NPC Natural Product Communications

2007
Vol. 2
No. 4
413 - 418

Reliable Identification of Terpenoids and Related Compounds by using Linear Retention Indices Interactively with Mass Spectrometry Search

Rosaria Costa[a], Maria Rosa De Fina[a], Maria Rita Valentino[a], Paola Dugo[b] and Luigi Mondello[a*]

[a]*Dipartimento Farmaco-chimico, Università di Messina, Viale Annunziata, 98168 Messina, Italy*

[b]*Dipartimento di Scienze degli Alimenti e dell'Ambiente, Università di Messina, salita Sperone 31, 98166 Messina, Italy*

lmondello@pharma.unime.it

Received: January 2nd, 2007; Accepted: January 27th, 2007

This paper is dedicated to Professor Yoshinori Asakawa for his 65th birthday.

An innovative tool in GC-MS peak assignment procedures is described. Besides considering the conventional spectra similarity matching, Linear Retention Indices are used as filters for enhancing reliability in compounds identification, especially in those cases ruled by ambiguity and uncertainty. This is a common issue when analyzing natural compounds (e.g. terpenoids) present in flavors, essential oils and fragrances that give rise to similar fragmentation patterns, and thus, to approximately identical spectra, making harder the unequivocal identification by the databases. Some applications of the method and its capabilities are reported.

Keyword: Terpenoids, GC-MS, database, LRI, Retention Index, essential oils, Spikenard, Hyssop.

The gas chromatography-mass spectrometry (GC-MS) technique has grown undisputedly through the years since its pioneer introduction, dated to 1913, and is today the elective technique employed in the analysis of natural products (e.g. essential oils) and their by-products [1-2]. The knowledge of the chemical composition of the plant material is relevant for several reasons; first of all, it helps in assessing the correct therapeutic/toxic-lethal dose and/or effect of a natural drug. GC-MS has contributed enormously to the discovery of new bioactive components and, when the latter was not new to the scientific community, it defined the best sources of them. It has not to be forgotten that the pharmaceutical activity is mainly inspired by nature, and most of the medicines available in the market are designed from natural product models. There is another facet related to natural products, drugs and their utilization. Whether it is a natural product or synthetic chemical, it is quite normal to find impurities in the composition of naturally derived products. Due to the possible biological activity of these impurities, the use of GC-MS in qualitative and quantitative control is, once again, extremely important. The nature and the amount of each substance considered as extraneous to the original composition might reveal the processing steps undergone by the product, inasmuch as extraction from natural sources and chemical synthesis give rise to different impurities [3]. As can be easily understood, the scientific literature lists thousands of papers based on the use of GC-MS for the characterization of volatiles occurring in essential oils. In fact, the success of the technique, if initially hindered by the high cost of the instrumental apparatus, is mainly due to the ease of use. Anyone who has experienced GC-MS knows that the most used method for identification of unknowns consists of the comparison of the target spectrum with reference spectra collected in a database commonly known as the MS library. Based on the similarity, the search function provides a list of the best matches

found in the library records. The bigger the similarity, the higher is the probability of correct identification. As is universally known, essential oils are complex matrices mostly characterized, as regards their volatile fraction, by the presence of terpenoids and their derivatives. Terpenoids are derived from condensation of isopentenylic units, and, after fragmentation, produce very similar spectra, often causing a loss of discrimination from the library. When an unambiguous identification is not obtained, the Linear Retention Index can be used as an additional criterion of searching [4-6]. The Retention Index system finds its origin in 1958 when Kováts proposed to relate the retention behavior of a substance to the retention properties of a series of reference standards, closely associated to each other by structure [7]. Reference standards used were *n*-alkanes, each of which was assigned a value 100 times the number of carbon atoms present in the molecule. Kováts findings can be summarized in the equation:

$$I = 100z + 100 [\log t'_{R(x)} - \log t'_{R(z)}] / [\log t'_{R(z+1)} - \log t'_{R(z)}]$$

where t'_R is the adjusted retention time, z the number of carbon atoms of the *n*-alkane eluting before the target compound (x), and z +1 the number of carbon atoms of the *n*-alkane eluting after the target compound (x). Based on this equation, Kováts found a direct relationship between the logarithm of the adjusted retention time of each member of a homologous series and the carbon number on a given column and temperature. Substantially, the Retention Index could be defined as a number equal to 100 times the carbon number of a hypothetical *n*-alkane having the same t'_R as the target compound.

