Chapter 2

Global Metabolic Profiling Using Ultra-Performance Liquid Chromatography/Quadrupole Time-of-Flight Mass Spectrometry

Yunpeng Qi, Yunlong Song, Haiwei Gu, Guorong Fan, and Yifeng Chai

Abstract

Currently, liquid chromatography-mass spectrometry (LC-MS) is one of the most important analytical technologies for detecting hundreds of metabolites in the field of metabolomics. A recent advance in LC that has impacted metabolomics is the development of UPLC (ultra-performance liquid chromatography). In this chapter, we describe the analytical methodologies for the global metabolic profiling of serum, urine, and tissue samples using UPLC-Q-TOF (quadrupole-time-of-flight)-MS. Aqueous metabolites are extracted after adding methanol/acetonitrile/acetone and then analyzed by UPLC-MS under positive and/or negative ionization mode. With the aid of multivariate statistical analysis, separation between various groups can be observed in the score plots, and biomarkers are screened in the loading/weight/VIP (variable importance in the projection) scatterplots. Furthermore, putative markers can be identified through comparison with the authentic standards based on tandem mass spectrometry (MS/MS) fragmentation pattern and LC retention. We expect that our protocol, with modifications if necessary, can be useful in many metabolomics studies and a wide range of research areas related to small molecules and LC-MS.

Key words Global metabolic profiling, Liquid chromatography, Mass spectrometry, Metabolomics, Quadrupole-time-of-flight

1 Introduction

Metabolomics aims to reveal various metabolic characteristics of external or internal perturbations to biological systems by profiling low-molecular-weight metabolites (typically <1,000 Da) in biosamples [1–9]. For example, the human metabolome currently comprises over 2,100 endogenous metabolites, including amino acids, amines, sugars, and organic acids [10]. These metabolites are the end products of cellular regulatory processes, and their levels can be regarded as the ultimate response of biological systems to genetic or environmental changes, from which important metabolic biomarkers may be discovered for disease diagnosis or
important biological mechanistic information [11]. Meanwhile, as a complementary field to genomics and proteomics, metabolomics can be integrated with other omics studies to achieve a comprehensive understanding of complicated biological systems. Over the last two decades, the field of metabolomics has rapidly developed with successful applications in various research areas including toxicology, disease detection, functional genomics, pharmacology, foods, and nutrition [1–8, 12–15].

Given the chemical diversity of metabolites, comprehensive and high-quality analysis is crucial for the global characterization of biological samples. Currently, $^1$H nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) remain the dominant analytical platforms in metabolomics [2, 4, 6, 16]. NMR is rapid and nondestructive, requires little or no sample preparation, and provides highly reproducible and quantitative results. Compared with NMR, the intrinsic high sensitivity (typically pg level) and selectivity of MS make it a powerful tool for detection, quantitation, and elucidation of small molecules in complex biological fluids. MS methods coupled with prior separation modalities such as gas chromatography (GC) and liquid chromatography (LC) provide enormous chemical information for metabolomics studies. In particular, LC-MS-based metabolic profiling can provide sensitive, accurate, and reproducible analysis of a wide range of metabolite classes [17–19]. An advantage of LC is that chemical derivatization is not required, making sample preparation and analysis relatively simple. In LC, reversed-phase (RP) columns, e.g., C18, are by far the most utilized in metabolic analysis.

In recent years, ultra-performance liquid chromatography coupled to mass spectrometry (UPLC-MS) has been often employed for the detection and characterization of small organic molecules in biological materials [20–24], offering improved resolution, greater sensitivity, and higher speed. UPLC uses smaller particles (sub-2 μm) and higher pressure (12,000–15,000 psi), which makes it feasible to carry out faster and more efficient separation [25]. In terms of the mass analyzer, a quadrupole-time-of-flight (Q-TOF) mass spectrometer is often coupled to UPLC to provide a large dynamic range for nontargeted data collection, and tandem mass capability for improved selectivity and unknown identification. Notably, electrospray ionization (ESI) is the most employed ionization method embedded in UPLC-MS. In metabolomics, UPLC-ESI-Q-TOF MS has been regarded as one of the premier tools for characterizing global metabolic profiles in complex biological samples (see Note 1). In addition, UPLC-MS can generate up to thousands of signals from hundreds of endogenous metabolites; therefore, multivariate statistical analysis has become an essential part of metabolomics to reduce the data dimensionality, differentiate similar spectra, build predictive models, etc. [26, 27].
This chapter focuses on describing a UPLC-MS-based global metabolomics protocol, with applications to serum, urine, and tissue samples. The major steps in metabolic profiling are introduced, including sample preparation, instrumental operation, data processing and analysis, and metabolite biomarker identification. We believe that our protocol is a good starting point to begin a metabolic profiling study, and with modifications (as necessary) it can be useful in many metabolite-related research areas.

