Journal of Chromatography A, 1217 (2010) 8308–8316

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Comparative evaluation of software for retention time alignment of gas chromatography/time-of-flight mass spectrometry-based metabonomic data

Yueting Koh, Kishore Kumar Pasikanti, Chun Wei Yap*, Eric Chun Yong Chan**

Department of Pharmacy, Faculty of Science, National University of Singapore, 18 Science Drive 4, Singapore 117543, Singapore

ARTICLE INFO

Article history:
Received 10 May 2010
Received in revised form
15 September 2010
Accepted 25 October 2010
Available online 31 October 2010

Keywords:
Retention time alignment
Peak alignment
Gas chromatography/mass spectrometry
Data processing
Metabonomics
Metabolomics

ABSTRACT

In chromatography-based metabonomic research, retention time (RT) alignment of chromatographic peaks poses a challenge for the accurate profiling of biomarkers. Although a number of RT alignment software has been reported, the performance of these software packages have not been comprehensively evaluated. This study aimed to evaluate the RT alignment accuracy of publicly available and commercial RT alignment software. Two gas chromatography/mass spectrometry (GC/MS) datasets acquired from a mixture of standard metabolites and human bladder cancer urine samples, were used to assess three publicly available software packages, MetAlign, MZmine and TagFinder, and two commercial applications comprising the Calibration feature and Statistical Compare of ChromaTOF software. The overall RT alignment accuracies in aligning standard compounds mixture were 93, 92, 74, 73 and 42% for Calibration feature, MZmine, MetAlign, Statistical Compare and TagFinder, respectively. Additionally, unique trends were observed for the individual software with regards to the different experimental conditions related to extent and direction of RT shifts. Conflicting performance was observed for human urine samples suggesting that RT misalignments still occurred despite the use of RT alignment software. While RT alignment remains an inevitable step in data preprocessing, metabonomic researchers are recommended to perform manual check on the RT alignment of important biomarkers as part of their validation process.

1. Introduction

Metabonomics involves the quantitative measurement of the dynamic multi-parametric metabolic response of living systems to pathophysiological stimuli or genetic modification [1]. With the advances in non-targeted analysis, many recent studies have demonstrated metabonomics as an effective tool to assess disease progression and diagnosis, biomarker screening and characterization of complex phenotypes [2–6]. Several analytical platforms such as nuclear magnetic resonance (NMR) spectroscopy [7], liquid chromatography/mass spectrometry (LC/MS) [8] and gas chromatography/mass spectrometry (GC/MS) [9] had been employed in metabonomic analyses. Compared to LC/MS analytical platforms, GC/MS provides higher chromatographic resolution, reproducibility and higher sensitivity, thus showing great potential in metabonomic research [9–11]. Similar to other analytical platforms utilized in metabonomic studies, spectral data generated from GC/MS has to be preprocessed prior to multivariate data analysis [12,13]. Multivariate data analysis generally involves interpretation of chemical variation among samples, using variables such as retention time (RT) and metabolite concentrations to differentiate and classify the samples. Currently, RT alignment during data preprocessing poses a challenge for the efficient and accurate profiling of biomarkers as chemometric techniques are inherently sensitive to RT precision [14].

In global metabonomic studies, whereby all metabolites in the samples are analyzed and quantified, RT should be reproducible to a high degree for accurate identification of metabolites [15]. However, RT variations are not uncommon in chromatography, especially when a large number of samples are analyzed. In GC/MS-based metabonomics, RT drifts may be induced by variation in column performance or column overloading with sample [16–18]. Hence, RT alignment during data preprocessing is important to align peaks originating from the same metabolite to an identical RT. Several algorithms have been proposed so far, which can be classified into linear and non-linear correction methods [12,18,19]. Recently, alignment algorithms such as correlation optimized warping (COW) [20], piecewise alignment [21] and dynamic time warping (DTW) [22] are gaining popularity and were implemented as web servers for alignment of chromatographic data. It was noted that despite detailed documentation on each algorithm, there is little information provided on the comparative perfor-

* Corresponding author. Tel.: +65 65165971; fax: +65 67791554.
** Corresponding author. Tel.: +65 65166137; fax: +65 67791554.
E-mail addresses: phayapc@nus.edu.sg (C.W. Yap), phaccye@nus.edu.sg (E.C.Y. Chan).

0021-9673/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2010.10.101

mance of these RT alignment algorithms with respect to a common set of metabonomic data. In addition, most of the RT alignment tools evaluated were based on LC/MS data [17,23], hence evaluation of the algorithms using GC/MS data becomes pertinent.

In metabonomic analysis, it is necessary to generate numerical data matrix that contain all metabolite concentrations and samples in an experiment prior to chemometric data analysis [13,24]. Therefore, it is of utmost importance to align area values of each metabolite accurately into appropriate RT window bins across all the samples. Although, it is desirable to restore the RT of misaligned peaks to their original RT values, it is not necessary to achieve 100% accuracy in the extent of RT alignment if the metabolites are accurately binned into appropriate RT bins [25]. Hence, the ultimate goal of any RT alignment software would be to generate a data matrix which comprises accurately aligned metabolites across all the samples in a metabonomic experiment. The primary objective of this study was to evaluate the RT alignment accuracy of data processing software where RT alignment accuracy was calculated based on percentage of samples that were correctly aligned across all the metabolites at different RT shifts and metabolite concentrations. The secondary objective aimed to assess the software packages on their interfaces and user-friendliness. Three freely available software and two commercially available data alignment applications were evaluated in this study. The freely available software include MetAlign [26], MZmine [27,28] and TagFinder [29] and the commercially available software applications include the Calibration feature and Statistical Compare of the ChromaTOF software (Version 4.21, LecoCorp). MetAlign and MZmine, which were initially designed for preprocessing LC/MS data, had been validated using GC/MS data as well [5,30]. The evaluation of comparative performance of selected software in this paper is presented from a perspective of an end-user utilizing RT alignment software for GC/TOFMS-based metabonomic experiments. An evaluation of software from a computational perspective is beyond the scope of this study. However, it is important to emphasize that each software utilizes a different algorithm for RT alignment. For example, MetAlign does not use any major published RT alignment algorithm and its algorithm and workflow mimic RT alignment performed manually by an expert user. On the other hand, MZmine utilizes the random sample consensus (RANSAC) algorithm for peak alignment [12,28]. Each algorithm has its unique advantages and limitations depending on the type of data and end-user requirements. Furthermore, the data preprocessing steps prior to RT alignment such as baseline correction, noise reduction, peak picking algorithm and peak area calculation are also distinct with regards to each software. Therefore, it is important to understand these caveats when interpreting the RT alignment data presented in this paper. The applications of MET-IDEA [31], MetaboliteDetector [32], MetaboAnalyst [33], MeltDB [34], ChromA [22], XCMS [35] and AnalyzerPro [36] were also attempted, but not selected for evaluation in this study due to difficulties in using them to process the GC/MS data. Unlike previous studies [26,27,29] that evaluated the algorithms individually, this study evaluated the five selected software packages using two similar sets of metabonomic data.