Although widely used in several databases collecting RIs, Kováts equation assumes that temperature program conditions are isothermal. It was in 1963 that Van den Dool and Kratz introduced the concept of PTRI (Programmed Temperature Retention Index) based on an approximately linear relationship between elution temperature of *n*-alkanes and their carbon number:

$$I = 100z + 100 [T_{R(x)} - T_{R(z)}] / [T_{R(z+1)} - T_{R(z)}]$$

where T_R is the elution temperature [8]. Since elution temperature and retention time are usually highly correlated in PT separations, the term T_R in the above reported equation can be substituted by t_R (retention time). As for retention time, the PTRI (so called "Linear") does not represent, by itself, an unequivocal system of identification: two different compounds might have the same LRI on the same column and at the same program temperature. However, it is very unlikely that two compounds would have the same retention time, LRI and mass spectrum. This means that, being equal the similarity score produced by a library search process, the LRI can work as a distinguishing item. Taking into account all the matters so far reported, this research group developed an alternative GC-MS library provided with LRI information for each compound. This innovative library, when used in its original GCMS solution (Shimadzu) format, makes possible the application of an LRI filter to search criteria. This filter consists of setting an LRI range within which the value of the unknown compound has to fall: this tool allows the shortening of the list of library matches given by the searching process, thus, getting closer to the best search result. For the creation of such a database, pure chemicals, essential oils and fragrances have been used. In order to collect all the LRIs, each batch of analyses of real samples has been preceded by the acquisition of a mix of *n*-alkanes ranging from C_7 to C_{30}. Due to the fact that LRIs are strongly dependent upon retention mechanism, their reliability on polar phases decreases significantly because of the interfacial adsorption sustained by the *n*-alkanes. For this reason, LRIs often show a big difference (up to one hundred units), when comparing values obtained in different laboratories. In the database here presented, named FFNSC (Flavour & Fragrance Natural & Synthetic Compounds), LRIs calculated on a non-polar stationary phase (5% diphenyl-95% polymethylsiloxane) are reported; this is the most suitable stationary phase ever utilized for the analysis of volatiles in flavors, fragrances and essential oils. Along with each spectrum, Chemical Abstract Service registered information is provided with respect to common name, IUPAC name, formula, CAS number, and molecular weight (Figure 1). Furthermore, all the LRIs experimentally determined have been verified with specific literature data, plus on-line dedicated databases. The LRI interactive system, once developed, has been applied to the analysis of real samples belonging to the flavor & fragrance field, as will be shown later in this manuscript.

In order to demonstrate the capabilities of the interactive LRI system in the applications here reported, the FFNSC library has been used,

alternatively, against another two well-known commercial MS libraries. Figure 2 shows the GC-MS chromatogram of a Spikenard oil (*Nardostachys jatamansi*) sample from Nepal (see table 1 for peak identification).

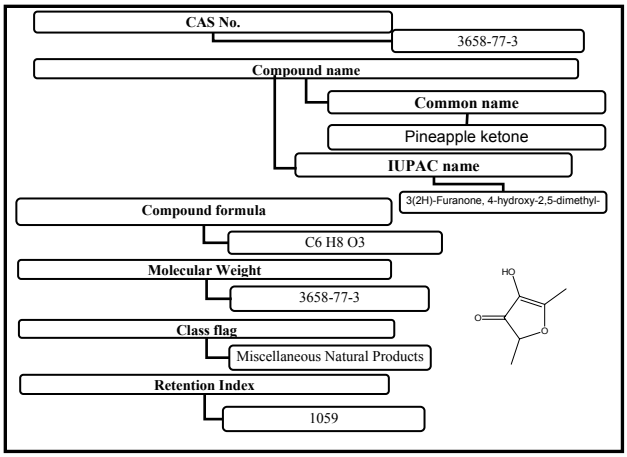

Figure 1: A typical example of the information provided by the FFNSC library for each compound.

