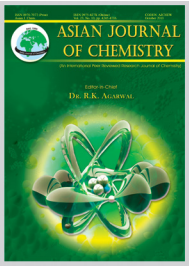
Asian Journal of Chemistry; Vol. 24, No. 9 (2012), 3814-3820

ASIAN JOURNAL OF CHEMISTRY

www.asianjournalofchemistry.co.in



Chemical Profile and Antibacterial Activity in Hydrodistilled Oil from Aerial Parts of *Prangos ferulacea* (L.) Lindl. and Prediction of Gas Chromatographic Retention Indices by Using Genetic Algorithm Multiple Linear Regressions

MAJID MOHAMMADHOSSEINI

Department of Chemistry, Shahrood Branch, Islamic Azad University, Shahrood, Iran

Corresponding author: Fax: +98 273 3394537; Tel: +98 273 3394530; E-mail: majidmohammadhosseini@yahoo.com

(*Received*: 20 May 2011; *Accepted*: 8 April 2012) AJC-11246

The essential oil obtained by hydrodistillation from aerial parts of *Prangos ferulacea* growing wild plant in the marginal brackish regions of Semnan province from Iran was analyzed by means of GC and GC/MS instrumentations. The main components were β-phellandrene (20.39 %), α-terpinolene (15.26 %), α-pinene (11.59 %), δ-3-carene (11.06 %), α-phellandrene (9.09 %), trans-β-ocimene (9.67 %). The oil shows antibacterial activity against *Escherichia coli* and *Staphylococcus saprophytacus* with minimum inhibitory concentration values of 3.27 and 8.19 μg/mL, respectively. Furthermore, quantitative structure-property relationships were developed using a genetic algorithm variable-selection approach. The correlation between the experimental and calculated Kovats indices (R^2_{train} = 0.95, REP=1.5; R^2_{test} = 0.9952, REP=0.9) as well as reliable cross validation results (Q^2_{LOO} = 0.98, Q^2_{LGO} = 0.96) demonstrate the high applicability of the proposed model to predict RIs of a diverse set of natural compounds.

Key Words: *Prangos ferulacea*, Essential oil, Antibacterial activity, Quantitative structure-property relationships, Genetic algorithm.

INTRODUCTION

Prangos genus of Umbelliferae family involves about 30 species[1]. Various species of this plant have been widely distributed in the brackish regions of Iran provinces. *Prangos* is frequently used as a powerful tropical medical plant and as an attractive alternative for chemical drugs with noxious side effects. The common pharmaceutical and clinical applications of various species of *Prangos ferulacea* are as emollient, carminative[2], tonic antiflatulent, anthelmintic, antifungal and antibacterial agents[3,4].

Hence, characterization and monitoring of each constituent component of the plant composition have a prime importance and a growing interest. Furthermore, some nutritional properties[5], antioxidant[6], antibacterial activities[7] and abortifacient effect[8] of this genus have been proved.

In recent years, Eshbakova *et al.*[9] reported the extraction of furocoumarins from *P. ferulacea*, while Nazemiyeh *et al.*[10] inspected isolation of coumarins from *P. uloptera* in their phytochemical investigations.

On the other hand, hyphenation of gas chromatography and mass spectrometry has created an exclusive technique for recognition and determination of volatile organic compounds in a wide variety of samples including different parts of plants[11-16].

A literature survey reveals that steam distilled from aerial parts and seeds[17] and hydrodistilled fruits of *P. ferulacea* have been the subject of some reports[18]. Moreover, Baser *et al.*[19] have reported the chemical composition of *P. ferulacea* (L.) Lindl. in Turkey. However, in the present work, the percentages and types of identified components have a considerable difference in comparison with similar reports. This may be most probably due to the district and climate of the sampling area.

One of the main criteria for recognition of the natural constituents found in the essential oils is their retention indices. This parameter is independent of chromatographic conditions and is directly proportional to 100 n where n implies the number of carbon atoms in the homologous paraffins co-injected with the target compound.

Hence, mathematical modeling of this parameter is of prime importance and helps chemists to find a model that can be used to obtain a deep insight about the mechanism of retention behaviour and to predict the RIs of new identified and/or even unsynthesized compounds[20].

Accordingly, quantitative structure-retention relationships (QSRR) models are constructed on the basis of the logical and interpretable correlations between the experimental values

of the RIs and descriptors reflecting the molecular structure of the compounds. In these strategies the critical points are: 1. Extraction the proper independent molecular descriptors; 2. Dividing the whole data to training and test sets, properly; 3. Description and justification the role of each variable in the proposed model; 4. Statistical evaluation of the respective parameters; 5. External and internal cross validation; 6. Plotting the prediction and residuals; 7. Chance correlation and standardization as complementary steps.

