Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Chromatography B, 828 (2005) 14–20
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Metabolic fingerprinting of rat urine by LC/MS Part 2. Data pretreatment methods for handling of complex data

Helena Idborg [a], Leila Zamani [a], Per-Olof Edlund [a], Ina Schuppe-Koistinen [b], Sven P. Jacobsson [a,*]

[a] *Department of Analytical Chemistry, Stockholm University, Svante Arrhenius väg 10-12, SE-106 91 Stockholm, Sweden*
[b] *Molecular Toxicology, Safety Assessment, AstraZeneca R&D, SE-151 85 Södertälje, Sweden*

Received 25 January 2005; accepted 16 July 2005
Available online 28 September 2005

Abstract

Metabolic fingerprinting of biofluids like urine is a useful technique for detecting differences between individuals. With this approach, it might be possible to classify samples according to their biological relevance. In Part 1 of this work a method for the comprehensive screening of metabolites was described [H. Idborg, L. Zamani, P.-O. Edlund, I. Schuppe-Koistinen, S.P. Jacobsson, Part 1, J. Chromatogr. B 828 (2005) 9], using two different liquid chromatography (LC) column set-ups and detection by electrospray ionization mass spectrometry (ESI-MS). Data pretreatment of the resulting data described in [H. Idborg, L. Zamani, P.-O. Edlund, I. Schuppe-Koistinen, S.P. Jacobsson, Part 1, J. Chromatogr. B 828 (2005) 9] is needed to reduce the complexity of the data and to obtain useful metabolic fingerprints. Three different approaches, i.e., reduced dimensionality (RD), MarkerLynx™, and MS Resolver™, were compared for the extraction of information. The pretreated data were then subjected to multivariate data analysis by partial least squares discriminant analysis (PLS-DA) for classification. By combining two different chromatographic procedures and data analysis, the detection of metabolites was enhanced as well as the finding of metabolic fingerprints that govern classification. Additional potential biomarkers or xenobiotic metabolites were detected in the fraction containing highly polar compounds that are normally discarded when using reversed-phase liquid chromatography.

Keywords: High-throughput profiling; Metabonomics; Metabolomics; Metabolite profiling; Metabolic profiling; Metabolic fingerprints; Biomarker; Biological fluids; Hydrophilic interaction; HILIC; LC/MS; ZIC-HILIC column; Data pre-treatment; Classification; Chemometrics

1. Introduction

Metabolic fingerprinting by LC/MS, e.g., as described in Part 1 of this work, generates a wealth of data [1]. The data structure is a matrix for each sample, i.e., retention time in one direction and mass spectra in the other direction (Fig. 1). There are thousands of data entries per sample, complicated by a vast amount of noise, artifacts, and redundancy in the data. Chemometric methods are therefore needed to reduce the large number of variables and obtain information-rich fingerprints suitable for pattern recognition and classification.

As it is difficult to visualize and handle two-way data, full-scan LC/MS data are often reported as total ion chromatograms (TICs), base peak intensity (BPI) chromatograms, or reconstructed ion chromatograms (RICs), i.e., the dimensionality has been reduced. With this approach, resolution and hence information will be lost. In addition to the problem of overlapping peaks, there are also problems of shifts in retention time and mass number, resulting in instability of sample fingerprints. TICs and similar approaches are one way of handling LC/MS data, although they will not, for instance, reduce the amount of noise or get rid of artifacts like electronic spikes. Furthermore, TICs might not give enough information to be used as a fingerprint and so alternative approaches might be needed. It would, of course, be attractive to use the information in the mass axes, i.e., the full data matrix with all m/z values at all time points. Data-analysis

DOI of related article:10.1016/j.jchromb.2005.07.031.
* Corresponding author. Fax: +46 8 156391.
E-mail address: sven.jacobsson@anchem.su.se (S.P. Jacobsson).

