Talanta 103 (2013) 116–122

Contents lists available at SciVerse ScienceDirect

Talanta

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

Identification of terpenoids from *Ephedra* combining with accurate mass and in-silico retention indices

Min He [a,1], Jun Yan [a,1], Dongsheng Cao [a], Shao Liu [b], Chenxi Zhao [c], Yizeng Liang [a,*], Yaping Li [d], Zhimin Zhang [a]

[a] Research Center of Modernization of Chinese Medicines, Central South University, Changsha 410083, PR China
[b] Department of Pharmacy, Xiangya Hospital, Central South University, Changsha 410008, PR China
[c] Department of Biological Engineering and Environmental Science, Changsha University, Changsha 410003, PR China
[d] Department of Neurology, Xiangtan Central Hospital, Xiangtan 411100, PR China

ARTICLE INFO

Article history:
Received 10 August 2012
Received in revised form
1 October 2012
Accepted 6 October 2012
Available online 12 October 2012

Keywords:
Terpenoids
Ephedra
Chemometric methods
Accurate mass
In-silico retention indices

ABSTRACT

Terpenoids in *Ephedra* were analyzed by gas chromatography–mass spectrometry (GC–MS), and their pure mass spectra (MS) were obtained using chemometric methods. NIST library search and retention indices (RI) were applied in further identification. The major bottleneck is the ambiguous NIST MS library search results, therefore, we used a simple external calibration method to obtain the accurate mass of molecular ion or key fragment, which included overlapped isotopes structures resolution and Gaussian fitting using Origin software. The calibration method was able to distinguish different molecular weights among a large number of known NIST MS library search results. In addition, a quantitative structure–retention relationship (QSRR) model with good predictive ability was established and the in-silico RI was applied in qualitative identification combined with NIST MS library search results. Two candidate compounds were found to have a moderate matching between the predicted RI values against the experimentally determined values, and incorrect formulas were excluded. Therefore, these auxiliary confirmation methods were able to reflect the analytical chemistry development trend.

1. Introduction

Terpenoids are the most abundant and diverse compounds in nature, including the hydrocarbons and their oxygenated derivatives which are composed of multiple isoprene units [1]. The structures of terpenoids are very complex, which can be divided into monoterpenes, sesquiterpenes, diterpenes, etc. Terpenoids may be chain, single-ring, bicyclic, or polycyclic. Their oxygenated derivatives can be alcohols, aldehydes, ketones, carboxylic acids and esters. Terpenoids are not only main flavor components in plants, resins and pigments [2], they are also an important class of pharmacological ingredients [3–8] in Chinese herbal medicine. Essential oils of *Ephedra* have been proven to be responsible for inducing diaphoresis, cardiovascular activity, which include many terpenoids, like alpha-terpineol [9,10], one of the main pharmacological constitutes [11–13].

The parent compounds of many drugs come from natural products. So, it is very important to identify the compounds that exist in the nature. One-dimensional gas chromatography mass spectrometry (GC–MS) has been used for the analysis of volatile and semi-volatile organic compounds [14]. Because of the complicated composition, it is difficult to completely separate all the compounds contained in a complex essential oil sample. In order to solve the co-eluting peaks problem, many classic chemometric approaches have been proposed, including window factor analysis (WFA) [15], heuristic evolving latent projections (HELP) [16,17], iterative target transformation factor analysis (ITTFA) [18,19] and automatic window factor analysis (AWFA) [20]. Finally, GC–MS compound identification could be accomplished through the resolved pure mass spectra (MS).

Accurate mass is usually obtained from pure MS, an experimentally determined mass of an ion according to an appropriate degree of accuracy and precision. It is expressed by an appropriate number of significant figures, which can deduce the elemental formula of the ion. Its importance for reliable identification has been repeatedly emphasized. In many GC–MS cases, multiple search compounds possess similar scores in NIST MS library, or some unknown compounds might not be found in a free or commercial database. These ambiguous NIST MS library search results could be removed using accurate mass measurement, which increased the certainty of elemental formula, and follows

---

[*] Corresponding author. Tel.: +86 731 88830824; fax: +86 731 88830831.
E-mail address: yizeng_liang@263.net (Y. Liang).

[1] First co-author, these authors contributed equally to this study.

0039-9140/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.talanta.2012.10.018

up the discrimination among multiple search compounds with different molecular ions and key fragments. There are a number of methods used for accurate mass measurement, some costly high-resolution MS instruments, including fourier transform mass spectrometry (FTMS) and others [21], were applied in the accurate mass detection. Low resolution instruments with a large systematic shift, including triple quadrupole mass spectrometry, could also be applied in such measurements, which need some mathematical calibration methods [22,23]. These methods were used in the experiment to make the mass deviation within the desired range. Wang and Gu, and Erve et al. [24,25] reported a concept of mass accuracy and spectral accuracy that can reduce the mass error for both high- and low-resolution instruments. This process improves accurate mass measurement and formula determination on a conventional quadrupole mass spectrometer or a high resolution instrument. Finally, it can provide accurate formula, and narrow the search range in the NIST MS library.

Retention indices (RI) are closely related to molecular structure [26,27], and under specific chromatographic conditions, with the column temperature and column properties. Therefore, RI combined with MS, can serve as an important tool for qualitative analysis [28], which was able to effectively improve the qualitative results. In particular, RI is useful for isomeric compounds that possess highly homologous MS, and for compounds whose corresponding MS are unavailable. So, RI was applied for further identification of the unknown compounds in a complex matrix natural sample by using different stationary phases. The amount of candidate compounds could be reduced through RI comparison between NIST library and experiment results while the RI values of some known or unknown compounds could not be found in the database or were commercially unavailable. For these compounds, the solution carries out the in-silico RI method using quantitative structure–retention relationship (QSRR), a technique that relates variations in RI with the variations of several molecular descriptors. There are different methods to develop a QSRR model with predictive or explanatory purpose [29–31], and their results have good statistical qualities with low prediction errors. Therefore, this approach offers an accessible and powerful tool for characterizing ambiguous identification results of essential oils.