In these studies, a regression model in the form ($y = Xb + e$) may be used to describe a set of predictor variables (X) with a predicted variable (y) by means of a regression vector (b).

In the present work, the author contributed the knowledge of the volatile oil from crushed dry aerial parts of *P. ferulacea* in Semnan province (Iran) by GC and GC-MS and its antibacterial activity. Also, a novel and reliable QSRR model is constructed for retention indices of the identified constituents in the oil.

EXPERIMENTAL

Aerial parts of the *P. ferulacea* were collected during the flowering stage in Shahmirzad Mountains located in the east of Semnan Province, Iran in April 2010 at an altitude of 1550 msl. A Voucher specimen was deposited at the Herbarium of the Research Institute of Forests and Rangelands (TARI), Tehran, Iran.

Separation of the essential oil: The air-dried aerial parts (150 g) of homogeneous coarse powder of *P. feulacea* were subjected to hydrodistillation (HD) by using a Clevenger-type apparatus for 3 h. The obtained volatile oil was dried over anhydrous sodium sulfate and kept under liquid nitrogen in a brown sealed vial at-10 ºC. The physical properties of the oil are given in Table-1.

TABLE-1
PHYSIOCHEMICAL SPECIFICATIONS OF THE OIL

w/w (%)	Physiochemical property		
	d^{20} (g/mL)	$[\alpha]^{20}$	$[n]^{20}$
0.6	0.951	-8.20	1.512

GC and GC/MS analyses: Gas chromatography analyses were performed on a Shimadzu 15 A gas chromatograph equipped with split/splitless injector (250 ºC) and flame ionization detector at 280 ºC. Nitrogen was used as carrier gas at a flow rate of 1 mL/min. A SE 30 column (50 m × 0.2 mm) with film thickness of 0.3 µm was employed for separation of the constituents. The column temperature was kept at 90 ºC for 3 min and then increased to 220 ºC with a 5 ºC rate and kept constant for 5 min at this temperature. Relative percentages were calculated considering peak area and using a Shimadzu C-R4A chromatopac.

GC-MS analyses of the plant were carried out on a Hewlett-Packard 6890/5973, GC-MS instrument under the recommendations of the manufacturers. The characteristics are given in Table-2.

Antibacterial activity: Determination of the antibacterial activity of the oil was performed by the agar well diffusion method[21] modified according to the present experimental conditions as follows. Mueller-Hinton was used as culture medium with the concentration of 34 g/L. The culture medium was inoculated either with *E. coli* (PTCC1330), *S. saprophytacus* (PTCC1113) and *B. sereus*. Except for *E. coli* and *S. saprophytacus*, *B. sereus* was pathogenic strains isolated from patients with different infections by passage. Each micro-organism had been previously suspended in physiological saline solution 0.9 w/v %. Six millimeter diameters wells were punched into the agar and impregnated with different dilutions of the essential oil (50 µL) and the solvent blanks were tested. Different dilutions of the oil in ethanol ranging from 0.2 up to 800 µg/mL were examined. The Petri dishes were put in an incubator with the incubation temperature of 30 ºC. Incubation times after culturing were 24, 24 and 48 h for *E. coli* (PTCC1330), *B. sereus* and *S. saprophytacus* (PTCC1113), respectively. The antibacterial activity was evaluated by measuring of the inhibition zone diameter.

TABLE-2
PERFORMANCE INSTRUMENTATION OF GC-MS

Conditions of GC	
Column type/film thickness (µm)	Fscc[a]: HP-5MS (30m L × 0.25 mm I.D.) / 0.25
Carrier gas/flow-rate (mL/min)	He/2
Temperature programming	60 ºC for 3 min. ramping to 220 ºC with a 5 ºC/min rate and remaining constant for 5 min in isothermal mode
Split ratio (mL/min)	1:50
Injector temperature (ºC)	220
Detector temperature (ºC)	270
Injection volume (µL)	0.2
Conditions of interface	
Interface type	Membrane separator
Interface temperature (ºC)	290
Conditions of MS 5972 selective detector	
Ion source	EI[b]
Mode	Full scan
Mass transfer line temperature (ºC)	290
Scan velocity (s/scan)	0.2

[a]Fscc: Fused silica capillary column; [b]EI: Electron impact (ionization)

Theoretical part

Computer hardware and software: A Pentium IV personal computer (CPU at 3.06 GHz) with the Windows XP operating system was used. The geometry optimization was performed with HyperChem. program (Version 8.0 Hypercube, Inc) by semi-empirical AM1 method[22].