1570-0232/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2005.07.049

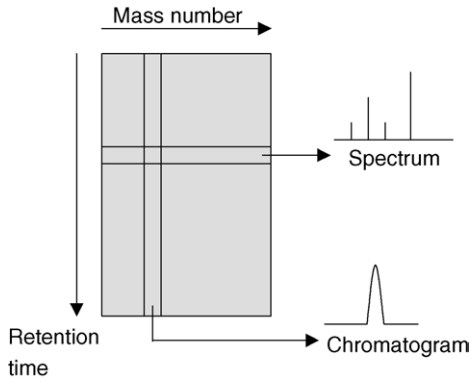

Fig. 1. The structure of LC/MS data is shown. A matrix is obtained for each sample.

methods that can handle three-way data, e.g., PARAFAC [2] and Tucker [3], were not used in this work as they are very demanding in terms of computer power and are thus not suitable for the direct data analysis of non-pretreated data.

In this second part of the work, three different methods/software packages have been used for data pretreatment:

- Reduction of dimensionality, i.e., reducing dimensionality by collapsing the mass axis (TICs) or by collapsing the time axis (Sum *m*/*z*).
- MarkerLynx™, i.e., a peak detection algorithm, where each mass number is analyzed separately in a search for peaks [4]. The area of these peaks will be given an identity of *m*/*z* and retention time and used as a fingerprint.
- MS Resolver™, i.e., two-way analysis by multivariate curve resolution (MCR) techniques [5]. Those are techniques that can recover response profiles, e.g., spectra, time profiles, or elution profiles, of components in an unresolved mixture. No prior information is needed about the composition of the mixture.

The results were compared in regard to their ability to give appropriate data for classifying urine samples using PCA [6] or PLS [7] discriminant analysis (PLS-DA) and differ between dosed and control rats. The question of interpretation and the ability to detect putative biomarkers/xenobiotic metabolites was also discussed.

2. Experimental

The workflow of the analysis is summarized in Part 1 of this work [1].

2.1. Urine samples

Urine from eleven male Wistar rats dosed with vehicle control (rats #1–5) or 130 mg of experimental drug X/kg/day (rats #6–11) was collected. It was collected on one occasion prior to dosing and on day 14 of the study. In a previous paper [8] a more detailed description of the study design is given, although the rats were dosed with different compounds.

2.2. Sample preparation and LC/ESI-MS analysis

Solid-phase extraction was used for sample preparation, as described in Part 1 of this work [1]. The wash fraction was injected onto a ZIC®-HILIC column and the eluate onto a C18 column. Both hydrophilic interaction liquid chromatography (HILIC) and reversed-phase liquid chromatography (RP-LC) have previously been described in [1].

An electrospray ionization (ESI) interface was used and the mass spectrometry (MS) detection was performed in full scan mode, alternating between positive and negative ion mode. Data obtained in profile mode from *m*/*z* 50 to 600 were recorded at a speed of 1 s/scan, as described in [1].

2.3. Data pretreatment

The data obtained are a matrix for each sample, i.e., retention time in one direction and mass spectra in the other direction. Each data file is a matrix of 551 *m*/*z* values for each scan (Fig. 1).

2.3.1. Reduction of dimensionality (RD)

LC/MS data obtained by Masslynx (Waters, Mass., USA) were converted to cdf files using the program DataBridge (Waters, Mass., USA). The data (.cdf) was then imported to MATLAB 6.5 [9]. The three-dimensional structure of each sample was then collapsed in the same way as when reporting total ion current (TIC) chromatograms, i.e., summing the intensities of all mass numbers at each retention time. Each sample is now a vector that can be used as the sample pattern (fingerprint).

With this approach, problems of retention time shifts between samples might occur. These may be reduced by bucketing [10], though at the expense of additional loss of information. In this work bucketing was not performed on this data; however, as all samples were analyzed on the same day, the retention time shifts are not that crucial, although larger retention shifts were observed when using HILIC compared to RP-LC. If needed, there are many methods available for peak alignment [11].

A second approach to reducing dimensionality of data involves collapsing the data in the other direction, i.e., the time axis. The intensities for one *m*/*z* value over time are summed (Sum *m*/*z*) and a row for each sample is obtained with intensities for each mass number. Our method differs from the method proposed by Bylund et al. [12] since it sums each mass channel, while Bylund's method uses the maximum for each *m*/*z* channel detected.