2. Materials and methods

2.1. Chemicals

MSTFA (N-methyl-N-(trimethylsilyl)trifluoroacetamide) with 1% TMCS (trimethylchlorosilane) were purchased from Pierce (Rockford, IL, USA). Urease of Sigma type III, alkane standard mixture (C10–C40), and sodium sulfate (anhydrous) were obtained from Sigma–Aldrich (St. Louis, MO, USA). Milli Q water (Millipore, Bedford, MA, USA) was used as blank sample.

Table 1
List of standard metabolites with the respective retention times (RT).

Peak	Standard metabolite	RT range	RT in control sample (s)[a]
1	Malonic acid	Early	500.0
2	Benzoic acid	Early	530.8
3	Glycerol	Early	573.0
4	Succinate	Early	584.9
5	Glycine	Early	596.7
6	Uracil	Early	604.3
7	Malic Acid	Mid	728.2
8	Adipic acid	Mid	728.4
9	Tartaric acid	Mid	838.5
10	Ribose	Mid	875.6
11	Xylitol	Mid	901.0
13	Ribitol	Mid	914.6
12	Isocitric acid	Late	952.3
14	Citric acid	Late	953.7
15	Glucose	Late	1019.4
16	Manitol	Late	1047.0
17	Fructose Peak 1	Late	1002.7
	Fructose Peak 2	Late	1008.0
18	Uridine	Late	1283.9

[a] RT refers to retention time of metabolites in control samples at high concentration.

2.2. Sample preparation

To investigate the performances of RT alignment software, two independent datasets were utilized. The first data set included a mixture of 18 standard metabolites (Table 1). The second data set was obtained from the urinary metabonomic study of human bladder cancer (BC) comprising 24 BC and 51 non-BC urine samples.

2.2.1. Standard metabolite mixture

The RT alignment accuracy was determined by evaluating the precision of RT alignment of spectral data obtained from the analyses of standard metabolite mixture at two concentration levels (low and high), in two RT shift directions (positive and negative) and at two levels of RT shift induction (one and two peak width). The selection of standard metabolites was based on three criteria. Firstly, the metabolites are physiologically relevant and are typically found in biological samples. Secondly, they cover a wide range of RT. Finally, the selection included a few pairs of metabolites that might be prone to RT misalignment due to chemical structural similarities. Eighteen standard metabolites (Table 1) were selected based on the above-mentioned criteria. The pairs of chemical structurally similar metabolites with RT in close proximity include malic acid/adipic acid, citric acid/isocitric acid and glucose/fructose. In addition, sugar metabolites such as glucose and fructose were included because they give rise to multiple derivatized peaks with a greater possibility of RT misalignment [37]. In total, the alignment of 19 derivatized peaks were analyzed.

Primary stock solution of each metabolite was prepared at 10 mg/mL in methanol. Subsequently, 200 μL of each metabolite was withdrawn, mixed and diluted to obtain a secondary stock solution at 200 μg/mL of each metabolite in the mixture. The mixture of metabolites were dried and subjected to 2 h of methoximation (MOX) and 1 h of MSTFA derivatization to form the TMS derivatives. Final concentrations obtained for the TMS derivatized products were 5 and 100 μg/mL for the high and low concentration replicates, respectively. Each set of the experiment was performed in triplicates. Subsequently, 1 μL of derivatized sample was utilized for gas chromatography/time-of-flight mass spectrometry (GC/TOFMS) analysis.

2.2.2. Bladder cancer and non-bladder cancer urine samples

Sample preparation and analysis of urine collected from 24 BC and 51 non-BC subjects was reported previously by Pasikanti

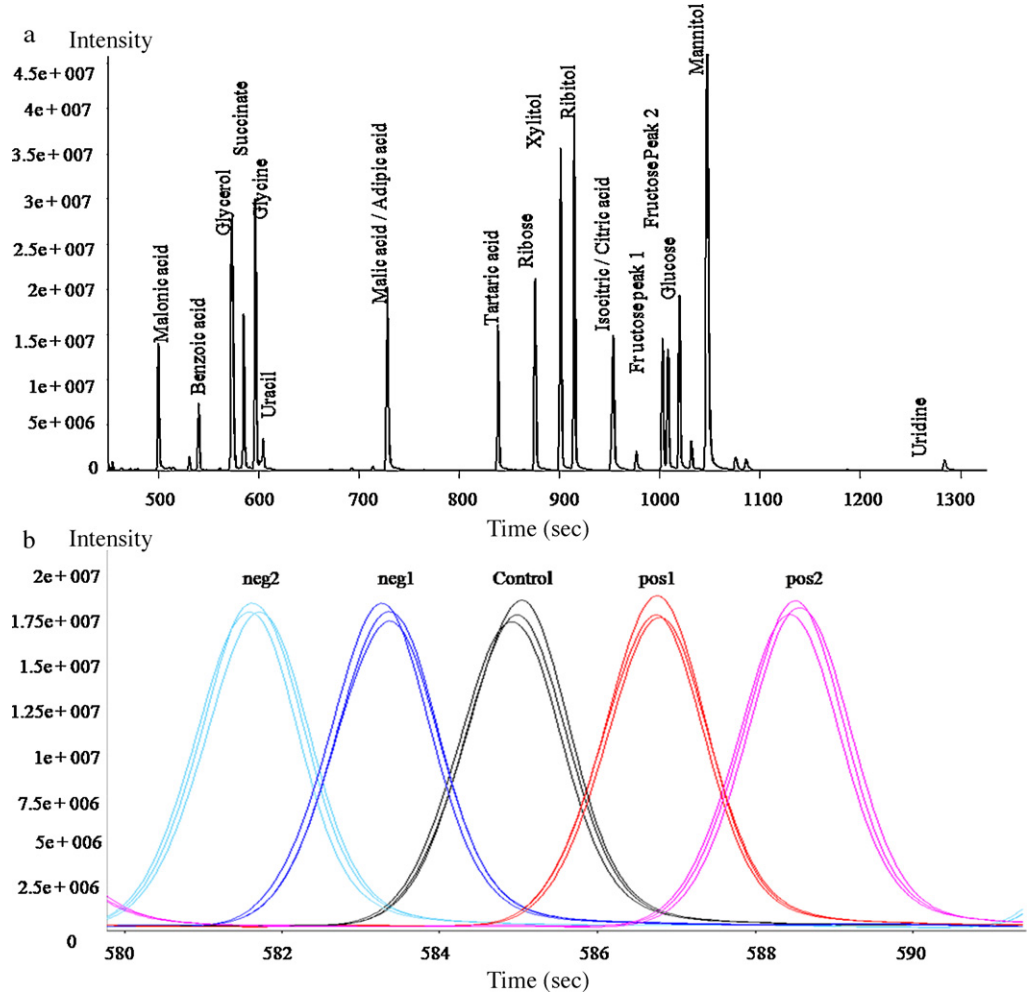

Fig. 1. (a) Total ion chromatogram inclusive of 19 derivatized metabolic peaks at high concentration. (b) Chromatogram of succinate peak at dual levels of positive and negative RT shift induction.

et al. [38]. These human urine samples were used to evaluate the different RT alignment algorithms in terms of their ability in preprocessing complex GC/TOFMS-based clinical samples. Complete analysis details are not provided for brevity. Briefly, 200 μL of urine samples were treated with urease, extracted using methanol, dried and derivatized first by MOX and subsequently by MSTFA. Derivatized samples were then subjected to GC/TOFMS analysis.

