Journal of Chromatography A, 1313 (2013) 245–252

Contents lists available at ScienceDirect

Journal of Chromatography A

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

# A comparative study of volatile components in green, oolong and black teas by using comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry and multivariate data analysis

Lei Zhang, Zhongda Zeng, Chunxia Zhao, Hongwei Kong, Xin Lu, Guowang Xu*

*CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China*

ARTICLE INFO

*Article history:*
Available online 19 June 2013

*Keywords:*
GC × GC–TOFMS
Structured chromatogram
Tea
Volatile components
Simultaneous distillation extraction

ABSTRACT

The difference of volatile components in green, oolong and black teas was studied by using comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry (GC × GC–TOFMS). Simultaneous distillation extraction was proved to be a suitable technique to extract the analytes with interest. A total of 450 compounds were tentatively identified with comparison to the standard mass spectra in available databases, retention index on the first dimension and structured chromatogram. 33 tea samples, including 12, 12 and 9 samples of green, oolong and black tea were analyzed by using GC × GC–TOFMS. After peak alignment, around 3600 peaks were detected. Partial least squares – discriminant analysis and hierarchical cluster analysis were used to classify these samples, then non-parametric hypothesis test (Mann–Whitney *U* test) and the variable importance in the projection (VIP) were applied to discover the key components to distinguish the three types of tea with significant difference amongst them. 74 differential compounds are defined to interpret the chemical differences of 3 types of tea. This study shows the power of GC × GC–TOFMS method combined with multivariate data analysis to investigate natural products with high complexity for information extraction.



## 1. Introduction

Tea is a kind of aromatic beverage with worldwide consumption. The attraction of tea comes from two aspects, one is the special taste and aroma, and the other is the benefits to health. The volatile components in tea play an important role to generate these characteristics, although the concentrations of most of them are not so high (about 100 mg $kg^{-1}$) [1]. In general, tea can be divided into the following six categories on the basis of the processing techniques, namely, white, yellow, green, oolong, black and post-fermented teas. Investigation of the volatile components in tea, and then origins of aroma may help to improve the quality of tea. Furthermore, it promotes the understanding and potential optimization of the manufacturing processes [2]. Although the study in this field was started from about 80 years ago [3], it is still active until now. Tea samples from different origins could be easily discriminated by the profiling of their volatile components [4]. Combining with other analytical technologies like high performance liquid chromatography and capillary electrophoresis, the volatile and nonvolatile components were compared in different types of tea [5], or among the homologous tea plants which grew in different environment [6], the results could assist in improving the flavor of tea.

In order to analyze the volatile components in tea, the first step is to find a suitable extraction technique [7]. Up to date, many extraction techniques have been developed to attain this goal including liquid–liquid extraction [3], steam distillation [8], solid phase micro extraction [9], supercritical fluid extraction [10], accelerated solvent extraction [11], purge and trap [12] and simultaneous distillation extraction (SDE) [13]. SDE merges common solvent extraction and vapor distillation extraction with a Likens–Nickerson apparatus, its main advantages are to simplify experimental procedures, save organic solvents and reduce loss of samples during the transfer process [7,14].

Traditionally, one-dimensional gas chromatography (1D GC) is a preferred tool to detect volatile components. However, its resolution is not enough to separate the mixtures including hundred or even thousands of components. Comprehensive two-dimensional gas chromatography (GC × GC) is a technique that allows samples to be separated by two independent columns coated with different stationary phases. In contrast to 1D GC, it can generate higher peak capacity, enhanced signal, and structured chromatograms [15]. TOFMS has a fast data acquisition rate, wide linear dynamic

* Corresponding author. Tel.: +86 411 84379530; fax: +86 411 84379530.
*E-mail address:* xugw@dicp.ac.cn (G. Xu).

0021-9673/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chroma.2013.06.022

range and full-range mass sensitivity [16]. Thus, GC × GC–TOFMS has been a powerful technique for analyzing complex mixtures, such as cigarette smoke condensate, wine, urine, plasma and yeast cell [16–18].

After obtaining raw data, data preprocessing including deconvolution and alignment of samples among different types of tea is another vital step to attain the research goals [19]. The commercial ChromaTOF software (LECO Corporation, St Joseph, MI, USA) can be used for deconvolution and alignment with the raw GC × GC–TOFMS data as input. To the analysis of complex peak clusters, parallel factor analysis has been proposed to resolve high-order data such as GC × GC–TOFMS [17]. In addition, the modified version of software MetAlign, namely, MetAlignID can be utilized to treat GC × GC–TOFMS data. It is capable of automated library-based identification and concentration estimation of target compounds [20]. For statistical analysis, the Fisher ratio method has been used in discovery-based metabolomics study because of the effectiveness to identify the peaks with classification ability in the two-dimensional (2D) chromatogram by using all mass detection channels [21]. Principal components analysis (PCA) and partial least-squares discriminant analysis (PLS-DA) are two widely used methods in chemistry and other related fields. Visual results can be provided to show samples distribution in 2D or 3D space (scores plot), and contribution of variables for sample classification or clustering (loadings plot). They have been the first choices to discover potential biomarkers to distinguish classes with different but related features [11,16,22–24].