2 Materials

2.1 Samples

1. 100 μL serum, plasma, or urine; 50 mg tissue (see Note 2).

2.2 Chemicals and Standards

All chemicals and solvents should be either of analytical or mass spectrometric grade. Deionized (DI) water (18.2 MΩ cm at 25 °C) can be purchased or produced in-house using one of a number of purification systems.

1. Protein precipitation solution: Methanol/acetonitrile (ACN)/acetone (1:1:1, v/v/v).
2. Chlorpropanamide (internal standard, 5 μM in the protein precipitation solution).
3. RP Solvent A: 0.1 % formic acid in DI H2O.
4. RP Solvent B: ACN containing 0.1 % formic acid.

2.3 UPLC-MS System

1. ACQUITY UPLC (Waters Corp., Milford, MA).
2. LC column: Reverse-phase 2.1 mm×50 mm ACQUITY UPLC BEH C18 1.7 μm column (Waters Corp.) (see Note 3).
3. Quadrupole-time-of-flight (Q-TOF) Premier mass spectrometer (Waters Corp.).

2.4 Other Instrumentation

1. Ultrasonication bath (Fritsch, Laborette 17, Idar-Oberstein, Germany).
2. Retsch Mixer Mill MM400 homogenizer (Retsch GmbH, Haan, Germany).
3. Vortexer.
5. Eppendorf vials (Eppendorf, Inc., Hauppauge, NY).
6. Tissue homogenizer, e.g., Retsch Mixer Mill MM400 homogenizer (Retsch GmbH, Haan Germany).

2.5 Data Processing Software

1. MassLynx software (Waters Corp.): This software is used to operate the UPLC-MS system and collect the LC-MS data.
2. MarkerLynx software (Waters Corp.): This software is used to process the collected LC-MS data. After processing, the
data are in a matrix format that is ready for further statistical analysis.

3. SIMCA-P V13.0 (Umetrics, Sweden): This software is used for multivariate statistical analysis.

3 Methods

The UPLC-MS-based global metabolomics protocol consists of steps including sample preparation, instrumental operation, data processing and analysis, and metabolite marker identification (see Fig. 1) [28]. Briefly, metabolites are extracted and proteins are precipitated after adding organic solutions, and then the samples are analyzed by UPLC-MS in positive (ESI+) and/or negative (ESI−) ionization mode. To avoid potential bias, it is important that the samples are randomized prior to sample preparation and analytical measurements.

3.1 Sample Preparation

3.1.1 Serum, Plasma, or Urine Samples

1. Take 100 μL of each sample into a 1.5 mL Eppendorf vial. Record the sample ID for each sample.
2. Add 400 μL of the protein precipitation solution containing 5 μM of the chlorpropamide internal standard to each of the samples.

Fig. 1 Workflow for UPLC-MS-based global metabolic profiling
3. After vigorous shaking for 1 min and incubation on ice for 10 min, centrifuge the mixture at 15,000 \( \times g \) for 15 min at 4 °C to precipitate proteins.

4. Collect the supernatant and transfer it to an Eppendorf tube for UPLC-MS analysis (see Note 4). Depending on preference, the same amount of supernatant, e.g., 475 \( \mu L \), can be collected, or the maximum amount of supernatant can be collected, which needs more experience and caution.

5. To evaluate reproducibility and stability of the UPLC-MS system, 20 \( \mu L \) of each sample is mixed to generate a pooled quality control (QC) sample, which is then treated using the above method and analyzed together with the real samples. Relative standard deviations (RSDs) of the same ten peak areas from the QC sample that is repetitively run every five injections are measured.

### 3.1.2 Tissue Samples

1. Weight approximately 50 mg of the tissue into a 1.5 mL Eppendorf vial, and record the sample ID.

2. Homogenize the tissue in 400 \( \mu L \) of the protein precipitation solution containing 5 \( \mu M \) of the chlorpropamide internal standard.

3. The sample is vortexed for 1 min and centrifuged at 15,000 \( \times g \) for 15 min at 4 °C to remove particles and precipitate proteins.

4. The supernatant (~370 \( \mu L \)) is collected and then used for UPLC-MS analysis (see Note 4).

5. Take 20 \( \mu L \) of each supernatant sample for the QC sample as discussed above in Subheading 3.1.1.

### 3.2 UPLC-MS Analysis

1. Prepare the UPLC-MS system according to the manufacturer’s guidelines. Prime and purge the LC system to exclude air bubbles. Warm up the MS detector for at least 15 min, until a stable baseline is obtained. Calibrate the MS detector, and a mass accuracy of <5 ppm should be achieved. Equilibrate the column with the initial mobile-phase composition (100 % Solvent A) for at least 15 min (this can be performed simultaneously with the MS preparation).

