Nonlinear Data Alignment for UPLC−MS and HPLC−MS Based Metabolomics: Quantitative Analysis of Endogenous and Exogenous Metabolites in Human Serum

Anders Nordström, Grace O'Maille, Chuan Qin, and Gary Siuzdak*

Department of Molecular Biology, Scripps Center for Mass Spectrometry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

A nonlinear alignment strategy was examined for the quantitative analysis of serum metabolites. Two small-molecule mixtures with a difference in relative concentration of 20−100% for 10 of the compounds were added to human serum. The metabolomics protocol using UPLC and XCMS for LC−MS data alignment could readily identify 8 of 10 spiked differences among more than 2700 features detected. Normalization of data against a single factor obtained through averaging the XCMS integrated response areas of spiked standards increased the number of identified differences. The original data structure was well preserved using XCMS, but reintegration of identified differences in the original data reduced the number of false positives. Using UPLC for separation resulted in 20% more detected components compared to HPLC. The length of the chromatographic separation also proved to be a crucial parameter for a number of detected features. Moreover, UPLC displayed better retention time reproducibility and signal-to-noise ratios for spiked compounds over HPLC, making this technology more suitable for nontargeted metabolomics applications.

Mass spectrometry has established itself as a useful tool for metabolomics analysis for its capability to measure compounds present at very low levels and at the same time provide structural information.[1,2] With separation systems such as gas or liquid chromatography (GC or LC) coupled with mass analysis, the information content can be dramatically increased due to reduced ionization suppression and temporal separation of isomers. Prior to any statistical or multivariate analysis of acquired data, it is a prerequisite that the data are aligned, i.e., in a GC− or LC−MS analysis, m/z (X) at retention time (Y) must be consistent throughout all observations. While drift in the m/z direction is fairly straightforward to correct by m/z calibration, the error in the retention time domain is more difficult to align. We have recently developed XCMS,[3] a software for nonlinear retention time correction of XC−MS data, where X denotes the possibility to use various types of chromatography prior to MS analysis. The software is freely available under an open source license from (http://metlin.scripps.edu/download/). XCMS reads CDF files from the mass spectrometer, finds peaks based on a matched filter whose coefficients are equal to a second-derivative Gaussian function. The peaks are reported in the form of "features", which are defined as a unique m/z at a unique time point. XCMS matches the features and performs nonlinear retention time correction through a local regression fitting method using peak groups that are initially well grouped. The output from XCMS can be visualized as a series of superimposed retention time corrected extracted ion chromatograms and also as a tab delimited table with columns denoting observations and a row for each feature aligned across the observations. The values populating the data matrix are computed areas representing the response from each feature. The table can be imported into Microsoft Excel, Matlab, or any preferred multivariate or statistical software for further analysis without additional processing, as the data are already aligned.

The aim of the present study was to evaluate the XCMS algorithm's capability to identify large and small differences in a complex biological matrix. A well-controlled set of LC−MS data was generated by spiking two sets of mixtures containing 19 endogenous and exogenous compounds with relative concentration differences (20−100%) for 10 of the compounds into human serum.

Several strategies to achieve data sets relevant for further statistical processing are found in the literature, and three main strategies can be distinguished;

(1) Alignment of Chromatograms. Alignment of the chromatograms without prior peak picking can be potentially very useful due to little or no operator intervention in setting criteria for peak finding. Furthermore, it enables direct differential analysis between entire sets of data matrices. An algorithm for aligning raw LC−UV chromatograms, named COW, was developed by Nielsen and co-workers.[4] The COW algorithm has been further adapted to make use of MS data for chromatogram alignment.[5]

* Corresponding author. E-mail: siuzdak@scripps.edu.

(1) Want, E. J.; Cravatt, B. F.; Siuzdak, G. *Chembiochem* **2005**, *6*, 1941−1951.
(2) Villas-Boas, S. G.; Mas, S.; Akesson, M.; Smedsgaard, J.; Nielsen, J. *Mass Spectrom. Rev.* **2005**, *24*, 613−646.
(3) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2005**, *78*, 779−787.
(4) Nielsen, N. P. V.; Carstensen, J. M.; Smedsgaard, J. *J. Chromatogr., A* **1998**, *805*, 17−35.
(5) Bylund, D.; Danielsson, R.; Malmquist, G.; Markides, K. E. *J. Chromatogr., A* **2002**, *961*, 237−244.

10.1021/ac060245f CCC: $33.50
Published on Web 04/11/2006

Another algorithm designed to align one-dimensional GC-FID data was developed by Johnson and co-workers[6]

(2) Summation or Binning of Chromatographic Data. Summation of chromatographic segments in LC− or GC−MS analysis is one way of reducing the problems associated with data alignment. Summation of m/z data across preset time windows will result in no loss of information and the alignment error will be confined to the edges of the windows or bins. After multivariate analysis, which reveals windows displaying significant difference and enables subsequent deduction of the m/z responsible for the difference, a reversion to the original chromatograms can confirm the differences[7,8]

