J. Basic. Appl. Sci. Res., 2(2)1615-1618, 2012

ISSN 2090-4304
Journal of Basic and Applied Scientific Research
www.textroad.com

Prediction of Relative Retention Indices of *Phlomis* Essential Oils: Use of QSRR Program

Afra Khosravi[1], Hamzeh Karimi[2], Hadi Nourizadeh[3]*Sajad Rezapasand[3] and Abbas Farmani[3]

[1]Departmentof Immunology, Faculty of Medicine, Ilam University of Medical Sciences, Ilam, Iran
[2]Faculty of Sciences, South Tehran Branch, Islamic Azad University, Tehran, Iran
[3]Department of Chemistry, Faculty of Science, Ilam Branch, Islamic Azad University, Ilam, Iran

ABSTRACT

Phlomis species from the Lamiaceae family are widely distributed in Turkey. Genetic algorithm and partial least square (GA-PLS) technique was used to investigate the correlation between relative retention indices (RRI) for essential oils of *Phlomis russeliana* (Sims.) Bentham and *Phlomis grandiflora* H.S. Thompson var. *grandiflora* which obtained by gas chromatography (GC) and gas chromatography–mass spectrometry (GC/MS). The applied internal (leave-group-out cross validation (LGO-CV)) and external (test set) validation methods were used for the predictive power of model. The results indicate that GA-PLS can be used as an alternative modeling tool for quantitative structure–retention relationship (QSRR) studies.

KEY WORDS: *Phlomis* sp.; Essential oil; gas chromatography–mass spectrometry; QSRR; Genetic algorithm-partial least squares.

INTRODUCTION

Plant extracts and essential oils constitute a natural source of antimicrobial mixtures or pure compounds for centuries. Essential oils and purified components are used as natural antimicrobials in food systems, as well as to prevent the growth of food borne bacteria and molds resulting in extension of the shelf life of processed foods [1]. Fruits, vegetables, grains and food constituents can be contaminated by various microorganisms and their hazardous toxic metabolites. Enterotoxins produced by *Escherichia coli, Staphylococcus pyogenes, Salmonella, Yersinia* and *Clostridium* species are responsible for toxicity in the intestinal tract causing vomiting, diarrhea, etc. Moreover, microorganisms are also associated with food spoilage causing economical loss. Research into more effective anti- microbial food agent's in particular natural antimicrobials such as essential oils received attention in the last decade [2].

Lamiaceae is an important economic plant source of essential oils and the genus *Phlomis* L. has more than 100 species distributed in Euro-Asia and North Africa. It is recently documented that the 52 taxa including 6 varieties, 12 natural hybrids and 34 endemic taxa are growing in Turkey. Both *Phlomis russeliana* and *Phlomis grandiflora var. grandiflora* species used in this study are endemic among the Turkish Flora and are characterized by yellow bilabiate corolla [3].

These essential oils are defined as very powerful aromatic and cyclic plants as well as strong natural antioxidant and were analyzed by means of gas chromatography–mass spectrometry (GC-MS) and assayed for their antimicrobial and antioxidant activities. Sage is one of the plants reported to show antioxidant activity. GC and GC–MS are the main methods for identification of these volatile plant oils. To increase the reliability of the MS identification, comprehensive two-dimensional GC–MS can be used. This technique is based on two consecutive GC separations, typically according to boiling point and polarity. The compounds are identified by comparison of retention indices with those reported in the literature and by comparison of their mass spectra with libraries or with the published mass spectra data [4]. Chromatographic retention for capillary column gas chromatography is the calculated quantity, which represents the interaction between the stationary liquid phase and gas-phase solute molecule. This interaction can be related to the functional group, electronic and geometrical properties of the molecule [4, 5].

Quantitative structure–retention relationship (QSRR) is statistically derived relationships between chromatographic parameters and descriptors related to the molecular structure of the analytes. A number of reports deals with QSRR retention calculation of essential oils compounds have been published in the literature [6-8].