No molecular symmetry constraint was applied; rather full optimization of all bond lengths and angles was carried out. All calculations were carried out at the restricted Hartree-Fock level with no configuration interaction. The molecular structures were optimized using the Polak-Ribiere algorithm until the root mean square gradient was 0.01 Kcal/mol. Geometry optimization was run multiple times with different starting points for each molecule and the lowest energy conformation was considered for the calculation of electronic properties. For the calculation of the molecular descriptors, the Dragon 2.1 software was used. The SPSS software package (version 14, SPSS, Inc.) was employed for the multiple linear

regression analysis, other calculations were performed in the MATLAB (version 7.1, Math Works, Inc.) environment.

Calculation of the molecular descriptors: After drawing all the chemical structures, the resulting geometries were transferred into the Dragon program package developed by Milano chemometrics and QSPR group to calculate descriptors in constitutional, topological, geometrical, charge, GETAWAY (geometry, topology and atoms-weighted assembly), WHIM (weighted holistic invariant molecular descriptors), 3DMoRSE (3D molecular representation of structure based on electrondiffraction), molecular walk count, BCUT, 2D-autocorrelation, aromaticity index, randic molecular profile, radial distribution function, functional group and atom-centered fragment classes. Molecular descriptors (1481) belonging to 18 different theoretical descriptors were calculated for each molecule. The calculated descriptors were first analyzed for the existence of constant or near constant variables. The detected ones were then removed. Correlation among descriptors with the activity of the molecules was examined and collinear descriptors (*i.e.* $r > 0.9$) were detected. Descriptors that contain a high percentage (> 90 %) of identical values were discarded to decrease the redundancy existing in the descriptor data matrix. Among the collinear descriptors, the one presenting the highest correlation with the activity was retained and others were removed from the data matrix. The dataset was splitted into two sets based on RI range; training set (77 %) with activity ranges from 936 to 1831 and test set (33 %) with activity ranges from 948 to 1566.

RESULTS AND DISCUSSION

Identification and quantitative monitoring of the components: By injection of the volatile oil sample to the HP-5 MS column a chromatogram was achieved, which was composed of 27 separated compounds. However, only 21 components were recognized constituting 96.23 % of the total aerial parts oil. The retention index of each component was then calculated and compared both in HP-5MS and in SE-30 columns. The results have been compared in Table-3. As can be seen, the results obtained by GC-MS and GC are in satisfactory agreement with each other. Linear retention indices of the entire ingredients of the oil were calculated after co-injection of the plant volatile oil with a solution containing homologous alkanes ranging between C_9 to C_{20}. Each composition component was thereafter identified *via* comparison and matching of its relative retention time and breakdown pattern with those reported in literature as well as tabulated information in the library of GC-MS instrument of known, certified and authentic samples[23,24].

Library searching was afforded by GC-MS library software and TARM library of essential oil constituents. Based upon the results given in Table-3, our attempt to identify the concerned active compounds resulted in isolation of several monoterpene hydrocarbons (MH), oxygenated monoterpene (OM), sesquiterpene hydrocarbons (SH), oxygenated sesquiterpene (OS) and non-terpenoid hydrocarbon (NH) from aerial plant parts of *P. feulacea* by hydrodistillation method. Accordingly, twelve monoterpene hydrocarbons (92.14 %), two sesquiterpene hydrocarbons (0.65 %), four oxygenated monoterpenes (1.83 %), one oxygen containing sesquiterpenes (1.1 %), one non-terpenoid hydrocarbon (0.36 %) and one diterpene hydrocarbon (0.15 %) contribute to the chemical profile. Hence, a ranking order of the groups of constituents as MH > OM > OS > SH > NH > DH is observed in the hydrodistillation method.

Relative percentage of each isolated ingredient was screened from the total chromatogram. Evaluation of the ingredients percentages indicates that major components are α-pinene, α-phellandrene, δ-3-carene, β-phellandrene, *trans*-β-ocimene and γ-terpinolene. Meanwhile, the MS fragmentations of the main constituents have been presented in Table-4.

Chemical profiles in other species of prangos genus: According to the literature, in the oils obtained from *P. uloptera* DC[25] (aerial parts, seeds, umbels and total); *P. latiloba*[11] (leaves and stems); *P. platychlaena* as well as *P. uechtritzii*[26], monoterpenes predominated over sesquiterpenes. However, in the essential oil separated from *P. heyniae*[27]; *P. uloptera*[25] (fruits); *P. latiloba*[11] (roots); *P. acaulis* (DC.) Bornm.[28] and *P. pabularia* Lindl.[29], sesquiterpenes and oxygenated sesquiterpenes were the most abundant components.

