JCHR (2018) 8(2), 117-126

Journal of Chemical Health Risks

www.jchr.org

ORIGINAL ARTICLE

A Robust Quantitative Structure-Property Relationship-Based Model for Estimation of Refractivity Indices of 101 Common Paraffin Derivatives Based on their Molecular Structures

Mehdi Nekoei

Department of Chemistry, College of Basic Sciences, Shahrood Branch, Islamic Azad University, Shahrood, Iran

(Received: 20 February 2016 Accepted: 20 April 2016)

KEYWORDS

QSPR;
Refractivity index;
MLR;
Molecular descriptors

ABSTRACT: A robust linear quantitative structure-property relationship (QSPR) model has been constructed to model and predict the refractivity indices of 101 organic compounds as common halo-derivatives of normal paraffin by application of the structural descriptors combined with multiple linear regressions (MLR) method. In the main part of this study, theoretical molecular descriptors were adopted from the original pool through the stepwise feature selection method. A simple model with low standard errors and promising correlation coefficients was obtained. MLR method could model the relationship between refractivity and structural descriptors, perfectly. The accuracy of the proposed MLR model was illustrated using cross-validation, validation through an external test set, and *Y*-randomization techniques. The linear techniques such as MLR combined with a successful variable selection procedure are capable of generating an efficient QSPR model for predicting the refractivity indices of different compounds. The constructed model, with high statistical significance (R^2_{train} = 0.926; F_{train} = 240.675; R^2_{test} = 0.947; F_{test} = 52.978; *REP (%)* = 1.219; Q^2_{LOO} = 0.914 and Q^2_{LGO} = 0.914), could be adequately used for the prediction and description of the affecting parameters on refractivity behavior of similar or even unknown compounds.

INTRODUCTION

The physical term of refractive index (RI or n) is a measure for how much the speed of light is reduced inside a distinct medium. The speed of light in a medium can be expressed as: V=C/n where V is the speed of light in the medium, C denotes the speed of light in air or vacuum (approximately 3×10^8 meters per second) while n refers to the refractive index [1-3].

Refractive indices (RIs) of liquids can be determined through several methods from experimental point of view. For instance, deviation refractometers use the angle of deviation [4-11].

Displacing the engaged beam, relating to a laser radiation which passes through a distinct medium (a liquid) accounting for the term refractive index. However, the interferometric refractometers are greatly influenced by changes in the optical path [12-21].

Interferometric methods are predominantly used for accurate measurements of the RIs of gases [18, 20] and so

---

* Corresponding author: m_nekoei1356@yahoo.com (M. Nekoei).

far relatively few experiments have been reported for liquids and solids. In most of these cases, the changes of n with temperature or pressure were determined [13, 15, 16 and 22-27] and only very few absolute measurements have been reported [17, 19 and 21].

Quantitative structure-property relationships (QSPR) represent statistical models, which quantify the relation between the structure of the molecule and the properties of the compound, allowing the prediction of the properties of the novel compounds [28-38]. QSPR models have been reported for different properties of organic compounds [29, 35-45].

The application of these techniques urgently requires the variable selection step for building well-fitted models. In this study, we employed the elimination selection-stepwise regression (ES-SWR) variable selection method. The result of this study was the development of a new linear and reliable QSPR-based model only containing 4 independent variables. The proposed methodology was validated using several strategies involving cross-validation, *Y*-randomization, and external validation using division of the entire data set into training and test sets.

The aim of this study was to search an efficient method to build an accurate quantitative relationship between the molecular structure and the refractive index of some organic compounds by using the elimination selection-stepwise multiple linear regression (ES-SW-MLR) approach.

MATERIALS AND METHODS

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 (Version 7.0 Hypercube, Inc). To calculate the molecular descriptors, the Dragon 2.1 software was used. The SPSS software (version 14, Chicago, IL, USA)) was employed for the subsequent MLR analysis. The other advanced calculations were performed in the MATLAB (Version 7.0, Math works, Inc) environment.

Data set

Experimental data of the refractivity indices of some organic compounds were taken from a well-known Handbook written by Dean [46]. The data set was randomly splitted into training and prediction (test) sets. In fact, the prediction set involving 20 compounds was randomly selected from the original data set including 101 compounds and the remaining compounds placed in the training set. The training set with 81 compounds, was used to establish and adjust the parameters of the model, while the test set was used to evaluate its predictive capability.