There is a trend to develop QSRR from a variety of methods. In particular, genetic algorithm (GA) is frequently used as search algorithms for variable selection in chemometrics and QSRR. GA is a stochastic method to solve the optimization problems defined by fitness criteria, applying the evolution hypothesis of Darwin and different genetic functions, i.e. crossover and mutation [9, 10]. Partial least square (PLS) is the

***Corresponding Author:** Hadi Nourizadeh, Department of Chemistry, Faculty of Science, Ilam Branch, Islamic Azad University, Ilam, Iran Tel.: + 989183452507 E-mail address: nourizadehadi@yahoo.com

most commonly used multivariate calibration method [11, 12]. In the present study, GA-PLS was employed to generate QSRR model that correlate the structure of *Phlomis* essential oils; with observed RRI.

MATERIALS AND METHODS

Table 1
The data set and corresponding observed RRI values for training set

No	Name	RRI	No	Name	RRI
	Training Set			Training Set	
1	p-Cymene	1280	43	2-Tetradecanone	1878
2	Tetradecane	1400	44	epi-Cubebol	1900
3	1-Octen-3-ol	1452	45	α -Agarofuran	1916
4	α -Cubebene	1466	46	Cubebol	1957
5	α -Copaene	1497	47	(E)- β -Ionone	1958
6	α -Campholene	1499	48	γ -Calacorene	1984
7	α -Bourbonene	1528	49	Caryophyllene oxide	2008
8	β -Bourbonene	1535	50	2-Pentadecanone	2036
9	β -Cubebene	1549	51	(E)-Nerolidol	2050
10	trans- α -Bergometene	1568	52	Ledol	2057
11	α -Cedrene	1577	53	Humulene epoxide-II	2071
12	β -Ylangene	1589	54	Elemol	2096
13	trans- β -Bergamotene	1594	55	Viridiflorol	2104
14	β -Caryophyllene	1612	56	10-epi- γ -Eudesmol	2127
15	β -Cedrene	1613	57	Hexahydrofarnesyl acetone	2131
16	β -Cyclocitral	1638	58	Rosifoliol	2144
17	γ -Elemene	1650	59	(Z)-3-Hexen-1-yl benzoate	2148
18	(E,E)-2,5-Epoxy-6,8- megastigmadiene	1654	60	5-epi-7-epi- α -Eudesmol	2157
19	c-Gurjunene	1659	61	β -Bisabolol	2170
20	(Z)- β -Farnesene	1668	62	3,4-Dimethyl-5-pentylidene- 2(5H)-furanone	2179
21	trans-Verbenol	1683	63	γ -Eudesmol	2185
22	α -Humulene	1687	64	Docosane	2200
23	γ -Muurolene	1704	65	Eremoligenol	2204
24	α -Terpineol	1706	66	T-Muurolol	2209
25	2-Dodecane	1718	67	ar-Turmerol	2214
26	Dodecanal	1722	68	δ -Cadinol	2219
27	Germacrene-D	1726	69	α -Eudesmol	2250
28	α -Muurolene	1740	70	α -Cadinol	2255
29	β -Selinene	1742	71	β -Eudesmol	2257
30	α -Cadinene	1743	72	Tricosane	2300
31	α -Selinene	1744	73	γ -Undecalactone	2300
32	Bicyclogermacrene	1755	74	Caryophylladienol	2316
33	d-Cadinene	1773	75	Farnesyl acetone	2384
34	c-Cadinene	1776	76	Caryophyllenol	2389
35	ar-Curcumene	1786	77	Caryophyllenol	2392
36	Nerol	1808	78	Pentacosane	2500
37	2-Tridecanone	1815	79	Hexacosane	2600
38	(E,E)-2,4-Decadienal	1827	80	1-Octadecanol	2607
39	β -Damascone	1830	81	Heptacosane	2700
40	(E)- β -Damascenone	1838	82	Octacosane	2800
41	Germacrene-B	1854	83	Nonacosane	2900
42	Geraniol	1857	84	Hexadecanoic acid	2931