With this approach, the problems of retention time shifts will disappear. This is similar to direct infusion except for the very important fact that the sample matrix effects are reduced, e.g., ion suppression effects. The difficulty lies in handling the problems of shifts in the mass axis. One straightforward and very simple approach to bucketing is to define an *m*/*z* axis in the same range as the scan range (50–600) and divide it into segments, in this case segments of 0.1 *m*/*z*. The intensities of each *m*/*z* value are then distributed to the appropriate segment.

By using these methods essential information may be lost, although the fingerprints obtained might be sufficient. This sacrifice might also be needed to enable information on putative

biomarkers to be obtained. If a putative marker is detected, it is possible to go back to the raw data and obtain reconstructed ion chromatograms for that particular ion. This *m/z* might then give rise to several peaks at different retention times that need to be evaluated, although this is much more feasible in this reduced data set.

2.3.2. *MarkerLynx*™

MarkerLynx™ Application Manager (version 1.0) is a software package from Waters (Mass., USA) for peak detection [4]. It uses ApexTrack™ peak detection to integrate peaks in LC/MS data, the identities of the peak areas being reported as retention time (RT) and *m/z*. Following peak detection, the peaks from different samples are aligned so that the same peaks (RT, *m/z*), i.e., most probably the same compound, are found in the same row for all samples. The retention times are allowed to differ by ±0.2 min and the *m/z* values by ±0.5 Thomson. The detected peaks were then exported to MATLAB v6.5 and processed as described in 2.4. Data obtained in negative mode resulted in 1105 peaks and in positive mode 844 peaks using HILIC. When using RP-LC, 226 peaks in negative mode and 297 peaks in positive mode were obtained.

If one compound is fragmented, the compound is reported as two different compounds, i.e., compound A as (RT, m/z_A) and compound B as (RT, m/z_B). However, this will not lead to problems for the classification, although it might make the results harder to interpret. The major advantage of using this method is that the software is easy to use and is very fast, even for calculations of high-resolution data.

2.3.3. *MS Resolver*™

Multivariate curve resolution (MCR) techniques are techniques that can recover response profiles, e.g., spectra, time profiles, or elution profiles, of components in an unresolved mixture. No prior information is needed about the composition of the mixture.

MS Resolver™ (Pattern Recognition Systems, Bergen, Norway) is a software package based on the *Gentle* algorithm [5,13] for multivariate curve resolution, which mathematically resolves the data into peaks. This method is an iterative approach to solving the equation $\mathbf{X} = \mathbf{CS}^T + \mathbf{E}$, where $\mathbf{X}$ is the data matrix obtained for one sample, $\mathbf{C}$ the true chromatographic profiles, $\mathbf{S}$ the spectra for all components, and $\mathbf{E}$ is the residuals. $\mathbf{s}_1$ is then the spectrum for the first peak, and if no fragmentation occurs, this vector will have numbers close to zero at all mass numbers (*m/z* 50–600) except for one *m/z*, which will be 1, corresponding most probably to the $[M + H]^+$ ion of that peak. $\mathbf{c}_1$ will contain information about the concentration at each retention time. When data from each sample have been resolved into peaks, the peaks are aligned using ReOrder™ (Pattern Recognition Systems, Bergen, Norway). ReOrder™ compares peaks between samples within a certain time interval, in this case ±30 s. Peaks are aligned so that if their spectra are similar (comparing the seven most intense fragments) within the current time interval, they are said to be the same compound. This method has been used in previous work and is also briefly described there [8,14].

Data obtained in negative mode resulted in 55 peaks and in positive mode 56 peaks using HILIC, and 45 from negative data and 62 peaks from positive data using RP-LC.

By using the MSResolver™ approach, the information in both the *m/z* direction and the retention time direction will be taken into consideration. In addition, problems of retention time shifts will be taken into account. Although this is an iterative method and there is no unique solution, the results are good enough for some pattern recognition problems. However, there are probably a lot of peaks that will not be resolved. In addition, the resolution of each sample is quite time-consuming (approximately 5 min per sample), although it is much faster than direct methods such as HELP [15]. If high-resolution data are to be analyzed, this method is not an option, as the calculation time will increase dramatically.