In this paper, terpenoids were identified from the "white gray black system" using chemometric methods and RI. Accurate mass of molecular ions was used for the confirmation of some compounds. In-silico RI of terpenoids were also applied to increase the confidence of identification of unknown compounds.

2. Materials and methods

2.1. Materials and the identification of crude drug

The crude herbs of *Ephedra sinica,* and *Ephedra equisetina* were collected in the fall of 2010 in Inner Mongolia, China. Alkane standard solutions of C8–C20 (mixture no. 04070) and C21–C40 (mixture no. 04071) were purchased from Fluka Chemika (Buchs, Switzerland). White camphor oil, cinnamon oil and cinnamon bark oil were purchased from Huabao Flavors & Fragrances Co., Ltd. The botanical origins of *Ephedra sinica,* and *Ephedra equisetina* were identified by Dr. Shao Liu. The chemical identification was done according to the HPLC-DAD detection of ephedrine alkaloids [32]. Crude drug chromatograms were compared with the chromatogram of ephedrine, pseudoephedrine reference substances.

2.2. Preparation of essential oil by hydrodistillation

After the air-dried *Ephedra* smashed, 30 g of samples were precisely weighed and mixed with 300 ml distilled water. The samples were kept for 3 h at room temperature, and subjected to hydrodistillation for 5 h in accordance to the standard extracting method described in the Chinese pharmacopoeia [33]. The obtained essential oils were recovered with *n*-hexane and dehydrated over anhydrous sodium sulfate, finally preserved in a sealed vial at 4 °C, and analysed within 6 h.

2.3. GC/MS analysis

GC–MS analysis was performed on a 7890A gas chromatograph (Agilent, Santa clara, CA, USA) equipped with a 5975 mass spectrometer (Agilent, USA). A HP-5 capillary column (30 m × 0.25 mm i.d., film thickness 0.25 μm, Agilent, USA) was used for separation. The following oven temperature program was used, initiated at 60 °C for 5 min, then increased to 150 °C at the rate of 4 °C min^{-1} for 2 min, finally increased to 250 °C at the rate of 10 °C min^{-1} for 10 min. The spectrometers were operated in electron-impact (EI) mode, scan range 35–500 amu and ionization energy 70 eV. The injector and ion source were kept at 250 and 230 °C, respectively. Split injection was 1 μl conducted with a split ratio of 1:50 and Helium carrier gas was used at a constant flow rate of 1.0 ml min^{-1}. The MS data were set to acquire in "Raw" and "Scan" mode.

2.4. The acquisition of pure MS and identified methods

HELP methods were implemented in C++ and MFC by our laboratory. Selective Ion Analysis (SIA) [34] and Alternative Moving Window Factor Analysis (AMWFA) [35,36] were coded in MATLAB 2010 (a) for windows. The identification of the chemical components in the essential oils was made by comparisons of their retention indices (RI) and MS from the NIST 08 data. The calculation of RI is based on a quasi-linear equation applicable to linear temperature-programmed GC operating conditions [37] using Eq. (1), where RI is the temperature-programmed retention indices; $T(x)$, $T(n)$ and $T(n+1)$ are the retention temperature of the measured composition and two adjacent standard *n*-alkanes containing *n* and $n+1$ carbons under the same operating conditions.

$$RI = 100\left(\frac{\mathrm{T}(x)-\mathrm{T}(n)}{\mathrm{T}(n+1)-\mathrm{T}(n)}+n\right) \tag{1}$$

2.5. Mass calibration methods of molecular ions

Origin V8.6 software (Origin Lab Company) was used for overlapped isotopes structures resolution and Gaussian fitting (Eq. (2)) of the "Raw" MS data. Seven Golden rules software [38] was free downloaded from http://fiehnlab.ucdavis.edu/projects/Seven_Golden_Rules/. Eight compounds whose molecular ions ranged from 99 to 182 were used for calibration. According to the statistical standard calibration curve, the calculated results were used to distinguish different molecular weights in a large number of search results.

$$y = y_0 + \frac{Ae^{(-4\ \ln\ (2)(x-x_c)^2)/w^2}}{w\sqrt{\pi/4\ \ln 2}} \tag{2}$$

2.6. In-silico RI measurement

Descriptor selection in QSRR modeling is important to ensure optimal predictive power of the models. Six types of molecular descriptors were calculated by our group, including constitutional descriptors, topological descriptors, electronic state indices, quantum chemical descriptors, the descriptors related to molecular properties and geometrical descriptors. Considering the energy-lowest structures, all chemical and geometrical descriptors were calculated based on the structural features optimized by MOPAC 2007 program. The specific steps can be seen in Reference [39], including molecular modeling, descriptors generation and cross model validation by measuring its accuracy in predicting the responses of the remaining data. The compound with known RI was measured, and the experimental RI was denoted as RI (experiment). The calculated in-silico RI using QSRR model was denoted as RI (prediction). The relative RI error was determined using formula (3) [29]. Most data analyses were mainly performed in Matlab 2010.

$$RI\ error = 100[\mathrm{RI\ (prediction)} - \mathrm{RI(experiment)}/\mathrm{RI(experiment)}] \tag{3}$$

3. Results and discussion

3.1. GC–MS analysis results of terpenoids

The essential oil samples were diluted using *n*-hexane, and chromatographic conditions were improved to obtain a good chromatographic profile. Samples were analyzed in triplicates, and the correlation coefficients of chromatograms exceeded 0.985. Finally, total ion chromatograms (TICs) of the essential oils from *Ephedra* were obtained. The quantitative analysis was performed with the overall volume integration method (OVI). Qualitative identification of all peaks with high signal-to-noise ratio accepted by comparisons of RI and MS from the NIST 08 data led to the identification of 33 terpenoids in 90 different compounds from *Ephedra*. The identified terpenoids are listed in Table 1 according to their elution order on a HP-5 capillary column.

