IsoMS: Automated Processing of LC–MS Data Generated by a Chemical Isotope Labeling Metabolomics Platform

ARTICLE *in* ANALYTICAL CHEMISTRY · APRIL 2014

Impact Factor: 5.64 · DOI: 10.1021/ac5009089 · Source: PubMed

4 AUTHORS, INCLUDING:

Tao Huan
The Scripps Research Institute
12 PUBLICATIONS **65** CITATIONS

SEE PROFILE

analytical chemistry

Technical Note

pubs.acs.org/ac

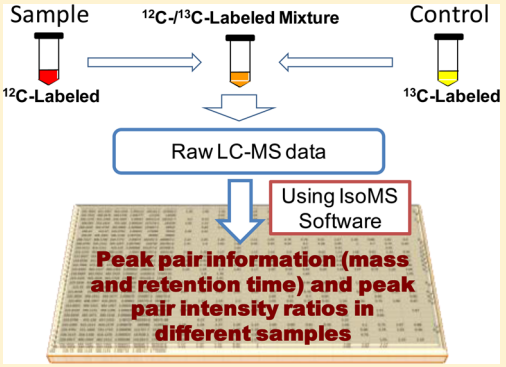
IsoMS: Automated Processing of LC-MS Data Generated by a Chemical Isotope Labeling Metabolomics Platform

Ruokun Zhou, Chiao-Li Tseng, Tao Huan, and Liang Li*

Department of Chemistry, University of Alberta, Edmonton, Alberta T6G2G2, Canada

Supporting Information

ABSTRACT: A chemical isotope labeling or isotope coded derivatization (ICD) metabolomics platform uses a chemical derivatization method to introduce a mass tag to all of the metabolites having a common functional group (e.g., amine), followed by LC-MS analysis of the labeled metabolites. To apply this platform to metabolomics studies involving quantitative analysis of different groups of samples, automated data processing is required. Herein, we report a data processing method based on the use of a mass spectral feature unique to the chemical labeling approach, i.e., any differential-isotope-labeled metabolites are detected as peak pairs with a fixed mass difference in a mass spectrum. A software tool, IsoMS, has been developed to process the raw data generated from one or multiple LC-MS runs by peak picking, peak pairing, peak-pair filtering, and peak-pair intensity ratio calculation. The same peak pairs detected from multiple samples are then aligned to produce a CSV file that contains the metabolite information and peak ratios relative to a control (e.g., a pooled sample). This file can be readily exported for further data and statistical analysis, which is illustrated in an example of comparing the metabolomes of human urine samples collected before and after drinking coffee. To demonstrate that this method is reliable for data processing, five $^{13}C_2$-/$^{12}C_2$-dansyl labeled metabolite standards were analyzed by LC-MS. IsoMS was able to detect these metabolites correctly. In addition, in the analysis of a $^{13}C_2$-/$^{12}C_2$-dansyl labeled human urine, IsoMS detected 2044 peak pairs, and manual inspection of these peak pairs found 90 false peak pairs, representing a false positive rate of 4.4%. IsoMS for Windows running R is freely available for noncommercial use from www.mycompoundid.org/IsoMS.



An isotopic analogue of a compound of interest as an internal standard is widely employed for LC-MS-based chemical quantification, because of its high accuracy and precision. However, due to limited availability of isotopic analogues of metabolites, this approach is not feasible for metabolomic profiling. Metabolic or *in vivo* stable isotope labeling can be used to generate isotopic analogues of metabolites, which has been combined with LC-MS for studying the dynamics of metabolic fluxes.[1,2] However, it requires cell culturing and thus cannot be readily applied for analyzing nonculturable samples (e.g., human biofluids). Alternatively, chemical derivatization can be used to introduce an isotope-coded mass tag to the metabolites that share a common functional group, such as amine, acid, phenol, etc. In this approach, a group-based submetabolome is labeled with a proper isotope labeling reagent for LC-MS analysis and the combined results of various submetabolomes represent a comprehensive profile of the whole metabolome. There are a number of isotope labeling reagents reported for labeling metabolites.[3−19] The use of a rational design of an isotope labeling reagent structure and simultaneous improvement in LC separation and MS detection, in addition to providing a mass tag, can be achieved to generate a more comprehensive profile of the submetabolome of a biological system, compared to conventional LC-MS methods.[11,12,20,21] Although several robust software tools have been developed for processing LC-MS data for metabolomics,[22−25] to our knowledge, there is no software available to handle the unique data sets generated by the chemical isotope labeling or isotope coded derivatization (ICD) LC-MS metabolomics platform. We also found that software tools used for quantitative shotgun proteomics involving differential isotope labeling of peptides are not useful for processing the metabolomics data. Adapting these tools for chemical isotope labeling metabolomics likely requires modifying the codes to accommodate the special needs of dealing with low mass and not identifiable analyte ions.

