Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

ANALYTICA CHIMICA ACTA 604 (2007) 89–98

available at www.sciencedirect.com

ScienceDirect

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

Mass spectral profiling: An effective tool for quality control of herbal medicines

Zhong-Da Zeng[a,b], *Yi-Zeng Liang*[b], *Foo-Tim Chau*[a,c,*], *Shuo Chen*[b], *Mok Kam-Wah Daniel*[a], *Chi-On Chan*[a]

[a] *Chemometrics and Herbal Medicine Laboratory, Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, PR China*

[b] *Research Center of Modernization of Chinese Medicines, College of Chemistry and Chemical Engineering, Central South University, Changsha 410083, PR China*

[c] *State Key Laboratory of Chinese Medicine and Molecular Pharmacology, Shenzhen 518057, PR China*

ARTICLE INFO

Article history:
Received 11 May 2007
Received in revised form
23 September 2007
Accepted 26 September 2007
Published on line 6 October 2007

Keywords:
Mass spectral profiling
Gas chromatography–mass spectroscopy
Houttuynia cordata Thunb
Herbal medicine
Chemometrics

ABSTRACT

Quality control of herbal medicines (HMs) is a big big headache because of the high complexity and unknown mechanism on disease treatment. In this work, mass spectral profiling, a new tool for data processing is proposed to help a lot in solving this problem as gas chromatography–mass spectroscopy (GC–MS) is used to detect both the active and non-active ingredients buried in HMs. The main idea of mass spectral profiling is employment of target *m/z* points of GC–MS data on the extraction of chromatographic profiles of pure and/or mixed compositions concerned. Further, the absolute or relative abundance at these *m/z* points can be utilized for results interpretation. With the help of this tool, the qualitative and quantitative information of chemical components within complicated HMs will be mined out effectively. It can then be recommended as reference indices to assess the importance of target compositions in HMs, such as efficacy evaluation on disease treatment of the active constituents. Mass spectral profiling with less data points significantly improves the possibility to get the rich information with no strong requirements of data preprocessing procedures, like alignment of shift of retention times among different chromatographic profiles. It is powerful for quality control of HMs coupled with pattern recognition techniques on high-throughput data sets. In this study, a commonly used herbal medicine, *Houttuynia cordata Thunb* and its finished injection products, were used to deliver the strategies. Absolutely, the working principles can be extended to the investigation of metabonomics with gas chromatography–time-of-flight–mass spectrometry (GC–MS–TOF). The good performance of mass spectral profiling shows that it can be a promising tool in the future studies of complex mixture systems.

1. Introduction

Herbal medicines (HMs) of traditional chinese medicine (TCM) and other folk medicines are well organized as very complicated multi-component systems with hundreds or even thousands of chemical components. Hence, quality control, discovery of active constituents in them, and proof of the efficacies have long been regarded as hard diffi-

* *Corresponding author at*: Chemometrics and Herbal Medicine Laboratory, Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China. Tel.: +852 3400 8672; fax: +852 2364 9932.
E-mail address: bcftchau@polyu.edu.hk (F.-T. Chau).

0003-2670/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2007.09.057

culties and very challenging tasks to scientists [1–4]. They have been the main drawbacks in the internationalization and globalization of HMs. In China, the governmental sectors have proposed a series of policies and measures on how to tackle these problems in recent years. The first and urgent job is to make HMs meet the globally acceptable criteria.

As the general guidelines of WHO for methodologies on research and evaluation of traditional medicines pointed out, "Despite its existence and continued use over many centuries, and its popularity and extensive use during the last decade, traditional medicine has not been officially recognized in most countries. Consequently, education, training and research in this area have not been accorded due attention and support. The quantity and quality of the safety and efficacy data on traditional medicine are far from sufficient to meet the criteria needed to support its use worldwide. The reasons for the lack of research data are due to not only to health care policies, but also to a lack of adequate or accepted research methodology for evaluating traditional medicine." [5]. Chromatographic fingerprint, a characteristic representation of the chemical components in HMs, some of which are pharmacologically active, has been developed to control and evaluate the quality of HMs. It is an acceptable technique both in theory and practice [1,3]. For example, Li et al. proposed local least square (LSS) and principal component analysis (PCA) to deal with the data sets of fingerprints, and further for quality evaluation of HMs with chromatographic profiles [6]. Gong et al. employed an effective definition in information theory, entropy, to assess the fingerprints obtained from chemical analysis of HMs [4]. It can be widely used to know chromatographic separation, determine the optimized analytical conditions, and so on. Yi et al. combined fingerprint technique with multivariate statistical analysis to discriminate the two kinds of HMs of Pericarpium Citri Reticulatae and Pericarpium Citri Reticulatae Viride [7]. In China, all finished injection products of HMs are forced to meet the state standards of fingerprints in very near future.

In terms of most reported studies on HMs, including the works in chemical analysis and fingerprint, it can be found that one of the important steps for quality control of HMs is extraction and discovery of the integrated information, as well as the sameness and difference of compositions hidden in large number of data sets, the same as processing of strong overlapping chromatographic profiles from high-throughput data in metabonomics [8–17]. However, how to get such rich information buried in complex data obtained from the modern hyphenated instruments like GC–MS with no lost of too much time and labor? Most conventional approaches can only help to certain extent because they cannot provide a clear picture of the studied systems with the main components unknown, which may or may not have synergic effects with one another. Many chemometric tools, including PCA [18], EFA [19,20], HELP [21–23], SFA [24,25], AMWFA [14,16,17], and so on, showed powerful abilities to find the information from coupled instruments and chromatographic fingerprints with the help of mathematics, statistics and computer science. These methods have been widely used in the analysis of HMs such as *Rhizoma Atractylodis*, *Pericarpium Citri Reticulatae*, *Cortex Magnoliae Officinalis*, *Radix Glycyrrhizae*, *Cordyceps sinensis*, drug pair *Rhizoma Ligustici Chuanxiong-Radix Paeoniae Rubra* (RLC-RPR), Chenpi of orange, and others [14,16,17,26,27]. But most of them only attempted to find the information of pure components of HMs, not regard them as a whole, like what they actually done on disease treatment. This makes full understanding of the "synergic effects" of multi-component be a difficulty both in the investigation of HMs and metabonomics. Moreover, the requirements of rapid, accurate, and online analysis in industry further challenge the scientists on the study of analytical chemistry and HMs [28,29]. At the same time, some necessary but time-consuming data pretreatment procedures also burden the analysis of gas chromatography–mass spectrometry (GC–MS) data acquired at the studies in these fields. For instance, alignment of chromatographic fingerprints or metabolite profilings is one of the common and unavoidable steps before continuing the next research like data comparison, discrimination analysis, and so on [30–32]. Nevertheless, how to do alignment of high-throughput data automatically, even semi-automatically with no consumption of too much time, and strong intervention of the users is still under investigation [33–35]. Furthermore, other procedures for data processing, including remove of instrumental background, filtering of homoscedastic and/or heteroscedastic noises, signal smoothing, normalization and pattern recognition also significantly influence the elucidation of the action mechanisms and "synergic effects" of bioactive ingredients in HMs [1,2,21–23]. It is not difficult to see that data processing of high-throughput data are very challengeable only with less costs. Some interesting works have been reported on the investigations of HMs, metabonomics, and others with effective data processing strategies [14–27,36,37]. Metabolite profiling/footprinting of metabonomics is always very complicated containing hundreds, or more small molecules [38,39]. It is very similar to the study of HMs on data processing.