2.4. *Classification*

The data (obtained as described in the Section 2.3, data pretreatment) were further processed using Matlab version 6.5. To prevent the largest peak from excessive influence on to the model, different scalings were evaluated. Four different processing methods were evaluated, i.e., mean centering, $\log_{10}$ transform + mean centering, fourth root + mean centering, or scaling of the matrix to mean zero (mean centering) and unit variance. When only mean centering was used the largest peak had the largest variance and thereby the largest influence on the model. For the last scaling method all peaks were given equal importance in the model and the two other methods generated intermediate scaling properties.

The data were analyzed using principal component analysis [6] to study data patterns. PLS-DA was then used for classification [7]. The classification was evaluated using measure of class separation (MCS) [16], which takes account of both the distribution within each class and the minimum Mahalanobis distance between the two classes, and a defined class boundary. An MCS > 1 means that the class boundary is >1 standard deviation from the nearest class center. MCS > 1.96 means that the classes are fully separated with a confidence interval of 95%, assuming that the population is infinite.

2.4.1. *Data fusion*

Data obtained from HILIC and RP-LC, respectively and pretreated using MS Resolver™ were concatenated to yield a larger matrix of 11 samples × 218 variables (peaks 1–45 from RP-LC negative mode, 46–107 from RP-LC positive mode, 108–163 from HILIC negative mode, and 164–218 from HILIC positive mode). The new matrix was $\log_{10}$ transformed and PLS-DA was performed. Concatenated data from MarkerLynx were also studied.

Hierarchical modeling was used as a second approach to data fusion [17]: a model was obtained for each of the four data sets, i.e. RP-LC positive, RP-LC negative, HILIC positive and HILIC negative, and these were nominated the lower level PLS models. Scores vectors from these models, based on the appropriate number of PLS components, were then used to obtain a new PLS model (higher level PLS). The number of PLS components

in lower level PLS was based on minimum root mean square error of prediction (RMSEP) and in higher level PLS the PLS components numbered 2.

Two rats, one from the control group and one dosed, were used as a test set to perform a brief evaluation of the models. However, this is not discussed further as the scope of this work was not to obtain predictive models but only to illustrate the advantages when using two different chromatographic set-ups.

The combination of different pretreatments and modes of data analysis gives a large number of different approaches that can be examined. Thus, in this study on the basis of experience and the literature, in the authors' best judgment, only the most obvious ones were examined. Although more fruitful approaches might, of course, exist, the aim was to demonstrate the improvements in results that could be achieved by multivariate data analysis. Furthermore, the main purpose was to show that additional analysis of the SPE wash fraction on a ZIC®-HILIC column improves the metabolic fingerprints.

3. Results and discussion

In Part 1 of this work, it was shown that by using a ZIC®-HILIC column to analyze highly polar compounds and a C18 column to analyze less polar compounds in urine, additional metabolites could be detected. In order to increase still further the number of reported metabolites and obtain sufficient metabolic fingerprints, three different procedures for data pretreatment were studied. These were compared in order to produce information-rich fingerprints suitable for the classification of dosed and control samples.

A straightforward method is to use the TICs as metabolic fingerprints. In a scores plot obtained from PLS-DA of TICs, the samples were assigned to two groups (Table 1). In this simple case where the rats had been dosed for 14 days, it was possible to classify dosed and control samples (Fig. 2). However, it is difficult to interpret the loadings plot if the aim is to identify putative biomarkers or xenobiotic metabolites. Variables with high absolute loadings in principal component 1 were studied as they correspond to variables that increased or decreased in respect to dosed or control samples. One of the variables with high absolute loadings corresponds to a retention time of 7.2 min. Spectra around this time were obtained and RICs for selected ions were studied, although it is difficult to suggest a putative biomarker. In addition to the problems of interpretation, there are also problems of shifts in retention time. As described in [1], the retention time shifts were larger when using a ZIC®-HILIC column than when using a C18 column, 0.2 min compared to 0.02 min. This indicates the need for proper peak alignment methods [11].