3.2. Pure MS acquisition of terpenoids

Single peaks, overlapping peaks and embedded peaks constructed a "white/gray/black system". In this system, pure MS could not be obtained using simple methods. The matrix data generated by chromatography hyphenated instrument including GC–MS contain both information of chromatogram and MS. Multivariate resolution methods based on matrix data form have satisfied the urgency to resolve the two-way data into pure chromatograms and pure MS, both quantitatively and qualitatively. HELP, SIA and AMWFA methods have been proven to be effective methods which can resolve some overlapped or embedded peaks developed by our lab. HELP is a classic chemometric method which uses selective information to resolve the overlapping peaks in two-dimensional chromatographic data matrix. Now, let us introduce the steps of HELP method. Fig. 1(a) shows a segment in the range of 12.70–13.10 min obtained from the sample of *Ephedra equisetina*. The baseline drift of GC–MS chromatogram will blur signals and deteriorate analytical results. Thus, background removal and curve fitting is necessary when we obtain pure MS. Here the ordinary manual linear deduction or the adaptive iteratively reweighted penalized least squares method [40] can satisfy our requirement. Three characteristic ions peaks represented with broad-brush, and the pure component regions marked with 1, 2, 3 and zero-component region of each component can easily be observed in Fig. 1(a), we can also use mathematical methods to prove the correctness. The data were determined using the evolving latent projective graph and rankmap on the basis of the eigenstructure tracking analysis (Fig. 1(b) and (c)). From the rankmap, we can see the rank in pure component regions equal to one, which means single component. At last, three pure MS were resolved by means of full rank resolution. Finally, Tetramethylpyrazine, α-Terpinolene and

Table 1
Identified terpenoids, their experimental RI and their corresponding category (train set or test set or candidate compounds) among the essential oils of *Ephedra* and other commercial samples.

RI	Name	Category	RI	Name	Category
926	α-Thujene	Train	1411	α-Gurjunene	train
972	2-Thujene	Train	1424	Caryophyllene	train
975	β-Pinene	Train	1457	Humulene	train
998	p-Menth-2-ene	Train	1463	(-)-Alloaromadendrene	train
1005	α-Phellandrene [a]	Train	1479	Germacrene D [a]	train
1014	Isocineole [a]	Train	1485	4-(2,6,6-Trimethylcyclohexa-1, 3-dienyl)but-3-en-2-one [a]	train
1016	α-Terpinene [a]	Train	1488	β-Ionone [a]	train
1028	Limonene [a]	Train	1523	Cadina-1(10),4-diene	train
1030	Cineole [a]	Train	1586	Caryophyllene oxide	train
1049	cis-β-Ocimene	Train	1606	Cedrol [a]	train
1058	γ-Terpinene [a]	Train	1639	γ-Eudesmol [a]	train
1090	4-Isopropenyltoluene [a]	Train	1654	δ-Cadinol, (-)-[a]	train
1121	cis-2-p-Menthen-1-ol [a]	Train	1700	Eudesm-7(11)-en-4-ol [a]	train
1134	1-Terpinenol [a]	Train	947	Camphene	test
1138	L-Pinocarveol	Train	991	β-Myrcene	test
1145	(-)-Camphor	Train	1024	p-Cymene [a]	test
1193	α-Terpineol [a]	Train	1039	trans-β-Ocimene	test
1198	γ-Terpineol [a]	Train	1088	α-Terpinolene [a]	test
1240	β-Citral	Train	1145	β-Terpineol [a]	test
1248	p-Menth-4-en-3-one [a]	Train	1164	δ-Terpineol [a]	test
1256	Geraniol [a]	Train	1177	4-Terpineol [a]	test
1273	trans-Citral	Train	1352	α-Longipinene	test
1275	Perilla alcohol [a]	Train	1385	Geranyl acetate	test
1351	α-Terpineol acetate [a]	Train	1414	α-Cedrene [a]	test
1354	1,1,6-Trimethyl-1,2-dihydronaphthalene [a]	Train	1581	(-)-Spathulenol [a]	test
1386	β-Damascenone [a]	Train	1258	p-Menth-2-en-7-ol, trans- [a]	compound
1409	1,4-Methanoazulene, decahydro-4,8,8-trimethyl-9-methylene-, [1S-(1a′,3aa′,4a′,8aa′)]-	Train	1284	2-Caren-10-al [a]	compound

[a] Identified compounds came from *Ephedra*.