In this work, a GC × GC–TOFMS method was developed for the analysis of the volatile components in green, oolong and black teas. With the help of available MS databases and retention index in the first dimension column ($^1I$), 450 compounds were tentatively identified. Then, 12, 12 and 9 samples of green, oolong and black tea were analyzed by using GC × GC–TOFMS method, the compounds with a significant difference were defined. These compounds were highly related with the characteristic aroma of every tea sample.

## 2. Experimental

### 2.1. Samples and chemicals

The tea samples were purchased from the market. In total, 12, 12 and 9 samples of green, oolong and black teas were respectively collected for the study, the details are given in the supplementary **Table S1**. All samples were first ground by a mortar device and stored at fridge with temperature under −20 °C before extraction. A quality control (QC) sample was prepared by mixing 15 g of tea powder of each sample. It was used to check the repeatability of experiments and reliability of analytical results.

All reagents were purchased from Sigma–Aldrich (Beijing, China), including dichloromethane (HPLC grade), biphenyl, benzeneacetaldehyde, linalool, methyl salicylate, (E)-geraniol, (E)-beta-damascenone, (E)-nerolidol, methyl jasmonate and (E)-phytol. Ultra pure Milli-Q water was manufactured by Millipore Elix advantage 5 (Billerica, MA, USA).

### 2.2. SDE process of samples

Fifteen grams of tea powder were transferred to a 1000 ml flask with 300 ml of ultra pure water, containing 0.1 mg of biphenyl as internal standard (IS), and 60 ml of dichloromethane was added to a 150 ml flask as extraction solvent. Then the two flasks were assembled to a Likens–Nickerson apparatus. Both the sample mixture and the dichloromethane solvent were heated to the boiling points and maintained for 2 h after refluxing. Then, the extraction solvent was cooled down to room temperature and dehydrated with 1.0 g of anhydrous sodium sulfate. Finally, the solvent was concentrated to 1.0 ml by using a rotary evaporator and stored at 4 °C for analysis. All real samples and 11 QC samples were processed with the same procedures.

### 2.3. GC × GC–TOF analysis

The extracts of teas and QC samples were respectively analyzed by using a LECO Pegasus 4D GC × GC–TOFMS instrument (LECO Corporation, St. Joseph, MI, USA) equipped with Agilent 6890N. A non-polar/mid-polar column set was optimized for GC × GC separation. The first dimension ($^1$D) column was 30 m × 250 μm × 0.25 μm DB-5MS (J&W Scientific, Folsom, CA) and the second dimension ($^2$D) one was 1.6 m × 180 μm × 0.20 μm RTX-200MS (Restek, Bellefonte, PA). The temperatures of GC inlet and transfer line were set at 280 °C and 270 °C, respectively. High purity helium (99.9995%) was used as carrier gas at a constant pressure mode. The pressure at the head of the column was 200 kPa. The cryogenic modulation was used with modulation period ($P_M$) of 5.0 s. An Agilent 7683B autosampler (Agilent, PaloAlto, CA, USA), was used with injection volume of 2.0 μl, and split ratio of 10:1. The oven temperature of the first column was held at 60 °C for 3 min, and then ramped to 290 °C (7 °C/min), and held for 5 min at the last temperature. The oven temperature of the second column was initially held at 70 °C for 3 min, and then following the same program of the first column. The total analysis time was 40.75 min.

For MS detection, the temperature of ion source was set to 220 °C. MS range was collected from $m/z$ 33 to 600 with 50 spectra per second. The solvent delay time was 120 s. The detector voltage was 1.60 kV and electron energy was −70 eV. A $C_{10}$∼$C_{20}$ $n$-alkanes series were analyzed to determine retention index in the $^1$D separation. Preliminary identification of compounds was based on similarity comparison of standard MS in NIST05 (National Institute of Standards and Technology, Gaithersburg, MD, USA) and Wiley (Wiley, New York, USA) libraries.

### 2.4. Data pretreatment

The raw data were pre-processed by LECO ChromaTOF$^{TM}$ workstation (version 4.44). The peaks with signal-to-noise ratio (S/N) larger than 50 were extracted, and then the corresponding peak areas were calculated by using an extracted ion chromatogram (EIC). The EIC of each peak was automatically determined by the software after background correction and deconvolution. Two important parameters, namely, $^1$D and $^2$D peak width may affect the number of peaks, they were set to 25 s (5 heart-cuts × 5 $P_M$ time) and 0.4 s, respectively. The peak merging was executed by the software with MS similarity of 65%, and minimum required S/N of 6 for all sub-peaks. This helps to produce a peak table with all slices of one analyte together. The features of $^1t_R$, $^2t_R$, peak area, S/N, and others were included in the table.

Automated retention time alignment was performed by statistical comparison function in the ChromaTOF software. The three groups of tea and QC samples were divided into four classes, then the data within and between the class(es) were aligned. After peak alignment, a new peak table was exported to .csv file, including the features of peaks introduced above. In order to evaluate the alignment performance, standard deviation (SD) of $^1t_R$ and $^2t_R$ for all peaks were calculated. If the SD of $^1t_R$ was larger than 2.5 s or SD of $^2t_R$ larger than 0.1 s, the corresponding peaks were manually inspected and corrected to guarantee the accuracy of results. The final peak table can then be applied as input to the third-part software for statistical analysis.

The strategies of metabolomics study were used to find components with significant difference among the tea classes. A modified "80% rule" was used for data pre-treatment, namely, a peak with