(3) Curve Resolution or Deconvolution. This strategy involves finding components (a component refers to a chromatographic peak representing a metabolite) that are subsequently matched over the different observations (data sets). With GC−MS electron ionization (EI) spectra, deconvolution using the free AMDIS software[9,10] or proprietary mass spectrometer manufacturer software can be done prior to matching of the components with software such as MSFACT (RTAlign).[11] However, AMDIS and other software primarily developed for EI spectra require the presence of several mass traces that converge at the same retention time as a prerequisite for deconvolution. With LC and electrospray ionization (ESI), this requirement is not always fulfilled. For LC−MS data, Windig and co-workers developed the component detection algorithm (CODA),[12] which reduced the number of spectra to be investigated by 1 order of magnitude. Component-resolving algorithms developed for LC−MS data include GENTLE[13,14] and MEND.[15] More recently, other integrated strategies for comparison of LC−MS data in metabolomics have been developed[16,17] including the MZmine software.[18] There are also commercial software available such as MS-resolver (Pattern Recognition Systems) and MarkerLynx (Waters), whose performance have been compared,[19] and metAlign (Plant Research International), which has been used for mining LC−MS data in a nontargeted fashion.[20] After peak finding, matching of peaks has to be performed. For this, several different matching strategies have been used. Examples include the use of time windows to assign components with specific retention time and intensity to a certain component group.[11] "Master chromatogram alignment", i.e., matching of all the peaks in a sample to a master peak list, is done using retention time and an m/z value to align components possessing similar m/z values[18] and are within time proximity of each other. Another approach is to calculate spectral similarity for adjacent components in a specified time segment and thereby matching components with high similarity within that particular time segment.[14]

Ultraperformance liquid chromatography (UPLC)[21,22] is a promising separation technique for metabolomics. The reduced particle size (1.4−1.7 μm) of the packing material offers increased separation through narrower chromatographic peaks over normal particle size (3.5−5 μm) HPLC, resulting in increased peak capacity, lower ion suppression and potentially better signal-to-noise ratio (S/N) for observed components. When analyzing complex mixtures with LC−MS, as often is the case in metabolomics investigations, UPLC can be of great advantage over regular microbore HPLC in that more components can be detected.[23]

In the present study, we have examined the quantitative aspects of the XCMS data alignment algorithm for metabolites, endogenous and exogenous, in human serum using UPLC−MS and HPLC−MS. The approach was shown to identify features with a concentration difference of only 20% between two sample classes among thousands of features that remained constant.

EXPERIMENTAL SECTION

The workflow is illustrated in Figure 1 and composition of the two spiking mixtures is shown in Table 1.

Chemicals and Sample Preparation. All solvents used were of HPLC grade (J.T. Baker, Philipsburg NJ). The serum used was human serum from clotted human male whole blood, sterile-filtered (Sigma, St. Louis, MO). The bovine serum albumin (BSA; Sigma) was subject to reduction, alkylation, and tryptic digestion.[24] $^2H_5$-Phenylalanine (98%) was obtained through Cambridge Isotope Laboratories (Andover, MA). All other chemicals for spiking mixtures A and B were obtained in high purity from Sigma. Protein precipitation of five aliquots (100 μL) of the human serum was performed with cold methanol.[25] The precipitated aliquots were dried, resuspended in 100 μL of acetonitrile/water, 5/95 v/v (0.1% formic acid) and subsequently pooled. From the pooled serum extract, 100 μL was aliquoted to four different HPLC vials. To vial one, 900 μL of acetonitrile/water, 5/95 v/v (0.1% formic acid) was added. To vial two, 87.7 μL of BSA digest (5.7 μM) and 812.3 μL of acetonitrile/water, 5/95 v/v (0.1% formic acid) were added, and to vial three and four, we added 87.7 μL of BSA digest (5.7 μM) and 808.3 μL pf acetonitrile/water, 5/95 v/v (0.1% formic acid) in addition to 4 μL of stock solution of mixtures A and B,