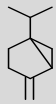
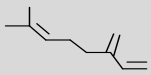
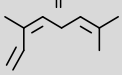
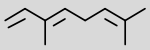
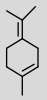
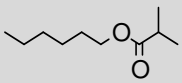
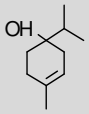
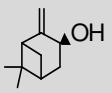
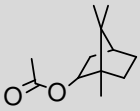
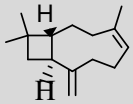
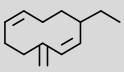
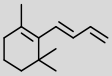
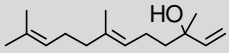
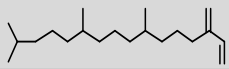
Antibacterial activity of the oil: The essential oil of the plant showed antibacterial activity against *E. coli* (PTCC1330) and *S. saprophytacus* (PTCC1113). However, the oil was found to be inactive against *B. sereus*. Antibacterial activity of the essential oil against *E. coli* (PTCC1330) and *S. saprophytacus* (PTCC1113) has been demonstrated in Table-5. Accordingly, MIC values for *E. coli* and *S. saprophytacus* are 3.27 and 8.19 μg/mL, respectively.

Genetic algorithm: Genetic algorithm approach, which is mainly based upon Darwin evolution hypothesis has been firstly proposed by John Holland in the early 1970s. However, the applicability domain of genetic algorithm was extensively developed since the1990s, when computers became much faster to process advanced calculations. Genetic algorithm is a stochastic method to solve the optimization problems and comprises various genetic functions including chromosomes, crossover, mutation and backward stepwise in distinct evaluations. The distinctive characteristic of a genetic algorithm is that it covers a wide variety of the possible solutions, each of which explores different regions in parameter space[30,31].

Compared with other feature selection approaches, genetic algorithm is robust, global and generally more straightforward to apply to situations in which little or no initial knowledge about the process to be controlled are available. It is capable to search the entire solution space with a greater probability to finding the global optimum[32].

In genetic algorithm, each individual of the population, defined by a chromosome of binary values as the coding technique, represented a subset of descriptors. The number of the genes at each chromosome was equal to the number of the descriptors. The population of the first generation was selected, randomly. A gene was given the value of one, if its corresponding descriptor was included in the subset; otherwise, it was given the value of zero. The genetic algorithm performs its optimization by variation and selection *via* the evaluation of the fitness function. Fitness function is used to evaluate alternative descriptor subsets that were finally ordered according to the predictive performance of related model by cross validation. The fitness function was proposed by Depczynski *et al.*[33]

TABLE-3
CHARACTERIZATION AND PERCENTAGE COMPOSITION OF CONSTITUENT COMPONENTS OF THE AERIAL PARTS OF *P. ferulacea*[a]

No	Compound	Structure	R_t[b] (min)	R_I[c]	Recognition methods	Percentage (%)
1	α-Pinene		5.76	936	GC-MS, R_I[b],R_I[c]	11.59
2	Camphene		6.05	948	GC-MS, R_I[b],R_I[c]	0.25
3	Sabinene		6.68	974	GC-MS, R_I[b],R_I[c]	4.38
4	Myrcene		7.11	991	GC-MS, R_I[b],R_I[c]	4.48
5	α-Phellandrene		7.53	1008	GC-MS, R_I[b],R_I[c]	9.09
6	δ-3-Carene		7.69	1014	GC-MS, R_I[b],R_I[c]	11.06
7	α-Terpinene		7.78	1018	GC-MS, R_I[b],R_I[c]	0.21
8	β-Phellandrene		8.23	1035	GC-MS, R_I[b],R_I[c]	20.39
9	Cis-Ocimene		8.34	1040	GC-MS, R_I[b],R_I[c]	2.36
10	Trans-β-Ocimene		8.67	1052	GC-MS, R_I[b],R_I[c]	9.67
11	γ-Terpinene		8.91	1062	GC-MS, R_I[b],R_I[c]	3.40
12	α-Terpinolene		9.69	1092	GC-MS, R_I[b],R_I[c]	15.26
13	Hexyl isobutyrate		11.12	1151	GC-MS, R_I[b],R_I[c]	0.36
14	Terpinene-4-ol		11.67	1170	GC-MS, R_I[b],R_I[c]	0.16
15	Cis-pinocarveol		11.90	1179	GC-MS, R_I[b],R_I[c]	0.38
16	Bornyl acetate		14.43	1290	GC-MS, R_I[b],R_I[c]	0.57
17	β-Carryophyllene		17.44	1423	GC-MS, R_I[b],R_I[c]	0.23
18	Germacrene-D		18.64	1480	GC-MS, R_I[b],R_I[c]	0.42
19	β-ionene		18.73	1484	GC-MS, R_I[b],R_I[c]	0.72
20	E-Nerolidol		20.34	1566	GC-MS, R_I[b],R_I[c]	1.10
21	Neophytadiene		25.20	1831	GC-MS, R_I[b],R_I[c]	0.15
		Total percentage				96.23

