

Short communication

Essential oils analysis. II. Mass spectra identification of terpene and phenylpropane derivatives

R. Oprean *, L. Oprean, M. Tamas, R. Sandulescu, L. Roman

*Analytical Chemistry Department, Faculty of Pharmacy, University of Medicine and Pharmacy “Iuliu Hatieganu”,
13 Emil Isac Street, RO-3400 Cluj-Napoca, Romania*

Received 16 May 2000; received in revised form 14 October 2000; accepted 27 October 2000

Abstract

Mass spectra are widely used in order to identify the peaks resulting from a chromatographic separation. The most common approach to solve the problem for unknowns on whom very little other structural information is available is the use of a retrieval algorithm and a reference mass spectra database. The wide variety of mass spectra recorded with different instruments under various experimental conditions can lead to erroneous results. In order to improve the accuracy of the results, we proposed earlier an identification algorithm, which combines the information obtained from both GC and MS fingerprints. This paper presents a new algorithm based on the comparison of the unknown mass spectra with several libraries (including Wiley and NIST) by using reverse and direct search algorithms respectively. The results of the comparisons were quantified with respect to the match quality and the interference compounds. A global match index for the comparison using all the above information was computed and the results were presented as the match probability. This index expresses more accurately the matches between unknown and all the available libraries mass spectra. In order to verify our algorithm, we tried to identify the compounds separated by GC-MSD from different species of *Acorus calamus L. (Araceae)* essential oils. The probability of the matches increases compared with the quality of matches resulting from Wiley and NIST libraries. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Gas chromatography-mass selective detector; Reference mass spectra libraries; Global match index; Terpenes; Phenylpropane derivatives; Essential oil analysis

1. Introduction

The chemical nature of essential oils makes them suitable for analysis by gas chromatography-mass selective detector (GC-MSD). The essential oils are mixtures of terpenes or

* Corresponding author. Fax: +40-64-197257.
E-mail address: roprean@umfcluj.ro (R. Oprean).

phenylpropane derivatives in which, the chemical and structural differences between compounds are minimal. The mass spectra of these compounds are very similar, peak identification being very difficult and sometimes impossible.

The most common approach to solve the problem for unknowns on whom very little other structural information is available, is the use of a retrieval algorithm and a reference mass spectra database. The capabilities of a retrieval algorithm depend directly on the quality and comprehensiveness of the reference database employed. Nowadays, two different mass spectral databases are commonly used as references: NIST/EPA/NIH and the Registry of Mass Spectral Data. The NIST/EPA/NIH mass spectral databases contain in the 1992 version some 62 000 mass spectra of an equal number of chemicals and about 12 500 replicas. The largest database is the Registry of Mass Spectral Data, hereafter called the Wiley database, in 1996 containing 275 000 different spectra. The last collection has resulted from the cooperation of many scientists and because the measurements have been made using a wide variety of instruments and experimental conditions, Wiley collection includes multiple copies of the mass spectrum of a chemical when the spectra are different.

In order to evaluate the composition of the essential oils by GC-MSD as well as possible and for increasing the reliability of the analytical results, we proposed earlier an algorithm that consider two parameters as identification criteria for an unknown peak [1]. The proposed parameters were the match index of the unknown mass spectrum with spectral libraries and the relative retention indices computed from the retention times of the unknowns relative to a mixture of *n*-alkanes. A new parameter that characterized the whole GC-MSD method was defined. The match index used for the calculations was obtained by using the reverse search algorithm (PBM) integrated in the ChemStation software and the Wiley library.

Because the uses of only one retrieval algorithm by using only one spectral library as reference database can led often to erroneous results, we are proposing the use of both reverse and direct search and more than one library in order to

compute a global match index that describes more reliable the concordance of the unknown with the references.

2. Materials and methods

2.1. Samples

The terpenes and phenylpropane derivative samples (Table 1) were purchased from Roth (Karlsruhe, Germany). The *Acorus calamus* L. essential oil was obtained by steam distillation of fresh rhizome of the Romanian plant.

2.2. Reagents

All solvents (hexane) were of chromatographic grade and were purchased from Merck (Darmstadt, Germany).

2.3. Solutions

The sample solutions of terpenes and phenylpropane derivatives were prepared dissolving 1 mg standard substance in 30 ml hexane. The sample solution of essential oils for GC-MS was prepared dissolving 5 mg essential oil in 20 ml hexane.

