burden of a subject than on the volume of the urine sample. For

optimization of the alignment, scaling, and normalization techniques, only day 2 (batch 2) of the “3d different-batches” data set was considered (30 am and pm samples each,  $n = 60$ ). Raw data were preprocessed as described above. Autoscaling was chosen to give equal weight to each of the variables. PCA and PLS-DA models were then generated in SIMCA-P after removal of the internal standard variables.

Two models were investigated, classification of gender (m\_f) and time point of urine sampling (am\_pm). These models served as evaluation tools for the optimization of data pretreatment. While, in general, the PCA models just showed slight trends in class separation, PLS-DA could easily discriminate male and female, respectively am and pm samples.

Twelve m\_f and am\_pm PLS-DA models were generated each, combining each alignment (no  $t_{R\text{-corr}}$ ,  $t_{R\text{-corr}}$ ), each scaling (no scaling, scaling to 3ISTs), and each normalization (no normalization, total volume normalization, creatinine normalization) option. Best models according to significant components and 7-fold cross-validation classification results were obtained, in both cases, using  $t_{R\text{-corr}}$  and 3IST scaling of the creatinine normalized data (data not shown). Therefore, these settings were taken as default for all consecutive models.

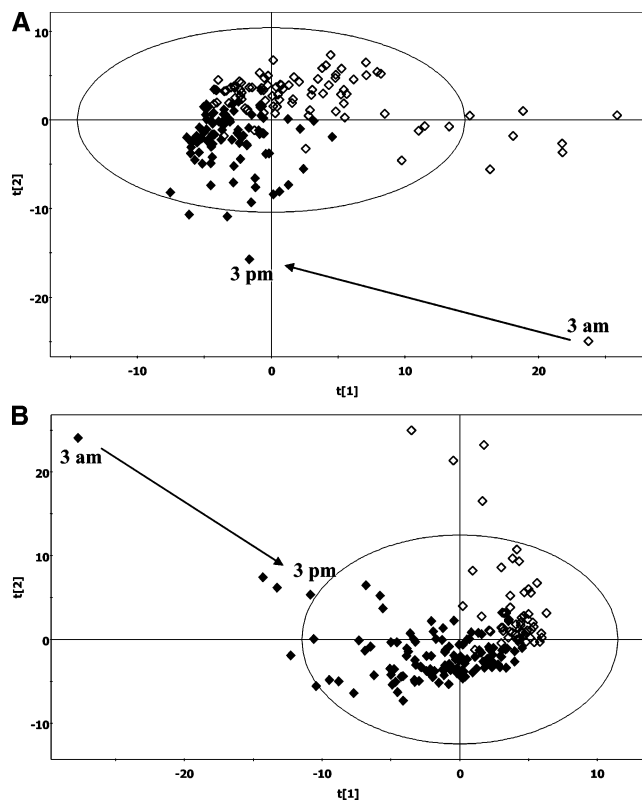
**Evaluation of Biological Variation in Urine Patterns.** These intermediate results suggested that using unsupervised chemometric tools like PCA for the investigation of normal human urine pattern will lead to poor clustering due to a high degree of biological variation. This phenomenon is known from the literature.<sup>15,39</sup> Compared to animal studies that are always carried out under defined conditions, no restrictions concerning diet or lifestyle activities were made in our study to get a better understanding of common human confounders. Dealing with this inherent diversity of human biofluid data sets will be the challenge of future metabonomic projects.

Gender (m\_f) and time of day (am\_pm) differences over the three consecutive days were investigated in the “3d one-batch” data set by generating PLS-DA and OPLS models. In an initial PCA, the QC samples were included and their behavior checked. QC specimens clustered very tightly together denoting stable chromatography and correct preprocessing parameters (data not shown). In order to avoid any impact of QC samples on the following models, a new data matrix was then generated that only contained samples of the study. Figure 3A shows the autoscaled PLS-DA model of am\_pm (day 1–3) classification after having removed the variables of the internal standards. For the discrimination of the two classes, a minimum of two components (PC1 vs PC2) was necessary. PC 1–3 were significant covering a cumulative (cum) variance of  $R^2X = 0.198$  and  $R^2Y = 0.687$  ( $Q^2(\text{cum}) = 0.444$ ), respectively. The “am” samples showed a wider scattering than the “pm” samples and also more outliers outside the hotelling  $T^2$  (0.95) ellipse. By visual inspection of the corresponding chromatographic runs and the data matrix, no inconsistencies could be detected for these outliers. Therefore statistical and biological rather than analytical or software reasons were responsible for their position.