(6) Johnson, K. J.; Wright, B. W.; Jarman, K. H.; Synovec, R. E. *J. Chromatogr., A* **2003**, *996*, 141−155.
(7) Plumb, R. S.; Stumpf, C. L.; Gorenstein, M. V.; Castro-Perez, J. M.; Dear, G. J.; Anthony, M.; Sweatman, B. C.; Connor, S. C.; Haselden, J. N. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1991−1996.
(8) Jonsson, P.; Gullberg, J.; Nordstrom, A.; Kusano, M.; Kowalczyk, M.; Sjostrom, M.; Moritz, T. *Anal. Chem.* **2004**, *76*, 1738−1745.
(9) Davies, T. *Spectrosc. Eur.* **1998**, 22−26.
(10) Stein, S. E. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 770−781.
(11) Duran, A. L.; Yang, J.; Wang, L. J.; Sumner, L. W. *Bioinformatics* **2003**, *19*, 2283−2293.
(12) Windig, W.; Phalp, J. M.; Payne, A. W. *Anal. Chem.* **1996**, *68*, 3602−3606.
(13) Shen, H. L.; Grung, B.; Kvalheim, O. M.; Eide, I. *Anal. Chim. Acta* **2001**, *446*, 313−328.
(14) Idborg-Bjorkman, H.; Edlund, P. O.; Kvalheim, O. M.; Schuppe-Koistinen, I.; Jacobsson, S. P. *Anal. Chem.* **2003**, *75*, 4784−4792.
(15) Andreev, V. P.; Rejtar, T.; Chen, H. S.; Moskovets, E. V.; Ivanov, A. R.; Karger, B. L. *Anal. Chem.* **2003**, *75*, 6314−6326.
(16) Radulovic, D.; Jelveh, S.; Ryu, S.; Hamilton, T. G.; Foss, E.; Mao, Y. Y.; Emili, A. *Mol. Cell. Proteomics* **2004**, *3*, 984−997.
(17) Jonsson, P.; Bruce, S. J.; Moritz, T.; Trygg, J.; Sjostrom, M.; Plumb, R.; Granger, J.; Maibaum, E.; Nicholson, J. K.; Holmes, E.; Antti, H. *Analyst* **2005**, *130*, 701−707.
(18) Katajamaa, M.; Oresic, M. *Bmc Bioinformatics* **2005**, *6*, 179−190.
(19) Idborg, H.; Zamani, L.; Edlund, P. O.; Schuppe-Koistinen, I.; Jacobsson, S. P. *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* **2005**, *828*, 14−20.
(20) Vorst, O.; Vos, C. H. R. d.; Lommen, A.; Staps, R. V.; Visser, R. G. F.; Bino, R. J.; Hall, R. D. *Metabolomics* **2005**, *1*, 169−180.
(21) Shen, Y. F.; Zhang, R.; Moore, R. J.; Kim, J.; Metz, T. O.; Hixson, K. K.; Zhao, R.; Livesay, E. A.; Udseth, H. R.; Smith, R. D. *Anal. Chem.* **2005**, *77*, 3090−3100.
(22) 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.
(23) Plumb, R.; Castro-Perez, J.; Granger, J.; Beattie, I.; Joncour, K.; Wright, A. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2331−2337.
(24) Nordstrom, A.; Apon, J. V.; Uritboonthal, W.; Go, E. P.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 272−278.
(25) Want, E. J.; O'Maille, G.; Smith, C. A.; Brandon, T. R.; Uritboonthai, W.; Siuzdak, G. *Anal. Chem.* **2005**, *78*, 743−752.