Instead of summing the intensities at all *m*/*z* values to obtain a TIC, the summation can be made in the other direction, i.e., summing the intensities of a particular *m*/*z* value at all time points. There are some difficulties in how to define the axis as the *m*/*z* values are float numbers, which results in that the axes between different samples cannot be compared. In addition, shifts in the *m*/*z* direction occur, although these are not as severe as for retention times. Different ways of defining the *m*/*z*-axis generated as profile data are undergoing evaluation, but here a simple and straightforward approach has been used, as described in Section 2.3.1. The scores plots from PLS-DA of this data also showed good separation between the two classes (Table 1, Fig. 2). Due to the shift in mass number of several variables, i.e., mass numbers, monitoring the same compound, this results in very complex data. This illustrates the difficulties with the fluctuations in the *m/z* direction and calls for methods of *m/z* standardization or more intelligent bucketing.

The classification of dosed and control samples based on data from MarkerLynx™ and MS Resolver™, respectively was also successful (Table 1, Fig. 2), although with these methods the results were easier to interpret when putative biomarkers and/or xenobiotic metabolites were considered. In both cases the variables correspond to a certain peak and both methods can be used to solve chromatographic coelution problems. As can be seen in Table 1, the complexity of the model, i.e., the number of variables, decreases using these methods. The question is whether or not important information is removed by preventing biomarkers from being detected.

When comparing the different data pretreatment methods, the ability to classify the two groups using PLS-DA is similar. The differences in classification when using different pretreatment methods are shown in Table 1 and in Fig. 2. PLS-DA might

Table 1
The quality of metabolic fingerprints obtained using different pretreatment methods was evaluated in regard to their power of classification

	Reduced dimensionality		MarkerLynx™	MSResolver™
	TIC	Sum *m*/*z*		
Number of variables[a]	857	5501	1105	55
PCA (MCS)[b]	0.9	0.9	1	0.07
PLS-DA (MCS)[c]	2	3	17	5
PCA (ExpV)[d]	72%	62%	30%	42%
PLS-DA (ExpV)[e]	*X*: 69%, *Y*: 78%	*X*: 60%, *Y*: 89%	*X*: 27%, *Y*: 99%	*X*: 34%, *Y*: 99%

Data obtained using ZIC®-HILIC in negative mode were used, together with $\log_{10}$ transformation prior to multivariate data analysis, i.e., PCA or PLS-DA. All eleven samples were used to build the model. The complexity of loadings was reported as number of variables.

[a] Number of variables corresponds to retention time, mass number, retention time and mass number, or retention time and spectrum.

[b] The measure of class separation (MCS) obtained using PCA.

[c] The measure of class separation (MCS) obtained using PLS-DA.

[d] The explained variance in X by two principal components.

[e] The explained variance in X and Y using two PLS components.