A PLS-DA model of male and female subjects is depicted in Figure 3B.  $R^2X(\text{cum})$  and  $R^2Y(\text{cum})$  after six significant compo-

(38) Jonsson, P.; Bruce, S. J.; Moritz, T.; Trygg, J.; Sjoestroem, M.; Plumb, R.; Granger, J.; Maibaum, E.; Nicholson, J. K.; Holmes, E.; Antti, H. *Analyst* **2005**, *130*, 701–707.

(39) Lenz, E. M.; Bright, J.; Wilson, I. D.; Hughes, A.; Morrisson, J.; Lindberg, H.; Lockton, A. J. *Pharm. Biomed. Anal.* **2004**, *36*, 841–849.



**Figure 3.** PLS-DA scores plot of all 3 study days ( $n = 180$ ), “3d one-batch” data; (a) am (◇) and pm (◆) time point classification, PC 1–3 are significant; (b) male (◇) and female (◆) classification, PC 1–6 are significant. The female who had taken acetaminophen in the am-collecting period on day 3 is highlighted (3 am). The arrow marks the “kinetics” of the corresponding urine pattern.

nents were 0.292 and 0.903 ( $Q^2(\text{cum}) = 0.645$ ), respectively. There are less outliers compared to the am\_pm model. Scattering of the female samples was about twice the scattering of the males. This was not surprising due to the fact that the oestrus cycle is an enormous biological confounder having impact on the metabolic pattern.<sup>40</sup> In this study, only mercapturic acid patterns were considered, leading to the assumption that there are oestrus cycle-specific mercapturic acids excreted in the urine of female subjects. Identification of these markers and their corresponding pathways is under investigation.

By taking a closer look at the chromatogram of the female sample in the lower right corner of the am\_pm model, it was seen that she had taken acetaminophen in the am-collecting period on day three (3 am). Acetaminophen forms two mercapturic acids<sup>41</sup> that had a strong impact on the model as seen in the corresponding loading plots. Since this is a commonly used drug that falls in the category “confounder”, the sample was not excluded. The same sample was detected as an outlier in the m\_f model (upper left corner of the scores plot). In both models the 3 pm score of this female almost returned to the main cluster as shown by the arrows (Figure 3A and 3B). This example indicates that it is possible to filter out samples that behave differently and thus do not fit the control models. In the future, the study could serve as

a reference data base for testing unknown samples and comparing their properties to the control samples. From the corresponding loading plots, variables responsible for the differences could then be filtered out.

In order to check whether there are shifts in metabolite patterns over the 3 days, OPLS was carried out on the same data set. OPLS is a projection tool that filters the  $X$ -matrix and models only information that is due to the discrimination of the different groups, i.e., the correlation between the  $X$ - and the  $Y$ -matrix is improved.<sup>42</sup> The am\_pm and m\_f OPLS models were generated analogous to the PLS-DA plots. In OPLS, the group discrimination is forced to the first component, and thus classification results improved enormously as shown in Figure 4A and 4B. Of special interest was the possibility of investigating the variation that was removed, called orthogonal variation, in a separate model of orthogonal components. In this way, information that is “behind” a data set but is not modeled can be made visible. By plotting the first versus the second orthogonal component of am\_pm and m\_f OPLS models, no trend concerning the three different days of the study (Figure 4C and 4D) could be seen. The same was true for any combination of the significant orthogonal components (five orthogonal components for the time point model and three for the gender model, respectively). In addition, independent PLS-DA failed in discriminating the study days. These results indicated that control urine mercapturic acid patterns remained relatively stable over the three study days, a good prerequisite for future metabolomic studies.