Determination of molecular descriptors

Molecular descriptors are defined as numerical characteristics associated with chemical structures. The molecular descriptor is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number applied to correlate physico-chemical properties of the organic compounds [47].

The Dragon software was used to calculate the descriptors in this research and a total of 1481 molecular descriptors, from 18 different types of theoretical descriptor, were calculated for each molecule. Since the values of many descriptors are related to the bonds length and bonds angles etc., the chemical structure of every molecule must be optimized before calculating its molecular descriptors. For this reason, chemical structure of the 101 studied molecules were first drawn with the Hyperchem software and immediately saved with the HIN extension. To optimize the geometry of these molecules, the AM1 geometrical optimization was applied. After optimizing the chemical structures of all compounds, the molecular descriptors were calculated using Dragon.

Elimination selection-stepwise multiple linear regression (ES-SW-MLR)

As mentioned earlier in the introduction section, the ES-SWR algorithm [47] was used to select the most impressive descriptors. ES-SWR is a popular stepwise technique that combines forward selection (FS-SWR) and backward elimination (BE-SWR). It is essentially a forward selection approach, but at each step it considers the possibility of deleting a variable as in the backward elimination approach, provided that the number of model variables is greater than two.

Model validation

The accuracy of the proposed MLR model was examined using the following evaluation techniques: leave-one-out (*LOO*) and leave-group-out (*LGO*) cross-validation procedures, validation through an external test set, and *Y*-randomization. Accordingly, proper statistical terms related to our constructed model confirm that it is a reliable approach to predict the numerical values of RIs of a broad spectrum of organic compounds.

Cross-validation test

Cross-validation is a popular technique used to explore the reliability of statistical models. Based on this technique, a number of modified data sets are created by deleting one or a small group (leave some-out) of objects in each case. For each data set, an input-output model is developed, based on the utilized modeling technique. The model is evaluated by measuring its accuracy in predicting the responses of the remaining data involving the ones that have not been utilized in the development of the model [48].

Y-randomization test

This technique ensures the robustness of a QSRR model and non-dependency of the implemented variables by chance [49, 50]. In this approach, the dependent variable vector (n) is randomly shuffled and a new QSPR model is developed using the original independent variables matrix. The new QSPR models being built should have low R^2 and Q^2 values over several repetitions and/or iterations. If the opposite happens, an acceptable QSPR model cannot be obtained for the specific modeling method and data processing.

RESULTS

The multiple linear regression method (MLR) is one of the most used modeling methods in diverse series of QSPR studies. As a matter of fact, MLR analysis has been carried out to derive the best QSPR model. A small set of molecular descriptors proposed by our team were used to establish a proper QSPR model. The MLR technique was performed on the molecules of the training set shown in Table 1.

Table 1. The data set and the corresponding observed and predicted values of refractivity indices (RIs) using the ES-SW-MLR approach for the training and test sets.

No.	Compound Training set	RI (Exp.)	ES-SW-MLR	R.E.(%) [a]
1	2-Bromobutane	1.436	1.44	0.267
2	1-Bromo-4-chlorobutane	1.488	1.467	-1.396
3	1-Bromo-2-chloroethane	1.492	1.437	-3.671
4	Bromo-chloro-fluoro-methane	1.414	1.408	-0.447
5	1-Bromo-3-chloro-2-methylpropane	1.481	1.465	-1.068
6	1-Bromo-3-chloropropane	1.485	1.461	-1.592
7	2-Bromo-2-chloro-1,1,1-trifluroethane	1.369	1.392	1.638
8	1-Bromodecane	1.456	1.465	0.636
9	2-Bromo-1,1-diethoxyethane	1.439	1.433	-0.383
10	2-Bromo-1,1-dimethoxyethane	1.445	1.449	0.311
11	1-Bromododecane	1.458	1.46	0.117
12	Bromoethane	1.424	1.437	0.912
13	2-Bromoheptane	1.447	1.452	0.358
14	1-Bromohexane	1.448	1.457	0.634