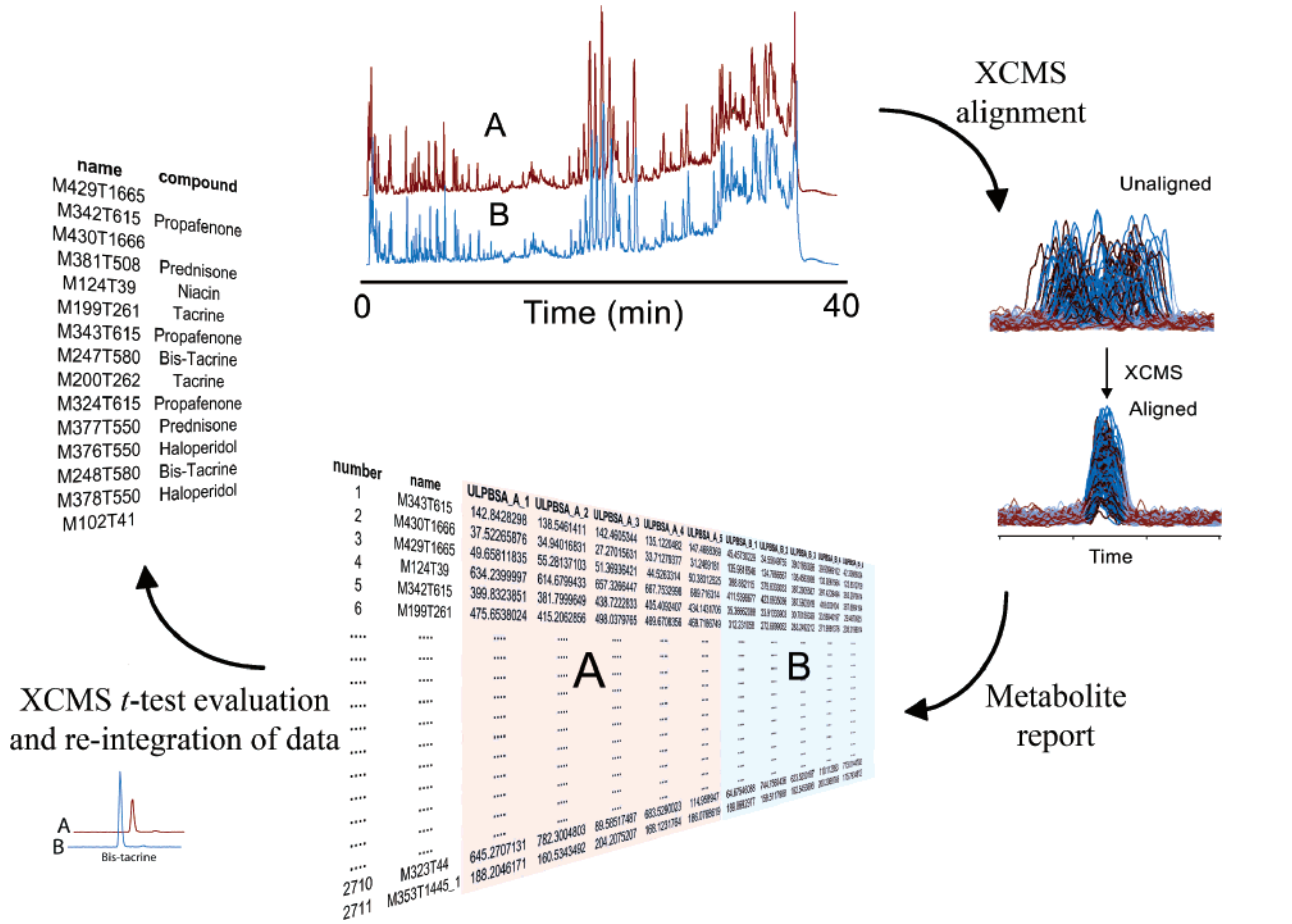

Figure 1. Workflow scheme, the acquired data were converted to CDF format. The CDF files were subsequently processed using XCMS. The output from XCMS contained aligned data that can be viewed as picture files and also as a resultant matrix where samples (observations) constitute the columns and the features (variables) constitute the rows. The features (a specific *m*/*z* at a unique time point), in this case 2711, were normalized and sorted according to *p*-value obtained through a *t*-test. Features that display significant difference ($p < 0.05$) between sample classes A and B were subject to reintegration in the raw data. The reintegrated areas were again normalized and sorted according to *p*-value, and compounds with $p < 0.05$ constituted the final table of metabolites that were different between samples A and B.

respectively. By keeping the volume constant (1000 μL), dilution variances in the serum were canceled.

LC−MS. The separation system used was a Waters Acquity UPLC (Waters, Millford, MA) coupled to Micromass Q-TOF Micro (Waters). The same system was used for both UPLC and HPLC experiments. Elution buffers were as follows: (A) water with 0.1% formic acid and (B) acetonitrile with 0.1% formic acid. Linear gradients from 5%B/95% A to 95%B/5% A over 10 or 30 min were used after 1-min fixed initial conditions. The gradient was kept at 95%B/5% A for three column volumes, and the column was subsequently reequilibrated with four column volumes. Flow rates used were 0.25 and 0.5 mL for HPLC and UPLC, respectively, and the flow was split 1:1 postcolumn in the UPLC experiments. We used a BEH chemistry C18 (2.1 × 100 mm) 1.7-μm particle size column for UPLC experiments and a Symmetry C18 (2.1 × 100 mm) 3.5-μm particle size column for HPLC experiments (both obtained from Waters). The column heater was set to 40 °C, and back pressures noted were around 8000 psi/1600 psi for UPLC/HPLC, respectively. Injection volume was set to 20 μL full loop with an overfill factor of 2.

The mass spectrometer was used with the regular ESI interface and calibrated prior to experiments. Data were collected in continuum mode between 100 and 1000 *m*/*z* with an acquisition

Table 1. Composition (Injected Amount) and $[M + H]^+$ of Spiking Mixes Used[a]

compound	$[M + H]^+$	mix A (pmol)	mix B (pmol)
nicotinic acid	124	80	40
phenylalanine $^2H_5$[b]	171	200	200
atenolol[b]	267	0.2	0.2
sulpiride	342	0.2	0.16
caffeine	195	0.2	0.2
tacrine	199	0.2	0.16
tetracycline	445	2	2
chlorotetracycline[b]	479	2	2
tenoxicam	338	0.2	0.1
propranolol[b]	260	0.2	0.2
prednisone	359	2	0
haloperidol	376	0.2	0.3
bis tacrine[c]	493	0.2	0.3
piroxicam	332	2	2.4
propafenone	342	0.2	0
mifepristone	430	0.2	0.24
verapamil[b]	455	0.2	0.2
ketoprofen	255	0.2	0.2
oleamide[b]	282	2	2

[b] Compounds used for normalization. [a] Compounds are arranged in relative elution order (UPLC) with top entry eluting first. [c] Bis tacrine was detected by its fragment at *m*/*z* 247.

rate of ~2 spectra/s, which produced at least 5 spectra/chromatographic peak (scan speed 0.52, interscan delay 0.1). Each of the four pools of serum was injected in five replicates with each chromatographic setup, giving a total of 80 injections.