We have developed a data processing method and a software tool, IsoMS, for processing LC-MS data by exploring a mass spectral feature unique to the chemical labeling approach, i.e., metabolites are detected as peak pairs with a fixed mass difference in a mass spectrum. In a typical workflow of the chemical isotope labeling metabolomics platform, an individual sample is labeled with a light-chain reagent and a pooled sample produced by mixing aliquots of individual samples is labeled

Received: March 11, 2014
Accepted: April 25, 2014
Published: April 25, 2014

   dx.doi.org/10.1021/ac5009089 | *Anal. Chem.* 2014, 86, 4675−4679

with a heavy-chain reagent. The mixture of the light-chain labeled individual sample and the heavy-chain labeled pooled sample is analyzed by LC-MS. A true metabolite should be detected as a pair of peaks with a defined mass difference governed by the mass difference of the heavy- and light-chain tags, while all the noises and background peaks are detected as singlet peaks. The intensity ratio of a given peak pair is reflective of the relative concentration of the metabolite in an individual sample to that in the pooled sample. Thus, the peak ratio values, along with the metabolite information (retention time and peak mass), can be directly exported into a data or statistical analysis program for further analysis.

In this paper, we describe how the data processing method was implemented in IsoMS, the procedure for processing chemical isotope labeled LC-MS data, and the performance of the method for analyzing dansyl labeled metabolites and human urine samples. All the scripts and instructions including a user manual can be downloaded from www.mycompoundid.org/IsoMS.

EXPERIMENTAL SECTION

Urine Sample Processing. The second morning urine sample was collected each day from a healthy volunteer over five consecutive days. On each day, one urine sample before drinking coffee and two urine samples 1 or 2 h after drinking coffee were collected. Within 1 h of urine collection, the urine sample was centrifuged at 4,000 rpm for 10 min. The supernatant was filtered twice by a 0.22 μM-pore-size Millipore filter (Millipore Corp., MA) and aliquoted into 0.5 mL vials. The urine samples were stored at −80 °C. A pooled urine sample was prepared by combining the aliquots of daily urine samples. The individual urine sample was derivatized by $^{12}C_2$-dansyl chloride (DnsCl), and the pooled urine sample was derivatized by $^{13}C_2$-DnsCl, according to the reported protocol.[21] An aliquot of the $^{12}C_2$-labeled individual sample was mixed with an aliquot of the $^{13}C_2$-labeled pooled sample, and the mixture was then analyzed by LC-MS in triplicate (i.e., 15 × 3 or 45 LC-MS runs).

LC-MS. For the urine metabolome profiling work, the LC-MS analysis was performed using an Agilent 1100 series binary system (Palo Alto, CA) connected with a Bruker Apex-Qe 9.4-T Fourier transform ion cyclotron resonance (FTICR) MS (Bruker, Billerica, MA) equipped with electrospray ionization (ESI). LC separation was performed on an Agilent Zorbax Eclipse Plus C18 column (2.1 mm × 100 mm, 1.8 μm particle size). The experimental setup conditions were the same as those reported elsewhere.[21] IsoMS was used to process the 45 LC-MS data sets to generate the metabolome profiles for comparison. This method was also tested on LC-MS data generated by a Bruker Maxis quadrupole time-of-flight (QTOF), Agilent 6220 TOF-MS, and Waters Primer QTOF-MS.