In this work, mass spectral profiling, the mass spectral profile acquired at certain retention time point of GC–MS data or the total (or mean) mass spectral profile of the whole coupled chromatographic data, is used to extract the components information in the analysis of HMs. It should be pointed out that the total intensities of mass spectra at different retention times are very useful in getting the integrated information of components of GC–MS data. With the mass spectral profiling at hand, the chromatographic profiles with the contribution of one or more interested components at certain mass-to-charge ratio (m/z) points can be used for data interpretations from the mass spectral direction. It is powerful in the analysis of high-throughput data obtained in the studies of HMs and metabonomics. Such strategy can make full use of the rich data information of hyphenated chromatographic instruments like GC–MS. It will be very useful for quality control of HMs, as well as discovery of biomarkers or biomarker patterns, and pharmacological and toxicological studies. A commonly used herbal medicine, *Houttuynia cordata Thunb* and its finished injection products analyzed by GC–MS are employed to deliver the advantages, strategies and good performance of mass spectral profiling in the analysis of mixture samples with high complexity.

2. Theory and methodology

2.1. Two-dimensional data and mass spectral profiling

As Lambert–Beer law shown in Eq. (1), every element x_{ij} ($i = 1, \ldots, m$; $j = 1, \ldots, n$) of a GC–MS data set $\mathbf{X}_{mxn}$ can be written as the common contributions of the N components included:

$$x_{ij} = \sum_{k=1}^{N} c_{ik} s_{kj}^{T} + eij \quad (1)$$

Here, denotation c_{ik} means the chromatographic concentration of the kth component at retention time point i, and s_{kj} means the spectral response of the kth component at the jth m/z point. Denotation e_{ij} is experimental noise at the corresponding position of x_{ij}. It is not difficult to see that the sum (or mean) of all rows in data set **X**, that is to say, the total (or mean) mass spectral profiling, collects all the contribution of components in data set **X** from the mass spectral direction. It is also why the total (or mean) mass spectral profiling can be used to represent the whole target mixtures for discrimination analysis, and then for quality control of HMs. In Fig. 1, the basic construction of data set **X** is shown.

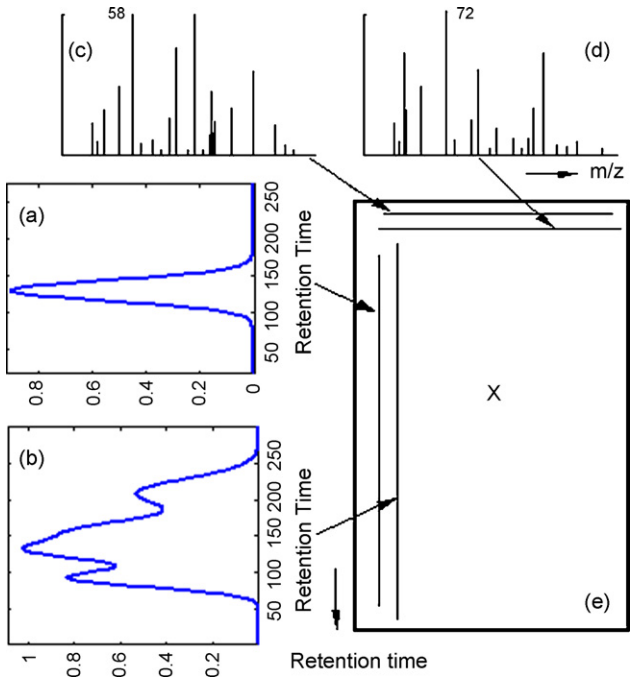

Fig. 1 – Schematic illustration of mass spectral profiling in two-dimensional GC–MS data set X. The mass chromatograms (MC) at certain *m/z* points and mass spectral profiling at certain retention time (s) are clearly shown in the figure. Also, the chromatographic profiles with pure and mixed chemical components, and selective *m/z* points at mass spectral profiling are presented for easy understanding purpose.

2.2. Working principles of mass spectral profiling

According to the statements described above, the ith row of data set **X**, namely, the mass spectral profiling, may be a pure or mixed contribution of only one or more interested compositions. If the m/z point of a certain mass spectral profiling is a selective one, in another word, the response produced by only one component, the mass chromatogram (MC) at this m/z point can be used to get the pure chromatographic profile of the target component. Also, its pure spectrum can then be computed by least square method. This has great significance both to qualitative and quantitative analysis of multi-component, as in quality control of HMs. In Fig. 1(a), the pure chromatographic profile can be extracted if the corresponding position of 58 in (c) is a selective m/z point. In the last decades, some chemometric resolution methods and selected ion monitoring-mass spectroscopy (SIM-MS) were developed for analysis of components embedded in overlapping peak clusters of complex mixtures based on the selective mass spectral information [40–42]. Wang et al. attempted to get the selective m/z point of GC–MS data using chemometric methods [40]. This kind of methods even proposed the strategies for the resolution of mixture systems with embedded overlapping components, in which some minor components are completely buried by a big one. But most of mass spectral profilings are total contribution of more than one component, as shown in Fig. 1(b) and (d). In Fig. 1(d), if the m/z point at 72 of mass spectral profiling is not a selective one, the MC at this position will be an overlapping peak cluster with several components, as Fig. 1(b) shows. In this case, the common m/z points and their abundance become effective indices to determine the importance of different components for disease treatment, and further used for quality control of HMs. They include quantitative information of the same type of chemical components, or interested in. Generally, the m/z points of mass spectral profiling are very abundant for the extraction of components information in complex mixtures. In most cases, it is not so difficult to find the interested m/z points from the mass spectral profiling to represent common contribution of more than one target components. Therefore, the researchers can employ these rich m/z points to discover the common qualitative information and quantitative variation rules of the chromatographic components. It will be very helpful in the evaluation of activities of constituents, and then quality control of HMs. For instance, the efficacies of HMs in treating diseases are mostly “synergic effects” of a few active ingredients. If one can find one or more common m/z points at mass spectral profiling of these components, the curative effects of HMs can probably be predicted using their abundance before pharmacological experiments. These m/z points and their intensities can also be used in modeling between the active constitutes and their activities. It is also easy to see they are informative variables collecting material information of the components. This will help to save much time, money and labor in analysis. So, mass spectral profiling is a very useful reference index to assess the curative effects of HMs. In contrast to the traditional chromatographic fingerprint, the possibility to extract the buried information of GC–MS data will be increased probably because of the discontinuity and selectivity of mass spectra (MS). There is more information included