Data Analysis. The files were converted to CDF format using Databridge (Waters) on a PC workstation (Pentium 4, 3.20 GHz, 1 GB RAM). The conversion for 80 LC−MS runs took ~4 h. The size of the converted CDF files ranged between 0.5 and 1 Gb, which required a fairly powerful computer for further processing. The publicly available XCMS software previously described[3] was downloaded (http://metlin.scripps.edu/download/) and installed on a computer with Linux operating system (dual processor 3.2 Ghz and 6 Gb RAM). The XCMS processing of 80 samples took ~24 h. Data were processed using XCMS's default settings (see documentation for XCMS, http://metlin.scripps.edu/download/) with the following exceptions: xcmsSet (profmethod="binlinbase"), retcor (p="m", f="s", missing=5, extra=5, span=0.2), group (bw=10 for HPLC and bw=5 for UPLC). The resulting table (TSV file) was viewed in Excel (Microsoft, Redmond WA). All *t*-tests performed in Excel were two-sided, unequal variance. The number of features originating from BSA was determined by comparing serum-only with serum spiked with BSA digest (five and five replicates), with *p*-value cutoff set to <0.01. The reintegrated areas were obtained by using QuanLynx (Waters) and by setting up an integration parameter file using the average m/z and retention time for features found to be significant (p <0.05) in normalized XCMS data. The normalization was performed by dividing each number in the XCMS output matrix with a scalar obtained through averaging the XCMS area value for the six compounds that remained constant between the spiked mixtures of A and B (Table 1). The reintegrated values were renormalized against a scalar obtained in the same manner from Quanlynx-integrated values. Apex peak tracking was used for defining peaks in Quanlynx. Reported false positives refer to cases where the *t*-test rejected null hypothesis when it was actually valid; that is, the *t*-test incorrectly reports that it has found a positive result where none really exists. The LC−MS data is available for download in CDF format from http://metlin.scripps.edu/download/.

RESULTS AND DISCUSSION

In a comprehensive metabolomics study, the aim is to compare the metabolomes of different samples in a nontargeted fashion. The result of such a comparison should list features/metabolites with quantitative information that are reflecting biochemical changes. The structural identification of unknown metabolites is a time-consuming process, and it is therefore desirable to make this list accurate, reflecting only the significant changes, both small and large, with as few false positives as possible.

Design of Experiment. The experiment was designed as splitting a pool of human serum that previously had been subjected to protein precipitation into four aliquots. One aliquot was kept as control. The second pool was spiked with a BSA digest to further increase the complexity. The third pool was spiked with a BSA digest and with a mixture (A) consisting of 19 nonendogenous compounds. The fourth aliquot was spiked with a BSA digest and with a mixture (B) that was a modified version of mixture A. The composition of mixtures A and B are shown in Table 1. While 9 compounds were kept constant, 10 concentration changes were made between (A) and (B). The changes included removal of two compounds, increased concentration in four compounds by 20 and 50%, and reduced concentration in four compounds by 20 and 50%, respectively. The strategy was to introduce a known set of differences between the two sample classes A and B, while keeping thousands of components constant, and follow how well our metabolomics strategy could identify these small differences. The design involved analyzing each of the four plasma pools with both UPLC and HPLC using long and short separation gradients. The primary motivation for this design was to evaluate the effect of the length of the chromatographic analysis on the number of detected components and on their retention and response reproducibility.

The outline of the examined metabolomics workflow is illustrated in Figure 1. After acquisition, the data were converted to CDF format and read into XCMS. The unaligned data were aligned and the output from XCMS can be visualized as aligned and superimposed extracted ion chromatograms or as a data matrix with observations as the columns and features (named by m/z and retention time) as rows aligned across the observations. The resulting table, in tab delimited format, can be readily exported to Microsoft Excel, Matlab, or other multivariate/statistical software. After normalization, sorting of the features according to *p*-value (cutoff p <0.05), calculated by performing a *t*-test between classes A and B, generated a list of significant features. The ions in the list were reintegrated with the raw data. Integration was manually inspected, and the integration report was exported to Excel where a new normalization was performed against areas from the six fixed compounds (integrated from the original data). A *t*-test was calculated on the resultant normalized data matrix, and the features were subsequently sorted according to *p*-value (cutoff p <0.05). The resulting list would be used for reverting back to raw data (or reanalyzis with other instrumentation) for identification of the most significantly varying features.

UPLC versus HPLC. After XCMS processing, the resulting tab delimited table was imported to Microsoft Excel. The number of detected features revealed that UPLC was superior to HPLC in that more features were detected. For both the 30- and 10-min gradients, UPLC generated roughly 20% more features than the HPLC analysis (Table 2). The length of the separation was also important for the number of detected features. This illustrates a potential for gradient length optimization by using XCMS to increase the gradient length until the increase in number of detected features plateaus. By searching the output from XCMS for the known added compounds, it was observed that the longer gradient enabled XCMS to extract more ions correctly (Table 2). The two compounds that were not found by XCMS with the 30-min gradient were tenoxicam and ketoprofen. With UPLC, these compounds could be manually identified in the chromatogram with an S/N close to the detection limit, whereas they remain undetected in the HPLC chromatogram even after manual inspection. A BSA digest was introduced to three of the serum aliquots to increase the complexity of the samples and to test whether XCMS could be used for finding quantitative differences among peptides and, thus, have applications in quantitative proteomics. In a comparison between the serum with and without spiked BSA digest, the 30-min UPLC gradient identified the highest number of significantly different features originating from the digest (Table 2). The XCMS report table contained minimum and maximum

Table 2. Comparison between UPLC and HPLC

separation and gradient (min)	total features found	spiked features found[a]	features from BSA[b]	retention time corrn[c] (s)	fraction (%) of features with RSD < 5% non-norm[d]	with RSD < 5% norm[e]	with 5% < RSD < 25% non-norm[d]	with 5% < RSD < 25% norm[e]	with RSD > 25% non-norm[d]	with RSD > 25% norm[e]
UPLC-30	2709	17	585	1.8 ± 2.0	6.9	25.2	84.1	64.4	9.0	10.4
HPLC-30	2125	17	494	10.1 ± 21.6	5.8	8.3	83.1	82.1	11.1	9.6
UPLC-10	2034	16	494	1.7 ± 1.7	51.0	52.1	47.0	45.8	2.0	2.1
HPLC-10	1619	16	501	2.9 ± 4.2	43.0	41.6	53.2	54.8	3.8	3.6