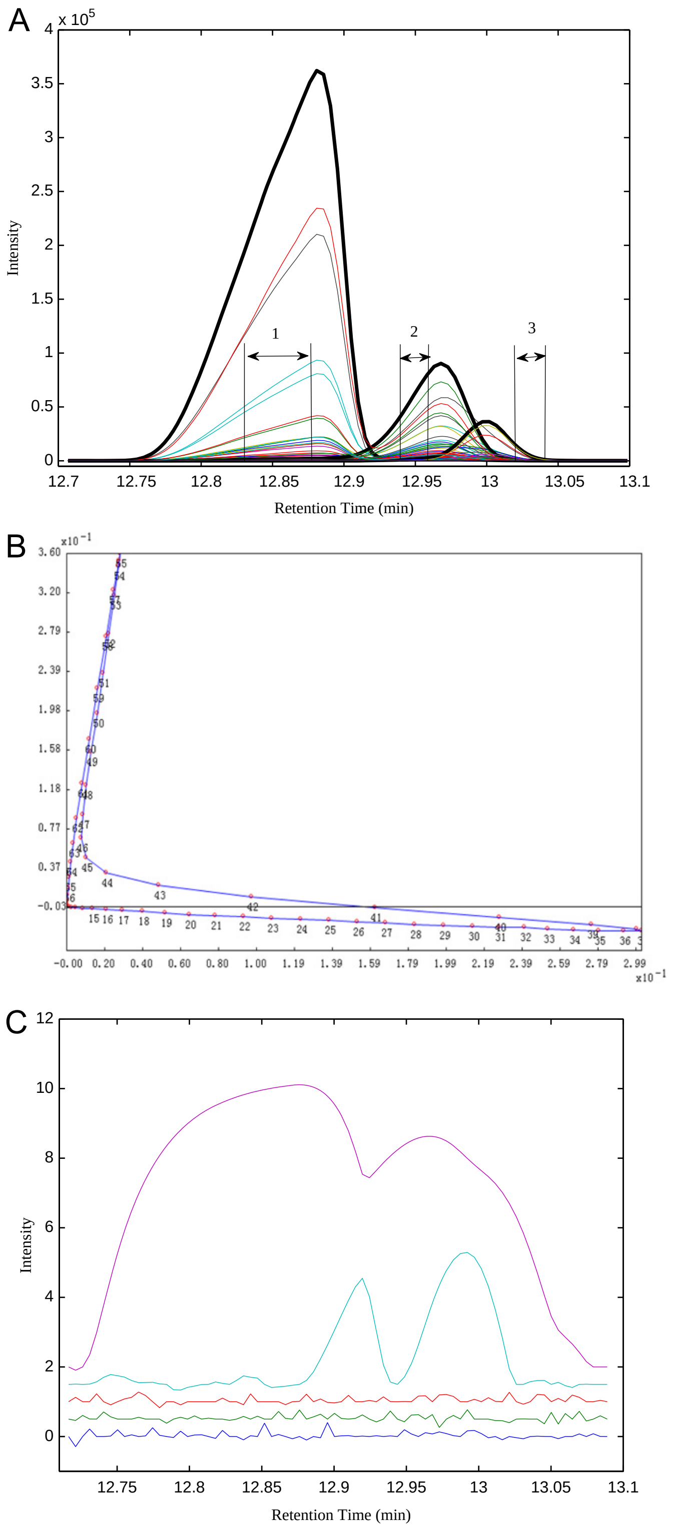

Fig. 1. (a) Overlapped peaks of the segment in the range of 12.70–13.10 min obtained from the sample of *Ephedra equisetina*, (b) the evolving latent projective graph and (c) the rankmap. (1). The pure component region of Tetramethylpyrazine; (2) the pure component region of α-Terpinolene and (3) the pure component region of Benzene, 1-methyl-4-(1-methylethenyl).

Benzene, 1-methyl-4-(1-methylethenyl)- were identified through NIST MS library search. In addition, AMWFA and SIA were also demonstrated effective in resolving overlapped peaks and embedded peaks. These chemometric methods, together with the oven temperature program, can effectively enhance the accuracy of quantitative and qualitative results [34–36]. The obtained pure MS can provide guarantee for the accurate mass determination of molecular ions or key fragments.

3.3. *Auxiliary confirmation using mass accuracy of molecular ions*

The ambiguous NIST MS library search results also emerged after using pure MS. Sometimes, candidate compounds with approximative molecular weight emerged in a lot of search results. Simultaneously, the final search results also needed confirmation of their molecular formulas. For the “Raw” mode GC–MS data, molecular ion contains information generated by the natural abundances of the various isotopes contained in the compound. Accurate mass measurement through sophisticated mathematical operations could determine elemental formulae to exclude the uncertainty. In recent years, some computational approaches based on the isotopes information, including Mass-works [41], have been applied in effective elemental composition elucidation.

A simple method to distinguish different molecular weights in a large number of search results was introduced below. Peak 1 and peak 2 in Fig. 1 are examples which were introduced in 3.2, both with a molecular weight of 136. Tetramethylpyrazine was verified by a reference substance as having a molecular weight of 136.1943; however, the molecular weight of peak 2 remained unknown. Firstly, the MS data was set to acquire in “Raw” mode. Secondly, a mixture of eight external calibration standards provided molecular ions of m/z 99.154, 108.1411, 121.1366, 125.191, 136.1943, 143.207, 178.2277 and 182.2179. Their molecular ions were extracted and isotopic profiles were fitted, like in Fig. 2. The errors were calculated by subtracting the measured value (accurate mass) from the true value (exact mass), in accordance with the term “accuracy” [42]. The obtained statistical standard calibration curve of the known masses against the calculated errors was linearly fitted (Fig. 3) and the correlations coefficient is 0.9517. Thirdly, the essential oil run in the GC–MS, and molecular ions of peak 1 and peak 2 were extracted. Their fine isotopes structures are typically overlapped with each other due to approximative masses, including M-H, M, M+1. The phenomenon often occurs in low-resolution instrument, and even in some high-resolution instruments. In these under-resolved overlapped isotopes structures, the information relevant to elemental composition is most involved. Using Peak Analyzer function in Origin software, the overlapped isotopes structures could be resolved. Additionally, a mathematical algorithm (Gaussian) fitting was applied to the raw spectrum to obtain the fully smooth and symmetrical spectrum using Origin software. The peak center masses (axis) and isotope ratio (height) were calculated using specific parameters (Fig. 2). Using the calculated values of the fitted isotopic profiles in combination with the correction 0.0097 between theory and fitting value (Fig. 3) applied to compensate for the mass errors, and the calculated molecular weight could be obtained. In order to elaborate the validity of this approach more effectively, isotopic profiles of molecular ion 136.1943 of peak 1 was used as an internal quality control compound to ensure the correctness of calibration data. The comparison between theory isotopic profiles calculated by elemental formula calculators (MWTWIN) and fitting isotopic profiles using Origin software can be seen in Fig. 4. Two compounds could be distinguished from their “Raw” mode of molecular ion isotopic profiles, even though their molecular weights are both 136, which could not be discerned from bar graphs obtained from “Scan” mode.