[a]Compounds have been sorted according to retention indices on HP-5 MS capillary column; [b]R_t: Retention time; [c]R_I: Retention Index in HP-5MS

TABLE-4
MS FRAGMENTS OF THE MAIN COMPONENTS OF THE OIL

Compound	m/z	Fragments
α-Pinene	136 (M^+, 0.5 %)	93.10 (100 %), 92.10 (37.93 %), 91.10 (26.73 %), 79.10 (26.7 %), 77.05 (14.83 %)
α-Phellandrene	136 (M^+, 0.1 %)	93.10 (100 %), 91.10 (35.27 %), 92.10 (29.6 %), 136.1 (24.1 %), 77 (19 %)
δ-3-Carene	136 (M^+, 1 %)	93 (100 %), 91 (28 %), 92 (27 %), 121 (24.1 %), 79 (34 %)
β-Phellandrene	136 (M^+, 0.5 %)	93 (100 %), 91 (25 %), 68 (27 %), 136 (26 %), 79 (24 %)
Trans-β-Ocimene	136 (M^+, 0.01 %)	93 (100 %), 92.1 (37 %), 79 (36 %), 41 (18 %), 105 (17 %)
γ-Terpinene	136 (M^+, 2 %)	93 (100 %), 136.1 (38 %), 121.10 (32 %), 79 (22 %)

TABLE-5
PRESENTATION OF ANTIBACTERIAL ACTIVITY AND MIC OF THE OIL

Essential oil concentration (μg/mL)		0.52	1.31	3.27	8.19	20.48	51.00	128.00	320.00	800.00
Diameter of inhibition (mm)	Tested bacteria									
	E. coli	0.00	0.00	0.00	6.23	8.34	9.02	10.33	11.00	14.00
	S. saprophytacus	0.00	0.00	0.00	0.00	6.65	8.43	9.10	10.27	11.00
MIC	*E. coli*			3.27						
	S. saprophytacus				8.19					

TABLE-6
DEFINITION OF THE SELECTED DESCRIPTORS IN GA-MLR MODEL

No.	Descriptor (type)	Definition
1	SMTIV (topological)	Schultz MTI by valence vertex degrees
2	GMTI(topological)	Gutman Molecular Topological Index
3	ATS5m (2Dautocorrelations)	Broto-Moreau autocorrelation of a topological structure - lag 4 / weighted by atomic masses
4	Mor 25 P (3D-MoRSE)	3D-MoRSE - signal 25 / weighted by atomic polarizabilities

based on the root-mean-square error (RMSE). The periodically evolutionary stages engaged in genetic algorithms are (i) exhibition of a chromosome by a binary bit string (ii) establishing an initial population of chromosomes in a random way (iii) evaluation of a value for the fitness function of each chromosome (iv) reproduction the chromosomes of the next generation by selection crossover and mutation operations according to the values of fitness function. To evaluate the most pertinent descriptors, this work was followed by simulation of population evolution with randomly selected first generation population[34-36].

For a typical run, number of chromosomes, the probability of cross over, the probability of mutation and the number of evaluation are 110, 0.05, 0.05 and 200, respectively. For each set of data, 100 routine runs were performed.

In genetic algorithm-multiple linear regression approach, the selected 4 descriptors were Mor 25 P belonging to 3D-MoRSE group, ATS5 m belonging to 2D autocorrelations class, SMTIV and GMTI belonging to topological, respectively. The negative sign of the coefficients of the descriptors denote their destructive influence on the retention process while the positive sign imply the constructive impact of each implemented feature in the model.

The best proposed model is as follows:

$$RI = 3314.18 (\pm 702.12) + 7102.60 (\pm 790.1) \times Mor\ 25\ P - 602.6 (\pm 7.3) \times ATS5m - 81.9 (\pm 6.2) \times SMTIV + 114.06 (\pm 12.1) \times GMTI$$

$N_{train} = 15$; $R^2_{train} = 0.9952$; $R^2_{adj} = 0.99$; MSE = 1261.50; F-statistical = 161.1; P value < 10^{-4}

In this model as a representative of a wide range of reliable genetic algorithm-multiple linear regressions, two topological (SMTIV and GMTI), one 2D autocorrelations (ATS5m) and one 3D-MoRSE (Mor 25 P) descriptors are observed. The 2D autocorrelations descriptors possess the negative sign while all topological and 3D-MoRSE features increase the retention indices. A brief description concerning these features is represented in Table-6. Basically, topological descriptors include valence and non-valence molecular connectivity indices calculated from the hydrogen-suppressed formula of the molecule, encoding information about the size, composition and the degree of branching of a molecule. The significant influence of this molecular descriptor type is reasonable since the size of the solutes penalizes the quality of physicochemical equilibria through the chromatographic capillary columns. However, 2D autocorrelations are spatial autocorrelations calculated on an H-depleted molecular graph weighted by atom physico-chemical properties. 2D autocorrelations are molecular descriptors which describe how a considered property is distributed along a topological molecular structure. Molecular property includes a set of heterogeneous molecular descriptors describing physicochemical and biological properties as well as some molecular characteristics obtained by literature models[37].