2.3. GC/TOFMS analysis

A Pegasus 4D GC × GC/TOFMS (LecoCorp., St. Joseph, MI, USA) was utilized in the GC/TOFMS mode for all analyses. The GC/TOFMS instrument was equipped with an Agilent 7890 gas chromatograph. A DB-1 20 m × 250 μm fused silica capillary column (Agilent J&W Scientific, Folsom, CA), with 0.25 μm film thickness, was used with open split interface. Helium was used as the carrier gas at 1.5 mL/min and the injector split ratio was set to 1:20.

2.4. Retention time shift induction

By altering the programmed temperature gradient, RT shifts equivalent to one and two peak widths (~2 and ~4 s, respectively) were induced in the positive and negative directions for each metabolite at both low and high concentrations. Induction of approximately one or two peak width RT shifts in the positive directions will be referred as pos1 and pos2 henceforth. Similarly, RT shifts in the negative direction will be referred as neg1 and neg2 shifts. In all the RT shift induction experiments, initial temperature of the oven was set to 70 °C and being held for 0.2 min. For the controls, the temperature was programmed to increase at 10 °C/min to 270 °C where it was held for 10 min. For the induction of pos1, temperature was programmed to increase at 9.95 °C/min to 170 °C and subsequently at 10 °C/min to 270 °C where it was held for 10 min. To induce pos2 RT shift, temperature was increased at 9.90 °C/min to 150 °C, followed by 10 °C/min increase to 270 °C where it was held for 10 min. RT shift for neg1 was induced by changing the temperature gradient to 10.05 °C/min until it reaches 170 °C, followed by 10 °C/min to 270 °C and held for 10 min. Finally for neg2, temperature gradient was set at 10.10 °C/min till reaching 150 °C and subsequently changed to 10 °C/min until 270 °C was attained and maintained for 10 min. The secondary oven was always maintained at 10 °C higher relative to primary oven temperatures for all the analyses. An example of the RT shift-induced chromatogram is shown in Fig. 1.

2.5. MS parameters

The mass spectrometry was operated in electron impact (EI) mode (70 eV). Data acquisition was performed in the full scan mode from *m/z* 40–600 with an acquisition rate of 20 Hz.

2.6. Data processing

Each chromatogram obtained from GC/TOFMS analysis was processed for baseline correction, noise reduction and deconvolution using the ChromaTOF software and subsequently exported in NetCDF (.cdf) format. The NetCDF files were imported into each software and further processed.

Except for the Calibration feature and Statistical Compare, performance characteristics of each software were evaluated using a 2.40 GHz, 32.0 GB RAM Intel(R) Xeon (R) processor running Windows Server 2008 R2 (64-bit) operating system. A number of experiments were performed to maximize the accuracy and performance of each software. For the Calibration feature and Statistical Compare, deconvoluted chromatograms were individually processed for library matching and peak area calculation using an identical data processing method created in ChromaTOF software. Area of each peak was calculated using the unique mass of each derivatized metabolite. Only peaks with signal to noise ratio (S/N) greater than 100 were retained for further analysis.

2.6.1. MetAlign

MetAlign is a RT alignment software developed originally for LC/MS-based plant metabonomic data but the application of this software has recently been extended to GC/MS data [30]. MetAlign interface is broadly divided into three sections, parts A, B and C, corresponding to data import, RT alignment and statistical analysis, respectively. All the files were imported into a single group and processed together (part A). The first file specified in group 1 was used as the reference for the alignment of peaks of subsequent chromatograms. Under part B, MetAlign provides two RT alignment algorithms, namely the rough and iterative mode alignment. In this study, only the rough alignment mode was employed, whereby a user defined time window was used throughout the time dimension of all data sets to be aligned. Within the specified RT window, each mass ion of a peak (also known as variables) was grouped together based on the amplitudes [26]. Other than utilizing the rough algorithm, no scaling was carried out using MetAlign and the minimum number of data files was set at 70% of total processing files. For the initial peak search criteria, maximum RT shift beginning and ending in the first region was set at 100 and 200 scans, respectively. The statistical analysis feature (part C) of MetAlign software was not explored since the focus of the study was to investigate RT alignment. A snap shot of MetAlign interface with the parameter settings utilized in our study is shown in Fig. S1 of Supplementary data.

2.6.2. MZmine

MZmine was implemented as a stand-alone open-source Java application toolbox, initially catered for LC/MS data. MZmine implements an alignment method which uses a master peak list generated from the sample files. Every peak from a sample file is mapped against the master peak list and assigned to the best matching existing peak. If the *m/z* and RT difference with the best matching existing peak are beyond the tolerance limit specified, the software appends a new peak to the master peak list [28]. MZmine offers two alignment methods, join aligner and RANSAC aligner. Several trials comparing these two alignment algorithms were performed in our study. The results showed that there were only minimal differences in the alignment quality of these two alignment algorithms. This is in concordance with the results obtained in a previous study [23]. The RANSAC aligner algorithm was used in this study and the other features of the software that were used included chromatogram builder and gap-filling. Chromatogram builder is a feature which detects the mass value within a spectrum and constructs a chromatogram within a specified time range. After aligning the peak lists, it is likely that the master peak list contains gaps as not every peak will be detected and aligned in each sample. The gap-filling feature thus serves to fill up these missing values. Under the alignment settings, the *m/z* tolerance was set at 0.1, RT tolerance at 10 s, minimum number of points at 0.2% and the threshold value at 10 s. Fig. S2 of Supplementary data shows the complete set of parameters used for MZmine processing.

2.6.3. TagFinder

TagFinder is a single user application for personal computer systems and is programmed using the Java™ programming language. A workspace has to be created for each batch of processing. In each workspace with the imported data, the mass fragments were sorted by mass and RT. Mass fragments of equal mass across all the sample files were binned and aligned into mass tags. Each mass tag was associated with its own specified RT window. Subsequent to the scanning of RT, clustering and correlation processes were implemented to generate a data matrix [29]. The time scan width was set at 1.0, tag time width 0–10 s, sample count being 70% of total files and a minimum 6 pairs clustering was used for the Pearson correlation method. The finalized parameters utilized for the TagFinder analysis are summarized in Fig. S3 of Supplementary data.