2.4. Instrumentation

GC-MSD analyses were performed on a Hewlett-Packard 5890 series II — 5972 MSD using a HP-5 MS (5%-Diphenyl-95%-dimethylpolysiloxane) column, 0.26 mm i.d. \times 30 m, 0.25 μ m coating thickness. The GC was operated under the following conditions: manual injection, split 1:20; injector temperature: 250°C; carrier gas: He; flow: 1 ml min⁻¹; linear velocity: 36.4 cm s⁻¹; oven temperature programmed from 60 to 240°C at 3°C min⁻¹; detector temperature: 280°C, time run: 60 min. The MSD was operated at 70 eV, scan range 41–300 amu, scan — TIC. MSD were tuned before each injection using perrfluorotributylamine as the tuning standard.

2.5. Computers and software

HP Vectra Pentium 75 computer equipped with Hewlett-Packard ChemStation B.02.02. Acquisition software was used. The peaks were integrated using ChemStation integrator.

Wiley and NIST libraries were used as reference databases. Library searches were performed using ChemStation B.02.02. Software for Wiley library and Mass Spectral Search Program 1.1.a for NIST libraries.

Further calculations were performed using Mi-

crosoft Excel 2000 and Microsoft VisualBasic for applications.

2.6. Abbreviations

Hereafter, the following abbreviations were used: Similarity index: SI; Position index: PI; Wiley mass spectra library: W; NIST/EPA/NIH mass spectral databases: N; Reverse search algorithm: R; Direct search algorithm: D. From the comparison algorithms results the following seven

Table 1
Indices computed using the proposed algorithm for the terpenes and phenylpropane derivatives

No.	Compound	SI		PI		Global indices		
		Wiley	NIST	Wiley	NIST	Wiley	NIST	Total
1	Asarone, alpha-	95.8	91.8	35.2	40.0	75.6	74.6	75.1
2	Asarone, beta-	96.4	90.4	49.2	47.4	80.7	76.1	78.4
3	Bornyl acetate	84.7	90.8	44.0	41.1	71.1	74.2	72.7
4	Caryophyllene, beta-	96.5	93.6	77.1	88.9	90.0	92.0	91.0
5	Carvone, d-	94.4	92.1	50.2	39.8	79.7	74.7	77.2
6	Cimene, p-	92.5	93.6	49.7	40.8	78.2	76.0	77.1
7	Cineole, 1,8-	90.1	92.3	100.0	99.0	93.4	94.5	94.0
8	Citronellal	79.7	89.2	90.7	66.1	83.4	81.5	82.5
9	Citronellol	59.7	89.5	27.7	46.3	49.0	75.1	62.1
10	Eugenol	95.9	91.8	83.4	52.5	91.8	78.7	85.3
11	Phellandrene, alpha-	78.2	91.0	91.4	73.6	82.6	85.2	83.9
12	Fenchone	81.6	91.1	85.1	77.8	82.8	86.7	84.8
13	Geranial	86.3	89.7	53.9	36.1	75.5	71.8	73.7
14	Geranyl acetate	70.7	92.5	74.7	41.7	72.0	75.6	73.8
15	Humulene, alpha-	90.0	86.8	71.9	62.4	84.0	78.7	81.4
16	Ionone, alpha-	93.6	91.3	80.5	78.8	89.2	87.1	88.2
17	Limonene	94.0	91.4	100.0	79.6	96.0	87.5	91.8
18	Linalyl acetate	86.6	88.4	78.5	24.7	83.9	67.2	75.6
19	Linalool	80.8	91.7	92.2	94.7	84.6	92.7	88.7
20	Myrcene, beta-	91.5	91.8	100.0	82.1	94.3	88.6	91.5
21	Neral	76.5	84.6	86.4	49.6	79.8	73.0	76.4
22	Neryl acetate	79.6	91.4	71.9	68.9	77.0	83.9	80.5
23	Ocimene, beta-	89.6	90.9	45.8	33.5	75.0	71.8	73.4
24	Pinene, alpha-	94.5	92.7	90.8	41.1	93.2	75.5	84.4
25	Pinene, beta-	94.1	93.7	84.2	93.3	90.8	93.6	92.2
26	Sabinil acetat	77.5	90.4	100.0	100.0	85.0	93.6	89.3
27	Terpinene, gamma-	94.5	92.9	85.5	86.7	91.5	90.8	91.2
28	Terpinen-4-ol	82.7	92.1	92.8	100.0	86.1	94.8	90.5
29	Terpineol, alpha-	85.4	90.0	83.8	81.1	84.9	87.1	86.0
30	Terpinolene	96.5	92.3	76.5	43.3	89.8	76.0	82.9
31	Thujone, alpha-	88.4	91.7	46.7	47.5	74.5	77.0	75.8
32	Thujone, beta-	93.5	86.9	54.7	31.0	80.6	68.3	74.5

factors: SI_{WR} , PI_{WR} , PI_{WD} , SI_{NR} , SI_{ND} , PI_{NR} , PI_{ND}

Observation: The ChemStation retrieval algorithm does not compute the similarity index (SI_{WD}) for direct search using Wiley database as reference library.