Furthermore, GETAWAY descriptors are based on a leverage matrix (molecular influence matrix) and easily calculated from the spatial coordinates of the atoms in a molecule. They are relatively new descriptors and were also developed by Consonni *et al.*[38].

To ensure of independent characteristic of the used variables, the correlation matrix was calculated for both approaches and the results are listed in Table-7. Based upon these results, the ability of the resulting QSRR regression models contributing to accurate forecasting of the Kovatz index is not related to co-linearity between the variables.

Predictive ability of the models: Fig. 1 and Fig. 2 show the plots of values predicted by genetic algorithm-multiple

linear regression against experimental values of the retention indices for training and prediction sets, respectively. The good agreement between the experimental and forecasted indices confirms the considerable potential of the proposed model for a wide range of natural compounds. Furthermore, the residuals (experimental RI- predicted RI) *versus* experimental RI values, obtained by genetic algorithm-multiple linear regression modeling are shown in Fig. 3. The distribution of the residuals on both sides of the zero line indicates there is no systematic error in both models.

TABLE-7
CORRELATION MATRIX FOR THE SELECTED DESCRIPTORS IN THE PROPOSED MODEL

	Mor 25 P	ATS5m	SMTIV	GMTI
Mor 25 P	1			
ATS5m	-0.105	1		
SMTIV	0.474	0.381	1	
GMTI	-0.027	0.341	0.288	1

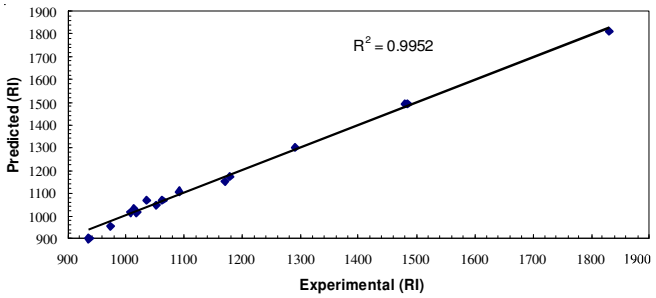


Fig. 1. Predicted RI values by the SW-MLR modeling vs. the experimental RI values for training set

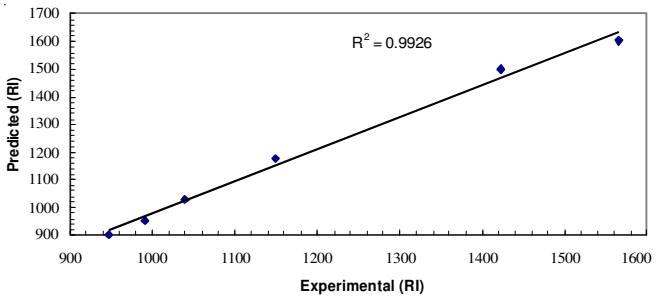


Fig. 2. Predicted RI values by GA-MLR for test sets vs. the experimental RI values

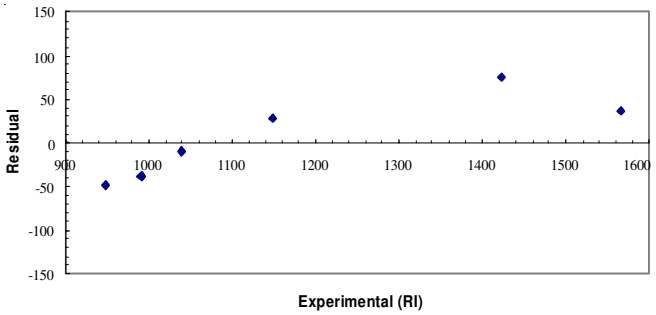


Fig. 3. Plot of the residuals against the experimental values of the retention indices in GA-MLR

Statistical parameters: To validate the suggested methods, statistical characteristics of the employed strategies were assessed and the corresponding results have been summarized in Table-7. In the present work, we have used 6 common statistical parameters to appraise the estimation ability of the constructed models. These parameters are square of determination coefficient (R^2), relative error of prediction (REP), root mean square error of prediction (RMSEP), standard error of prediction (SEP), relative standard error of prediction (RSEP) and mean absolute error (MAE).