2.6.4. Statistical Compare

Statistical Compare is a recently introduced feature in the ChromaTOF software (Version 4.21) to align peak information obtained from multiple chromatograms. Statistical Compare used a mass spectral match criterion of 60% when aligning the multidimensional peak data comprising sample names, metabolites, RT, mass and integrated peak area. Quantitation mass for each peak in the data table was selected from the unique mass that was most common to all matching peaks within the RT window (two times of peak width). Subsequently, this unique mass was used to calculate the peak areas of each metabolite. In addition, the best quality peak from matching peaks was then selected, whose name was used for the analyte in the data table. The resulting data table comprised observations where each observation was described by variables (peak intensities) aligned according to their RT/unique mass pairs as identifiers. The data table was exported as a .csv file.

2.6.5. Calibration feature

A calibration method was created in ChromaTOF software to align peak information obtained from multiple chromatograms. One of the control samples (where no RT shift was induced) at high concentration range was used as a reference. Analyte peaks were identified from the reference chromatogram corresponding to each standard metabolite. In the calibration method, three metabolites (glycerol, malic acid and ribitol) corresponding to the early, mid and late RT ranges were specified as RT markers. Subsequently, relative retention times (RRT) of remaining analytes were calculated by linear extrapolation with respect to the RT of the nearest RT markers. A mass spectral similarity match of 70% and RRT deviation of 2 s were used as criteria to identify the peaks in the sample. All the remaining chromatograms were then processed to identify and calculate area values of the peaks specified in this calibration method. For each metabolite, the calculated area values from all the chromatograms listed in the calibration table were copied and pasted into a Microsoft Excel worksheet manually.

2.7. Evaluation

2.7.1. Standard metabolite mixture

The alignment performance of each software was assessed using the alignment accuracy percentage, which was defined as the percentage of samples in which a particular metabolite was accurately identified and aligned across all RT shifts and at both concentration levels. An unique mass for each analyte peak was selected using

peak finding algorithm of ChromaTOF software to perform peak integration. A mass was considered unique if it was present in the spectrum of the selected analyte but not the coeluting analytes. In the case where all the masses of the analyte were comparable to those of the coeluting peaks, a deconvolution algorithm would select the unique mass based on various criteria including S/N ratio and the degree to which the mass was unique. Individual unique mass identified for each peak by ChromaTOF software was subsequently used as a marker for investigating the RT alignment accuracy for all the alignment software. In the case where the unique mass was not available when the data was processed using freely available software, the mass with the highest intensity was utilized instead. Mapping of the unique masses across all the five software tools were performed manually with the aid of Microsoft Excel 2007 functions. The metabolite unique mass with area values outside 40% range of mean area of the three control replicates was considered incorrectly identified (Table S1, Supplementary data).

2.7.2. Bladder cancer and non-bladder cancer urine samples

The urine data were processed using the same set of software parameters used for the standard mixtures. However, the unique masses were not mapped across software used in the study. Aligned sample data were manually normalized using the total area normalization. The normalized data matrix, consisting of a random training set of 20 BC and 41 non-BC samples, was used to construct the orthogonal projections to latent structure-discriminant analysis (OPLS-DA) models using SIMCA-P+ software (Umetrics, Kinnelon,NJ, USA). An independent OPLS-DA model was constructed for each alignment software package using the same training dataset. The alignment accuracy performance of the software was assessed by comparing the overall cross-validation coefficient, Q^2, and the prediction accuracy of external prediction set. Each external prediction set comprised four BC and ten non-BC samples. The Q^2 value is defined as the fraction of total variation of X or Y that can be predicted by a component, as estimated by seven-fold cross validation. Q^2 measures goodness of fit of an OPLS-DA model and a value greater than 0.5 indicates good model. Generally, higher Q^2 indicates better predictability of the model and values above 0.2 are commonly observed and considered acceptable in clinical metabonomics [39,40]. Prediction accuracy was defined as the percentage of external samples in the prediction set that were predicted correctly when cross-validated using the corresponding OPLS-DA model.

3. Results

3.1. Standard metabolite mixture

Although the standard mixture contained only 18 metabolites, none of the software tools evaluated in this study was able to achieve 100% RT alignment accuracy. The heatmap in Fig. 2 shows the trends in RT alignment accuracy (percentage) observed for the different software packages. The heatmap is divided into 5 columns with regards to the average RT alignment accuracy (percentage) of the five software applications. Each column is further separated according to the direction and extent of RT shifts while the rows represent the two levels of concentration of metabolites. The results pertaining to each software are elaborated below.