[a] Serum spiked with mixture A. [b] Significance cutoff at p <0.01. [c] Average of observed [maximum − minimum] retention time for each feature before correction ± standard deviation. [d] Feature areas as obtained by XCMS without normalization (5 observations/feature). [e] Feature areas as obtained by XCMS with normalization (5 observations/feature) (see Experimental Section).

retention time for each feature as measured prior to correction. By subtracting minimum from maximum for each feature, an estimate of retention time correction was obtained. These differences were averaged over all observations for the different separation schemes and are reported in Table 2. Interestingly, the average required retention time corrections were much lower using UPLC. This might reflect an individual column difference but could potentially imply higher retention time reproducibility using an ultra-high-pressure separation system. Response reproducibility was similar between UPLC and HPLC as shown in Table 2. The largest fraction of features showed an RSD between 5 and 25%. Taken together, these results confirm that UPLC offered an advantage in terms of more detected ions, which has been suggested in previous work.[23] Surprisingly, the retention time did appear more reproducible with UPLC compared to HPLC. This might in part be due to better peak shapes obtained with UPLC and the fact that faster chromatography (10-min gradient) displayed higher reproducibility (Table 2). However, faster chromatography resulted in less detected features overall and is therefore not a good option for untargeted comparisons. Signal-to-noise values comparing UPLC and HPLC are displayed in Figure 2A. For a majority of the spiked compounds, UPLC gave higher S/N; this is in agreement with previous research.[22] Shown in Figure 2B is a more detailed examination of selected EICs and corresponding spectra combined over elution profiles from representative spiked compounds. One reason for better signal-to-noise ratios is the narrower elution profiles obtained with UPLC. Furthermore, it is evident by comparison of the spectra in Figure 2B that the spectral purity of the chromatographic components is higher with UPLC. Narrower peaks will result in better separation and therefore less suppression during ionization.

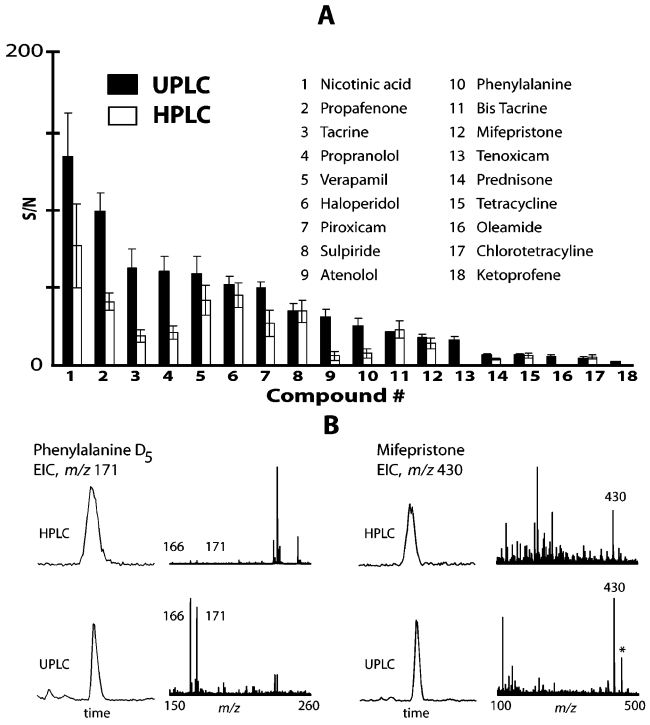

Figure 2. (A) S/N for the spiked compounds. Black and white bars are S/N measured with UPLC and HPLC, respectively. Error bars illustrate standard deviation (n = 5). (B) Extracted ion chromatograms (EIC) and spectra for selected spiked compounds. The EICs are shown over 1 min. Asterix indicates the $[M + Na]^+$ of mifepristone, which was significant using UPLC.

Normalization and Reintegration. In the literature, several normalization strategies are found.[8,26−28] It has also been demonstrated that using nanospray infusion MS, the need for normalization is minimized.[29] In the present study, normalization was performed by calculating the average XCMS integrated area for six of the compounds that were kept constant between mixes A and B and then dividing each feature area value in the separate observations with this average. This method of normalization resulted in a response factor matrix of the same size compared to the original matrix, and the response factor matrix was subsequently further processed. We chose to normalize with this approach because averaging the responses of several spiked components appearing over the entire chromatographic range gave good results even for small differences (Figure 3) and could be easily performed. However, as Table 3 indicates, we found that normalization increased the number of spiked features that were significant (i.e., p-values <0.05). Also shown in Table 3 are the number of significant features (p <0.05) before and after normalization of XCMS data, as well as the number of significant features observed after reintegration of the raw data, before and after

(26) Fiehn, O.; Kopka, J.; Dormann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. *Nat. Biotechnol.* **2000**, *18*, 1157−1161.
(27) Wang, W. X.; Zhou, H. H.; Lin, H.; Roy, S.; Shaler, T. A.; Hill, L. R.; Norton, S.; Kumar, P.; Anderle, M.; Becker, C. H. *Anal. Chem.* **2003**, *75*, 4818−4826.
(28) Shurubor, Y. I.; Paolucci, U.; Krasnikov, B. F.; Matson, W. R.; Kristal, B. S. *Metabolomics* **2005**, *1*, 75−85.
(29) Boernsen, K. O.; Gatzek, S.; Imbert, G. *Anal. Chem.* **2005**, *77*, 7255−7264.