The first parameter is r^2 which indicates the quality of fitness or divergence of the points toward a straight line can be calculated as:

$$r^2 = 1 - \frac{\sum_{i=1}^{n}(y_{pred} - y_{mean})^2}{\sum_{i=1}^{n}(y_{exp} - y_{mean})^2} \quad (6)$$

The second and third statistical characters are REP and RMSEP. REP (eqn.7) represents the predictive ability of each individuals and its low value are desired while RMSEP (eqn. 8) connects with the mean discrepancy between predicted and experimental values. It could be inspected as the average prediction error having the same units with the original response values.

$$REP(\%) = \frac{100}{\bar{y}}\left[\frac{1}{n}\sum_{i=1}^{n}(y_{pred} - y_{exp})^2\right]^{0.5} \quad (7)$$

$$RMSEP = \left[\frac{1}{n}\sum_{i=1}^{n}(y_{pred} - y_{exp})^2\right]^{0.5} \quad (8)$$

Third and fourth parameters are SEP and RSEP according to the eqns. 9 and 10:

$$SEP = \left[\frac{\sum_{i=1}^{n}(y_{pred} - y_{exp})^2}{n-1}\right]^{0.5} \quad (9)$$

$$RSEP(\%) = 100\left[\frac{\sum_{i=1}^{n}(y_{pred} - y_{exp})}{\sum_{i=1}^{n}(y_{exp})^2}\right]^{0.5} \quad (10)$$

Standard error of prediction and RSEP are general ways used to evaluate the predictive applicability of a regression model. The final figure of merit is MAE which is a statistical term dealing with the average distance of predicted from their exact values and its formula is represented in eq. 11.

$$MAE(\%) = \frac{100}{n}\left[\sum_{i=1}^{n}\left|(y_{pred} - y_{exp})\right|\right]^{0.5} \quad (11)$$

In these equations y_{pred}, y_{exp}, $\bar{y}$ and n imply the predicted retention index, the experimental value of the RIs, the mean or average of experimental RI in the test set and number of samples in the training or test sets, respectively.

Validation of the models: To inspect the cross validation of the model, leave-one-out (LOO) and leave-group-out (LGO) cross-validation processes were conducted. For leave-one-out cross-validation, a data point is removed from the set and the model is recalculated. The predicted property for that point is then compared with its actual value. This is repeated until each

data point is omitted once. For leave-group-out, 20 % of the data points are removed from the dataset and the model is refitted, the predicted values for those points are then compared with its experimental values. Again, this is repeated until each data point has been omitted once. The cross-validation correlation coefficient (Q^2) is 0.96 for leave-group-out and 0.98 for leave-one-out. This denotes that the obtained regression model has a good internal and external-predictive power.

Conclusion

In this report, chemical composition of the hydrodistilled volatile oil separated from the aerial parts of *P. ferulacea* has been identified and quantified. The oil was characterized manily by presence of highly amounts of monoterpene hydrocarbons. Moreover, genetic algorithm variable subset selection method was used for the selection of the most relevant descriptors from the pool of remaining descriptors. It was found that, two topological (SMTIV and GMTI), one 2D autocorrelations (ATS5m) and one 3D-MoRSE (Mor 25 P) descriptors contribute to the retention behaviour of the oil components from theoretical point of view.

ACKNOWLEDGEMENTS

The author is grateful for the financial supports provided by Islamic Azad Universitiy of Shahrood. Furthermore, special thanks are also due to Dr. Mozaffarian (Research Institute of Forests and Rangelands, Teheran, Iran) for botanical identification and authentication of the plant sample.