Table 1. Continued

15	Bromo-methane	1.423	1.447	1.691
16	1-Bromo-3-methyl butane	1.441	1.449	0.548
17	1-Bromo-2-methyl propane	1.436	1.442	0.389
18	2-Bromo-2-methyl propane	1.425	1.429	0.262
19	1-Bromooctane	1.452	1.464	0.839
20	1-Bromopentane	1.444	1.45	0.375
21	3-Bromopentane	1.445	1.444	-0.002
22	1-Bromopropane	1.437	1.437	-0.005
23	2-Bromopropane	1.429	1.433	0.312
24	1-Bromotetradecane	1.46	1.447	-0.888
25	1-Bromotridecane	1.459	1.454	-0.371
26	1-Bromoundecane	1.456	1.463	0.444
27	1-Chloro-3-iodopropane	1.546	1.514	-2.07
28	Chloromethane	1.371	1.399	2.056
29	Dibromochloromethane	1.547	1.529	-1.138
30	1,2-Dibromo-2-chloro-1,1,2-trifluoroethane	1.428	1.416	-0.838
31	1,10-Dibromodecane	1.491	1.515	1.617
32	1,2-Dibromo-1,1-difluoroethane	1.446	1.421	-1.733
33	1,2-Dibromo-3,3-dimethylbutane	1.505	1.497	-0.52
34	1,1-Dibromoethane	1.538	1.498	-2.622
35	1,2-Dibromoethane	1.539	1.498	-2.62
36	1,2-Dibromohexafluoropropane	1.361	1.363	0.154
37	Dibromomethane	1.542	1.515	-1.739
38	1,8-Dibromooctane	1.498	1.516	1.174
39	1,4-Dibromopentane	1.509	1.501	-0.508
40	1,5-Dibromopentane	1.509	1.506	-0.198
41	1,3-Dibromopropane	1.523	1.499	-1.568
42	1,2-Dibromotetrafluoroethane	1.367	1.394	1.95
43	1,2-Dibromo-1,1,2-trifluoroethane	1.419	1.394	-1.74
44	1,4-Dichlorobutane	1.457	1.445	-0.769
45	1,1-Dicholoroethane	1.416	1.429	0.876
46	1,2-Dichloroethane	1.445	1.434	-0.734
47	Dichlorofluoromethane	1.372	1.337	-2.602
48	1,6-Dichlorohexane	1.457	1.457	0.011
49	1,5-Dichloropentane	1.455	1.451	-0.299
50	1,2-Dichloropropane	1.439	1.432	-0.469
51	1,3-Dichloropropane	1.449	1.438	-0.704
52	1,1-Dichlorotetrafluoroethane	1.309	1.345	2.758
53	1,1-Difluorotetrachloroethane	1.413	1.418	0.376
54	1,2-Difluorotetrachloroethane	1.413	1.413	0
55	1,4-Diiodobutane	1.621	1.645	1.443
56	Diiodomethane	1.743	1.725	-0.986
57	Fluorotrichloromethane	1.382	1.406	1.723
58	Hexachloropropane	1.548	1.569	1.345
59	1-Iodobutane	1.5	1.509	0.592
60	2-Iodobutane	1.499	1.52	1.369
61	2-Iodododecane	1.484	1.496	0.781
62	Iodoethane	1.513	1.534	1.365
63	1-Iodoheptane	1.49	1.503	0.869
64	1-Iodohexadecane	1.481	1.458	-1.523
65	2-Iodo-2-methylpropane	1.492	1.51	1.206
66	1-Iodononane	1.487	1.503	1.063
67	1-Iodooctane	1.489	1.506	1.126