Seven golden rules algorithm correctly changed accurate mass to elemental compositions for compounds consisting of the elements C, H, N, S, O, P, F, Cl and Br at a high probability, and can automatically constrain the possible candidate compounds.

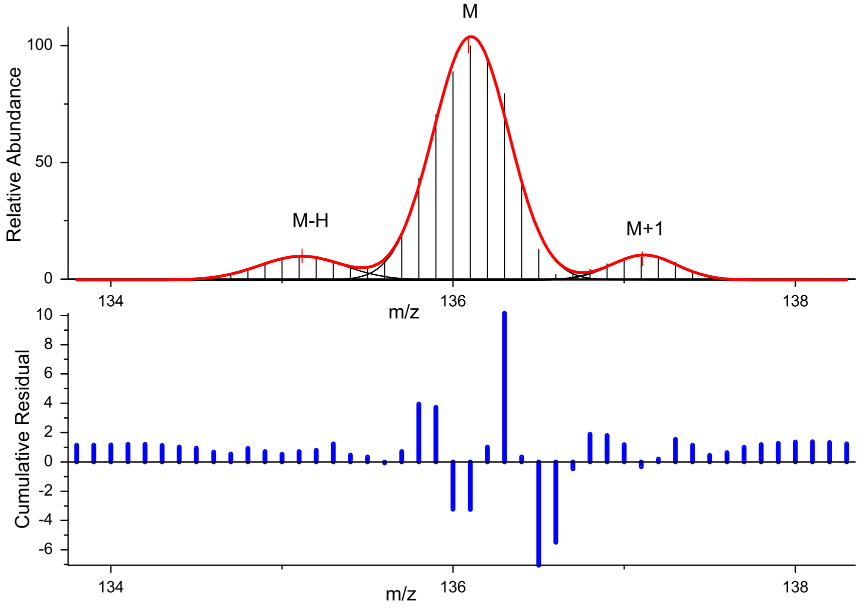

Fig. 2. A fitted isotopes profile and cumulative residual of *m*/*z* 136.1943 using Origin software, which came from tetramethylpyrazine.

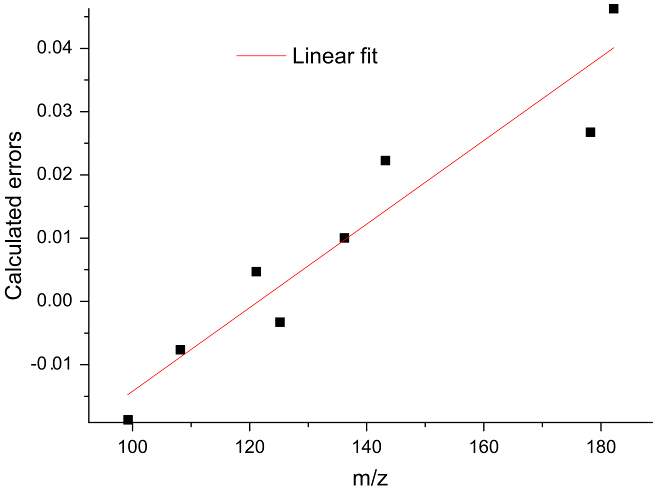

Fig. 3. A statistical standard calibration curve between the known masses and calculated errors.

In the test, the software yields the elemental formula within the top hits according to the calibration molecular weight of peak 2 (136.2344), and selects the most likely molecular formulas. Integrated with NIST MS library search results, C10H16 (136.2340) is the correct formula (molecular weight). Using this simple method, we also distinguished C9H10O (134.1751) from C8H9NO (135.1632) for the GC–MS peaks located in the 15.39 min obtained from the sample of *Ephedra equisetina*. These candidate compounds possessed high similarity ($>$ 90%), and could not be differentiated using ordinary methods because parts of RI could not be found. The final calculated molecular weight using the method is 134.1773 through the method; thus, the answer is C9H10O. For some weak molecular ions, this method also seems to be effective, e.g. the GC–MS peaks located in the 4.32 min obtained from the sample of *Ephedra sinica*. The ambiguous candidate compounds were C9H14 (122.2075) and C8H10O (122.1644), and their similarity is 93% and 86% respectively. The calculated molecular weight is 122.1938 using the incomplete molecular ion isotopic profiles, and C9H14 was selected because its relative error is lower.

Co-eluting compounds make it difficult to obtain the corresponding pure MS. While many identified compounds also existed in the essential oils, their molecular weights were very close to each other. Numerous ingredients of essential oil also produce a favorable side for the molecular weight confirmation of compounds. There is no urgency to add pure internal standard compounds; the identified components in the essential oils can also serve as internal standards. Mass accuracy is influenced by the signal strength, noise, isotope interferences etc, and the identification and structure elucidation of unknown compounds remain challenging using this approach. However, it is sufficient to distinguish different molecular weights in a large number of known search results.