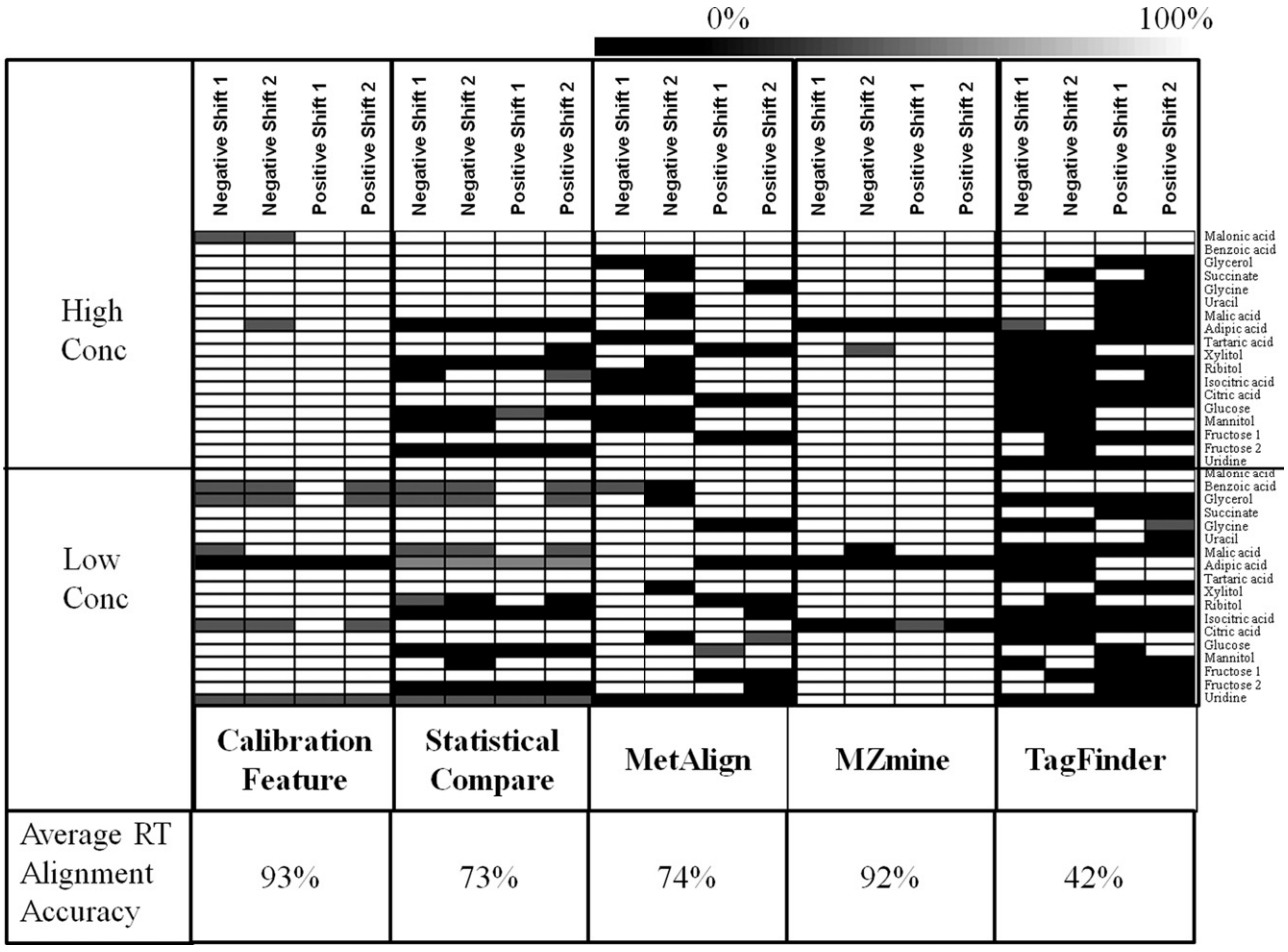

Fig. 2. Heatmap showing RT alignment accuracy (percentage) across five processing methods. Bright green and black colored boxes indicate 0 and 100% accuracy, respectively.

Table 2
RT alignment accuracy (percentage) under different extent of RT shifts.

	Calibration feature	Statistical Compare	MetAlign	MZmine	TagFinder
One peak width shift	94	76	81	93	45
Two peak width shift	93	70	66	92	38

3.1.1. MetAlign

The results derived from MetAlign showed an interesting trend, which was specific to the software. For the high concentration samples, RT alignment accuracy of 58% in negative RT shifts was noted which was significantly lower compared to 82% accuracy obtained in positive RT shifts. On the other hand, for the low concentration samples, the opposite trend was observed where 88 and 73% RT alignment accuracies were observed for the negative and positive RT shift induction, respectively.

3.1.2. MZmine

MZmine showed the best performance among the three publicly available software packages with an alignment accuracy of 92% (Fig. 2). It was noted that the metabolites that failed to achieve 100% RT alignment accuracy resided within the middle RT range.

3.1.3. TagFinder

TagFinder had the lowest RT alignment accuracy of all the software evaluated. In addition, five metabolites were not detected at each concentration level. Specifically, isocitric acid and uridine were not detected at both sample concentrations and out of the 19 derivatized peaks; 13 peaks had RT alignment accuracy of approximately 50%. Extensive optimization of software parameters was performed but failed to improve the RT alignment accuracy.

3.1.4. Statistical Compare

Statistical Compare attained an overall RT alignment accuracy of 73%. The relatively low accuracy value was mainly caused by six peaks, namely adipic acid, xylitol, ribitol, glucose, mannitol and fructose peak 2, which were poorly aligned. These metabolite peaks showed poor alignment accuracy at low concentration when the Calibration feature was used. All the other peaks showed RT alignment accuracy of over 80%.

3.1.5. Calibration feature

The Calibration feature of ChromaTOF software generated the highest RT alignment accuracy of 93%. Although alignment of metabolites was accurate at high concentration, a few metabolites were not integrated at low concentrations based on the S/N ratio criterion.

3.1.6. General observations

The overall RT alignment accuracies in aligning standard compounds mixture were 93, 92, 74, 73 and 42% for Calibration feature, MZmine, MetAlign, Statistical Compare and TagFinder, respectively. Calibration feature and MZmine demonstrated consistently higher RT alignment accuracies across all the samples, except for a few metabolites. In contrast, TagFinder showed a relatively low accuracy across the various tested conditions, with no particular trend observed. As for MetAlign, the RT alignment accuracy was reduced when the metabolites utilized in our study were present at higher concentration and when RT was shifted in the negative direction. An opposite trend was observed for the low concentration samples.

For all the five software tools, RT alignment accuracy was consistently lower when challenged with two peak width shift induction as compared to one peak width shift induction for the 19 metabolite peaks investigated (Table 2). For example, the RT alignment accuracies of MetAlign were 81 and 66% with regards to one and two peak width shifts, respectively. Analysis of the relationship between metabolite concentration and RT alignment accuracy showed no consistent trend for standard compounds. Both MetAlign and TagFinder showed higher RT alignment accuracy for low concentration samples, while Calibration feature, Statistical Compare and MZmine demonstrated higher RT alignment accuracy for high concentration samples.

3.2. Bladder cancer and non-bladder cancer urine samples

The standard mixture contained a small number of metabolites and might not adequately represent the complex peak-shift artifacts commonly encountered in the clinical metabonomics [15]. Hence, in this study, the results from the standard metabolites were supplemented with findings generated using human urine samples, in order to comprehensively assess the RT alignment efficiency in relation to biomarker discovery application. Samples in the training dataset were used to build OPLS-DA models for each software (Fig. 3). The models were subsequently tested with samples in the external prediction set which consisted four BC and ten non-BC samples to determine prediction accuracy of each OPLS-DA model. For TagFinder, despite using a Java memory space of 32 Gb, only 32 of the urine sample files could be imported into the workspace. This observation suggested the possible limitation of the software's import function. Considering the wide variance of metabolites among different samples, validation by OPLS-DA prediction modeling was not performed for data generated by TagFinder. Performance statistics of the OPLS-DA models and prediction accuracy (percentage) constructed using the other four RT alignment software packages are listed in Table 3.

4. Discussion

4.1. Standard metabolite mixture

Our results showed that the performance of the RT alignment algorithms was clearly limited by the extent of RT shift where a larger shift expectedly resulted in poorer RT alignment performance. In large scale metabonomic studies, the extent of RT shifts

Table 3
Comparison of R^2, Q^2 and prediction accuracy (percentage) of the OPLS-DA prediction models.