at different *m/z* points, which can be employed for analysis in detail. Similarity index among vectors, such as mass spectral profilings and chromatographic fingerprints, was used to partially assess their advantages and disadvantages in this work. It is defined as the following equation:

$$\mathrm{SI} = \frac{\mathbf{x}^{\mathrm{T}} \cdot \mathbf{y}}{\mathrm{norm}(\mathbf{x}) \cdot \mathrm{norm}(\mathbf{y})} \quad (2)$$

where denotation SI means the similarity. The symbol "·" represents dot product of two vectors of chromatographic profiles **x** and **y**, and "norm" is the operation of vector norm.

2.3. *Advantages of mass spectral profiling*

The main advantages of the strategy of mass spectral profiling can be summarized as follows: (1) Total (or mean) mass spectral profiling of GC–MS data shows the total contribution of all the chromatographic components in mixture samples, which is a new point of view to consider the mixture systems with less data points. In most cases, it is also very helpful to know what components, at least what kinds of components presenting in a target system like HMs. (2) In general, several metabolites with similar molecular structures always contribute to the same *m/z* points. Therefore, extraction and mining of the positions and abundance of such *m/z* points will be a very good choice to evaluate the integrated qualitative and quantitative variation of the important target metabolites. (3) Selective *m/z* points of components are informative and useful to extract the buried knowledge of components in complex mixture samples. Some conventional theoretical and instrumental methods were proposed on the basis of these *m/z* points. It is one of the main considerations of mass spectral profiling in the analysis of mixtures in contrast to chromatographic profiles (fingerprints). The discovery and extraction of information of selective *m/z* point of mass spectral profiling is probably a key point to solve the present problems. (4) Data preprocessing of mass spectral profiling is completely easier than chromatographic profiles with only hundreds of data points. In general, the number of points of the former is 400–1000, but the latter one more than 10,000, always. Another competitive advantage is no shift of *m/z* points of mass spectral profiling, unlike the usual shift of retention times in chromatographic profiles. Correction of retention time shift of high-throughput data is still a great difficulty. Most reported methods couldn't calibrate the shift of large number of chromatographic data automatically, even semi-automatically. (5) The common patterns, i.e., the common features of components in GC–MS data of mixtures can be easily constructed with mass spectral profiling. Furthermore, the researchers can conveniently extract one or more special *m/z* points of mass spectral profiling or the chromatographic profiles at these positions to further study the components exactly, if necessary. This will make good use of the information of two-way GC–MS data, no losing information as the conventional one-way fingerprint technique. Even it is possible to study the complex mixtures with information only from mass spectrometer without any chromatographic separation. All these operations are timesaving to model analysis of complex systems, such as pattern recognition of HMs collected from different origins, or quality good and fake products. How to differentiate these samples is the main factors in quality control of HMs. Mass spectral profiling can be a promising tool in the analysis of high-throughput mixture systems with very high complexity.

3. Experimental

In this work, *Houttuynia cordata Thunb* and its finished injection products are employed as examples to deliver the strategies of mass spectral profiling of GC–MS data. In China and many other Asian countries, it has been a commonly used herbal medicine for anti-inflammation and anti-bacterial purpose. *Houttuynia cordata Thunb* is one of the six HMs in preventing and treating severe acute respiratory syndrome (SARS) recommended by the State Administration of Traditional Chinese Medicine of the People's Republic of China. There are more than 40 pharmaceutical corporations to manufacture injection products using this medicine in China. But the efficacies are different in some degree because of different planting geographic places or production procedures. It is clear that quality control of this medicine can be an example with great interests and importance both in academic and practical significance.

3.1. *Experimental GC–MS data for the analysis of medicinal materials of Houttuynia cordata Thunb*

Quality control of medicinal materials is one of the key points to guarantee the good quality of final products. Many Chinese research units have paid much time and labor on the establishment of state standards of some HMs with the help of fingerprint technique. In this work, mass spectral profiling tries to consider the GC–MS data from a different point of view. It can be an effective and necessary supplement of the conventional tools. Thirty-four samples of fresh medicinal materials of *Houttuynia cordata Thunb* collected from four different areas will be employed to show the strategy of mass spectral profiling. The total ion current (TIC) chromatograms of every GC–MS data are shown in Fig. 2.

3.2. *Investigation of GC–MS data for quality control of finished injection products of Houttuynia cordata Thunb*

In Chinese pharmacopeia (version 2005), Only the fresh medicinal materials of *Houttuynia cordata Thunb* are allowed to produce finished injection products by the State Food and Drug Administration (SFDA) of the People's Republic of China. But most factories in North China have to use dried medicinal materials for manufacture because of the lack of fresh herbs and long-distance transportation problem. Then, whether is it possible to determine the products made with fresh and dried medicinal materials? Are there apparent differences between the two kinds of injections? Can the activities be evaluated with the mass spectral profiling of the GC–MS chromatographic data? On June 2006, SFDA prohibited the production, marketing and use of the injection products of *Houttuynia cordata Thunb* because of the several anaphylactoid and lethiferous cases. That led to nationwide disputation and censure of the widely used herbal medicine. In this investigation, 82