3.4. Auxiliary confirmation using in-silico RI

NIST MS library search results usually possess the same molecular weight, and some molecular ions peaks are low or undetectable in GC–MS data. The method of mass accuracy is useless when facing the situations described above. RI is an important qualitative indicator, which is closely related to molecular structure. The identification may resort to RI because the literature data are available for inspection. There are four NIST MS library search results against peak 2 (Fig. 1), with similarity over 93%, but only that of α-Terpinolene is identical with the experimental RI. This approach could identify some known compounds. Unfortunately, RI of many compounds cannot be found in the commercial or free database, hence, the only resort is in-silico RI using QSRR method.

RI is a standardized parameter in reporting GC data, while its experimental measured value is affected by a number of factors such as column type, gradient temperature program, etc. To obtain more RI (experiment) data of terpenoids under the same conditions, fifty-five terpenoids from essential oils of *Ephedra*, white camphor oil, cinnamon oil, cinnamon bark oil were determined by GC–MS after successive injections to the same instrument. Terpenoids were selected to cover a relatively large range of RI, from 900 to 1700 (Table 1). Topological, constitutional, quantum chemical and geometrical descriptors were used to describe molecular structure. Multivariable linear regression (MLR) was employed to map the descriptors to the RI. For a more

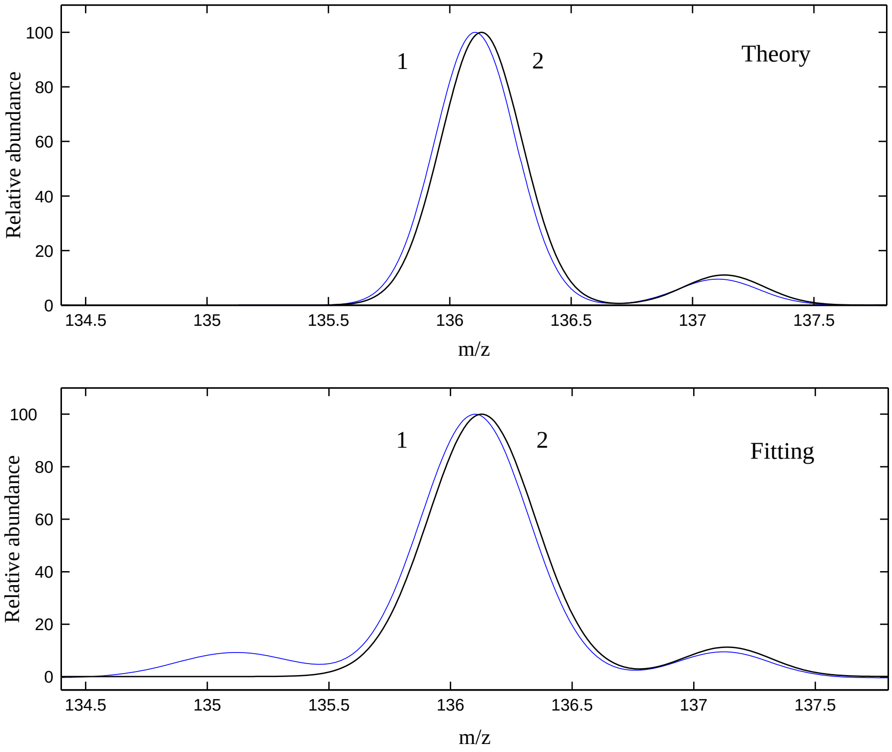

Fig. 4. Comparison between theory isotopic profiles calculated by elemental formula calculators (MWTWIN) and fitting isotopic profiles using Origin software. (1) *m*/*z* 136.1943 came from Tetramethylpyrazine; (2) *m*/*z* 136.2340 came from α-Terpinolene.

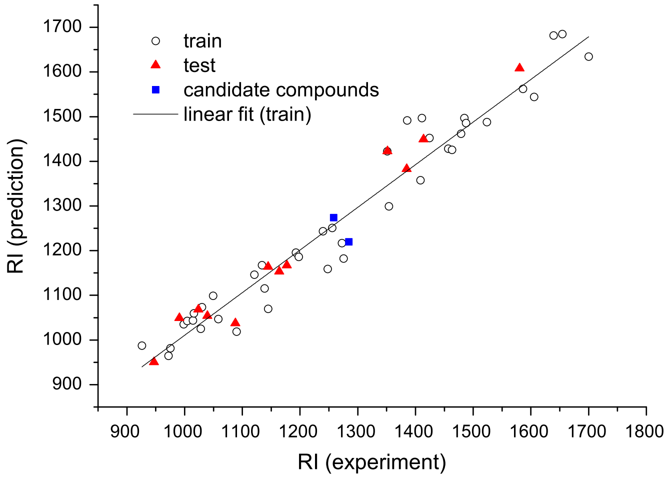

Fig. 5. Plot of RI (prediction) vs. RI (experiment) for train set, test set and two candidate compounds.