Table 3. Number of Features, Total and Spiked, above Significance Cutoff (p <0.05) as a Function of Normalization and Reintegration

separation and gradient (min)	XCMS before normalization		XCMS after normalization		integrated before normalization		integrated after normalization		false positives[a]
	total	spiked	total	spiked	total	spiked	total	spiked	
UPLC-30	41	6	70	8	9	6	17	8	9
HPLC-30	19	4	26	6	4	8	13	5	8
UPLC-10	77	6	183	8	15	7	31	8	23
HPLC-10	222	7	138	8	21	7	46	8	38

[a] False positives: the *t*-test rejected null-hypothesis that was valid; i.e., the *t*-test incorrectly reports that it has found a positive result where none really exists.

Table 4. Concentration and Area Ratios for Spiked Features Obtained through XCMS, Manufacturer's Software, and the Described Protocol for Normalization and Reintegration[a]

compound	feature name (m/z − s)	[B]/[A]	XCMS raw area B/A	integration raw area B/A	full protocol area B/A
tacrine	M199T261	0.8	0.59	0.51	0.54
prednisone	M381T508	0	0.19	0.04	0.01
propafenone	M342T615	0	0.08	0.00	0.00
bis tacrine	M247T580	1.5	1.73	1.85	1.92
haloperidol	M376T550	1.5	1.25	1.34	1.40
nicotinic acid	M124T39	0.5	0.62	0.59	0.62
mifepristone	M430T608	1.2	1.11	1.13	1.18
sulpiride	M342T131	0.8	0.85	0.84	0.88
tenoxicam	nd[b]	0.5	nd[b]	0.89	nd[b]
piroxicam	M332T550	1.2	1.07	1.07	no diff[c]
phenylalanine $^2H_5$	M171T76	1	0.99	0.97	no diff[d]
atenolol	M267T117	1	0.84	0.82	no diff[d]
caffeine	M195T193	1	1.03	1.00	no diff[d]
tetracycline	M445T274	1	0.97	0.94	no diff[d]
chlorotetracycline	M479T378	1	1.04	1.07	no diff[d]
propranolol	M260T471	1	0.91	0.89	no diff[d]
verapamil	M455T633	1	1.04	1.05	no diff[d]
ketoprofen	nd[b]	1	nd[b]	0.95	no diff[d]
oleamide	M282T1419	1	1.02	1.01	no diff[d]

[a] Data obtained with UPLC 30-min gradient. [b] Feature not detected by XCMS. [c] Feature detected by XCMS but no significant difference ($p > 0.05$). [d] Feature spiked at same concentration in A and B, not detected as a difference.

normalization. As noted in Table 3, the normalization initially caused an increase in the number of false positives. However, a large reduction of false positives was achieved when the data were reintegrated and renormalized. The RSD distribution was kept fairly constant after normalization except for UPLC 30 min, where a shift toward smaller RSD values was noted (Table 2). This might reflect better chromatographic peak shapes, which potentially could be integrated more coherently and subsequently better-normalized. In Figure 3, a representative example of the effect of initial normalization on the XCMS output is shown for a feature that represents a compound (Mifepristone) that was increased 20% between (A) and (B). The feature M430T608 (Mifepristone) becomes significant only after normalization. An interesting aspect of our data processing protocol is that the known differences first detected as features by XCMS remained throughout the procedure whereas the number of false positives was reduced (Table 3). The initial increase of significant features is reduced when subjected to reintegration of the original data. After renormalization against the average integrated area for the same six standards previously used, there is again an increase in the number of false positives. But in all cases, more of the spiked differences are found after renormalization. The largest amounts of spiked compounds paired with the smallest amount of false positives are found in the case of running UPLC with the longer gradient (30 min). This observation makes UPLC a good choice for metabolomics.

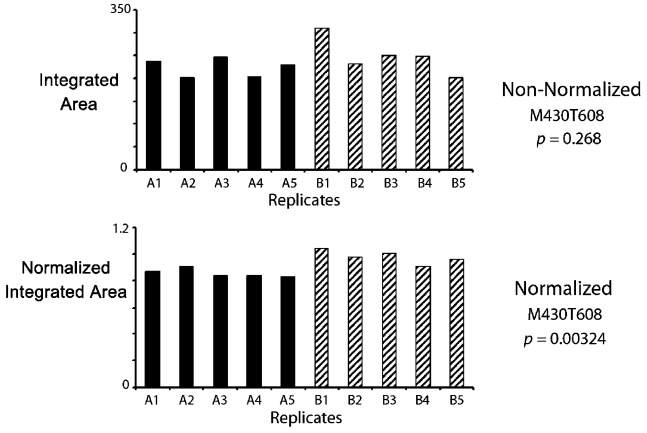

Figure 3. Effect of normalization. The top graph displays areas for feature M430T608 (Mifepristone) as integrated by XCMS. Letters A and B and the numbers correspond to the injection replicates of the respective spiking mixture. The bottom graph shows the normalized area values for the same feature. The concentration of mifepristone is increased 20% in mixture B (the dashed bars); see Table 1 for details.