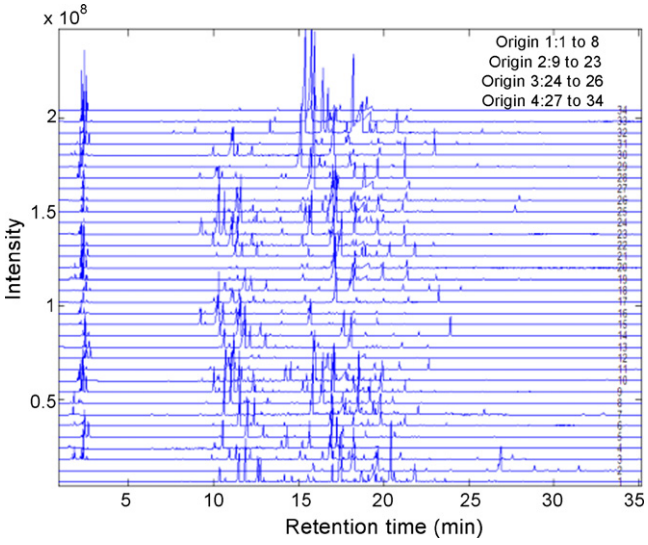

Fig. 2 – The total ion current (TIC) chromatograms of 34 GC–MS data of fresh medicinal materials of *Houttuynia cordata Thunb* collected from four representative planting areas in China.

injection samples were studied with GC–MS technique and the proposed tool. These samples were collected from four factories located in four representative areas of China. Two factories manufacture the products with fresh medicinal materials and other two with dried ones.

3.3. *Extraction of volatile chemical components in Houttuynia cordata Thunb and its finished injection products*

The volatile chemical components are main pharmacological active constitutes in *Houttuynia cordata Thunb*. To the extraction of these components in medicinal materials, 600 g of fresh herb was sliced to 0.5–1 cm first and added into a flask (2000 mL) with 1000 mL distilled water together. Then, the mixture was heated and extracted for about three hours with circumfluence speed about 20 drops per minute. Finally, the primrose organic liquid on the above layer was collected and stored in a refrigerator at 4 °C before use. The experimental devices and procedures of the injections were very similar to extraction of the medicinal materials. But four milliliter of hexane need be added into the extractor to collect the volatile components. The circumfluence time is 45 min with a speed about 14 drops per minute. The above organic layer of hexane is transferred to a pear-shape bottle with capacity of 100 mL for decompression distillation. Finally, the liquid was concentrated to 0.5 mL for GC–MS analysis.

3.4. *Instruments and experimental conditions for the analysis of volatile chemical components in Houttuynia cordata Thunb*

Analysis of the volatile components in medicinal materials was performed on a GC-17A gas chromatographic instrument hyphenated with a QP-5000 mass spectrometer from Shimadzu (Tokyo, Japan). The chromatographic column is an OV-1 capillary column (30 m × 0.25 mm i.d., film thickness 0.25 μm). In chromatographic analysis, the column temperature was set at 50 °C and held for 6 min first. Then, the temperature was programmed to 230 °C at a rate of 10 °C min^{-1} and held for 25 min. The temperatures of injection port and interface were both kept at 280 °C. The carrier gas was Helium with a constant flow-rate of 0.7 mL min^{-1}. One microlitre of the sample was injected for analysis under 10:1 split mode. To the experimental conditions of mass spectrometer, electron impact (EI^+) mass spectra were used to record at 70 eV. Ionization energy in full scan model was in the ranges of 20–350 atomic mass units (amu) with 5 scans s^{-1}. The ionization source temperature was set at 230 °C.

3.5. *Instruments and experimental conditions for the analysis of volatile chemical components in the injection products*

The difference of conditions for the analysis of volatile components in injection products with the above statements is described as follows. GC–MS analysis was performed on a Shimadzu GC-2010 (Kyoto, Japan) gas chromatography instrument coupled to a Shimadzu QP2010 mass spectrometer. In chromatographic analysis, the column temperature was set at 50 °C and held for 6 min at first. Then, the temperature was programmed to 230 °C at a rate of 10 °C min^{-1}, but held for 16 min at the last temperature. The flow rate of carrier gas was 1.0 mL min^{-1} and the volume of sample injection was 2 μL with the same split ratio in the above section. As to the conditions of mass spectrometer, the only one difference was the ionization energy in the ranges of 20–450 atomic mass units (amu).

3.6. *Data analysis*

All computer programs used in this study were coded in MATLAB 6.5 and all calculations were performed on a Pentium IV processor based ASUS compatible personal computer with 256 M RAM memories.

4. Results and discussion

4.1. *Results from the GC–MS data of medicinal materials of Houttuynia cordata Thunb*

After correction of shift of retention time, the TIC profiles of the fresh medical materials used for analysis were clearly shown in Fig. 2. The number of the samples was 34, totally. They were collected from the four main areas in China to plant *Houttuynia cordata Thunb*, namely, Fujian (including Jiangxi), Yunan, Hunan and Sichuan provinces with a little difference of conditions for storage and transportation. In this work, the numbers of samples from the four areas were 8, 15, 3 and 8. The conventional chromatographic fingerprints technique for quality control attempted to establish a common state standard of the products obtained from different manufacturers. It can uniquely represent the medicine and differentiate from others. Then, the common pattern of good quality can

be differentiated from the fakes of medicinal material. But this will be unavoidable to lose most useful information of two-way GC–MS data because of the deduction of data dimension. From the TIC profiles, it is easy to see there is significant qualitative and quantitative difference among most chemical components in these samples, as shown in Fig. 2. The similarities among the original profiles and their mean profiles are only from 0.1922 to 0.5198. Most of them are less than 0.40.

Therefore, the common features (patterns) of components in these samples are very difficult to be found out for quality control. In practice, the evaluation of sameness and difference of components in the same medicine only with different origins are not so easy. Additionally, data processing of these samples, such as correction of retention time shifts, challenges the conventional chemometric methods. Then, it is even difficult to get the relationships of presence and absence of target components among these samples unless to obtain all the components information using the existing chemometric methods. For succinct description purpose, only mean of the total mass spectral profiling (after normalization operation) acquired at the four different areas is shown in Fig. 3. Obviously, the four mass spectral profiling is very similar to each other. The similarity of the 34 mass spectral profiling with their mean mass spectral profiling is between 0.8151 and 0.9909. The five minimum similarities also attain to 0.8151, 0.8692, 0.8735, 0.8809 and 0.8828, respectively. Most other similarities are more than 0.9500. So, the common information of mass spectral profiling among different samples, namely, the common features of medicinal materials of *Houttuynia cordata Thunb*, can be extracted out using mass spectral profiling. Furthermore, the special interested *m/z* points at mass spectral profiling are very useful to get the chromatographic profiles of components in mixture samples. Such information is very important for quality control of HMs. One can define the intensities of the interested *m/z* points, or the MC (s) at different *m/z* points, if necessary. Also, these extracted *m/z* points and chromatographic profiles are crucial in the study of metabonomics, such as the discovery of biomarkers or biomarker patterns on disease diagnosis. The main advantage of the strategy of mass spectral profiling for quality control of HMs is the effective employment of spectral information of all the components in GC–MS data, including the major and minor ones. It considers the important volatile components in *Houttuynia cordata Thunb*, such as terpenes and the ramification, and aliphatic compounds as a whole. With the help of the additional information of conventional chromatographic fingerprint, mass spectral profiling will assist one to get the comprehensive and integrated information of components in complex GC–MS data for experimental and industrial applications.