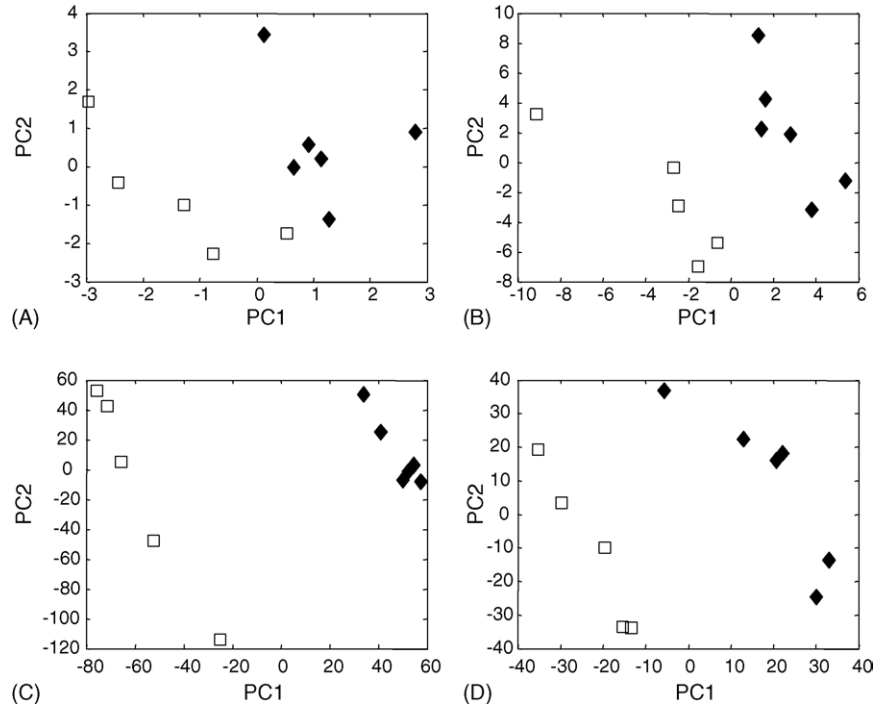

Fig. 2. Data from HILIC in negative mode was used and scores plots from the four different pretreatment methods are shown: upper plots were obtained using PLS on data pretreated by reduced dimensionality, TICs to the left (A) with a MCS of 1.8 and Sum *m/z* to the right (B) with a MCS of 2.6. The lower left plot (C) was obtained by PLS on Markerlynx™ data (MCS = 16.6) and the lower right (D) PLS on MS Resolver™ data (MCS = 5.3). Control samples (non-filled squares) and dosed samples (filled rhombs) were well separated in all cases with plot A (TICs) as the only exception.

overestimate the classification power (poor predictions), but is very useful in this case as the aim is to find metabolites that vary in accordance with the two groups. When comparing the measure of class separation, the separation was increased when using data from both ZIC®-HILIC and RP-LC data (Table 2). To illustrate what a measure of class separation (MCS) of six means, a scores plot is shown in Fig. 3. The best MCS was obtained when using Markerlynx™ as a pretreatment method and data from RP-LC positive, RP-LC negative, HILIC positive and HILIC negative were concatenated (Table 2). However, all results reported in Table 2 showed sufficient class separation.

The difficulty in the interpretation of loading plots also varies when using different pretreatment methods, as mentioned earlier. The reasons were primarily the difference in number of variables generated from the various pretreatments, and the nature of the variables, thus loading displayed. For all four methods, each variable in the loading plot corresponds to a metabolite in some way: the total ion intensity at a particular retention time (TICs), the intensity of a particular *m/z* value over time (Sum *m/z*), the area for a peak at a particular retention time

Table 2
Comparison of different PLS-DA models based on separate models (C18 or ZIC®-HILIC data and positive or negative ions) and a model based on concatenated data and finally a model obtained using hierarchical modelling

MCS[a] (PLS-DA)	MarkerLynx™	MSResolver™
C18 neg	45	7
C18 pos	36	6
ZIC®-HILIC neg	20	5
ZIC®-HILIC pos	13	5
Concatenate	57	12
Hierarchical modelling	40	27

All models were based on nine samples as two rats were used as the test set. Measures of class separation for dosed and control samples are reported.

[a] The measure of class separation (MCS) obtained using PLS-DA.

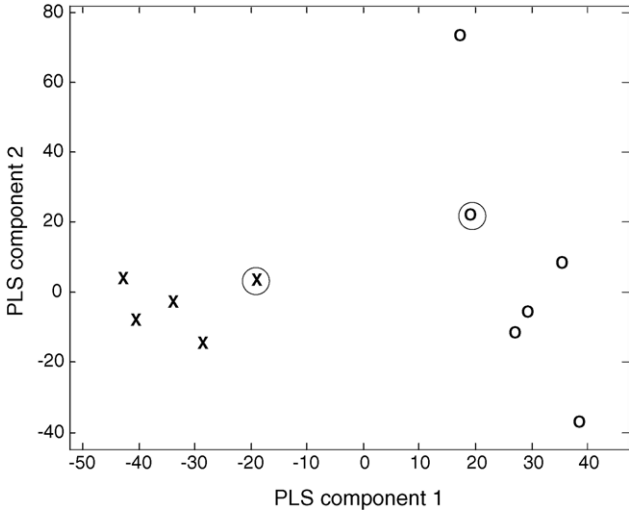

Fig. 3. The above scores plot is obtained from PLS of data obtained using RP-LC in positive mode and MS Resolver™ for pretreatment. The model is based on four control samples (X) and five dosed samples (O) and the measure of class separation was 6. In addition, the test set (one dosed and one control rat) is projected onto this plot and marked with circles.