Data set

Relative retention time of the 108 compounds in *Phlomis* essential oils were taken from literature [13] is shown in Table 1 and Table 2. *Phlomis* essential oils were analysed by GC using a Hew-lett-Packard 6890 (Sem Ltd., Istanbul, Turkey) system and an HP Innowax FSC column (60 m × 0.25 mm, with 0.25 μ m film thickness) was used with nitrogen at 1 ml/ min. Initial oven temperature was 60 °C for 10 min, and

increased at 4 °C/min to 220 °C, then remained constant at 220 °C for 10 min and increased at 1 °C/min to 240 °C. GC/MS analysis was performed with a Hewlett-Packard GCD (Sem Ltd., Istanbul, Turkey), system and Innowax FSC column (60 m ×0.25 mm, 0.25 μ m film thickness) was used with helium. The data set was randomly divided into two groups including training set (calibration and prediction sets) and external (test) sets, which consists of 84 and 24 molecules, respectively.

Computer hardware and software

All calculations were run on a HP Laptop computer with AMD Turion64X2 processor with windows XP operating system. The optimizations of molecular structures were done by the HyperChem 7.0 (AM1 method) and descriptors were calculated by Dragon Version 3.0 software's. Cross validation, GA-PLS and other calculation were performed in the MATLAB (Version 7, Mathworks, Inc.) environment.

Table 2

The data set and corresponding observed RRI values for test set

No	Name	RRI
	Test Set	
1	Decanal	1506
2	Linalool	1553
3	Isocaryophyllene	1589
4	(Z,E)-2,5-Epoxy-6,8-megastigmadiene	1627
5	Muurola-4,11-diene	1674
6	Geranyl formate	1715
7	Verbenone	1725
8	Sesquicineole	1747
9	Octadecane	1800
10	trans-Carveol	1845
11	(E)-Geranyl acetone	1868
12	Tetradecanal	1933
13	Isocaryophyllene oxide	2001
14	Salvial-4(14)-en-1-one	2037
15	Heneicosane	2100
16	Spathulenol	2144
17	Nor-Copaonone	2179
18	T-Cadinol	2187
19	Hinesol	2210
20	Selin-11-en-4a-ol	2273
21	Caryophylladienol	2324
22	Tetracosane	2400
23	Phytol	2622
24	Pentadecanoic acid	2822

Cross validation technique

Cross validation is a popular technique used to explore the reliability of statistical model. Based on this technique, a number of modified data sets are created by deleting in each case one or a small group (leave-some-out) of objects. For each data set, an input–output model is developed, based on the utilized modeling technique. Each model is evaluated, by measuring its accuracy in predicting the responses of the remaining data (the ones or group data that have not been utilized in the development of the model) [14]. In particular, the leave group out (LGO) procedure was utilized in this study.