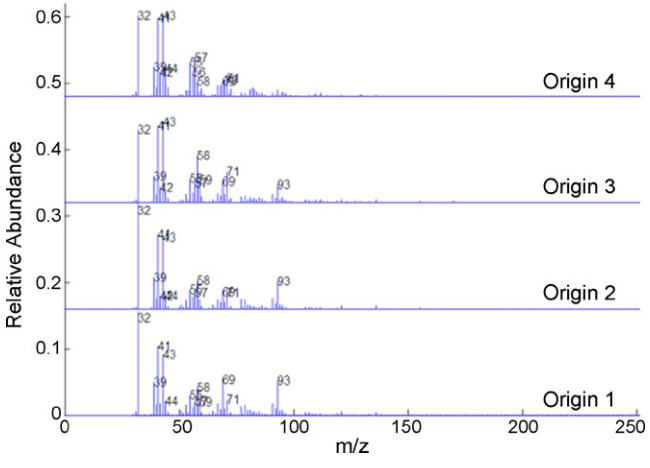

Fig. 3 – The mean mass spectral profiling of the same GC–MS data shown in Fig. 2 of fresh medicinal materials after sum and normalization operations.

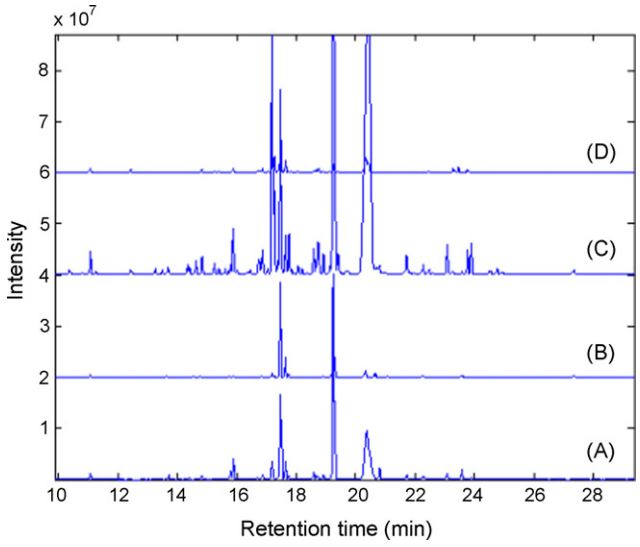

Fig. 4 – The four average TIC chromatograms of the samples of final injection products collected from four different manufacturers after alignment of the shift of retention times.

4.2. *Results from the GC–MS data of finished injection products of Houttuynia cordata Thunb*

In this section, the finished injection products collected from four factories were studied in detail. Their average TIC profiles were shown in Fig. 4. The total number of samples was 82, which is largely more than the number of the medicinal materials analyzed in the above section. It is really a time-consuming and challengeable task if only the conventional methods are used for alignment of shift of retention times, and then, other treatments. These samples were respectively collected from Zhengqing, Yaan, Lingrui and Ange corporations for HMs production. The front two factories locate in South China using fresh medicinal materials, and the latter two in North China using dried ones. The main task for quality control of such products is how to find the sameness and difference with discrimination analysis techniques, as well as the special interested components. Whether the dried medicinal materials can be used in production is a vexed question in China. In this work, the new technique may do a favor on solution of such problem. The numbers of samples for analysis of the above four factories were respectively, 13, 18, 45 and 6. The four average chromatograms shown in Fig. 4 are mean profiles of the samples extracted from the same manufacturers. From this figure, the main qualitative information

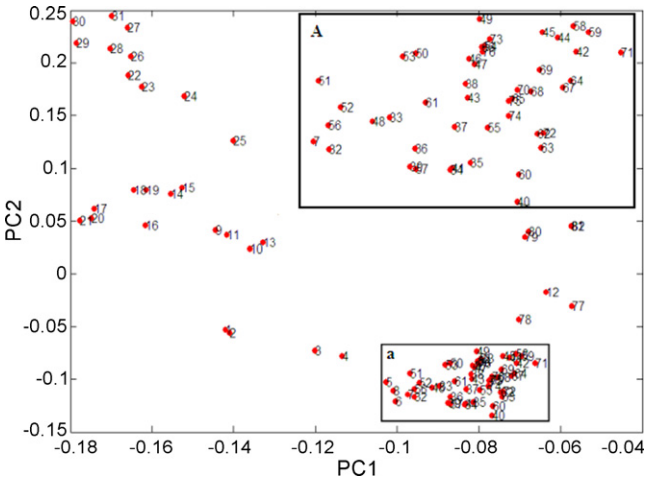

Fig. 5 – The clustering results of the samples in Fig. 4 with principal components analysis (PCA) method. Only the first two principal components (PC1 and PC2) are used in the figure. The sub-plot (A) shown in the left above corner is an enlarged figure corresponding to the circled region (a).

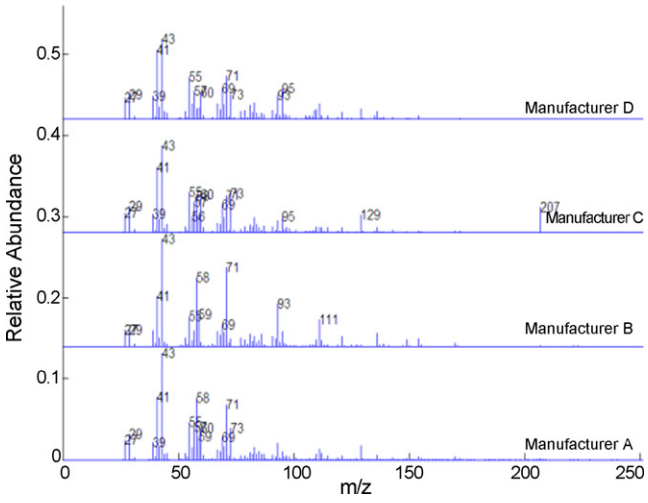

Fig. 6 – The mean mass spectral profiling of the same GC–MS data in Fig. 4 of final injection products of *Houttuynia cordata Thunb* after the sum and normalization operations.