RESULTS AND DISCUSSION

Figure 1 shows the overall workflow for processing LC-MS data generated by a chemical isotope labeling metabolomics platform. IsoMS can be used to process data generated from different MS instruments, as long as they provide adequate resolving power to resolve the peaks within an isotope peak pair. As Figure 1 shows, there are generally three steps involved in data processing. The first step is to convert the raw LC-MS data into a centroid peak list file that contains information on

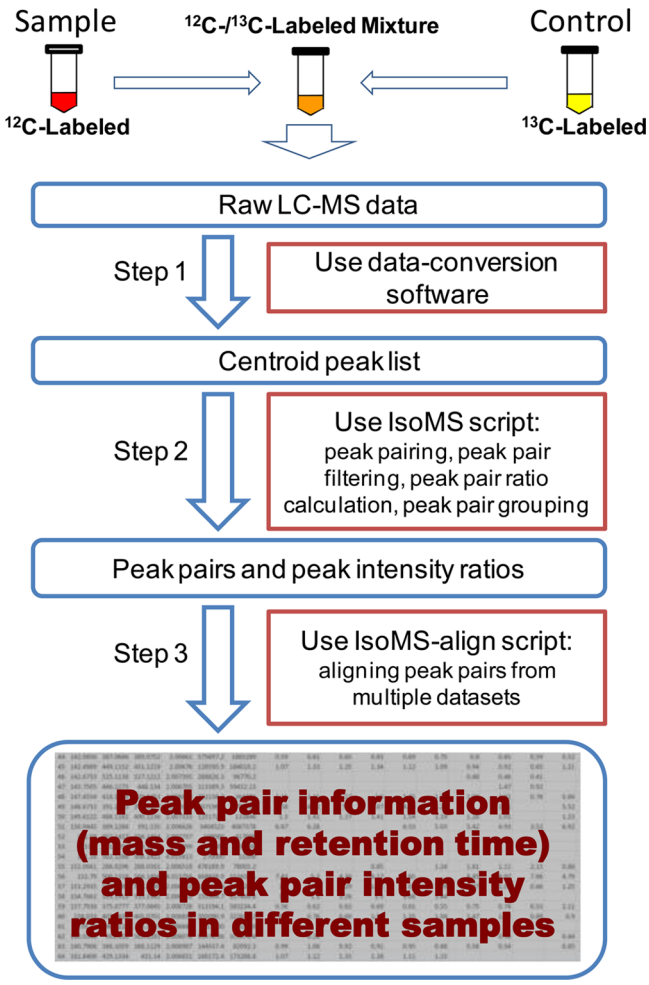


Figure 1. Workflow for IsoMS data processing.

retention time, m/z, and peak intensity. An example of the peak list file is shown in Figure S1, Supporting Information. All mass peaks with a user-defined intensity threshold are retained in the peak list file. This conversion can be done using the data analysis software of the manufacturer. The instruction (and a script for Bruker data) for converting the Bruker, Waters, or Agilent data can be downloaded from the MyCompoundID Web site. For Bruker FTICR-MS data, the peak list file also contains information on signal-to-noise ratio and peak width of a mass peak (see Figure S2, Supporting Information, as an example), which is useful for removing noisy spectra (see below). While there are software tools available for extracting and aligning peaks, IsoMS uses the raw peak list data as the starting point for further data processing. This should provide a more uniform approach in dealing with the chemical isotope labeling metabolome data set, independent of any second party software tools which may extract or align peaks differently from one software to another.

