Analytica Chimica Acta 990 (2017) 96–102

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

An alignment algorithm for LC-MS-based metabolomics dataset assisted by MS/MS information

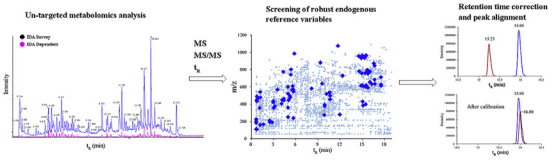
Lili Li [a, c, 1], Weijie Ren [b, 1], Hongwei Kong [a], Chunxia Zhao [a], Xinjie Zhao [a], Xiaohui Lin [b, **], Xin Lu [a, *], Guowang Xu [a]

[a] CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China
[b] School of Computer Science & Technology, Dalian University of Technology, Dalian 116024, China
[c] University of Chinese Academy of Sciences, Beijing 100049, China

HIGHLIGHTS

- A MS/MS-based peak alignment method for LC-MS metabolomics data was developed.
- A rigorous strategy for screening endogenous reference variables was proposed.
- MS/MS data were used to screen rigorous endogenous reference variables and in further peak alignment.
- The developed method had good performance, especially for metabolomics data with larger retention time drift.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 3 May 2017
Received in revised form
18 July 2017
Accepted 28 July 2017
Available online 4 August 2017

Keywords:
Liquid chromatography-mass spectrometry
Metabolomics
Peak alignment
Tandem mass spectrometry
Metabolic profiling

ABSTRACT

Liquid chromatography-mass spectrometry (LC-MS) is an important analytical platform for metabolomics study. Peak alignment of metabolomics dataset is one of the keys for a successful metabolomics study. In this work, a MS/MS-based peak alignment method for LC-MS metabolomics data was developed. A rigorous strategy for screening endogenous reference variables was proposed. Firstly, candidate endogenous reference variables were selected based on MS, MS/MS and retention time in all samples. Multiple robust endogenous reference variables were obtained through further evaluation and confirmation. Then retention time of each metabolite feature was corrected by local linear regression using the four nearest neighbor robust reference variables. Finally, peak alignment was carried out based on corrected retention time, MS and MS/MS. Comparing with the other two peak alignment methods, the developed method showed a good performance and was suitable for metabolomics data with larger retention time drift. Our approach provides a simple and robust alignment method which is reliable to align LC-MS metabolomics dataset.

1. Introduction

Metabolomics is a rapid developing field in post-genome era, which is an indispensable part of systems biology studies [1,2]. It has been widely used for disease studies [3], drug discovery [4] and

* Corresponding author.
** Corresponding author.
E-mail addresses: datas@dlut.edu.cn (X. Lin), luxin001@dicp.ac.cn (X. Lu).
[1] Lili Li and Weijie Ren contributed equally to this paper.

http://dx.doi.org/10.1016/j.aca.2017.07.058
0003-2670/

plant researches [5]. Liquid chromatography-mass spectrometry (LC-MS) is a most popular tool for metabolomics analysis [6,7] because of its good reproducibility, high resolution and wide metabolome coverage [8,9]. The primary step after LC-MS data acquisition is data deconvolution, peak detection and alignment before further statistical analysis. During an analytical batch, many unexpected factors may cause the drift of retention time and mass (m/z), such as environmental temperature, pressure and humidity fluctuation, changes in mobile phase pH, chromatographic column running time, sample matrix and ion suppression [10–12]. Shift of mass during an analytical batch is within an acceptable range with the rapid improvement of mass spectrometry technique, but drift of retention time is usually non-linear and different for each sample, which has a larger negative effect on further data analysis [13].

Many alignment methods have been developed, which can be classified into two major types, profile-based and feature-based methods [10,14,15]. The profile-based methods perform alignment before peak detection and use the raw eluting chromatograms [16,17]. The feature-based methods perform alignment after peak detection. XCMS includes two types alignment methods, obiwap and peakgroups, for alignment of profile matrix and features (peak groups), respectively [18,19]. Tsai et al. [20] proposed a Bayesian alignment model which integrated the information of both internal standards and clustered chromatograms. Amsrpm [21] implements robust point matching algorithm for both chromatograms and LC/MS peak list alignment without pre-specified reference variables. For most of feature-based alignment methods, reference variables were often used as landmarks for further retention time correction. Reference variables include exogenous internal standards and endogenous metabolites [20,22,23]. For a typical LC-MS metabolomics analysis, exogenous internal standards are spiked into all samples for process control [12]. A screen of suitable exogenous internal standards is very time-consuming, which increases the complexity of the operation. Stable-isotope-labeled exogenous internal standards are most convenient for LC-MS analyses, but they are less commercial available and more expensive. And often limited number of exogenous internal standards is added in metabolomics analysis [24,25]. Therefore, endogenous metabolites are more preferred as reference variables.