#### Fusion of “Different-Batch” Data to “One-Batch” Data.

Up to now, fusion of LC-MS metabolomic data is very difficult to perform due to variation in chromatography, mass and intensity accuracy of the mass spectrometer, and many other factors that decrease reproducibility compared to  $^1\text{H}$  NMR techniques. Here, a new approach is presented for the combination of LC-MS “different-batch” data to a single “one-batch” data set, namely the “3d different-batches” data to a new “3d one-batch” set of data. The subsequent comparison of the original “3d one-batch” data set with the new “3d one-batch” data demonstrated the success of the approach.

As a first step, the 21 QC samples of the “3d different-batches” measurements were investigated. But even with  $t_R$ corr and scaling to the three internal standards a clear shift (day 1  $\rightarrow$  day 2  $\rightarrow$  day 3) was visible in the PCA plot (data not shown).

Therefore, a new approach was tested to remove this kind of shift: autoscaling of each day (batch) separately. Using this method leads to the calculation of a separate mean and standard deviation value for each day, i.e., day-specific detector fluctuations and differences in absolute metabolite levels are removed. After this batch-wise scaling procedure, the shift in the new PCA scores plot had completely disappeared and the QC samples showed normal distribution (data not shown).

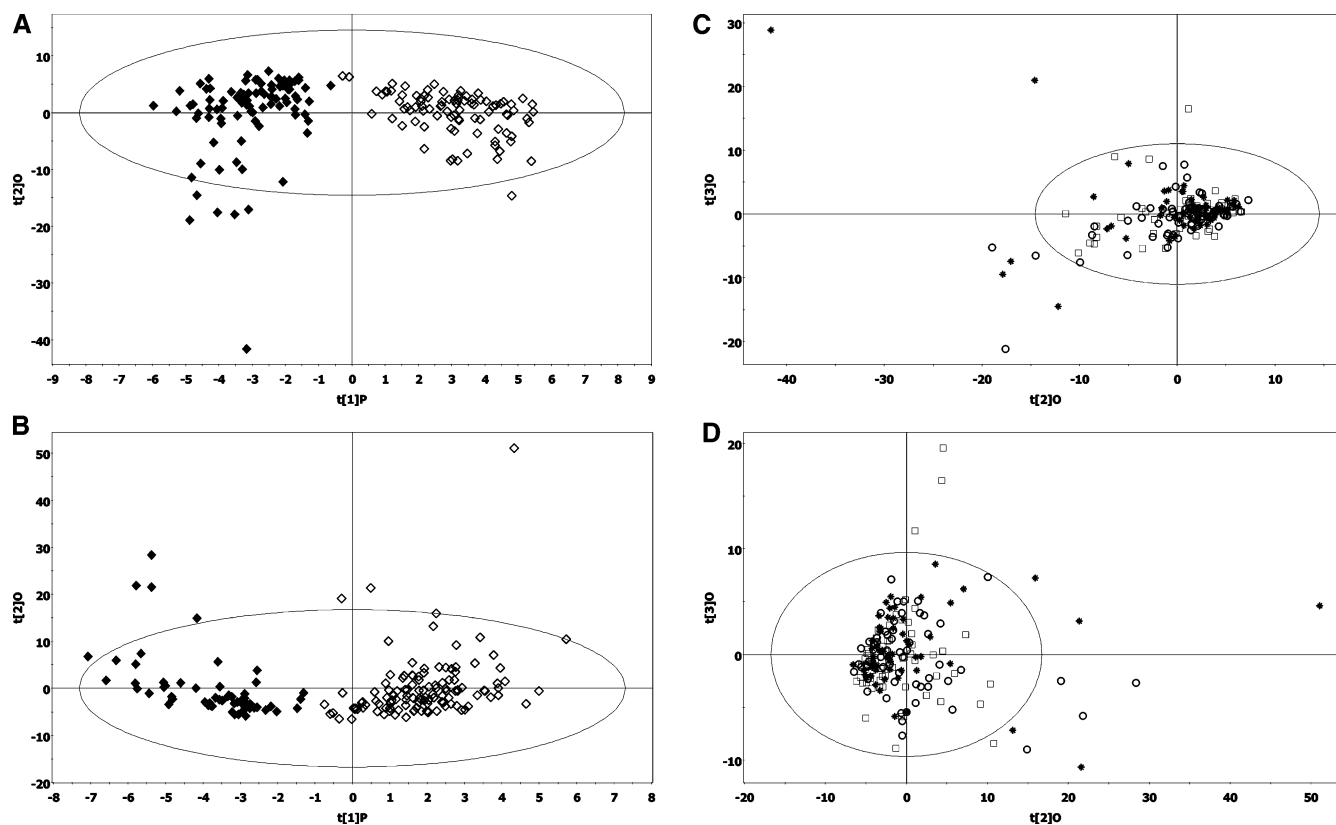
On the basis of these results, the same was done with the “3d different-batches” LC-MS data. To check the success of the approach, the following procedure was carried out: four sets of data were generated, “3d one-batch” with scaling all 3 days together, “3d one-batch” with scaling each day separately, “3d different-batches” with scaling all 3 days together, and “3d different-batches” with scaling each day separately. The set “3d