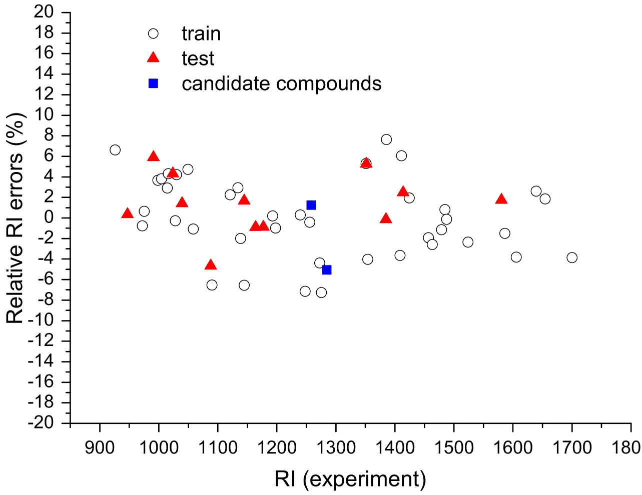

Fig. 6. Plot of relative RI errors for train set, test set and two candidate compounds.

detailed modeling process, refer to [39]. Forty terpenoids were used as the training set, 12 terpenoids were used as the test set, and two candidate compounds without NIST library values were used as samples. The RI (prediction) or relative RI errors against RI (experiment) are shown in Fig. 5 and Fig. 6. The predictive power of the model has been validated by cross model validation and external test. Statistical parameters were used for assessing the statistical significance of the QSRR model. The square of the correlation coefficient (R^2), square of correlation coefficient for cross validation (Q^2) and square of correlation coefficient for external test (Test^2) were 0.961, 0.953 and 0.954, respectively. Finally, a model with a low prediction error and a good correlation coefficient was obtained. This model could be used for the auxiliary confirmation of some compounds whose NIST library data is unavailable.

Despite relatively large errors of predicted RI, the combination of in-silico RI and pure MS provides a powerful means to increase the confidence of the identification. Two GC–MS peaks obtained successful results from the test, their experimental RI were 1258, 1284 respectively, and their reference RI values could not be found in the NIST library. After treating the candidate compounds of two GC–MS peaks with chemometric methods, similarity and RI values in the NIST library can be compared (Table 2). RI values of highly similar compounds are unknown, that is, identified results could not be further confirmed. Using in-silico RI, relative RI errors of two highly similar compounds are less than 5.10%.

Table 2
Candidate compounds of two GC–MS peaks, similarity, RI values in the NIST library and predicted RI values.

No	RI (experiment)	Candidate compounds	Similarity (%)	RI (NIST)	Predicted RI (Relative RI errors)
1	1258	p-Menth-2-en-7-ol, trans-	92	Unknown	1220 (−5.0%)[a]
		Bicyclo[4.1.0]heptane, 7-(1-methylethylidene)-	88	1003	
2	1284	2-Caren-10-al	89	Unknown	1274 (1.3%)[a]
		Bicyclo[3.1.1]hept-2-ene-2-carboxaldehyde, 6,6-dimethyl-	80	1196	

[a] a moderate agreement between the predicted and experimental RI values was found.

The predicted and experimental RI values agree to some extent (Fig. 5 and Fig. 6). Certainly, final results have been verified, RI values of trans-p-Menth-2-en-7-ol, the stereoisomers of p-menth-2-en-7-ol, trans- and that of 2-Caren-10-al could be found in References [43,44].

4. Conclusion

The paper has shown how to apply pure MS, NIST MS library search, RI, accurate mass and in-silico RI constraints to determine the chemical composition of complex mixtures, so as to improve the accuracy of compounds identification based on GC–MS data. After background removal, curve fitting and chemometric resolution, pure MS were obtained. Together with Kovats RI, thirty three terpenoids from *Ephedra* were mainly identified through NIST MS library search and RI matching. In addition, accurate mass determination was developed to distinguish different molecular weights in a large number of search results, using a simple external calibration method, Origin software and seven golden rules algorithm. Moreover, RI predictions were able to provide a powerful tool to increase the confidence of the identification, as evidence, the predicted and experimental RI values of two candidate compounds agreed to some extent. All in all, these mathematical approaches and statistical methods can be used to obtain relevant information about a material system under study, and it can enhance the separation and identification at large.

Novelty statement

Many ambiguous search results kept us confused despite pure mass spectra and retention indices were applied in further identification. In the paper, we developed a simple external calibration method to produce accurate mass using Origin software. In addition, in-silico RI were also applied to increase the confidence of identification of unknown compounds.

Acknowledgments

This work is financially supported by the National Nature Foundation Committee of PR China (Grant nos. 20875104, 20975115 and 10771217), the international cooperation project on traditional Chinese medicines of ministry of science and technology of China (Grant no. 2007DFA40680), and Central South University for special support of the basic scientific research project (No. 2010QZZD007). The studies meet with the approval of the university's review board.