Table 5. Resulting Table Displaying Differences Significant at $p < 0.05$ for UPLC 30-min Gradient

feature name (m/z − s)	p-value	compound
M199T261	1.35×10^{-7}	tacrine
M381T508	2.11×10^{-7}	prednisone
M342T615	3.59×10^{-7}	propafenone
M247T580	5.09×10^{-7}	bis tacrine
M429T1665	6.94×10^{-6}	unknown 1
M324T615	8.24×10^{-6}	propafenone[a]
M267T117	1.9×10^{-5}	atenolol
M376T550	1.93×10^{-5}	haloperidol
M124T39	5.88×10^{-5}	nicotinic acid
M455T633	0.0005	verapamil
M430T608	0.0012	mifepristone
M342T131	0.0016	sulpiride
M452T608	0.0033	false positive 1
M420T110	0.0102	false positive 2
M344T134	0.0177	false positive 3
M671T1636	0.0209	false positive 4
M217T31	0.0440	false positive 5

[a] In source fragment of propafenone.

Identified Differences. We performed a detailed trace of the fate of the spiked compounds in the UPLC 30-min experiment (Table 4). From these data, we concluded that XCMS preserves the aspect ratio of data as well as the manufacturer's software in the integration of raw data. The final area ratios reported obtained by XCMS and reintegration represented accurately the actual differences between (A) and (B). The final result of the entire processing protocol is displayed in Table 5. The fact that tacrine (concentration reduced by only 20%) gave the lowest p-value compared to prednisone and propafenone (both subjected to complete removal) illustrates that factors such as chromatography and S/N (Figure 2A) would be reflected in smaller standard deviations and hence, lower p-values. Nine out of the 12 top features identified as differences between (A) and (B) corresponded to spiked alterations when ranked according to p-value. For instance, the feature M324T615 is a source fragment from propafenone and the feature M429T1665 is found to be an actual difference with all separation setups. The feature M429T1665, however, is not a spiked component and does not originate from column bleed. It is also not a contaminant from any of the standards as it was not detected when standards were run individually. The spiked but nonchanged compounds atenolol and verapamil also appeared as false positives. The remaining hits in Table 5, i.e., false positives 1−5, were manually inspected in the original data. They all represented real chromatographic peaks. If a significance criterion is set to a lower p-value, some of these features would disappear together with some of the spiked compounds present only at a 20% concentration difference. This reflects a general problem with detecting differences in relative concentration levels that approximate the precision limit of the analytical protocol. An estimation of the analytical precision is shown in Table 2. A majority of the features shows an RSD in the range of 5−25%, and hence, this is a theoretical limit for detection of differences.

To address the issue of correcting for the occurrence of false positives, multiple test correction strategies can be used. These p-value correction calculations adjust the individual p-value for each feature to keep the overall error rate (or false positive rate) to less than, or equal to, the specified p-value cutoff. For example, the false discovery rate[30−32] has been helpful for setting p-value thresholds in gene expression studies. A more stringent correction is the use of Bonferroni multiple test correction, where the corrected p-value equals the initial p-value times the number of observed features.

CONCLUSION

The quantitative comparison of sample classes in a nontargeted fashion represents a challenging problem in metabolomics and proteomics.[5,8,15,16,18] For the determination of significant differences from multiple sample classes, the screening of features is best obtained using data mining such as ANOVA, partial least-squares projection, or latent structures discriminant analysis. Since all variables are aligned over the observations, nonlinear alignment offers preprocessing of the data that can be readily applied to these types of computations, ultimately allowing for more accurate quantitative analysis.

It is also evident that UPLC is a useful tool for LC−MS-based metabolomics since more features were detected at a higher S/N, which provides a better basis for peak finding, integration, and further statistical evaluation. The proposed protocol using UPLC−MS and XCMS with normalization/reintegration has been proven to detect 8 out of 10 spiked differences, some only differing at a relative concentration of 20%. Moreover, few false positives were identified, and a list was generated that represented compounds worthy of further identification efforts.

ACKNOWLEDGMENT

Work was supported by NIH grant MH62261 and DOE grant AC02-05CH11231. A.N. is supported by a postdoctoral fellowship from the Swedish Research Council (VR).

(30) Benjamini, Y.; Hochberg, Y. *J. R. Stat. Soc., Series B* **1995**, *57*, 289−300.
(31) Storey, J. D.; Tibshirani, R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9440−9445.
(32) Reiner, A.; Yekutieli, D.; Benjamini, Y. *Bioinformatics* **2003**, *19*, 368−375.

Received for review February 6, 2006. Accepted March 17, 2006.

AC060245F