Spikenard oil, commonly named *Jatamansi*, is native to the Himalayas and is used for anticonvulsant and anti-stress conditions in Ayurvedic formulations. In spite of its wide use in various fields of medicine and cosmetics, analytical data available on its composition are scarce [9-10]. The use of the FFNSC library in the present paper has led to one of the few fingerprints available in the literature of the volatile fraction of Jatamansi. For some constituents (unknown peaks at 36.9, 41.0, 45.2 and 46.1 min) the identification could not be carried out with a high level of reliability, although mass spectral interpretation highlighted the possibility of these compounds being oxygenated sesquiterpene hydrocarbons with a MW equal to 222, except for the last eluting peak, which shows a MW of 248. As can be seen in table 2, when library searching for peak 27 is performed, data processing reports, at the same

Table 1: Qualitative composition of the volatile fraction of Spikenard oil.

Compounds	LRI on SLB-5MS	Compounds	LRI on SLB-5MS
1 -Thymol methyl ether	1230	21 -10-βH-Cadina-1(6),4-diene	1472
2 -Carvacrol methyl ether	1239	22 -γ-Muurolene	1478
3 -Undecan-2-one	1294	23 -(E)-β-Ionone	1490
4 -Myrtenyl acetate	1326	24 -β-Selinene	1492
5 -δ-Elemene	1335	25 -Valencene	1492
6 -α-Cubebene	1349	26 -α-Bulnesene	1505
7 -Cyclosativene	1367	27 -γ-Cadinene	1512
8 -α-Copaene	1375	28 -δ-Cadinene	1518
9 -β-Patchoulene	1383	29 -Zonarene	1526
10 -β-Cubebene	1392	30 -(±)-Naphthalene, 1,2,3,4,4a,7-hexahydro-	1536
11 -Cyperene	1407	1,6-dimethyl-4-(1-methylethyl)-, (1α,4β,4aβ)-	
12 -β-Maaliene	1415	31 -Spathulenol	1576
13 -Isocaryophyllene	1424	32 -Viridiflorol	1594
14 -Aristola-1(10),8-diene	1431	33 - Carotol	1606
15 -Calarene	1434	34 -Epicubenol	1631
16 -α-Guaiene	1438	35 - τ-Muurolol	1651
17 -6,9-Guaiadiene	1444	36 -Cadin-4-en-10-ol	1659
18 -Seychellene	1445	37 - Eudesma-4(15),7-dien-1βol	1671
19 -α-Humulene	1454	38- Valeranone	1675
20 -9-*epi*-(*E*)-Caryophyllene	1464	39- Valerenal	1706
		40- Cyclocolorenone	1761

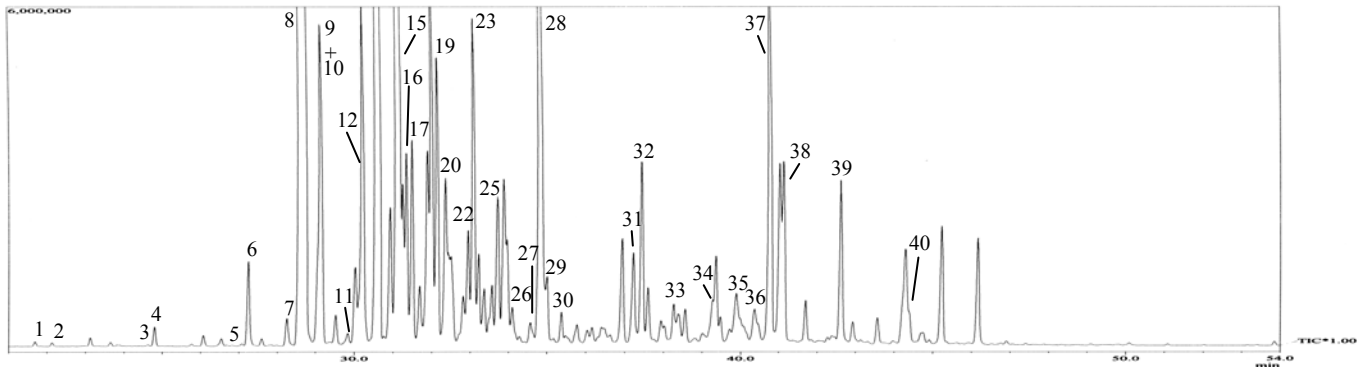

Figure 2: GC-MS chromatogram of Spikenard oil (*Nardostachys jatamansi*).