REFERENCES

1. W.C. Evans, Trease and Evans Pharmacognocy, Bailliere Tindall, London, edn. 13th (1989).
2. A. Zargari, Medical Plants, Tehran University publications, Tehran, vol. **2** (1988).
3. K.H.C. Baser, B. Demirci, F. Demirci, E. Bedri, P. Weyerstahl, H. Marschall, H. Duman, Z. Aytac and M.T. Hamann, *Planta Med.*, **66** 674 (2000).
4. A. Ulubelen, G. Topcu, N. Tan, S. Olcal and S. Tamer, *J. Ethnopharmacol.*, **45** 193 (1995).
5. M.M. Ozcan, N. Dursun and D. Arslan, *Int. J. Food Sci. Nut.*, **58**, 162 (2007).
6. K. Kogure, I. Yamauchi, A. Tokumura, K. Kondou, N. Tanaka, Y. Takaishi and K. Fukuzawa, *Phytomedicine*, **11**, 645 (2004).
7. H. Durmaz, E. Sagun, Z. Tarakci and F. Ozgokce, *Afr. J. Biotechnol.*, **l5**, 1795 (2006).
8. T. Kazerooni, K. Mousavizadeh, A. Abdollahee, M. Sarkarian and A. Sattar, *Contraception*, **73**, 554 (2006).
9. K.A. Eshbakova, A.I. Saidkhodzhaev, K.H.C. Baser, H. Duman, A.D. Vdovin and N.D. Abdullaev, *Chem. Nat. Comp.*, **42**, 102 (2006).
10. H. Nazemiyeh, A. Delazar, R. Hajiboland, M.M. Rahman, S. Gibbons, L. Nahar, S.M. Razavi and S.D. Sarker, *Phytochem. Lett.*, **1**, 159 (2008).
11. S.H. Akhlaghi and P. Hashemi, *Chem. Nat. Compd.*, **41**, 542 (2005).
12. R. Vila, B. Milo, F. Tomi, J. Casanova, E.A. Ferro and S. Canigueral, *J. Ethnopharmacol.*, **76**, 105 (2001).
13. N. Tabanca, N. Kirimer and K.H.C. Baser, *Turk. J. Chem.*, **25**, 201 (2001).
14. G.E. Meral, S. Konyalioglu and B. Ozturk, *Fitoterapia*, **73**, 716 (2002).
15. C.J. Xu, Y.Z. Liang and F.T. Chau, *Talanta*, **68**, 108 (2005).
16. M. Mohammadhosseini, A. Pazoki and H. Akhlaghi, *Chem. Nat. Comp.*, **44**, 127 (2008).
17. F. Sefidkon, M.S. Khajavi and B. Malackpour, *J. Essent. Oil Res.*, **10**, 81 (1998).
18. A. Menghini, M.R. Cagiotti, L. Montanarella, F.C. Fischer and R. Bos, *Essenze Derivati Agrumari*, **57**, 34 (1987).
19. K.H.C. Baser, N. Ermin, N. Adiguezel and Z. Aytac, *J. Essent. Oil Res.*, **8**, 297 (1996).
20. R. Kaliszan, Structure and Retention in Chromatography-A Chemometric Approach. Harwood Academic Publishers, Amsterdam, pp. 43-45 (1997).
21. C. Perez, M. Pauli and P. Bazerque, *Acta Biol. Med. Exp.*, **15**, 113 (1990).
22. M.J.S. Dewar, E.G. Zeoblish, E.F. Healy and J.J. Stewart, *J. Am. Chem. Soc.*, **107**, 3902 (1985).
23. R.P. Adams, Identification of Essential Oils Components by Gas Chromatography/Quadrapole Mass Spectrometry, Allured Publishing Corporation, Illinois, USA, (2001).
24. R.P. Adams, Identification of Essential Oil Components by Gas Chromatography/Mass Spectrometry, Allured Pubishing Co., Carol Stream, IL., USA, (2007).
25. F. Sefidkon and M.N. Navaii, *J. Essent. Oil Res.*, **13**, 84 (2001).
26. A. Uzel, T. Dirmenci, A. Celik and T. Arabaci, *Chem. Nat. Comp.*, **42**, 169 (2006).
27. K.H.C. Baser, T. Ozek, B. Demirci and H. Duman, *Flav. Fragr. J.*, **15**, 47 (2000).
28. A. Rustaiyan, H. Mazloomifar, S. Masoudi and Z. Aghjani, *J. Essent. Oil Res.*, **18**, 682 (2006).
29. G. Ozek, T. Ozek, G. Iscan, K.H.C. Baser, E. Hamzaoglu and A. Duran, *South Afr. J. Bot.*, **73**, 563 (2007).
30. W. Cai, B. Xia, X. Shao, Q. Guo, B. Maigret and Z. Pan, *J. Mol. Struct. (Theochem).*, **535**, 115 (2001).
31. H.M. Cartwright, Applications of Artificial Intelligence in Chemistry, Oxford University Press, Oxford, (1993).
32. M. Jalali-Heravi and A. Kyani, *Eur J. Med. Chem.*, **42**, 649 (2007).
33. U. Depczynski, V.J. Frost and K. Molt, *Anal. Chim. Acta*, **420**, 217 (2000).
34. J. Hunger and G. Huttner, *J. Comput. Chem.*, **20**, 455 (1999).
35. S. Ahmad and M.M. Gromiha, *J. Comput. Chem.*, **24**, 1313 (2003).
36. C.L. Waller and M.P. Bradley, *J. Chem. Inf. Comput. Sci.*, **39**, 345 (1999).
37. R. Todeschini and V. Consonni, Handbook of Molecular Descriptors, Weinheim, Germany, Wiley-VCH, (2000).
38. V. Consonni, R. Todeschini and M. Pavan, *J. Chem. Inf. Comput. Sci*., **42**, 682 (2002).