Table 3
Variables with high absolute loadings in PLS component 1 obtained using data from RP-LC and HILIC in both positive and negative ion mode are reported

Putative biomarkers[a]	Rank[b]
	RP-LC (±) + HILIC (±)[c]
Var 48, RT = 3.4 min, 76 *m/z*	1 (C18 +)
Var 79, RT = 9.9 min, 477 *m/z*	2 (C18 +)
Var 130, RT = 10.7 min, 159 *m/z*	3 (ZIC®-HILIC −)
Var 184, RT = 10.8 min, 159 *m/z*	4 (ZIC®-HILIC +)
Var 168, RT = 3.8 min, 188 *m/z*	5 (ZIC®-HILIC +)
Var 91, RT = 11.4 min, 488 *m/z*	6 (C18 +)
Var 93, RT = 11.7 min, 424 *m/z*	7 (C18 +)
Var 121, RT = 6.8 min, 166 *m/z*	8 (ZIC®-HILIC −)
Var 175, RT = 5.8 min, 261 *m/z*	9 (ZIC®-HILIC +)
Var 86, RT = 10.5 min, 170 *m/z*	10 (C18 +)

In both cases MS Resolver™ and $\log_{10}$ transformed data were used. Additional putative biomarkers were detected using the combined method.

[a] Putative biomarkers are reported as retention time (RT) and *m/z* value.

[b] The variable with the highest absolute loading in PLS component 1 is ranked number 1. The LC/MS procedure used to obtain this peak, i.e., variable, is shown in parenthesis.

[c] Results from PLS-DA on concatenated data where variable nos. 1–45 are from C18 negative mode, 46–107 from C18 positive mode, 108–163 from ZIC®-HILIC negative mode, and 164–218 from ZIC®-HILIC positive mode.

and *m/z* value (MarkerLynx™) or the peak area for a peak at a particular retention time with a particular mass spectrum (MS Resolver™).

3.1. Putative biomarkers

By using the proposed method, additional putative biomarkers can be detected compared to when only using RP-LC. In Table 3, the putative biomarkers found by using both RP-LC and HILIC (concatenated data) are reported. Reconstructed ion chromatograms (RICs) of the two highest ranked putative biomarkers found using ZIC®-HILIC are shown in Fig. 4. The results from the concatenated data using MarkerLynx for pretreatment are not shown as the ten highest ranked variables were found to originate from the same compound. Of course, more of the highly ranked variables could be studied.

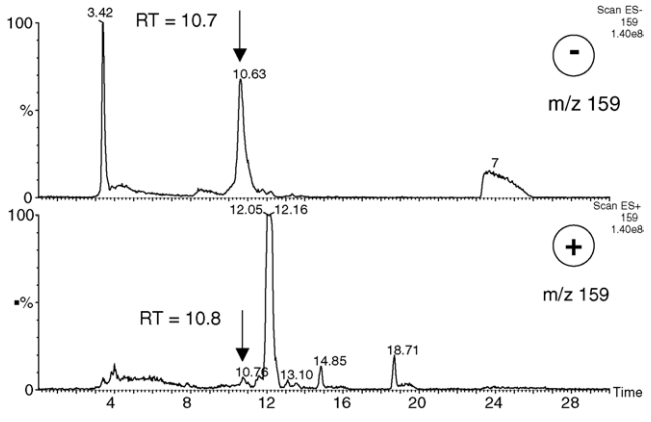

Fig. 4. Reconstructed ion chromatograms (RICs) from potential biomarkers are shown to illustrate the chromatography of the ZIC®-HILIC column. The third highest ranked variable from the concatenated data (Table 3) was obtained using HILIC and negative ion mode (upper RIC), while the fourth highest was obtained in positive mode using the same ZIC®-HILIC column (lower RIC).