After the file conversion, all the peak list files from multiple LC-MS runs are placed in one folder, which are then automatically processed by IsoMS (i.e., Step 2 in Figure 1) to generate the CSV files containing information on sample identifier, retention time and m/z's of individual peak pairs, and peak ratio. An example is shown in Figure S3, Supporting Information, for the urine metabolome profiling work. The IsoMS algorithm processes the peak list data in the following manner.

dx.doi.org/10.1021/ac5009089 | Anal. Chem. 2014, 86, 4675−4679

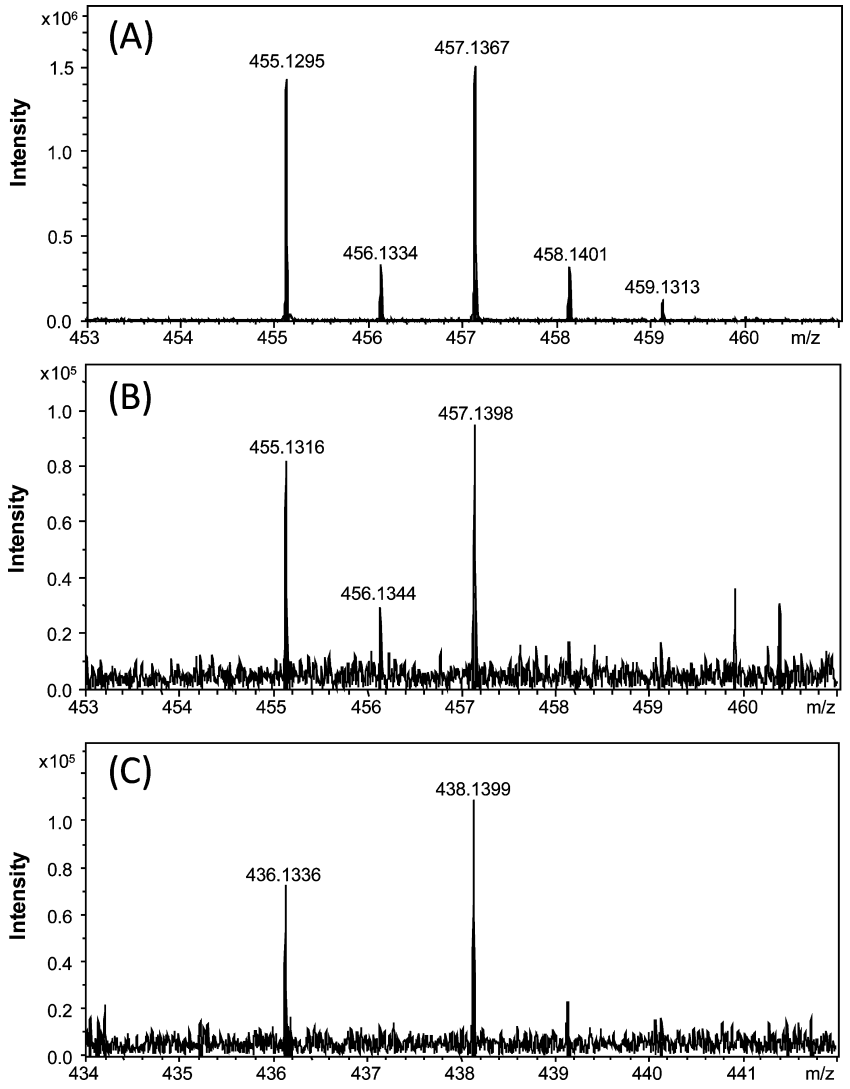


Figure 2. Mass spectra showing (A) Level 1 peak pair (both natural isotope peaks, 456.1334 and 458.1401, are detected in the peak pair of 455.1295 and 457.1367), (B) Level 2 peak pair (one of the natural isotope peaks, 456.1344, in the peak pair of 455.1316 and 457.1398, is detected), and (C) Level 3 peak pair (none of the natural isotope peaks in the peak pair of 436.1336 and 438.1399 is detected).

Peak Pairing. IsoMS determines the charge state and isotope distribution of each ion in a mass spectrum. Two ions having a user-defined mass difference (e.g., 2.0067 for $^{12}C_2/^{13}C_2$-dansyated metabolites) within a tolerance window (<5 ppm) are then paired to each other, if they have the same charge and isotopic pattern (Figure 2A). They are groups as Level 1 peak pairs. However, for low intensity peaks, one of the natural isotope peaks in the peak pair may be missing (Figure 2B). IsoMS detects and classifies them as Level 2 peak pairs. For very low abundance peaks, neither the natural isotope peaks within a peak pair is detected (Figure 2C). These ions are paired by IsoMS as Level 3 peak pairs. They are not as reliable as Levels 1 and 2 pairs, as the chance of random matches of two adjacent peaks increases significantly; in some cases, as high as 30% of these pairs were found to be false. In our current work, we only use Levels 1 and 2 peak pairs for further analysis, which usually gives less than 5% false positive rate (FPR) as determined by manual inspection of the peak pairs found (see an example shown below).