Selection of endogenous reference variables is a critical step for retention time correction. Several selection methods have been reported. XCMS (peakgroups method) binned “meta-peaks” with similar mass and retention time by kernel density estimator. Then alignment was performed based on peak groups (features) as temporary reference variables which can be found in most/all samples of an experiment [19]. COW-CODA developed component detection algorithm (CODA) to select “high quality” mass chromatogram with low noise and low background by the similarity between the smoothed and original data [26]. MetAlign firstly aligned mass peaks with the user-defined time window on the basis of amplitude and then selected mass peaks present in all data sets as reference variables [27]. MET-COFEA grouped molecular ion, adduct ions and isotope ions for each compound through retention time and peak shape similarities, and aggregated these annotated compound groups across all samples, then selected the highest intensity compounds as the reference variables in the middle of the alignment window to align those unaligned compounds [11]. Most of current algorithms select the endogenous reference variables based on molecular ion and retention time. Although numerous algorithms have been developed, the accuracy and adaptability of endogenous reference variables still don't meet the requirements of complex LC-MS dataset. Metabolites could be recognized according to their specific MS/MS fragmentation patterns. With the development of the high-resolution (HR) mass spectrometry, simultaneous data acquisition of HR precursor ion (MS) and fragment ion spectrum (MS/MS) data could be achieved [28,29]. The MS/MS data have been used for identification of peptides from database of protein, and then the identified peptides were aligned in proteomics dataset [30,31]. This method is not suitable for metabolomics data alignment due to the limitation of LC-MS metabolome databases. To the best of our knowledge, few MS/MS-based alignment methods have been applied for LC-MS metabolomics data.

In the study, a method for peak alignment of LC-MS metabolomics data was proposed. Metabolomics analysis was performed by LC coupled with HR MS with simultaneous data acquisition of MS and MS/MS data. Multi-endogenous reference variables were rigorously screened based on retention time, MS and MS/MS in all samples. Then, retention time corrections were carried out by local linear regression of multiple robust endogenous reference variables. The developed method was evaluated and compared with the other two peak alignment softwares (XCMS and MarkerView).

2. Experimental sections

2.1. Reagents and chemicals

Acetonitrile and methanol with HPLC grade were purchased from Merck (Darmstadt, Germany). Ultrapure water was obtained by a Milli-Q water system (Millipore, Billerica, USA). Formic acid was purchased from Sigma-Aldrich (St. Louis, MO, USA). Standards chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA), J&K chemicals (Shanghai, China), Avanti polar lipids Inc. (Alabaster, AL, USA).

2.2. Sample preparation

Rice leave powders of 20 mg were weighed into Eppendorf tube. Extraction solvent of 1 mL methanol/water (v/v, 4:1) was added and then vibrated for 6 min. The mixture was then centrifuged for 15 min and 400 μL supernatant was taken out and lyophilized. The lyophilized samples were redissolved by solvent of 80 μL acetonitrile/water (v/v, 1:4) and then waited for injection in sample bottles. Quality control (QC) sample was prepared by mixing equal weight samples to be analyzed.

2.3. Instrument analysis

Samples were analyzed by ACQUITY ultra performance liquid chromatography (UPLC) (Waters, Milford, MA, USA) coupled with TripleTOF 5600 mass spectrometer (AB SCIEX, Framingham, USA). ACQUITY UPLC BEH C_8 column (2.1 mm × 100 mm, 1.7 μm, Waters, Ireland) was used for separation. The mobile phase A was water with 0.1% formic acid. The mobile phase B was acetonitrile with 0.1% formic acid. The gradient elution was as follows: started at 5% B and maintained for 1 min, to 35% B at 8 min, to 60% B at 9 min, to 85% B at 13 min, to 100% B at 17 min and kept for 5 min, to 5% at 22.1 min and kept for 3 min. For MS conditions, the m/z range was 50–1200 Da. Insource temperature was 500 °C. Ionspray voltage floating was 5500 V. MS was operated in an Information Dependent Analysis (IDA) mode. An IDA cycle consisted of a 250 msec TOFMS survey scan, from which the top 12 ions were selected for auto MS/MS experiment with 25 msec accumulation time. The total cycle time was 0.6 s. The collision energy and the collision energy spread were 30 eV and 10 eV, respectively. Samples were acquired with electrospray ionization (ESI) in the positive mode.