of the components between retention time 16 and 21 min is very similar among them. But the difference in some other retention time regions can be found, especially to the third one (manufacturer C). The similarities of the 82 TIC profiles with their mean profiles are from 0.5826 to 0.9377. Most of them are from 0.7 to 0.9. In Fig. 5, the clustering results of all the 82 mixture samples using the chromatographic profiles and PCA are shown with the first two principal components, titled PC1 and PC2, respectively. It is apparent to see the discrimination results are not very satisfactory for quality control of the products from the four factories. First, the samples from Nos. 1 to 13 of manufacture A are very dispersed and cannot differentiate from B (from Nos. 14 to 31) and C (from Nos. 32 to 76) very well. The samples of manufacturer D (from Nos. 77 to 82) are nearly mixed together with the samples of manufacturer C. Clustering of the twelfth sample of manufacturers A is false, in fact. The main sameness and difference of the samples produced by fresh and dried medicinal materials cannot be clearly distinguished in the projection loadings plot. Then, the discrimination results are unacceptable used for quality control. On the other hand, the mean of total mass spectral profiling of the four different factories after normalization is shown in Fig. 6. Obviously, the main feature of these spectra at different *m*/*z* points of mass spectral profiling is very similar to each other. As described above, the chromatographic profiles of the pure and/or mixed components at the *m*/*z* points can be further extracted and studied in detail. For example, if one is interested in the *m*/*z* point 207 at mass spectral profiling of the second manufacturer B, chromatographic profile of the component at this *m*/*z* point can be easily extracted from the original GC–MS data for validation. Then, the efficacies of the products from different factories can be possibly elucidated with mass spectral profiling. The similarities of the 82 mass spectral profiling with their mean profiling are between 0.8854 and 0.9937. Most of them are more than 0.9500.

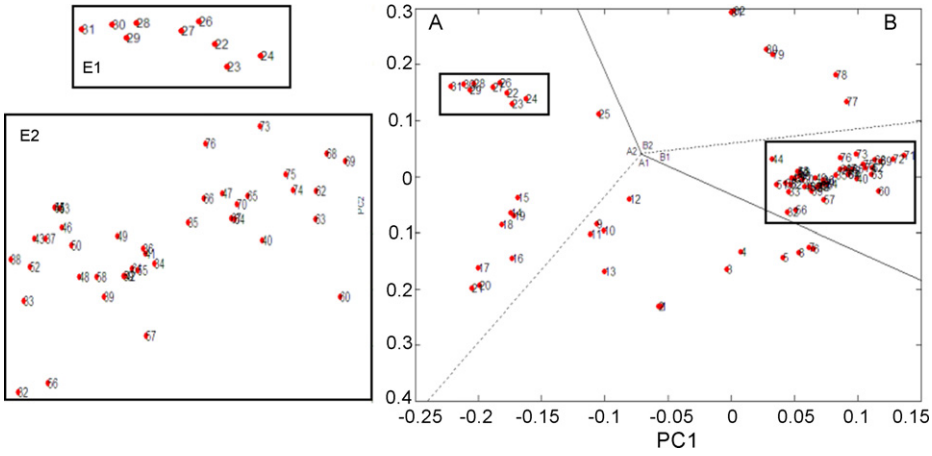

Fig. 7 – The discrimination results of the mass spectral profiling after deduction of mean of all the 82 samples in Fig. 6 with first 2 principal components PC1 and PC2 by PCA analysis. The two sub-plots (E1 and E2) shown in the left are enlarged figures corresponding to the circled parts of regions A2 and B1, respectively.

Obviously, the relative intensities of spectra at *m/z* points 58 and 71 in the manufacturers A and B using fresh medicinal materials are significantly larger than them with dried ones. The main reason is the degradation of some volatile components in the processes of drying, storage and production. The similarities between the samples of manufacturers A and B, C and D are respectively from 0.9113 to 0.9981, and from 0.9231 to 0.9990. Most of them attain to 0.9900. Compared with the similarities obtained from chromatographic fingerprints, mass spectral profilings of the products acquired at the same kinds of medicinal materials (fresh or dried) are more similar than the different kind. After deduction of the mean of mass spectral profiling (see Eq. (3)), the results of all the 82 samples with PCA analysis are shown in Fig. 7:

$$\mathbf{x}_{\text{new}} = \mathbf{x} - \text{mean}(\mathbf{x}) \tag{3}$$

Here, denotations $\mathbf{x}$, mean($\mathbf{x}$) and $\mathbf{x}_{\text{new}}$, respectively mean the original mass spectral profiling $\mathbf{x}$, mean of $\mathbf{x}$ and the new vector after the operation. It is very clear to see that the samples collected from the four different manufacturers can be discriminated with very good performance. The four discrimination regions A1 and A2, B1 and B2 in the figure can be used to divide the samples from different factories into four parts very well. Furthermore, the samples manufactured by fresh (in A1 and A2) and dried medicinal materials (in B1 and B2) also clearly distributed in two different sides. Therefore, mass spectral profiling with the same medicinal materials has more common features. It is crucial for the quality control of HMs.

On the other hand, the chromatographic profiles of components at the interested *m/z* points of mass spectral profiling can be used for further extraction of buried information in GC–MS data. In Fig. 8, eight chromatographic profiles with different complexities at *m/z* points of mass spectral profiling are shown. The possible pure profiles with only one component are plotted in the first three profiles. At the same time, the molecular structures and pure spectra of the three components A, B and C are shown. The three components are very important bioactive constituents of *Houttuynia cordata Thunb*. With the help of mass spectral profiling, the qualitative and quantitative information of these components can be obtained at the target *m/z* points. So, the three compounds are very important for evaluation and improvement of the efficacy of *Houttuynia cordata Thunb*. Of course, some conventional methods such as EFA [19,20], HELP [21–23], and FSMWEFA [43], etc. can be further used for purity test of these components, if necessary. Other six mixed chromatographic profiles with two, three, four and more compsitions are shown from the forth to ninth profiles. It can be found that all the above three components A, B and C are included in the sixth and ninth chromatographic profiles. Therefore, the *m/z* points with contribution to these profiles are very helpful to provide possible efficacy information of *Houttuynia cordata Thunb*. All the nine chromatographic profiles in Fig. 8 are obtained from different *m/z* points. Most of them have more than two target *m/z* points. Of course, the sum of chromatographic profiles of different *m/z* points can be used for evaluation when the number of *m/z* points is more than one to a certain profile. These *m/z* points of mass spectral profiling are very important reference indices to evaluate the qualitative and quantitative information or the variation rules of the active constituents. In general, it is not difficult to find out the interested *m/z* points at different mass spectral profiling for information extraction of one or more target components in HMs or metabolites in metabonomics. The *m/z* points with common contributions of different components are very rich in mass spectral profiling.