Table 1. Continued

68	1-Iodopentane	1.495	1.501	0.386
69	2-Iodopropane	1.499	1.521	1.441
70	Pentachloroethane	1.503	1.528	1.671
71	1,1,2,2-Tetrabromoethane	1.636	1.619	-1.028
72	Tetrachloro-1,2-difluoroethane	1.413	1.413	0
73	1,1,2,2-Tetrachloroethane	1.491	1.504	0.887
74	Tetrachloroethylene	1.506	1.499	-0.455
75	Tetradecafluorohexane	1.252	1.227	-1.997
76	Tribromomethane	1.601	1.574	-1.627
77	1,1,1-Trichloroethane	1.438	1.459	1.492
78	1,1,2-Trichloroethane	1.471	1.469	-0.139
79	Trichlorofluoromethane	1.384	1.406	1.583
80	1,2,3-Trichloropropane	1.485	1.472	-0.911
81	1,1,2-Trichlorotrifluoroethane	1.356	1.377	1.587
	Test set			
1	Bromochloromethane	1.48	1.47	-0.689
2	Bromodichloromethane	1.497	1.501	0.285
3	1-Bromoheptane	1.451	1.46	0.663
4	2-Bromo-2-methyl butane	1.442	1.436	-0.442
5	2-Bromopentane	1.44	1.444	0.28
6	Bromotrichloro methane	1.506	1.529	1.521
7	Chloro-2,2,2-trifluoroethane	1.309	1.336	2.07
8	Dibromodifluoromethane	1.402	1.421	1.387
9	1,6-Dibromohexane	1.507	1.511	0.281
10	1,2-Dibromopropane	1.52	1.497	-1.505
11	1,1-Dichloro-3,3-dimethylbutane	1.439	1.44	0.064
12	Dichloromethane	1.425	1.439	0.992
13	1,2-Dichloro-1,1,2,2-tetrafluoroethane	1.309	1.349	3.022
14	1,3-Diiodopropane	1.642	1.627	-0.948
15	1-Iododecane	1.485	1.503	1.182
16	1-Iodo-2-methylpropane	1.496	1.511	0.97
17	1-Iodopropane	1.506	1.506	0.002
18	1,1,1,2-Tetrachloroethane	1.481	1.499	1.231
19	1,2,3-Tribromopropane	1.584	1.559	-1.584
20	1,1,1-Trichlorotrifluoroethane	1.37	1.391	1.539

[a] Relative error percentage

After regression analysis, a few suitable models were obtained among which the best model was selected and presented in eq. (1). A limited number of the proposed molecular descriptors (MAXDP, H0p, X1sol and Au) were used to establish the QSPR model. Additional validation was performed on an external data set consisting of the RIs of 20 organic compounds. Finally, the simple MLR analysis provided a useful equation that can be used to predict the parameter n of the mentioned compounds. The best equation obtained for the RIs of these compounds appears as follows:

$$\mathbf{n} = 1.3275 - 0.0223\times\mathbf{MAXDP} + 0.0186\times\mathbf{H0p} + 0.0388\times\mathbf{X1sol} - 0.0052\times\mathbf{Au} \qquad \text{(eq.1)}$$

N = 81; $R^2_{train} = 0.926$; $F_{train} = 240.675$; $R^2_{test} = 0.947$; $F_{test} = 52.978$; REP(%) = 1.219; $Q^2_{LOO} = 0.914$; $Q^2_{LGO} = 0.914$

In this equation, N is the number of compounds; R^2 is the squared correlation coefficient, Q^2_{LOO} and Q^2_{LGO} are the squared cross-validation coefficients for *LOO* and *LGO* respectively, REP is the relative error for prediction set and *F* is the Fisher statistic term. From the above equation, it can be concluded that according to the ES-SWR algorithm the most significant descriptors are maximal electropological positive variation (MAXDP), H autocorrelation of lag0/weighted by atomic polarizabilities (H0p), salvation connectivity index chi-1 (X1sol) and A total size index/unweighted (Au).

DISCUSSION

Table 2 presents the correlation matrix, where it is clear that the four selected descriptors are almost completely uncorrelated. This is an essential condition by which we can claim that the selected variables behave, independently. A brief explanation of the descriptors that were selected in the final linear model is as follows.

Table 2. Correlation matrix for the four selected descriptors in the linear model

	MAXDP	**H0p**	**X1sol**	**Au**
MAXDP	1			
H0p	0.64	1		
X1sol	0.51	0.34	1	
Au	0.68	0.46	0.25	1

The first descriptor is MAXDP, which is one of the geometrical descriptors. Its effect on the n was negative, indicating that the n is inversely related to this descriptor.

The second descriptor of this model was H autocorrelation of lag0/weighted by atomic polarizabilities (H0p). It is one of the GETAWAY descriptors. Its effect on the n was positive. Another descriptor of this model was X1sol that had a positive effect on the n. It is among the topological descriptors.

The final descriptor is Au which was among the WHIM descriptors. This descriptor exerted a negative effect on n. Thereby, by increase in Au descriptor; a perceptible decrease in the dependent variable (RI) is noted.