2.7. Protocols

Standard solutions of 32 terpenes and phenylpropane derivatives were analyzed by GC-MSD under given conditions. Each mass spectra correspondent to a chromatographic peak was subjected to comparison with NIST and respectively Wiley mass spectra libraries (W) by using both reverse (R) and direct (D) search algorithms delivered with the above-mentioned software. Each search routine assigns significance to each of the peaks in the unknown spectrum and uses these to find the most probable matches in the reference libraries. The results are presented as a classification of the reference library spectra that matches the unknown and there are sorted descendent by using the similarity index (SI) as the sort criterion. The SI was considered as the primary criterion in order to compute the global match index (GI).

If the reference library contains more than a spectrum for one compound, this can be placed on different positions in the hierarchy. Often, between these compounds are placed other substances, hereafter, called interferences. We consider that the interferences affect the results so, in order to compute the GI for a comparison with reference databases, it is necessary to consider a second index that take on count the interferences. This index was called position index (PI) and refers to the ranks of the presumed identity of an unknown relative to the databases and the presence of the interferences.

3. Results and discussion

In order to compute a GI for an unknown mass spectrum by using commercially available retrieval algorithms delivered with the acquisition software, we considered that there are two factors that influence the final result: the SI and the PI. Each of the lately mentioned indices could be computed using both of R and D.

The global similarity index (G_{SI}) express the influence of each individual similarity indices computed by using retrieval algorithms over the final result. For Wiley and NIST databases, respectively, we proposed the following equations:

$$\text{Wiley: } SI_W = \frac{\sum_{i=1}^n SI_{WR}(i)}{n} \quad (1)$$

$$\text{NIST: } SI_N = \frac{\sum_{i=1}^n \left[\frac{SI_{NR}(i) + SI_{ND}(i)}{2} \right]}{n} \quad (2)$$

Similarity indices are computed according to Eqs. (1) and (2) for each of the Wiley and NIST databases. For the Wiley mass spectra database, only the SI_{WR} (SI of the reverse search, so called match quality/Qual) is available, thus the SI for this database is the average of the individual SI_{WR} for the presumed identity of the unknown. For the NIST mass spectra database, the SI is computed as the average for each direct (SI_{ND}) and respectively reverse search (SI_{NR}). The values obtained for all this parameters are directly correlated to the quality of the unknown mass spectrum and the diversity of the mass spectra of the correspondent compounds from the databases.

The G_{SI} was defined as the average of the two above-defined indices and express the correlation of all the similarity indices for the above-motioned databases. The G_{SI} value can vary in large ranges, depending on the individual SI values. For an ideal compound and for ideal databases, the G_{SI} value is 100 and values over 70 show a good correlation between the search algorithms and the quality of the mass spectra from different mass spectra databases.

In order to calculate the PI, we are proposing the following algorithm:

1. For the last position occupied by an presumed match compound it is attributed value of 1.
2. The above value is increasing (step 1) to the top position.
3. It is computed the sum of the ranks occupied of the presumed match compound.
4. It is computed the sum of the ranks occupied of all compounds till the level of the first step.

5. PI is computed as the probability for which a certain compound is present in the hierarchy.

Notations: R : rank in hierarchy; R_{\max} : last rank of the presumed compound; N_{\max} : number of the same compounds presents in the hierarchy

In Eq. (3), it is computed the influence of the position over the GI:

$$PI = \frac{(PI_{\max} + 1)N_{\max} - \sum_{i=1}^{P_{\max}} PI(i)}{\sum_{i=1}^{N_{\max}} i} \times 100 \quad (3)$$

For each search algorithm and each database is computed a PI factor, thus we have four position factors (two for Wiley library and two for NIST library, for each of them considering the reverse and the Ds, respectively). The values of each individual PI factor can also vary in large ranges. A value of 100 for PI signifies that there is no interferent in the hierarchy resulted from comparing the databases. A lower value for PI shows that in the databases are different compounds with similar mass spectra and the retrieval algorithms cannot discriminate them. The global position index (G_{PI}) express the influence of each individual position indices over the final result.