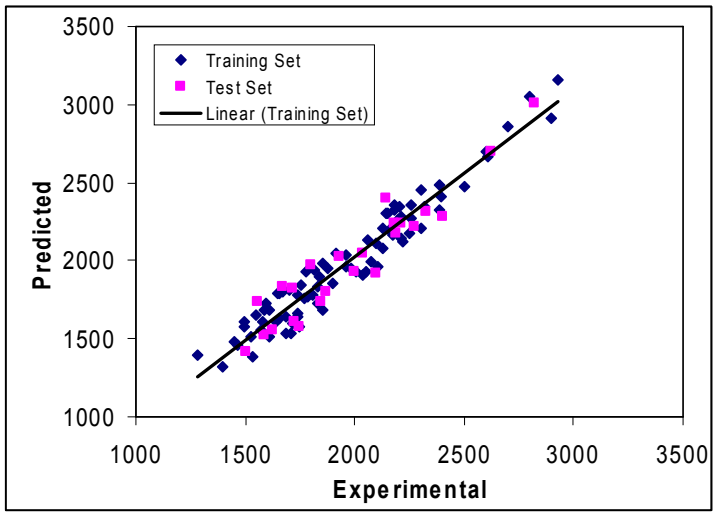
RESULTS AND DISCUSSION

Results of the GA-PLS model

To reduce the original pool of descriptors to an appropriate size, the objective descriptor reduction was performed using various criteria. Reducing the pool of descriptors eliminates those descriptors which contribute either no information or whose information content is redundant with other descriptors present in the pool. The remained descriptors were employed to generate the model with the GA-PLS program. The best model is selected on the basis of the highest square correlation coefficient (R^2) and relative error (RE) of prediction and simplicity of the model. These parameters are probably the most popular measure of how well a model fits the data. The best GA-PLS model contains 24 selected descriptors in 11 latent variables space. The RE for training and test sets was (5.02, 6.39), respectively. For this in general, the number of components (latent variables) is less than number of independent variables in PLS analysis. The PLS model uses higher number of descriptors that allow the model to extract better structural information from descriptors to result in a lower prediction error. Inspection of the results reveals a lowers RE value parameter

for the training and test sets GA-PLS. The GA-PLS linear model has good statistical quality with low prediction error. Plots of predicted RRI versus experimental RRI values by GA-PLS for training and test set are shown Fig. 1. Obviously, there is a close agreement between the experimental and predicted RRI and the data represent a very low scattering around a straight line with respective slope and intercept close to one and zero.

Fig.1. Plot of predicted RRI obtained by GA-PLS against the experimental values



Conclusion

In this research, an accurate QSRR model for estimating the retention time of *Phlomis* essential oils compounds was developed by employing the GA-PLS technique. This model has good predictive capacity and excellent statistical parameters. It is easy to notice that there was a good prospect for the GA-PLS application in the QSRR modeling. It can also be used successfully to estimate the RRI for new compounds or for other compounds whose experimental values are unknown.

REFERENCES

[1] Burt, S. (2004). International Journal of Food Microbiology, 94, 223–253.

[2] Draughon, F. A. (2004). Food Technology, 58, 20–28

[3] Demirci, B., Baser, K. H. C., & Dadandi, M. Y. (2006). Journal of Essential Oil Research, 18, 328–331

[4] Baricevic D, Bartol T (2000) Harwood Academic Publishers, Amsterdam, The Netherlands 143-184.

[5] Ong VS, Hites RS (1991) Anal. Chem 63: 2829-2837.

[6]Chen, J.; Yang, T.; Cramer, S. M. *J. Chromatogr. A*. 2008, **1177**, 207.

[7] Noorizadeh, H.; Farmany, A. *Chromatographia*. 2010, **72**, 563.

[8] H. Noorizadeh, A. Farmany, M. Noorizadeh, Quim. Nova, 34 (2011) 242-249.

[9] Hemmateenejad, B.; Javadnia, K.; Elyasi, M. *Anal. Chim. Act*, 2007, **592**, 72.

[10] J. Aires-de-Sousa, M.C. Hemmer, J. Casteiger, Prediction of H-1 NMR chemical shifts using neural networks, Anal. Chem. 74 (2002) 80–90.

[11] Riahi, S.; Pourbasheer, E.; Ganjali, M. R.; Norouzi, P. *J. Haz. Mat.* 2009, **166**, 853.
[12] H. Noorizadeh, S. Sobhan Ardakani, T. Ahmadi, S. S. Mortazavic, M. Noorizadehd, Drug Test Anal, 2011, in press.

[13] F. Demirci, K. Guven, B. Demirci, M.Y. Dadandi, K.H.C. Baser, Food Control 19 (2008) 1159–1164

[14] H. Noorizadeh, A. Farmany, Drug Test Anal, 2011, in press.