Chinese Chemical Letters 25 (2014) 1331–1335

Contents lists available at ScienceDirect

Chinese Chemical Letters

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

Original article

A quick and effective multivariate statistical strategy for imaging mass spectrometry

Fei Tang [a], Yi Chen [a], Tie-Gang Li [b], Jiu-Ming He [b], Zeper Abliz [b], Gang Huang [c], Xiao-Hao Wang [a,*]

[a] State Key Laboratory of Precision Measurement Technology and Instruments, Department of Precision Instrument, Tsinghua University, Beijing 100084, China
[b] State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050, China
[c] National Key Laboratory of Human Factors Engineering, China Astronaut Research and Training Center, Beijing 100094, China

ARTICLE INFO

Article history:
Received 18 February 2014
Received in revised form 24 March 2014
Accepted 9 April 2014
Available online 2 May 2014

Keywords:
Imaging mass spectrometry
Multivariate statistical analysis
Potential biomarkers
Quick and effective

ABSTRACT

A new multivariate statistical strategy for analyzing large datasets that are produced by imaging mass spectrometry (IMS) techniques is reported. The strategy divides the whole datacube of the sample into several subsets and analyses them one by one to obtain the results. Instead of analyzing the whole datacube at one time, the strategy makes the analysis easier and decreases the computation time greatly. In this report, the IMS data are produced by the air flow-assisted ionization IMS (AFAI-IMS). The strategy can be used in combination with most multivariate statistical analysis methods. In this paper, the strategy was combined with the principal component analysis (PCA) and partial least square analysis (PLS). It was proven to be effective by analyzing the handwriting sample. By using the strategy, the m/z corresponding to the specific lipids in rat brain tissue were distinguished successfully. Moreover the analysis time grew linearly instead of exponentially as the size of sample increased. The strategy developed in this study has enormous potential for searching for the m/z of potential biomarkers quickly and effectively.

1. Introduction

In recent years, imaging mass spectrometry (IMS), a new field in mass spectrometry research, has developed rapidly. IMS images contain information on hundreds or even thousands of different molecules in the samples, and both molecular and spatial information can be obtained through a single sample analysis [1,2]. IMS can be used to track the spatial distribution of specific molecules in diseased and normal tissue. It plays an important role in various applications, such as medical diagnostics, physiological and pathological investigation, and biomarker assays [3–5].

Many IMS technologies have been developed, including matrix-assisted laser desorption ionization IMS (MALDI-IMS) [6,7], desorption electrospray ionization IMS (DESI-IMS) [8], and air flow-assisted ionization IMS (AFAI-IMS) [9], among others. Since the development of IMS, its pixel resolution and mass resolution have continuously improved, resulting in a large amount of mass spectrometry data. As a result, extracting, classifying, and sorting the effective information from the original mass spectrum data has become a focus of current research [10,11]. Due to the large size of mass spectrum data, it is difficult to determine particular m/z of biomarkers through manual screening and checking. Statistical analysis, data dimensionality reduction and extraction, image processing, and analysis are necessary before using IMS data for various applications.

Multivariate statistical analysis is a suitable method to achieve this purpose [12–16]. Multivariate statistics is a general statistical method used to process more than one variable. There are many different models of it, such as principal component analysis (PCA) [17,18], hierarchical cluster analysis (HCA) [19] and partial least square discriminate analysis (PLS-DA) [20]. Good results have been obtained when applying these methods to IMS data processing.

Most of the current methods for multivariate statistical analysis are used to analyze the complete mass spectrometry datacube of the entire sample at one time. However, this is a difficult and

* Corresponding author.
E-mail address: xhwang@mail.tsinghua.edu.cn (X.-H. Wang).

http://dx.doi.org/10.1016/j.cclet.2014.04.028
1001-8417/

time-consuming strategy because the whole datacube is too large [21,22].

In this paper, we propose a new strategy that can be used in combination with most of the multivariate statistical analysis methods. When applying this strategy, the whole mass spectrometry datacube of the sample was first divided into several subsets. Each of the subsets then was analyzed one by one by a particular multivariate statistical method (*e.g.* PCA) to get the initial results. The final multivariate statistical analysis result of the whole datacube was obtained by adding the entire initial results together. In doing so, the specific m/z reflecting the composition distribution of sample can be effectively determined.

PCA and PLS algorithms from MATLAB have been well developed and hence can be applied in most IMS technologies. These two methods were adopted in this paper as multivariate statistical methods to analyze the treated data. The strategy used in combination with these two multivariate statistical methods was validated by analyzing handwriting samples. The analysis results clearly show that the m/z of the two components in the sample was correctly extracted. The strategy was preliminarily applied for samples of rat brain tissue slices, and the m/z corresponding to tissue slice contours and particular lipids were determined. More importantly, the strategy reduced the analysis time drastically. By using this strategy, the analysis time grew linearly instead of exponentially as the amount of data to be analyzed increased.

This strategy has enormous potential for searching for the m/z of potential biomarkers both quickly and effectively. It can facilitate the research on the large and whole-body IMS technology, clinical applications, cancer diagnosis, *etc*. It extends the application fields of IMS technology.