	R^2X	R^2Y	Q^2	Prediction accuracy (%)	
				BC	Non-BC
MetAlign	0.316	0.993	0.439	50	100
MZmine	0.233	0.980	0.185	75	70
Calibration feature	0.416	0.887	0.300	75	80
Statistical Compare	0.252	0.883	0.436	75	60

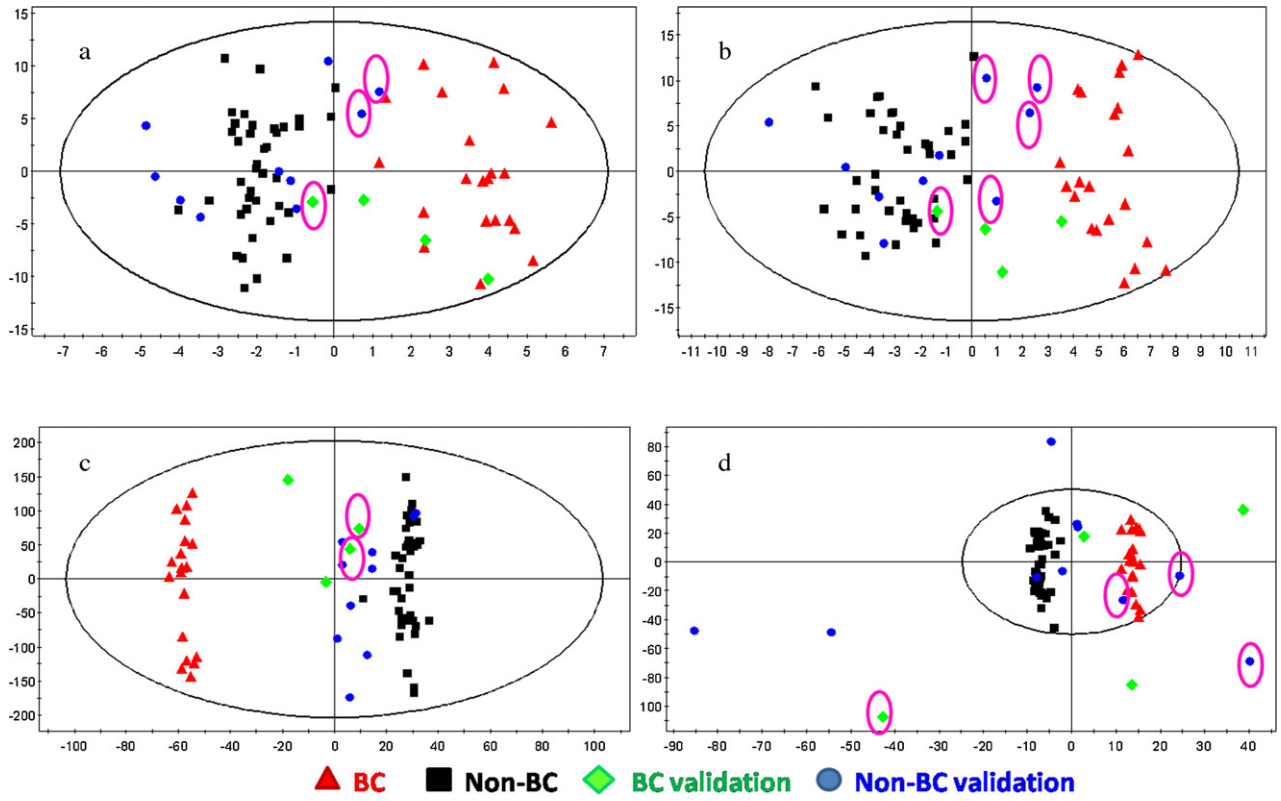

Fig. 3. OPLS-DA score plots obtained from the RT aligned data tested with external prediction set of four BC and ten non-BC samples. (a) Calibration feature (b) Statistical Compare (c) MetAlign (d) MZmine. Incorrectly predicted samples are highlighted by circles.

is variable and might be greater than those induced in the current study, emphasizing the need for these alignment algorithms to be more flexible in accommodating a wider range of RT shifts. For instance, the RT shifts induced in our study were systematic and not frequently encountered in reality. On the other hand, in some instances, RT shifts may occur frequently and randomly which are beyond the control of the electronic pressure control (EPC) system of the GC instrument. In such cases, the alignment of peak information is dependent on subsequent data preprocessing.

Unlike the extent of RT shift, the effect of concentration of metabolites on RT alignment performance was not as clearly defined. As column overloading was unlikely in our controlled study where the detector was not saturated, our results underscored the possibility of poor RT alignment due to variation in concentration of endogenous metabolites. This inconsistency in RT alignment performance highlighted a concern in chromatography-based biomarker discovery as human biological samples are diverse in metabolite concentration and the direction and extent of RT shifts are difficult to control experimentally. Therefore, it is essential to select the appropriate RT alignment software according to the known characteristics of the biological data. For example, if the metabolites are estimated to be present at higher concentrations (>100 μg/mL), Calibration feature, MZmine or Statistical Compare may be more suitable than MetAlign or TagFinder. The Calibration feature method demonstrated the highest RT alignment accuracy in alignment of standard compounds but one major drawback was the human intervention required to manually identify analyte peaks from the reference chromatogram. The process was time-consuming and tedious.

While Statistical Compare is a convenient platform since it is an integral part of ChromaTOF software, its main limitation is that the user cannot specify and optimize alignment-related parameters such as RT window and m/z similarity. Thus, it may not be possible to optimize its alignment efficiency if larger RT shifts are encountered. Moreover, data generated from other GC/MS instruments cannot be processed using the Statistical Compare feature of ChromaTOF software.

MetAlign had been used in a number of LC/MS-based metabonomic studies and recently, Peters et al. evaluated the effect of parameter optimization on MetAlign RT alignment efficiency [30]. The poorer alignment accuracy of MetAlign observed in our study compared to earlier studies suggested the importance to re-evaluate software that were originally developed for LC/MS data preprocessing prior to utilizing them in GC/MS metabonomic experiments. As our experiments were performed using the rough mode alignment, it was possible that the RT alignment accuracy might be higher if the alternative iterative mode alignment was used instead. In view of this, several runs were performed using the iterative mode alignment. Nonetheless, no conclusive result was obtained as the alignment process could not be completed whenever more than 14 data files were imported.

While MZmine appeared to be promising based on the urine metabonomic study, it was found to be deficient in aligning closely eluting peaks, such as adipic acid/malic acid and isocitric acid/citric acid metabolic pairs when the metabolite standards were analyzed. Although the raw data were deconvoluted by ChromaTOF before preprocessing, deconvolution by ChromaTOF alone might not be adequate in resolving these close-eluting peaks since adipic acid was also not detected at low concentration using the ChromaTOF Calibration feature method.