(40) Bollard, M. E.; Holmes, E.; Lindon, J. C.; Mitchell, S. C.; Branstetter, D.; Zhang, W.; Nicholson, J. K. *Anal. Biochem.* **2001**, *295*, 194–202.

(41) Andrews, R. S.; Bond, C. C.; Burnett, J.; Saunders, A.; Watson, K. J. *Int. Med. Res.* **1976**, *4*, 34–39.

(42) Trygg, J.; Wold, S. J. *Chemom.* **2002**, *16*, 119–128.





**Figure 4.** OPLS models of all 3 study days ( $n = 180$ ), “3d one-batch” data; (a) am (◆) and pm (◇) time point classification; first principal component PC1 ( $t[1]P$ ) vs the first orthogonal component ( $t[2]O$ ); (b) male (◆) and female (◇) classification;  $t[1]P$  vs  $t[2]O$ ; (c) first vs second orthogonal component ( $t[2]O$  vs  $t[3]O$ ) of time point classification, labeled by study day: day 1 (□), day 2 (○), day 3 (\*); (d) first vs second orthogonal component ( $t[2]O$  vs  $t[3]O$ ) of gender classification, labeled by study day: day 1 (□), day 2 (○), day 3 (\*).

**Table 1. PLS-DA Predictions Dependent on Analysis Protocol and Scaling Procedure**

	trainings set <sup>a</sup> (any 2 days)		test set <sup>b</sup> (remaining day)	
	am_pm	m_f	am_pm	m_f
“3d one-batch”				
scaling all 3 days together	95.8 ± 2.2	95.3 ± 2.9	83.9 ± 6.3	81.1 ± 2.6
scaling d1, d2, d3 separately	97.5 ± 0.8	97.2 ± 2.5	88.9 ± 7.5	83.9 ± 3.5
“3d different-batches”				
scaling all 3 days together	95.0 ± 2.2	96.7 ± 2.5	73.5 ± 12.8	81.1 ± 2.6
scaling d1, d2, d3 separately	94.2 ± 3.6	96.9 ± 1.7	83.4 ± 7.6	84.4 ± 4.2

<sup>a</sup> Training set: values are given as right classification results (mean ± SD in %) of three training operations. <sup>b</sup> Test set: values are given as right prediction results (mean ± SD in %) of three predictive operations.

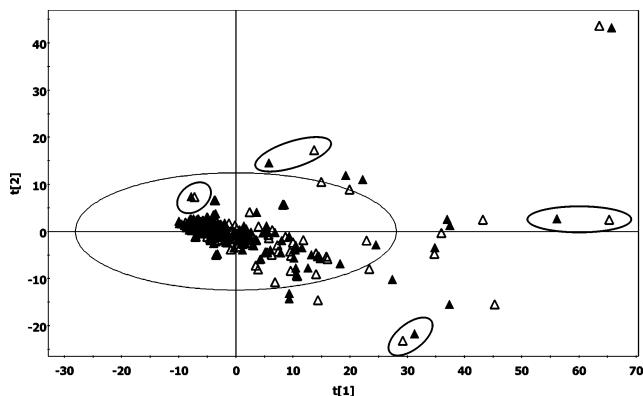
one-batch” with scaling each day separately was added to account for different metabolite levels on the three study days that are automatically removed in the “3d different-batches” set when scaling day-wise.