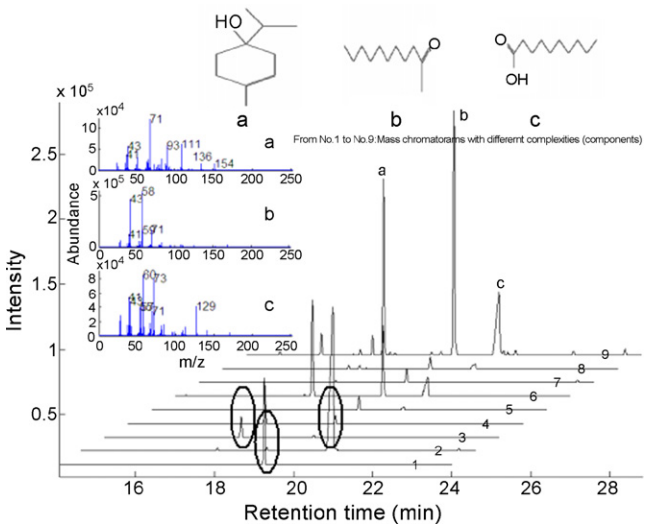

Fig. 8 – Nine chromatographic profiles with different complexity at the interested *m/z* points of mass spectral profiling. The three approximately pure profiles with only one chemical component are shown in the first three profiles (1–3). The molecular structures and pure spectra in the figure are information of the three components a–c. Other five mixed chromatographic profiles with two, three, four and more components are extracted from *m/z* points of the mass spectral profiling.

In Fig. 9, eight chromatographic profiles with two groups (one from Nos. 1 to 4 and another from Nos. 5 to 8) were extracted from *m/z* points 58 and 71 of the four manufacturers, respectively. It seems that the relative intensities of manufacturers A and B (with fresh herbs) at the two *m/z* points are significantly larger than the relative intensities in manufacturers C and D (with dried herbs) in Fig. 6. Then, the chromatographic profiles at these *m/z* points can be extracted for further study. The relative quantitative information of components a1 and a2 in manufacturers C and D are smaller than the same components a3 and a4 in A and B (at *m/z* point of 58). The similar results of components b1, b2, b3 and b4 can also be obtained at *m/z* point 71 to show the quantitative variation rules of these components among different samples. It can be found that the qualitative and quantitative information of components in mass spectral profiling and their *m/z* points is very rich for the analysis of complex mixture samples. Such information is also useful to elucidate the mechanism of difference of active ingredients among different but related samples like HM products of different manufacturers. Mass spectral profiling of GC–MS is a powerful tool to further understand the integrated structures of contaminated

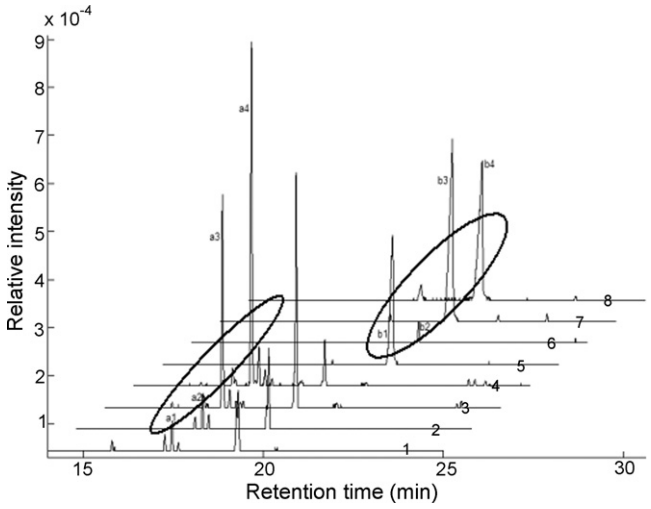

Fig. 9 – Eight chromatographic profiles in two groups (one from 1 to 4, and another from 5 to 8) were acquired at *m/z* points 58 and 71 of the GC–MS data collected from four different manufacturers, respectively. The quantitative variation information of the target components (such as components a1, a2, a3 and a4; and b1, b2, b3 and b4) of the injection products at the two *m/z* points can be obtained from the figure.

mass spectral data and their corresponding chromatographic profiles of chemical components. It can be an effective strategy in the study of HMs, as well as metabonomics and system biology.

5. Conclusions

Mass spectral profiling is proposed to dig out the chemical information hidden in GC–MS data in the analysis of complex systems like HMs. Not only the selective *m/z* points, but also the common or non-common ones of the mass spectral profiling have great significance to interpret the efficacies of active constitutes of HMs on disease treatment, as well as quality control. Most useful information buried in the contaminated data sets can be mined out in a simple and convenient way. The study of *Houttuynia cordata Thunb* is a good example used in this work. No requirement of correction of shift of retention times among chromatographic profiles with a great deal of data points, the dried and fresh medicinal materials collected from different origins, as well as the final injection products from four manufacturers, were discriminated with good performance. Up to the present, information extraction of high-throughput data obtained from hyphenated instruments like GC–MS is one of the main procedures with great challenges in the study of HMs, and other complicated mixture systems. Mass spectral profiling can be a possible tool to resolve the problems in these scientific fields. It can be an effective and necessary supplement to the conventional techniques. This tool is also a good choice on data processing in the study of metabonomics and biomarker discovery.

Acknowledgements

This research work is financially supported by National Natural Science Foundation of China (Grant Nos. 20235020 and 20475066), the Cultivation Fund of the Key Scientific and Technical Innovation Project, Ministry of Education of China (No. 704036), and The Innovation and Technology Fund, Innovation and Technology Support Programme of Guangdong, Hong Kong Technology Cooperation Funding Scheme (ITF-TCFS) Platform Research Scheme (Modernization of Chinese Medicine: From Clinical Efficacy to Drug Production, Two Innovative Formulae for Comprehensive Research, from May 2006 to April 2008).