Similarity Index (S. I.) level (95%), γ-muurolene and γ-cadinene by the FFNSC library; and isoledene at 91% of S.I. by the commercial library. Of course, the last match cannot be accepted as true, either for the lower S.I. or for the LRI value (1419 vs. 1516 of the unknown compound). Correct peak assignment is, therefore, not so easy to be carried out but, after selecting the LRI filter function (see figure 3), the highest S.I. is produced by γ-cadinene, followed by other compounds with S.I. ≤87%. Indeed, the correct choice is γ-cadinene, which presents a difference of 4 units in between the LRI calculated by the software and the one reported in the FFNSC library (1516 vs. 1512). This result is confirmed by the fact that another peak (22) is due to γ-muurolene, which is the compound rejected in the previous search procedure, showing an S.I. equal to 100%. Figure 4 reports the TIC chromatogram of Hyssop oil (*Hyssopus officinalis*), the composition of which is reported in table 4. If peak 46 is considered for identification, the list of library matches reported in Table 3 shows an S.I. of 97% for α-*cis*-bergamotene and 96% for α-*trans*-bergamotene. The commercial library reports a generic α-bergamotene with an S.I. of 95%. In such a case, the choice of the correct result is somehow troubling, since the two values of S.I. are close to each other and both good in the same manner. At this point, the activation of the LRI filter becomes useful and necessary; a Retention Index allowance (see Figure 3) of +/- 5 units leads to only one acceptable match: α-*trans*-bergamotene, having the closest LRI value to the one of the unknown compound (1432 vs. 1435). The activation of the LRI filter produces, as best match, the *trans* isomer of α-bergamotene, rejecting all the other matches previously characterized by high degree of similarity, with the second possible match being *cis*-thujopsene, but with only 83% of S.I. In conclusion, the use of an interactive Linear Retention Index system for GC-MS correct peak assignment has been successfully demonstrated. When analyzing complex matrices, such as essential oils and plant derived products, even performing an optimal GC separation, identification only by mass spectrometry becomes challenging. The database here presented (FFNSC), besides being a high quality reference for GC-MS users, strongly enhances the capability of the GC-MS technique due to the powerful tool of the LRI filter.

Table 2: GC-MS search results obtained for peak 27 of Spikenard oil

Compounds	Library	S.I.*	LRI of unknown	LRI reported in the library
γ-Muurolene	FFNSC	95	1516	1478
γ-Cadinene	FFNSC	95	1516	1512
Isoledene	C.L.**	91	1516	1419

*Similarity Index; **Commercial Library.

Table 3: GC-MS search results obtained for peak 46 of Hyssop oil.

Compounds	Library	S.I.*	LRI of unknown	LRI reported in the library
α-*cis*-Bergamotene	FFNSC	97	1435	1416
α-*trans*-Bergamotene	FFNSC	96	1435	1432
α-Bergamotene	C.L.**	95	1435	-

*Similarity index; **Commercial Library

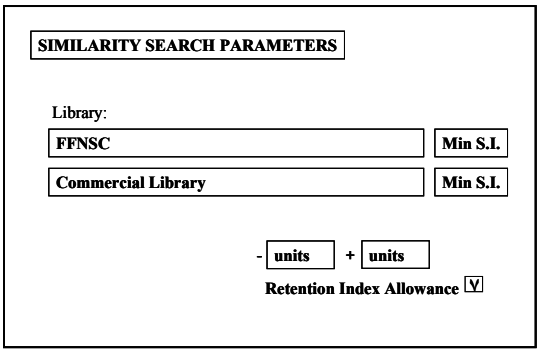

Figure 3: Scheme of the LRI filter tool inside the library software.