2. The UPLC gradient is as follows: start with 100 % Solvent A for 0.5 min, increase to 100 % B over the next 7.5 min, and return to 100 % A during the last 2 min. A typical chromatogram is shown in Fig. 2.

3. Set the column temperature to 60 °C.

4. Set the flow rate to 0.400 mL/min.

5. Maintain the capillary voltage and cone voltage at 3 kV and 20 V, respectively.
6. Set the source temperature and desolvation temperature at 120 °C and 350 °C, respectively.

7. The Q-TOF mass spectrometer is operated for positive (for amino acids, amines, sugars, etc.) and/or negative (for organic acids, etc.) ionization in full-scan mode with \( m/z \) ranging from 65 to 1,000.

8. After setting up the UPLC-MS conditions, execute a sample work list starting with a blank injection, a QC sample, and five completely randomized samples, and continuing with this order until the end of experiments.

9. Set the injection volume to be 2 μL (the injection volume can be adjusted depending on the criteria such as the number of detectable metabolites and LC elution).

10. For tandem mass spectrometry (MS/MS) experiments for better selectivity or for identifying unknown species, create scheduled precursor ion lists in the data acquisition software. In general, collision energies of 10, 25, and 40 V can be used to collect and monitor MS/MS spectra for both positive and negative ionization.

3.3 Data Processing and Analysis

There are many algorithms and software packages for effective in silico data preprocessing, including peak picking, peak deconvolution, peak alignment, etc. [29]. Generally, the instrument manufacturer will provide software packages to process raw data and generate a list of ion intensities for each detected peak.

1. Chromatographic and spectral data can be deconvoluted using MarkerLynx software. A multivariate data matrix containing information on sample identity, ion identity (retention time and \( m/z \)), and ion abundance is generated after centroiding, deisotoping, filtering, peak recognition, and integration (see Fig. 3).
MarkerLynx parameters, such as retention time (RT) and m/z ranges, are set to match those from MS detection. The mass tolerance is set to 0.01 Da, the RT window is 0.15 min, and isotopic peaks are excluded from the analysis. For integration, peak width at 5 % height is automatic and 6 s, intensity threshold is 50 counts for positive detection and 25 counts for negative detection, and the noise elimination level is set to 6.00. Notably, the parameters in data processing software packages can be optimized with different algorithms such as using internal standards [29–32].

2. The intensity of each ion is normalized by the total ion count integrated over the whole chromatogram. This method can at least partially eliminate the variation caused by overall fluctuations of the MS signals. There are numerous other normalization methods that exist, including using internal standard(s), normalizing to the corresponding metabolites detected in QC samples, etc.

3. The data matrix is further exported into SIMCA-P software and preprocessed using Pareto scaling and mean centering which increases the importance of low-abundance ions without
significant amplification of noise. Multivariate statistical analysis and pattern recognition analysis include principal components analysis (PCA), partial least squares-discriminant analysis (PLS-DA), orthogonal projections to latent structures-discriminant analysis (OPLS-DA), logistic regression, and support vector machines (SVMs) (see Note 5).

4. Separation between various groups can be observed in the PCA/PLS score plots (see Fig. 4a for an example of a PLS-DA score plot), and biomarkers are screened by analyzing ions in the loadings/weight/VIP (variable importance in the projection) plots (see Fig. 4b for an example of a PLS-DA loading plot),

Fig. 4 Illustration of a score plot (a) and a loading plot (b) from a PLS-DA model generated from SIMCA-P software. In the loading plot the ions scattering away from the origin (circled by a red ellipse) are regarded as the marker ions which contribute significantly to the clustering of the groups in (a).
in which the ions located away from the origin contribute significantly to the clustering of the groups and are regarded as the biomarker candidate ions.

3.4 Metabolite Identification and Quantitation

1. Metabolomics databases, e.g., METLIN (http://metlin.scripps.edu/index.php), Madison Metabolomics Consortium Database (http://mmcd.nmrfam.wisc.edu/), and HMDB (http://www.hmdb.ca/) can be searched to find potential candidates using the retention time and \( m/z \) information of the biomarker candidate ions.

2. Authentic standards are compared with the metabolite candidates based on MS/MS fragmentation pattern and retention time to confirm the identities of putative markers (see Fig. 5).

3. Concentrations of the metabolites are determined based on standard calibration curves using authentic standards. For better accuracy, standard addition or internal standards can be used for quantification.