Peak Pair Filtering. A labeled metabolite with a purposely designed tag, such as a dansyl group, is often detected in LC-MS in the form of MH^+. However, other forms, including adduct ions from Na^+, K^+, or NH_4^+, dimers and other multimers, and in-source fragment ions (usually minus CO_2), can sometimes be detected, particularly when the labeled metabolite concentration is high. Fortunately, these ions are detected within the same mass spectrum as that of MH^+ and thus can be readily filtered out after peak pairing. It should be noted that, if a chemical labeling reaction produces a labeled metabolite that forms MH^+ plus relatively high abundant adduct ions in a mass spectrum, filtering out the adduct ions may affect the quantification accuracy. In this case, MH^+ can still be used for the initial quantification. However, in the final result, one may use the summed area of the MH^+ and adduct ion peaks for quantifying the significant metabolites in comparative samples. IsoMS also allows a user to remove any peak pairs deemed to be from the background or blank. This can be done by entering the peak masses to be removed in a CSV file, isMZBackground (see User Manual for instruction). For FTICR-MS data, some noisy spectra with much reduced resolution are occasionally acquired (up to 15 or ~2.5% spectra in one LC-MS run), likely due to space charge issue related to ion detection (see an example in Figure S4, Supporting

dx.doi.org/10.1021/ac5009089 | Anal. Chem. 2014, 86, 4675−4679

Information). IsoMS has a filter to remove these noisy spectra based on the comparison of peak widths (resolutions) of a noisy spectrum to the neighboring spectra.

Peak Pair Grouping. After peak pair filtering to retain the only MH^+ peak pair from a metabolite, the intensity ratio of the light- and heavy-chain labeled peaks can be calculated in each mass spectrum. Thus, IsoMS can be used to determine the relative concentration of a labeled metabolite even when it only shows up in one spectrum, which can happen for low abundance analytes detected at a limited spectral acquisition speed, during the LC-MS run. However, many metabolites are repeatedly detected in several spectra. IsoMS groups the peak pairs found in adjacent spectra according to the light-chain labeled metabolite mass with a user-defined m/z tolerance window (e.g.,10 ppm). The peak ratio and retention time of the most intense peak pair within a group are retained. Note that IsoMS does not correct for natural isotope abundance in calculating the peak ratio. In general, for low mass metabolites (e.g., <800), the peak intensity contribution of the natural abundance $^{13}C_2$ peak of the $^{12}C_2$-labeled metabolite to that of the $^{13}C_2$-labeled metabolite peak is usually less than 10%, which is much less than the threshold value used to call for a significant metabolite concentration change (e.g., 1.5- or 2-fold change).

Peak Pair Alignment and Final Result. For each LC-MS run, IsoMS generates a list of m/z values of peak pairs, retention times, and peak ratios of individual pairs as a CSV file (an example is shown in Figure S3, Supporting Information). This is the final result, if only one run is performed. For a metabolomics study involving the analysis of multiple samples, the same peak pairs found in two or more runs are aligned using the IsoMS-align script (i.e., Step 3 in Figure 1). This is done according to their mass and retention time within user-defined tolerance windows to produce the final result as a CSV file. For example, in the 45 LC-MS runs of the urine metabolome profiling work, the final CSV result is shown in Table T1, Supporting Information, in Excel. This file can be uploaded to a data or statistics tool for further analysis. As an example, Figure S5, Supporting Information, shows the two plots generated by principle component analysis (PCA) and partial least-squares discriminant analysis (PLS-DA) of the urine metabolomes using the MetaboAnalyst program. The separation of the two groups (i.e., before and after drinking coffee) can be readily seen.