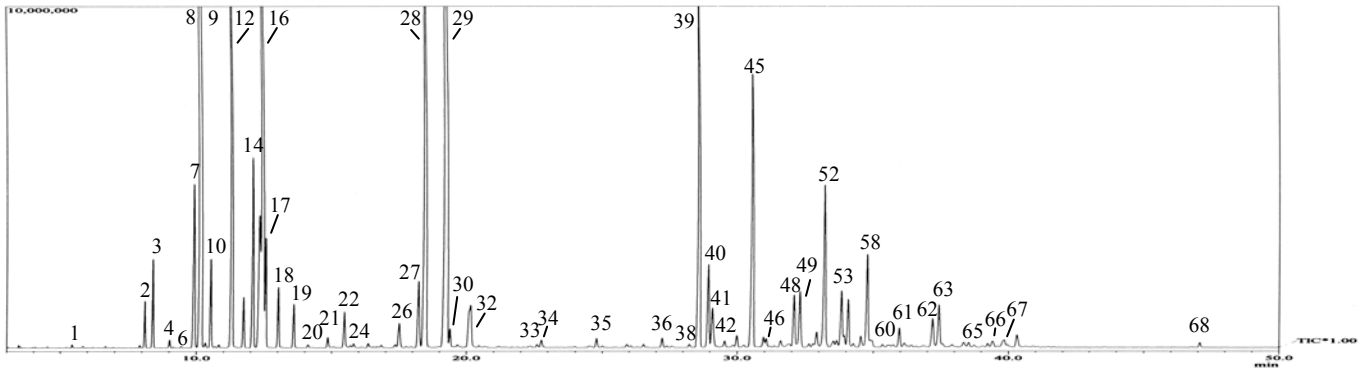

Figure 4: GC-MS profile of Hyssop oil (*Hyssopus officinalis*).

Table 4: Qualitative composition of the volatile fraction of Hyssop oil.

No.	Compounds	LRI on SLB-5MS	no.	Compounds	LRI on SLB-5MS
1	3-Methylbutanol	729	35	Phellandral	1277
2	*cis*-3-Hexenol	853	36	α-Cubebene	1349
3	α-Thujene	927	37	Eugenol	1357
4	α-Pinene	933	38	Cyclosativene	1367
5	Camphene	953	39	α-Copaene	1375
6	Thuja-2,4(10)-diene	953	40	β-Bourbonene	1382
7	Benzaldehyde	964	41	β-Cubebene	1392
8	Sabinene	972	42	Methyleugenol	1403
9	β-Pinene	978	43	α-Gurjunene	1406
10	3-Octanone	986	44	β-Maaliene	1415
11	Myrcene	991	45	Isocaryophyllene	1424
12	3-Octanol	999	46	α-trans-Bergamotene	1432
13	α-Phellandrene	1007	47	Aromadendrene	1438
14	α-Terpinene	1018	48	α-Humulene	1454
15	*p*-Cymene	1025	49	9-*epi*-(*E*)-Caryophyllene	1464
16	Limonene	1030	50	10-βH-Cadina-1(6),4-diene	1472
17	Eucalyptol	1032	51	γ-Muurolene	1478
18	(*Z*)-β-Ocimene	1035	52	Germacrene D	1480
19	(*E*)-β-Ocimene	1046	53	Bicyclogermacrene	1497
20	γ-Terpinene	1058	54	α-Muurolene	1497
21	*cis*-Sabinene hydrate	1069	55	(*E*,*E*)-α-Farnesene	1504
22	Terpinolene	1086	56	β-Bisabolene	1508
23	Linalool	1101	57	γ-Cadinene	1512
24	Nonanal	1107	58	δ-Cadinene	1518
25	α-Thujone	1110	59	Zonarene	1526
26	β-Thujone	1118	60	(±)-Naphthalene, 1,2,3,4,4a,7-hexahydro-1,6-dimethyl-4-(1-methylethyl)-, (1α,4β,4aβ)-	1536
27	*trans*-Pinocarveol	1141	61	α-Elemol	1546
28	*trans*-Pinocamphone	1160	62	Spathulenol	1576
29	*cis*-Pinocamphone	1176	63	Caryophyllene oxide	1587
30	Terpinen-4-ol	1180	64	Viridiflorol	1594
31	Cryptone	1187	65	Humulene epoxide II	1613
32	Myrtenol	1202	66	1-*epi*-Cubenol	1631
33	Carvotanacetone	1249	67	τ-Muurolol	1651
34	2-Hydroxypinocamphone	1252	68	Phytone	1841