REFERENCES

[1] Y.Z. Liang, P.S. Xie, K. Chan, J. Chromatogr. B 812 (2004) 53.
[2] Y.L. Wang, Y.Z. Liang, Y. Hu, B.Y. Li, Z.D. Zeng, Y.K. He, Chemom. Intell. Lab. Syst. 82 (2006) 229.
[3] P.S. Xie, S.B. Chen, Y.Z. Liang, X.H. Wang, R.T. Tian, R. Upton, J. Chromatogr. A 1112 (2006) 171.
[4] F. Gong, Y.Z. Liang, P.S. Xie, F.T. Chau, J. Chromatogr. A 1002 (2003) 25.
[5] WHO: General Guidelines for Methodologies on Re search and Evaluation of Traditional Medicines, World Health Organization, Geneva, 2000, p. 1.
[6] B.Y. Li, Y. Hu, Y.Z. Liang, P.S. Xie, Y.P. Du, Anal. Chim. Acta 514 (2004) 69–77.
[7] L.Z. Yi, D.L. Yuan, Y.Z. Liang, P.S. Xie, Y. Zhao, Anal. Chim. Acta 588 (2007) 207–215.
[8] J.K. Nicholson, J. Connelly, J.C. Lindon, E. Holmes, Nat. Rev. Drug Dis. 1 (2002) 153.
[9] J.K. Nicholson, J.C. Lindon, E. Holmes, Xenobiotica 29 (1999) 1181.
[10] V.D.G. Jan, S. Paul, V.D.H. Rob, Curr. Opin. Chem. Biol. 8 (2004) 559.
[11] A.C. Servais, J. Crommen, M. Fillet, Electrophoresis 27 (2006) 2616.
[12] A. Gomiero, D.M. Pampanin, A. Bjornstad, B.K. Larsen, F. Provan, E. Lyng, O.K. Andersen, Aquat. Toxicol. 78 (2006) s34.
[13] H. Jungnickel, E.A. Jones, N.P. Lockyer, S.G. Oliver, G.M. Stephens, J.C. Vickerman, Anal. Chem. 77 (2005) 1740.
[14] Z.D. Zeng, Y.Z. Liang, Y.L. Wang, L.M. Liang, X.R. Li, C.X. Zhao, Q.S. Xu, B.Y. Li, F.T. Chau, J. Chromatogr. A 1107 (2006) 273.
[15] C.A. Bruckner, B.J. Prazen, R.E. Synovec, Anal. Chem. 70 (1998) 2796.
[16] L.Z. Yi, Y.Z. Liang, Z.D. Zeng, P. Wang, D.L. Yuan, Chem. J. Chin. U. 27 (2006) 1626.
[17] L.Z. Yi, Y.Z. Liang, Z.D. Zeng, D.L. Yuan, P. Wang, Chem. J. Chin. U. 27 (2006) 2052.
[18] I.T. Jolliffee, Principal Component Analysis, Springer Verlag, New York, 1986.
[19] M. Maeder, Anal. Chem. 59 (1987) 527.
[20] M. Maeder, A. Zilian, Chemom. Intell. Lab. Syst. 3 (1988) 205.
[21] O.M. Kvalheim, Y.Z. Liang, Anal. Chem. 64 (1992) 936.
[22] Y.Z. Liang, O.M. Kvalheim, H.R. Keller, D.L. Massart, P. Kiechle, F. Erni, Anal. Chem. 64 (1992) 946.
[23] Y.Z. Liang, O.M. Kvalheim, A. Rahmani, R.G. Brereton, J. Chemom. 7 (1993) 15.
[24] R. Manne, H.L. Shen, Y.Z. Liang, Chemom. Intell. Lab. Syst. 45 (1999) 171.
[25] H.L. Shen, R. Manne, Q.S. Xu, D.Z. Chen, Y.Z. Liang, Chemom. Intell. Lab. Syst. 45 (1999) 323.

[26] F. Gong, Y.Z. Liang, H. Cui, F.T. Chau, B.T.P. Chan, J. Chromatogr. A 909 (2001) 237–247.
[27] F. Gong, Y.G. Peng, H. Cui, Y.Z. Liang, A.K.M. Leung, F.T. Chau, Chem. J. Chin. U 20 (1999) 199–203.
[28] J.C. Lindon, E. Holmes, J.K. Nicholson, Anal. Chem. 75 (2003) 384A.
[29] G.H. George, H.L. Roxanne, N.C. Greg, C. Gary, G. Royston, F.M. Jane, P.L. James, E.G. Patricia, A.R. Robert, Toxicol. Lett. 146 (2004) 197.
[30] F. Gong, Y.Z. Liang, Y.S. Fung, F.T. Chau, J. Chromatogr. A 1029 (2004) 173.
[31] N.P.V. Nielsen, J.M. Carstensen, J.J. Smedsgaard, Chromatogr. A 805 (1998) 17.
[32] C.A. Smith, E.J. Want, G.O. Maille, R. Abagyan, G. Siuzdak, Anal. Chem. 78 (2006) 779.
[33] A.L. Duran, J. Yang, L.J. Wang, L.W. Sumner, Bioinformatics 19 (2003) 2283.
[34] K. Mikko, M. Jarkko, O. Matej, Bioinform. Appl. Note 22 (2006) 634.
[35] K. Mikko, O. Matej, BMC Bioinform. 6 (2005) 179.
[36] M. Alexandra, J. Jochen, B. Steffi, S. Annette, M. Armin, C.S. Adrian, B. Hans, J. Chromatogr. B 792 (2003) 269.
[37] S. Eisaburo, S. Naoko, I. Kentaro, S. Akemi, S. Kenji, H. Shin-Ichiro, N. Kohei, N. Kei, Lung Cancer 48 (2005) 77.
[38] O. Fiehn, J. Kopka, R.N. Trethewey, L. Willmitzer, Anal. Chem. 72 (2000) 3573.
[39] W. Weckwerth, O. Fiehn, Curr. Opin. Biotech. 13 (2002) 156.
[40] Z.G. Wang, Z.P. Chen, F. Gong, H.L. Wu, R.Q. Yu, Analyst 127 (2002) 623.
[41] H.S.M. Ahmida, P. Bertucci, L. Franzo, R. Massoud, C. Cortese, A. Lala, G. Federici, J. Chromatogr. B 842 (2006) 43.
[42] J.G. Krabbe, F. Gao, J. Li, J.E. Ahlskog, H. Lingeman, W.M.A. Niessen, H. Irth, J. Chromatogr. A 1130 (2006) 287.
[43] H.R. Keller, D.L. Massart, Anal. Chim. Acta 246 (1991) 379.