Despite efforts to optimize the parameters, the low RT alignment accuracy of TagFinder was possibly related to errors incurred by inappropriate parameter settings. TagFinder is a relatively new software designed specifically for GC/MS data and little information has been provided. Although m/z 73 or 147 were recognized by other software as the highest intensity ion for more than 10

peaks, TagFinder recognized m/z 73 or 147 as the highest intensity ion only for four peaks. The lower RT alignment accuracy might also be due to the high dependency of the alignment algorithm on retention index (RI) calculation. The RI calculation step was omitted in this study so that the software would align data using RT. The omission of RI calculation appeared to have a significant effect on the alignment accuracy of TagFinder. Further exploration of the software is required to evaluate its utility.

4.2. Bladder cancer and non-bladder cancer urine samples

For the processing of 106 files, MetAlign required a processing time of 10 h 30 min while MZmine took 21 h 20 min for complete processing. The processing time of Calibration feature and Statistical Compare were not compared as they could not be operated on the same computer system.

Calibration feature, Statistical Compare and MetAlign demonstrated Q^2 of 0.3, 0.436 and 0.439, respectively, and thus were considered as acceptable prediction models. The number of latent variables and R^2 values of these three models did not show any likelihood of being over-fitted. For MZmine, the Q^2 of 0.185 suggested that the prediction model was relatively poor in terms of prediction performance. This poor result was in contrast to its good RT alignment performance in the standard metabolite experiment. In addition, there was a poor correlation between the performance indicator, Q^2, and the prediction accuracies of the external validation samples. This observation could be due to the fact that not all the functions related to each software were utilized in this study. For example, MZmine provided other processing options like peak deconvolution and normalization which were not utilized in the current study. These options were not adopted to ensure consistency in comparison with other software. In this study, all processing options and settings were kept consistent with the standard metabolite method. However, considering the metabolic complexity of human urine samples, the more extensive processing functions of each software should be investigated in future studies.

The mapping of unique mass was not performed for the clinical metabonomic dataset due to two main reasons. Firstly, each software generated a different number of total metabolite peaks. Therefore, the unique masses could not be mapped across the five software investigated in the study. Secondly, the dataset comprised a large number of variables (mass ions) and observations which prevented manual comparison of unique masses between software. The differential picking of mass ions for peak integration could influence the predictive performance of each software tool in addition to RT alignment. RT alignment accuracy depended on the ability of each software to accurately perform peak picking, noise reduction, mass spectral matching and resolution of mass ions. Comparison of Q^2 values of OPLS-DA models enabled an evaluation of the overall performance of each software with respect to parameters such as noise reduction, peak picking, peak integration and RT alignment.

Based on standard compound and urine analyses, differences in the performance of RT alignment software were observed. Comparatively, higher RT alignment accuracies were observed by Calibration feature and MZmine in the alignment of standard compounds. As inferred from a computational perspective, this could be due to the use of RT markers and RRT to align the metabolite peaks. While Calibration feature specified RT markers manually, MZmine selected RT markers using a normalization algorithm that scouted for common peaks to perform RT normalization. In addition, RANSAC and gap filling algorithm utilized in MZmine might have resulted in its higher RT alignment accuracy. For instance, RANSAC algorithm corrected both linear and non-linear RT deviations while gap filling algorithm performed an additional step of searching missing peaks retrospectively. However, the poor Q^2 value observed by MZmine for the clinical dataset might be due to difficulty in choosing appropriate common peaks via the RT normalization algorithm. On the other hand, RT markers were specified manually in Calibration feature even for the clinical data. Therefore, the performance of Calibration feature in RT alignment of the clinical data was not altered. In MetAlign, user defined time window was utilized to align each mass trace. Subsequently, an average RT shift was calculated to determine the RT shift correction estimate. It might be possible that the RT shift correction estimate was influenced by the concentration of analyte and direction of RT shift which in turn determined the observed accuracy of RT alignment in MetAlign. As the first chromatogram was used as a reference to align peaks in MetAlign, spectral purity of peaks in the reference chromatogram might also influence the RT alignment outcome. In summary, while the evaluation of the software from a computational perspective was not within the scope of the current work, one has to be mindful of the varied computational factors that contributed to the accuracy of RT alignment.

4.3. Evaluation on user-friendliness of external software

Each external software (MetAlign, MZmine and TagFinder) evaluated in this study provided varying performance and ease of use. All three software tools supported the import of .cdf files, which are generated by most of the GC/MS instruments, thus reducing the need for file conversion. In this study, MetAlign required the shortest processing time for the 106 files, and was almost twice as fast as MZmine. MZmine offered the ability to distribute the computation workload to increase the processing speed and thus multiple processor system was utilized for our operation of MZmine [12]. For first-time users, MetAlign is a suitable choice as it displays an easy-to-use interface, in which the processes are listed as sequences to guide the user. To allow graphical viewing of the aligned peaks, MetAlign provides an additional application, GM2MS, which can convert the .csv output to a .cdf file. Graphical viewing of the aligned peaks enables the visualization of individual peak shifting by the alignment algorithm. This provides a clearer indication on how well the algorithm aligned the peaks. However, despite several attempts using different input files from different samples, both low and high concentration samples showed the same peaks with exactly the same intensities. This suggested that GM2MS is probably only applicable for LC/MS data currently. Both MetAlign and TagFinder offer a direct total processing function, allowing automated complete processing of chromatographic peaks if the parameters have been optimized and confirmed. As for MZmine, it provides an advantage of graphical visualization of the chromatograms, which can be used for optimizing parameter settings. For the processing of 106 files of approximately 100 Mb each, MZmine requires a Java memory of 20 Gb to complete all the steps listed in Section 2.6.2. Details of the comparison on the user-friendliness of the software packages are listed in Table 4.

Although each RT alignment software was comprehensively evaluated, it is important not to overlook the limitation of the current study. Firstly, not all parameters available in some software were utilized. For instance, the deconvolution feature was not utilized in MZmine which might have compromised its applicability. Secondly, the RT shifts evaluated in our study were linear and all the metabolites were shifted in positive or negative directions uniformly. However, in metabonomics analysis, RT shifts may be more complex and non-linear. Finally, the performance of RT alignment software might be influenced by high concentration saturated peaks that were more challenging to align as they were represented as multiple peak maxima during peak picking.

Table 4
Comparison on user-friendliness of the three publicly available software.

	MetAlign	MZmine	TagFinder
Processing time	10 h 30 min	21 h 20 min	NA
Input file formats	netCDF MassLynx Xcalibur Agilent Chemstation	netCDF Xcalibur mzDATA MZXML mzML	netCDF .smp (PEG files)
Output file	.csv .cdf .txt	.csv XML	.tab NIST.msp
Processing pipeline coverage	Baseline correction RT alignment Normalization Univariate analysis	RT alignment Deconvolution Normalization Peak identification	Peak table alignment RI calculation Correlation, clustering Univariate analysis Peak identification
Graphical visualization	No	Yes	No
Ability to export intermediate results	Yes	Yes	No

5. Conclusion

The inconsistent RT alignment results obtained for both standard metabolites and human urine samples suggested that the existing software algorithms require further improvement before they can be considered as ideal platforms for GC/MS data alignment. While RT alignment remains an inevitable step for data preprocessing, metabonomic researchers should take note of the possible misalignments of metabolites which might occur due to the varied concentrations and complexity of the metabolome. In the near future, metabonomic scientists are recommended to perform manual checks on the RT alignment of important biomarkers as part of the validation process.