Experimental

Chemicals were purchased mainly from Sigma-Aldrich (Italy). Other companies, such as Bedoukian Res. Inc. (USA), IFF GMbH (Germany), Givaudan (Switzerland), Firmenich (France), Acros (Belgium), and Citrus & Allied (USA) contributed to the supply of the material for the library. Essential oils were purchased from Sigma-Aldrich, Essential Oil University (USA) and from local producers of Citrus oils.

Chemicals, essential oils and fragrances were daily injected into a GCMS-QP2010 (Shimadzu) instrument under the following experimental conditions: 30 m x 0.25 mm I.D. x 0.25 μm d_f SLB-5MS column (Supelco); oven temperature program: 50°C to 300°C at 3°C/min, held 5 min; injection mode: split, with a split ratio of 50:1; injector temperature: 300°C; carrier gas: helium, at 32.4 cm/s; injection volume: 1.0 μL (chemicals diluted 1:50, oils 1:10 in the compatible solvents). MS parameters: ion source temperature set at 200°C; interface temperature at 250°C; detector voltage: 0.9 kV with a threshold of 100; scan interval: 0.25 s; scan speed: 1666 amu/s; mass range: 40-400 amu/s. In order to calculate LRIs, a C7:C30 *n*-alkanes mix (Supelco, Italy) was injected for the first analysis prior to the acquisition of the real samples. The analysis parameters for *n*-alkanes were the same as those used for the samples. Data handling was performed by means of the software GCMS solution version 2.5 (Shimadzu).

Acknowledgments – We wish to acknowledge Bedoukian Research Inc. for its valuable support in providing samples. Also, Shimadzu Corporation is thanked for the loan of the instrumentation and for the productive exchange of ideas for software development.

References

[1] Masada Y. (**1976**) *Analysis of essential oils by gas chromatography and mass spectrometry*. John Wiley & Sons, New York.

[2] Zamureenko VA, Kluyen NA, Dmitriev LB, Grandberg II. (**1984**) Gas-liquid chromatography-mass spectrometry in the analysis of essential oils. *Journal of Chromatography, 303*, 109-115.

[3] Leung AY, Foster S. (**2003**) *Encyclopedia of common natural ingredients used in food, drugs, and cosmetics*. John Wiley & Sons, Hoboken.

[4] Mondello L, Dugo P, Basile A, Dugo G. (**1995**) Interactive use of linear retention indices, on polar and apolar columns, with a MS-library for reliable identification of complex mixtures. *Journal of Microcolumn Separations, 7*, 581-591.

[5] Bicchi C, Binello A, D'Amato A, Rubiolo P. (**1999**) Reliability of Van den Dool retention indices in the analysis of essential oils. *Journal of Chromatographic Science, 37*, 288-294.

[6] Zhao C-X, Liang Y-Z, Fang H-Z, Li X-N. (**2005**) Temperature-programmed retention indices for gas chromatography-mass spectroscopy analysis of plant essential oils. *Journal of Chromatography A, 1096*, 76-85.

[7] Kováts E. (**1958**) Gas-chromatographische Charakterisierung organischer Verbindungen. Teil 1: I. Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone *Helvetica Chimica Acta*, *41*, 1915-1932.

[8] Van den Dool H, Kratz PD. (**1963**) A generalization of the retention index system including linear temperature programmed gas-liquid partition chromatography. *Journal of Chromatography, 11*, 463-471.

[9] Chatterje A, Basak B, Saha M, Dutta U, Mukhopadhyay C, Banerji J, Konda Y, Harigaya Y. (**2000**) Structure and stereochemistry of nardostachysin, a new terpenoid ester constituent of the rhizomes of *Nardostachys jatamansi*. *Journal of Natural Products, 63*, 1531-1533.

[10] Bagchi A, Oshima Y, Hikino H. (**1991**) Neolignans and lignans of *Nardostachys jatamansi* Roots. *Planta Medica, 57*, 96-97.