In using IsoMS for processing LC-MS data, several parameters in the scripts (see User Manual) can be fine-tuned, depending on the experimental conditions used. This is useful for optimizing the sensitivity and specificity for peak pair detection and peak ratio calculation for different instruments. In our experience, after the initial adjustment of the parameters, there is no need to perform any further adjustment.

The performance of the IsoMS method can be examined using two testing samples. In one test, five standards, 7-hydroxy-coumarin, tryptamine, phenol, tryptophan, and *N*-methyl-D-phenylalanine, labeled by $^{12}C_2$-/$^{13}C_2$-dansyl chloride were analyzed by Bruker LC-FTICR-MS. Figures S6–8, Supporting Information, show the ion chromatogram, a peak pair, and the final result obtained by IsoMS, respectively. IsoMS was able to pick the expected peak pairs correctly.

In another test, a differentially dansyl labeled human urine was analyzed by Bruker LC-QTOF-MS. The ion chromatogram is shown in Figure S9, Supporting Information. IsoMS detected 2044 peak pairs. We then manually inspected these peak pairs and found 90 false ones. Thus, the false positive rate was 4.4%, indicating that good specificity (<5%) could be obtained by IsoMS. Among the 2044 peak pairs, 1388 pairs belonged to Level 1 peak pairs and 736 pairs belonged to Level 2 peak pairs. Out of the 90 false peak pairs, 24 and 77 pairs were found in Level 1 and Level 2 peak pairs, respectively. Most of the false pairs had low signal intensities. Positive identification of the detected peak pairs is beyond the scope of this work. However, on the basis of the accurate mass information (<5 ppm), we searched the Human Metabolome Database (HMDB)[26] and the Evidence-based Metabolome Library (EML)[27] using MyCompoundID[27] to generate a list of putative metabolites (see Tables T2 and T3, Supporting Information). Among the 2044 peak pairs detected, 908 matched the metabolites in HMDB and 657 matched the predicted metabolites in EML with one metabolic reaction. Thus, 76.6% of the peak pairs could be matched.

CONCLUSIONS

We have developed a method based on peak pair picking and filtering in mass spectra to process LC-MS data generated by a chemical isotope labeling or isotope coded derivatization (ICD) metabolomics platform. This method has been implemented in a software tool, IsoMS, that can be freely downloaded from www.mycompoundid.org/IsoMS. IsoMS can be used to process the raw LC-MS data sets of multiple samples into a CSV file that can be readily exported for further data and statistical analysis. From the analysis of $^{12}C_2$-/$^{13}C_2$-dansyl labeled metabolite standards and human urine, we demonstrated that IsoMS is a reliable tool for processing chemical isotope labeling LC-MS data. We will continue to develop IsoMS by implementing other functionalities, such as automated statistical analysis and metabolome database search for compound identification.

ASSOCIATED CONTENT

Supporting Information

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: Liang.Li@ualberta.ca.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was funded by the NSERC, CIHR, CRC, Genome Canada, and Alberta Innovates. We thank Dr. Xianmin Hu, Yiman Wu, Dr. Jun Peng, Ronghong Li, Chunqu Tang, Professor Guohui Lin, and other Li group members for their contributions to this work.

REFERENCES

(1) Creek, D. J. *Bioanalysis* **2013**, *5*, 1807–1810.
(2) Bennett, B. D.; Yuan, J.; Kimball, E. H.; Rabinowitz, J. D. *Nat. Protoc.* **2008**, *3*, 1299–1311.
(3) Bruheim, P.; Kvitvang, H. F. N.; Villas-Boas, S. G. *J. Chromatogr., A* **2013**, *1296*, 196–203.
(4) Guo, K.; Ji, C. J.; Li, L. *Anal. Chem.* **2007**, *79*, 8631–8638.
(5) O'Maille, G.; Go, E. P.; Hoang, L.; Want, E. J.; Smith, C.; O'Maille, P.; Nordstrom, A.; Morita, H.; Qin, C.; Uritboonthai, W.;