Conflicts of interest

The authors indicated no potential conflicts of interest.

Acknowledgements

This study was supported by the National University of Singapore (NUS) Grant R-148-000-100-112 and National Medical Research Council grant R-176-000-119-213 provided to E.C.Y. Chan, the NUS President Graduate Fellowship to K.K. Pasikanti and the NUS Department of Pharmacy Final Year Project grant (R-148-000-003-001 to Y.T. Koh). GC/TOFMS was kindly sponsored by the NUS grant R-279-000-249-646.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.chroma.2010.10.101.

References

[1] J.K. Nicholson, J.C. Lindon, E. Holmes, Xenobiotica 29 (1999) 1181.
[2] Q. Zhang, G. Wang, Y. Du, L. Zhu, A. Jiye, J. Chromatogr. B: Anal. Technol. Biomed. Life Sci. 854 (2007) 20.
[3] H. Wu, R. Xue, C. Lu, C. Deng, T. Liu, H. Zeng, Q. Wang, X. Shen, J. Chromatogr. B: Anal. Technol. Biomed. Life Sci. 877 (2009) 3111.
[4] H. Wu, R. Xue, L. Dong, T. Liu, C. Deng, H. Zeng, X. Shen, Anal. Chim. Acta 648 (2009) 98.
[5] T. Kind, V. Tolstikov, O. Fiehn, R.H. Weiss, Anal. Biochem. 363 (2007) 185.
[6] H.J. Issaq, O. Nativ, T. Waybright, B. Luke, T.D. Veenstra, E.J. Issaq, A. Kravstov, M. Mullerad, J. Urol. 179 (2008) 2422.
[7] R. Powers, Magn. Reson. Chem. 47 (Suppl. 1) (2009) S2.
[8] C.K. Lim, G. Lord, Biol. Pharm. Bull. 25 (2002) 547.
[9] K.K. Pasikanti, P.C. Ho, E.C. Chan, J. Chromatogr. B: Anal. Technol. Biomed. Life Sci. 871 (2008) 202.
[10] H.J. Issaq, E. Abbott, T.D. Veenstra, J. Sep. Sci. 31 (2008) 1936.
[11] O. Fiehn, Trends Anal. Chem. 27 (2008) 261.
[12] M. Katajamaa, M. Oresic, J. Chromatogr. A 1158 (2007) 318.
[13] P. Jonsson, A.I. Johansson, J. Gullberg, J. Trygg, J.A.B. Grung, S. Marklund, M. Sjostrom, H. Antti, T. Moritz, Anal. Chem. 77 (2005) 5635.
[14] K.J. Johnson, B.W. Wright, K.H. Jarman, R.E. Synovec, J. Chromatogr. A 996 (2003) 141.
[15] K.M. Aberg, E. Alm, R.J. Torgrip, Anal. Bioanal. Chem. 394 (2009) 151.
[16] N. Etxebarria, O. Zuloaga, M. Olivares, L.J. Bartolome, P. Navarro, J. Chromatogr. A 1216 (2009) 1624.
[17] A.M. van Nederkassel, M. Daszykowski, P.H. Eilers, Y.V. Heyden, J. Chromatogr. A 1118 (2006) 199.
[18] A.L. Duran, J. Yang, L. Wang, L.W. Sumner, Bioinformatics 19 (2003) 2283.
[19] M.D. Robinson, D.P. De Souza, W.W. Keen, E.C. Saunders, M.J. McConville, T.P. Speed, V.A. Likic, BMC Bioinform. 8 (2007) 419.
[20] S.Y. Wang, T.J. Ho, C.H. Kuo, Y.J. Tseng, Bioinformatics (2010).
[21] K.M. Pierce, B.W. Wright, R.E. Synovec, J. Chromatogr. A 1141 (2007) 106.
[22] N. Hoffmann, J. Stoye, Bioinformatics 25 (2009) 2080.
[23] E. Lange, R. Tautenhahn, S. Neumann, C. Gropl, BMC Bioinform. 9 (2008) 375.
[24] J. Trygg, E. Holmes, T. Lundstedt, J. Proteome Res. 6 (2007) 469.
[25] P. Jonsson, J. Gullberg, A. Nordstrom, M. Kusano, M. Kowalczyk, M. Sjostrom, T. Moritz, Anal. Chem. 76 (2004) 1738.
[26] A. Lommen, Anal. Chem. 81 (2009) 3079.
[27] M. Katajamaa, J. Miettinen, M. Oresic, Bioinformatics 22 (2006) 634.
[28] M. Katajamaa, M. Oresic, BMC Bioinform. 6 (2005) 179.
[29] A. Luedemann, K. Strassburg, A. Erban, J. Kopka, Bioinformatics 24 (2008) 732.
[30] S. Peters, E. van Velzen, H.G. Janssen, Anal. Bioanal. Chem. 394 (2009) 1273.
[31] C.D. Broeckling, I.R. Reddy, A.L. Duran, X. Zhao, L.W. Sumner, Anal. Chem. 78 (2006) 4334.
[32] K. Hiller, J. Hangebrauk, C. Jager, J. Spura, K. Schreiber, D. Schomburg, Anal. Chem. 81 (2009) 3429.
[33] J. Xia, N. Psychogios, N. Young, D.S. Wishart, Nucleic Acid Res. 37 (2009) W652.
[34] H. Neuweger, S.P. Albaum, M. Dondrup, M. Persicke, T. Watt, K. Niehaus, J. Stoye, A. Goesmann, Bioinformatics 24 (2008) 2726.
[35] C.A. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, Anal. Chem. 78 (2006) 779.
[36] M.P. Styczynski, J.F. Moxley, L.V. Tong, J.L. Walther, K.L. Jensen, G.N. Stephanopoulos, Anal. Chem. 79 (2007) 966.
[37] A. Fox, J. Chromatogr. A 843 (1999) 287.
[38] K.K. Pasikanti, K. Esuvaranathan, P.C. Ho, R. Mahendran, R. Kamaraj, Q.H. Wu, E. Chiong, E.C. Chan, J. Proteome Res. 9 (2010) 2988.
[39] L. Zheng, D.G. Watson, B.F. Johnston, R.L. Clark, R. Edrada-Ebel, W. Elseheri, Anal. Chim. Acta 642 (2009) 257.
[40] X. Zhang, Y. Wang, F. Hao, X. Zhou, X. Han, H. Tang, L. Ji, J. Proteome Res. 8 (2009) 5188.