In a parallel study, rats were dosed with citalopram [8] instead of, as in this case, substance X. Both substances were found to induce phospholipidosis [18,19], and by studying data from both compounds it might be possible to find putative biomarkers for phospholipidosis without running the risk of assigning xenobiotic metabolites as putative biomarkers. Naturally, the markers need to be identified to confirm that they are not monitoring a different phenomenon due to the dosage instead of phospholipidosis.

4. Conclusions

The classification of samples for diagnosis of a particular disease is based on suitable metabolic fingerprints. It is therefore important to obtain relevant data, i.e., fingerprints reflecting the discrimination. In Part 1 of this work, data were obtained by LC/MS using two chromatographic systems (RP-LC and HILIC). However, just as important as collecting suitable data is having the right tools to extract the information from the data and transform the data into useful fingerprints.

The metabolic fingerprints were improved by using data from both ZIC®-HILIC and RP-LC as the separation of dosed and control samples increased. It was shown that MarkerLynx™ and MSResolver™ were two good data pretreatment methods that gave interpretable results. Peaks at a particular retention time and mass number were suggested as putative biomarkers, although further investigations are needed to assign them biological significance.

Acknowledgements

The authors would like to thank Olav Kvalheim and Pattern Recognition Systems in Bergen, Norway, for providing the software of MS Resolver™. We would also like to thank Sigurd Hermansson at Waters Sweden for analysis using MarkerLynx. We are grateful to Fredrik O. Andersson and Ralf J.O. Torgrip for helpful discussions regarding Matlab.

References

[1] H. Idborg, L. Zamani, P.-O. Edlund, I. Schuppe-Koistinen, S.P. Jacobsson, J. Chromatogr. B 828 (2005) 9.
[2] R. Bro, Chemometr. Intell. Lab. Syst. 38 (1997) 149.
[3] L.R. Tucker, Psychometrika 31 (1966) 279.
[4] www.waters.com.
[5] R. Manne, B.-V. Grande, Chemometr. Intell. Lab. Syst. 50 (2000) 35.
[6] J.E. Jackson, A Users Guide to Principal Components, Wiley, New York, 1991.
[7] S. Wold, Chemometr. Intell. Lab. Syst. 58 (2001) 109.
[8] H. Idborg-Björkman, P.-O. Edlund, O.M. Kvalheim, I. Schuppe-Koistinen, S.P. Jacobsson, Anal. Chem. 75 (2003) 4784.
[9] The MathWorks, 2002.
[10] R.S. Plumb, C.L. Stumpf, M.V. Gorenstein, J.M. Castro-Perez, G.J. Dear, M. Anthony, B.C. Sweatman, S.C. Connor, J.N. Haselden, Rapid Commun. Mass Spectrom. 16 (2002) 1991.
[11] J. Forshed, R.J.O. Torgrip, M.K. Åberg, B. Karlberg, J. Lindberg, S.P. Jacobsson, J. Pharma. Biomed. Anal. 38 (2005) 824.

[12] D. Bylund, J. Samskog, K.E. Markides, S.P. Jacobsson, J. Am. Mass Spectrom. 14 (2003) 236.
[13] H. Shen, B. Grung, O.M. Kvalheim, I. Eide, Anal. Chim. Acta 446 (2001) 313.
[14] H. Idborg, P-O. Edlund, S.P. Jacobsson, Rapid Commun. Mass Spectrom. 18 (2004) 944.
[15] O.M. Kvalheim, Y.Z. Liang, Anal. Chem. 64 (1992) 936.
[16] M. Åberg, Variance reduction in analytical chemistry, thesis, Stockholm University, Stockholm, 2004.
[17] S. Wold, N. Kettaneh, K. Tjessem, J. Chemometr. 10 (1996) 463.
[18] U.P. Kodavanti, H.M. Mehendale, Pharma. Rev. 42 (1990) 327.
[19] H. Halliwell, Toxicol. Pathol. 25 (1997) 53.