References

[1] E. Realdon, Boll. Chim. Farm. 99 (1960) 52–58.
[2] L. Caputi, E. Aprea, Recent Pat. Food Nutr. Agric. 3 (2011) 9–16.
[3] J.J. Lu, Y.Y. Dang, M. Huang, W.S. Xu, X.P. Chen, Y.T. Wang, J. Ethnopharmacol. 143 (2012) 406–411.
[4] A.K. Ibrahim, Nat. Prod. Res. (2012). (Epub ahead of print).
[5] T. Goto, N. Takahashi, S. Hirai, T. Kawada, PPAR Res. (2010). (Epub ahead of print).
[6] S. Fukumoto, E. Sawasaki, S. Okuyama, Y. Miyake, H. Yokogoshi, Nutr. Neurosci. 9 (2006) 73–80.
[7] I.C. Sun, Y. Kashiwada, S.L. Morris-Natschke, K.H. Lee, Curr. Top Med. Chem. 3 (2003) 155–169.
[8] K. Shimizu, M. Ozeki, A. Iino, S. Nakajyo, N. Urakawa, M. Atsuchi, Jpn. J. Pharmacol. 86 (2001) 223–229.
[9] L. Ji, Z. Xu, G. Pan, G. Yang, Zhongguo Zhong Yao Za Zhi 22 (1997) 489–492.
[10] L.L. Ding, S.S. Shi, J. Cui, S.C. Wang, Z.T. Wang, Zhongguo Zhong Yao Za Zhi 31 (2006) 1661–1664.
[11] M.G. de Oliveira, R.B. Marques, M.F. de Santana, A.B. Santos, F.A. Brito, E.O. Barreto, D.P. De Sousa, F.R. Almeida, D.J. Badauê-Passos, A.R. Antoniolli, L.J. Quintans-Júnior, Basic Clin. Pharmacol. Toxicol. 111 (2012) 120–125.
[12] L.J. Quintans-Júnior, M.G. Oliveira, M.F. Santana, M.T. Santana, A.G. Guimarães, J.S. Siqueira, D.P. De Sousa, R.N. Almeida, Pharm. Biol. 49 (2011) 583–586.
[13] R. Souza, M. Cardoso, C. Menezes, J. Silva, D. De Sousa, J. Batista, Daru 19 (2011) 277–281.
[14] B. Mendes, J. Gonçalves, J.S. Câmara, Talanta 88 (2012) 79–94.
[15] R. Edmund, Malinowski, J Chemometr. 6 (1992) 29–40.
[16] O.M. Kvalheim, Y.Z. Liang, Anal. Chem. 64 (1992) 936–946.
[17] Y.Z. Liang, O.M. Kvalheim, Chemometr. Intell. Lab. Syst. 20 (1993) 115–125.
[18] L. Zhang, W. Zhang, D. Cao, M. Zeng, Y. Liang, O.M. Kvalheim, J. Chromatogr. A. 1218 (2011) 7219–7225.
[19] B.G.M. Vandeginste, W. Derks, G. Kateman, Anal. Chim. Acta 173 (1985) 253–264.
[20] E.R. Malinowski, J. Chemometr. 10 (1996) 273–279.
[21] C.L. Chitescu, E. Oosterink, J. de Jong, A.A. Stolker, Talanta 88 (2012) 653–662.
[22] H.P. Benton, E.J. Want, T.M. Ebbels, Bioinformatics 26 (2010) 2488–2489.
[23] H.C. Köfeler, M.L. Gross, J. Am. Soc. Mass Spectrom. 16 (2005) 406–408.
[24] Y. Wang, M. Gu, Anal. Chem. 82 (2010) 7055–7062.
[25] J.C. Erve, M. Gu, Y. Wang, W. DeMaio, R.E. Talaat, J. Am Soc, Mass Spectrom. 20 (2009) 2058–2069.
[26] M.H. Fatemi, Z. Ghorbannezhad, J. Chromatogr. Sci. 49 (2011) 476–481.
[27] S. Bouzouane, M. Righezza, A. Touabet, J. Chromatogr. Sci. 50 (2012) 137–144.
[28] L.X. Zhang, B.B. Tan, M.M. Zeng, H.M. Lu, Y.Z. Liang, Talanta 88 (2012) 311–317.
[29] V.V. Mihaleva, H.A. Verhoeven, R.C. de Vos, R.D. Hall, R.C. van Ham, Bioinformatics 25 (2009) 787–794.
[30] F. Liu, Y. Liang, C. Cao, N. Zhou, Talanta 72 (2007) 1307–1315.
[31] V.K. Gupta, H. Khani, B. Ahmadi-Roudi, S. Mirakhorli, E. Fereyduni, S. Agarwal, Talanta 83 (2011) 1014–1022.
[32] M. He, Y.Z. Liang, Z.M. Zhang, Y.P. Li, Z.D. Zeng, D.S. Cao, Y.H. Yun, J. Yan, J. Liq. Chromatogr. Related Technol. (2012). (Epub ahead of print).
[33] National Commission of Chinese Pharmacopoeia, Pharmacopoeia of the People's Republic of China, Publishing House of People's Health, Beijing, 2000.
[34] B.B. Tan, Y.Z. Liang, L.Z. Yi, Metabolomics 6 (2010) 219–228.
[35] M. He, Y.P. Li, J. Yan, D.S. Cao, Y.Z. Liang, J. Chromatogr. Sci. (2012). (Epub ahead of print).
[36] Z.D. Zeng, Y.Z. Liang, Y.L. Wang, X.R. Li, L.M. Liang, Q.S. Xu, J. Chromatogr. A 1107 (2006) 273–285.
[37] E. Kováts, Helv. Chim. Acta 41 (1958) 1915–1932.
[38] T. Kind, O. Fiehn, BMC Bioinformatics 8 (2007) 105.
[39] J. Yan, D.S. Cao, F.Q. Guo, L.X. Zhang, M. He, J.H. Huang, Q.S. Xu, Y.Z. Liang, J. Chromatogr. A. 1223 (2012) 118–125.
[40] Z.M. Zhang, S. Chen, Y.Z. Liang, Analyst 135 (2010) 1138–1146.
[41] K. Liu, B. Ma, Y.D. Wang, X.Y. Chen, D.F. Zhong, Yao Xue Xue Bao 42 (2007) 1112–1114.
[42] A.G. Brenton, A.R. Godfrey, J. Am. Soc. Mass Spectrom. 21 (2010) 1821–1835.
[43] Q.B. Wang, Y. Yang, X.M. Zhao, B. Zhu, P. Nan, J.Y. Zhao, L. Wang, F. Chen, Z.J. Liu, Y. Zhong, Food Chem. 98 (2006) 52–58.
[44] J.F. Cicció, C. Chaverri, Rec. Nat. Prod. 6 (2012) 371–375.