The GI is computed as the well balanced average of the G_{SI} and the G_{PI} , considering that SI is the primary criterion for the identification of the unknown:

$$GI = \frac{2(G_{SI}) + G_{PI}}{3} \quad (4)$$

The results for the 32 terpenes and phenylpropane derivatives are presented in Table 1. The computed GI expresses the match probability of the unknown with reference databases by using all the above-mentioned indices (similarity and position indices, respectively). As results from Table 1, in some cases there are consistent differences between Wiley and respectively NIST databases. The differences between similarity indices are not significant, except the case of citronellol ($SI_W = 59.7$, $SI_N = 89.5$). More obvious appear the differences of the PI values. This can be explained only by taking in count the number of replicates from each database, Wiley database

having at least two times more replicates than the NIST library. The values of the GI (computed according to Eq. (4)) are lower than the similarity indices. A GI value above 70 express a relative good correlation between the position and the similarity indices, despite the lower values for the former indices. In the case of citronellol, the GI was under 70 meaning that for the given compound we cannot conclude that citronellol is a match even we used a standard. In this cases it is necessary to apply a supplemental comparison algorithm that compares the unknown chromatographic retention indices with a database, as were described in [1].

The above-mentioned algorithm was applied in order to analyze the Romanian *Acorus calamus* essential oil. Table 2 presents the 24 compounds that were identified using the above algorithms and the computed indices. For some compounds, the identification was performed comparing the unknown only with NIST library because the SIs obtained by using the retrieval algorithm for Wiley database were under 40, value that cannot be accepted as a match. Even the SI values were high enough in order to consider a match between the unknown and the databases, the presence of interferences, expressed by the value of the PI, decrease the final results. Thus, the incertitude of the identification of the sesquiterpenic derivatives is greater than that for monoterpenes or their derivatives. A good match between libraries were considered for values greater than 70 and GI between 50 and 70 are considered corresponding to possible matches. The obtained GI can be used for further calculations as were described in [1], by combining the retention indices with the information obtained from the mass spectra comparison with databases, for a given chromatographic peak.

4. Conclusions

The proposed algorithm improved the analytical results of the GC-MSD analyses for the components identification and indicate, which compound from the list offered by using retrieval

Table 2

Identification of the *Acorus calamus* essential oils compounds and computed indices for the presumed matches

Peak no.	Compound	SI		PI		Global indices		
		Wiley	NIST	Wiley	NIST	Wiley	NIST	Total
3	Linalool	82.4	89.9	95.5	92.9	86.8	90.9	88.9
4	Alloocimen	83.0	75.3	100.0	26.4	88.7	59.0	73.9
5	Camphor	96.2	93.4	100.0	100.0	97.5	95.6	96.6
11	Bornyl acetate	78.0	73.2	100.0	53.8	85.3	66.7	76.0
12	Isoterpinolene	76.0	82.5	54.5	100.0	68.8	88.3	78.6
16	Beta-Elemene	67.0	87.5	63.3	62.9	65.8	79.3	72.6
18	Beta-Caryophyllen	71.3	87.1	42.9	61.8	61.8	78.6	70.2
21	Calarene	87.0	90.9	70.4	100.0	81.5	94.0	87.8
22	Alpha-Bergamotene	66.8	88.3	57.5	100.0	63.7	92.2	78.0
26	Alpha-Humulene	85.0	89.8	67.4	70.9	79.1	83.5	81.3
27	Farnesene	63.5	79.7	50.8	25.6	59.3	61.6	60.5
28	Germacrene D	55.8	75.6	69.3	21.7	60.3	57.6	59.0
31	Germacrene B	97.0	80.8	100.0	6.7	98.0	56.1	77.1
33	Isoshyobunone		63.81		25.85		51.2	51.2
35	Aromadendrene	77.67	71.1	83.34	34.97	79.6	59.1	69.4
37	Delta-Cadinene	88.8	77.48	63.37	8.98	80.3	54.6	67.5
40	Alpha-Calcorene	70.5	71.08	83.93	24.53	75.0	55.6	65.3
42	Beta-Calcorene	41.75	54.28	54.17	2.59	45.9	37.1	41.5
44	Cuparene	47	68.43	21.21	51.47	38.4	62.8	50.6
47	Shyobunone		57.79		14.9		43.5	43.5
48	Calamenene	47.5	78.08	34.56	41.67	43.2	65.9	54.6
49	Beta-Asarone	96.1	90.02	75.53	42.38	89.2	74.1	81.7
57	Azulen-6-one drivative	96	77.03	100	8.88	97.3	54.3	75.8
58	Acorenone	80.8		100		87.2		87.2

algorithms and reference mass spectra libraries is the most probable match. The 'interference' of other compounds that are proposed as possible matches by the comparison algorithms were minimized by introducing the PI and by computing the global index for each compound. This global index, for all the available databases and for all the retrieval algorithms reduces the difference be-

tween mass spectra libraries and the search algorithms.

References

- [1] R. Oprean, M. Tamas, R. Sandulescu, L. Roman, J. Pharm. Biomed. Anal. 18 (1998) 651–657.