Apon, J.; Moore, R.; Garrett, J.; Siuzdak, G. *Spectrosc.: Int. J.* **2008**, *22*, 327−343.
(6) Shortreed, M. R.; Lamos, S. M.; Frey, B. L.; Phillips, M. F.; Patel, M.; Belshaw, P. J.; Smith, L. M. *Anal. Chem.* **2006**, *78*, 6398−6403.
(7) Yang, W. C.; Regnier, F. E.; Sliva, D.; Adamec, J. *J. Chromatogr., B* **2008**, *870*, 233−240.
(8) Shanaiah, N.; Desilva, M. A.; Gowda, G. A. N.; Raftery, M. A.; Hainline, B. E.; Raftery, D. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 11540−11544.
(9) Abello, N.; Geurink, P. P.; van der Toorn, M.; van Oosterhout, A. J. M.; Lugtenburg, J.; van der Marel, G. A.; Kerstjens, H. A. M.; Postma, D. S.; Overkleeft, H. S.; Bischoff, R. *Anal. Chem.* **2008**, *80*, 9171−9180.
(10) Giavalisco, P.; Kohl, K.; Hummel, J.; Seiwert, B.; Willmitzer, L. *Anal. Chem.* **2009**, *81*, 6546−6551.
(11) Guo, K.; Li, L. *Anal. Chem.* **2009**, *81*, 3919−3932.
(12) Guo, K.; Li, L. *Anal. Chem.* **2011**, *82*, 8789−8793.
(13) Armenta, J. M.; Cortes, D. F.; Pisciotta, J. M.; Shuman, J. L.; Blakeslee, K.; Rasoloson, D.; Ogunbiyi, O.; Sullivan, D. J.; Shulaev, V. *Anal. Chem.* **2010**, *82*, 548−558.
(14) Tang, Z. M.; Guengerich, F. P. *Anal. Chem.* **2010**, *82*, 7706−7712.
(15) Huang, Y. Q.; Liu, J. Q.; Gong, H. Y.; Yang, J.; Li, Y. S.; Feng, Y. Q. *Analyst* **2011**, *136*, 1515−1522.
(16) Wang, H.; Wang, H. Y.; Zhang, L.; Zhang, J.; Guo, Y. L. *Anal. Chim. Acta* **2011**, *690*, 1−9.
(17) Dai, W. D.; Huang, Q.; Yin, P. Y.; Li, J.; Zhou, J.; Kong, H. W.; Zhao, C. X.; Lu, X.; Xu, G. W. *Anal. Chem.* **2012**, *84*, 10245−10251.
(18) Zhang, S. J.; You, J. M.; Ning, S. J.; Song, C. H.; Suo, Y. R. *J. Chromatogr., A* **2013**, *1280*, 84−91.
(19) Tayyari, F.; Gowda, G. A. N.; Gu, H. W.; Raftery, D. *Anal. Chem.* **2013**, *85*, 8715−8721.
(20) Peng, J.; Li, L. *Anal. Chim. Acta* **2013**, *803*, 97−105.
(21) Zhou, R. K.; Guo, K.; Li, L. *Anal. Chem.* **2014**, *85*, 11532−11539.
(22) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779−787.
(23) Melamud, E.; Vastag, L.; Rabinowitz, J. D. *Anal. Chem.* **2010**, *82*, 9818−9826.
(24) Bueschl, C.; Kluger, B.; Berthiller, F.; Lirk, G.; Winkler, S.; Krska, R.; Schuhmacher, R. *Bioinformatics* **2012**, *28*, 736−738.
(25) Chokkathukalam, A.; Jankevics, A.; Creek, D. J.; Achcar, F.; Barrett, M. P.; Breitling, R. *Bioinformatics* **2013**, *29*, 281−283.
(26) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M. A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; MacInnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. *Nucleic Acids Res.* **2007**, *35*, D521−D526.
(27) Li, L.; Li, R. H.; Zhou, J. J.; Zuniga, A.; Stanislaus, A. E.; Wu, Y. M.; Huan, T.; Zheng, J. M.; Shi, Y.; Wishart, D. S.; Lin, G. H. *Anal. Chem.* **2013**, *85*, 3401−3408.