![Fig. 5 Identification of a potential biomarker ion ([M + H]\(^+\), \( m/z = 205.0977 \)](image)

(a) Extracted ion chromatogram (EIC) of \( m/z 205.0977 \) (RT = 4.16 min); (b) MS/MS spectrum of the ion; (c) MS/MS spectrum of a commercial standard L-tryptophan
Metabolic pathway interpretation is an important step in metabolomics for connecting different omics domains in systems biology; however, it is highly dependent on the study purposes. Besides biological experience, a number of software packages are available to assist pathway interpretation, including IPA (Ingenuity Systems Inc., Redwood City, CA), MetaboAnalyst (www.metaboanalyst.ca/MetaboAnalyst/faces/Home.jsp), Genespring (Agilent Technologies, Inc., Santa Clara, CA), and Cytoscape (http://www.cytoscape.org/).

4 Notes

1. Different from targeted analyses, the global metabolic profiling described here aims to detect as many metabolites as possible, so as to characterize the overall metabolic profile of the biological system of interest. However, many factors may affect the outcomes of LC-MS-based metabolic profiling, including the complexity of biological systems, analyte coelution, ion suppression, in-source fragmentation, and adduct formation.

2. In a typical comparative metabolomics study, samples are from different groups, e.g., disease and healthy control. The volume/weight of samples can be adjusted, mainly depending on the sensitivity of UPLC-MS system; however, the volume/weight for all the samples of different groups should be at least roughly the same to avoid introducing bias.

3. Currently, most metabolite separations are performed on C18 reversed-phase (RP) columns to retain and separate medium-polar and nonpolar metabolites. C18 columns, although offering excellent resolution for hydrophobic metabolites, are not particularly good for the separation of hydrophilic metabolites. For very polar metabolites such as sugars, nucleosides, and hydrophilic amino acids, hydrophilic interaction liquid chromatography (HILIC) can provide a complementary method for metabolic profiling. Therefore, for the aqueous extracts, after initial RPLC profiling, it is beneficial to reanalyze the samples by HILIC to obtain improved coverage of more polar metabolites. For HILIC, the initial chromatographic condition can be a high proportion of organic solvent (<95 %), which may require adjustment of the sample solvent to ensure analyte retention. It should be noted that additional or modified sample preparation may be needed to ensure reproducible HILIC results. The coupling of two or more different columns could lead to much better separation and more comprehensive profiling [28, 33, 34].

4. The samples from this step can be directly used for LC-MS injection. Notably, many studies further dry the samples and reconstitute them into the LC-matched solvent prior to
LC-MS analysis. While it may improve LC elution, this method may cause more interference/degradation and is more laborious. Sample preparation is extremely important in metabolomics studies [35]; however, it is impossible to evaluate and validate a global metabolite profiling method for each metabolite present in the samples. Herein, we are providing an initial protocol for UPLC-MS-based metabolomics, and we encourage examination and optimization of the sample preparation protocol to ensure the reliability of metabolite profiles derived from the users’ samples of interest.

5. A detailed discussion of multivariate statistical analysis in metabolomics is beyond the scope of this chapter. Briefly, the statistical analysis approaches can be categorized as either unsupervised or supervised, and PCA (unsupervised) and PLS (supervised) are the two most commonly used methods in metabolomics. Unsupervised methods do not require information about sample membership (such as disease or control) prior to analysis. On the other hand, supervised methods analyze the spectra with the knowledge of class membership of the biological specimens. Supervised methods have the unique ability to build predictive models. The model is first built based on the training sample set and then applied to a test sample set to classify unknown samples (such as healthy control or disease). It is important to note that rigid cross validation is required for supervised methods such as PLS-DA, because supervised methods tend to achieve too optimistic class clusters when large numbers of variables are involved. In addition, univariate data analysis (such as the Student’s t-test) can assist variable selection prior to multivariate statistical analysis.

Acknowledgments

This work was supported by the National Science and Technology Major Project of the Ministry of Science and Technology of China (2013ZX09103-001), the National Natural Science Foundation of China (No. 21365001), the Chinese National Instrumentation Program (No. 2011YQ170067), and Li-Shi-Zhen Young Scientist Project of School of Pharmacy at the Second Military Medical University, Shanghai, China.

References

5. Gu H, Gowda GAN, Raftery D (2012) Metabolic profiling: are we en route to better diagnostic tests for cancer? Fut Oncol 8:1207–1210
overview of MS data handling. J Sep Sci 33:290–304


Mass Spectrometry in Metabolomics
Methods and Protocols
Raftery, D. (Ed.)
2014, XVI, 360 p. 71 illus., 52 illus. in color., Hardcover
A product of Humana Press