PLS-DA training and test sets for each of the four approaches were generated. For both am\_pm and m\_f discrimination, the training set consisted of any 2 days and the model was taken as a basis for the remaining third day that was predicted (test set). This resulted in three test sets for gender and time of day classification, respectively. The quality of the three training sets was expressed as the mean and SD of right classification results. The same was done with the test set predictions. Only significant components were allowed and accounted for the am\_pm and m\_f PLS-DA training and test set models. Table 1 summarizes the results. By first looking at the “3d one-batch” results, it can be

seen that scaling each day separately yielded the best models and predictions for training and test sets with 97.5% (am\_pm) and 97.2% (m\_f) and 88.9% (am\_pm) and 83.9% (m\_f) right classifications, respectively. This was not very surprising since the scaling procedure has put every day on an equal footing, and thus model and predictive ability improved compared with the overall scaling. The corresponding values for scaling over all 3 days were 95.8% and 95.3%, and 83.9% and 81.1%, respectively. This was only a slight, but remarkable, difference.

The same was true for the “3d different-batches” data. While providing similar training sets, in both cases (scaling over all 3 days and scaling separately) of 95.0% and 94.2% (for am\_pm) and 96.7% and 96.9% (for m\_f), respectively, predictive ability improved significantly when scaling each day separately. Thus, correct classification of am\_pm test set samples improved from 73.5% to





**Figure 5.** Combined PCA scores plot of the original ( $\blacktriangle$ ) and the new ( $\triangle$ ) “3d one-batch” data set. Each day was separately scaled in both sets prior to PCA. Discrimination of the two data sets by supervised PLS-DA was not possible.

83.4% and for m\_f samples from 81.1% to 84.4%, respectively. Regarding the relative standard deviations, it is remarkable that gender predictions showed in general less variation than time point predictions while having similar SDs for the trainings sets.

In summary, applying batch-wise, i.e., day-wise, scaling to the “3d different-batches” data resulted in a data set comparable to the original “3d one-batch” data concerning model and predictive capacity; hence, it was possible to generate a new “3d one-batch” data set out of the three different batches. This is a great step forward for the routine application of human biofluid metabolomics in, for example, clinical practice where biofluids of the patients are sampled at an arbitrary time point and are then compared to a database of urine patterns analyzed perhaps months ago.

**Proof of Concept.** For a proof of concept, the consistency of the original “3d one-batch” data set (originally analyzed in one batch) and the new “3d one-batch” data (originally measured in three batches, one for each day) was checked. Each day was separately scaled in both sets, and then PCA was carried out (Figure 5). Most of the scores of the two sets of data were grouped closely together, denoting that the investigated differences between the original and new data set were rather small. Discrimination of the original and the new “one-batch” set by supervised PLS-DA and OPLS techniques failed. No significant components could be computed due to a correlation between  $X$  (matrix) and  $Y$  (original or new data set) of almost zero. This ultimately demonstrated the power of the approach and showed that a step forward has been taken for future projects toward personalized medicine and nutrition.

## ACKNOWLEDGMENT

We are grateful to the Deutsche Forschungsgemeinschaft for their financial support of this project (VO 860/2-1). The authors thank Applied Biosystems/MDS Sciex (Concord, Canada) for their helpful correspondence concerning the data preprocessing optimization process and for providing the new MarkerView software 1.2.0.0. Theresa Ehrlich and Thomas Fischer are acknowledged for their assistance in sample preparation and measurement and Eva Kopp for providing the acrylamide- $d_3$  mercapturic acid standard.

Received for review November 15, 2006. Accepted January 19, 2007.

AC062153W