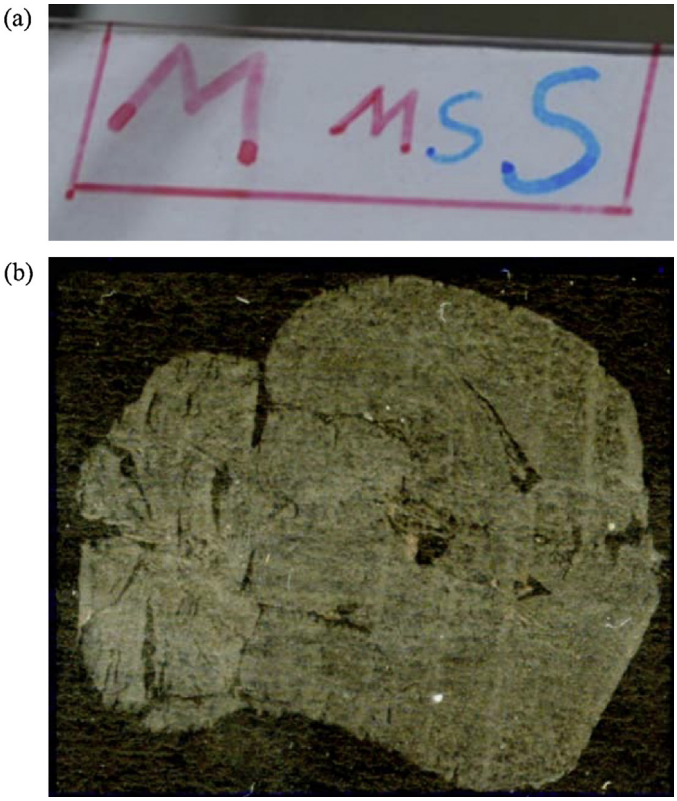

Fig. 1. Samples to be analyzed: (a) sample with character written in red and blue inks. (b) Rat brain tissue slices.

2. Experimental

2.1. Instruments

The IMS technology used in this paper was based on the AFAI-IMS. The experiments were performed using the QTRAP 5500 and QSTAR Elite QTOF mass spectrometers (AB SCIEX Foster City, CA, USA). The ion sources of the mass spectrometers were replaced with the AFAI source, which includes an ESI spray needle, an ion transport tube, a mass spectrometer interface, and a pump [23].

The ESI spray gas was N_2 and had a flow rate of 2 L/min. The voltage of the ESI needle was 5000 V, and the spray solution was the mixture of methanol and water (4:1, v/v) with 0.1% formic acid [24]. The spray solution was delivered to the needle with an Agilent LC pump with a liquid flow rate of 10 μL/min. The assisting air flow rate of the AFAI source was 40 L/min [25].

The data processing program was implemented using MATLAB, which integrated the imaging and multivariate analysis functions. The PCA and PLS algorithms were included in the program, including the import of mass spectrometry data, data reconstruction, and multivariate statistical analysis. The results of PCA and PLS analysis were given by scatter plots, with each data point representing one m/z. The value of the m/z corresponding to a specific point can be obtained by selecting it directly from the program.

2.2. Sample preparation

As shown in Fig. 1, sample (a) was used to validate the strategy and to compare the results of the PCA and PLS analysis. The letters "M" and "S" were written on a glass slide with either red or blue ink. The main components of the two inks were Rhodamine B (m/z = 443.2) and Basic Blue 7 (m/z = 478.4). The size of the glass slide was 150 mm × 50 mm, and analysis area was 100 mm × 30 mm, as shown in Fig. 1a. The range of the m/z was 100–500 when the sample (a) was analyzed.

Sample (b) was a rat brain tissue slice that contained some lipids, for which the m/z of the potential biomarkers was unknown. The rat was euthanized by ether overdose, frozen entirely in dry ice/isopentane, and then prepared for slicing. The Leica CM3600 cryomacrotome was used to obtain 20 μm rat brain tissue slices. The sample size was 15 mm × 14 mm and the slice was pasted on the glass slide. The range of the m/z was 500–999 when the sample (b) was analyzed.

2.3. The strategy

In this paper, the mass spectrometer collected data at regular intervals. The distance between two adjacent sample points analyzed by the mass spectrometer was 200 μm. In other words, the spatial resolution of AFAI-IMS was 200 μm. One mass spectrometry data corresponded to one sample point and contained the ion intensity of the m/z. The data of the whole sample formed a three-dimensional datacube (X, Y, m/z), as shown in Fig. 2. Each element of the datacube was the ion intensity corresponding to one particular m/z at one particular position (X, Y).

To make the analysis easier and reduce the analysis time, the strategy reported in this paper did not analyze the whole datacube of the sample at one time. Instead, the datacube was divided into several subsets at first, as shown in Fig. 2. In this study, a subset was simply defined as a dataset with the same X or the same Y. The data with the same X or Y corresponded to the sample points that were in the same column or row of the sample. So, one subset corresponded to one sample row or column.

All of the subsets were analyzed one by one by the specific multivariate statistical methods (*e.g.* PCA) to get the initial results.