Equation (1) was used to predict the RIs for the prediction (test) set. The data set and the corresponding experimental and predicted values of n for all of the molecules studied in this work are summarized in Table 1. Figure 1 shows the plot of the predicted values by the ES-SW-MLR against the experimental values of the RIs for both the training and the prediction sets. The residuals (experimental RI− predicted RI) versus experimental values of n, obtained by the ES-SW-MLR modeling are shown in Figure 2. Accordingly, the normal and rationale distribution of the residuals on both sides of the zero line indicates there is no systematic error in the built and developed ES-SW-MLR model. In summary, the results obtained altogether explicitly emphasize once more that the linear MLR technique combined with a successful variable selection procedures are adequate to generate efficient QSPR-based models for predicting of the term n for a wide variety of organic compounds.

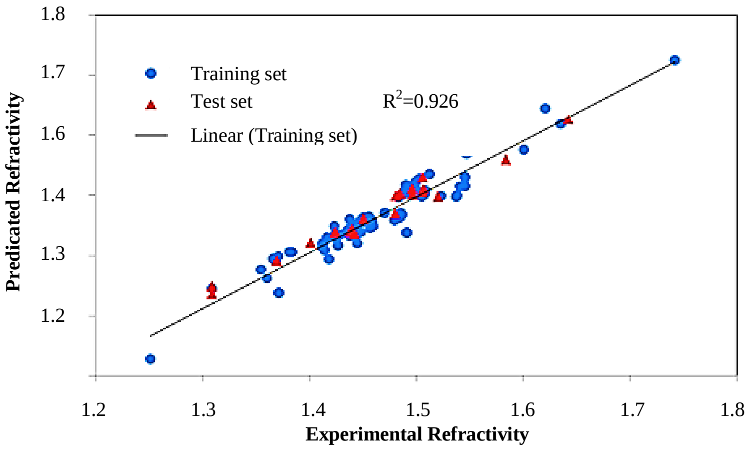

Figure 1. The predicted RI values by the MLR modeling vs. the experimental values of refractivity indices (RIs)

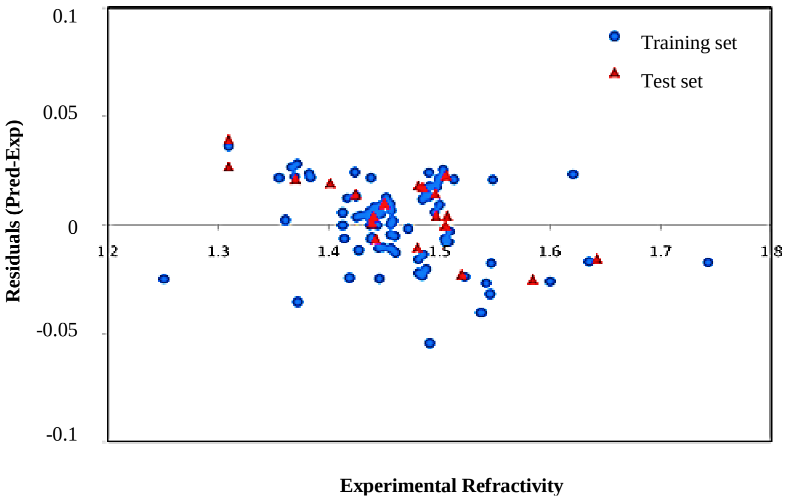

Figure 2. Plot of the residuals against the experimental values of the refractivity indices (RIs)

For a more exhaustive testing of the predictive power of the model, validation of the model was also carried out using the *LOO* and the *LGO* cross-validation techniques on the training set of compounds. For *LOO* cross-validation, a data point is removed from the set, and the model is recalculated. The predicted n for that point is then compared with its actual value. This is repeated until each data point has been omitted once. On the other hand, for *LGO*, 20% of the data points are removed from the data set and the model was refitted; the predicted values for those points were then compared with the experimental values. Again, this is repeated until each data point has been omitted once. The results produced by the *LOO* (Q^2_{LOO} = 0.914) and the *LGO* (Q^2_{LGO} = 0.912) cross-validation tests illustrated the quality of the obtained model.

The model was further validated by applying *Y*-randomization approach. Several random shuffles of the *Y* vector (n) were performed and the low R^2 and Q^2 values that were obtained showing that the good results in the original model use not due to a chance correlation or structural dependency of the training set. The results of the *Y*-randomization test are presented in Table 3. In view of the success of the given linear model over external and internal validations and closeness of the predicted and experimental values of RI, the use of our model in further complementary investigations seems highly justified. In other words, the proposed method could be a useful aid to the costly and time-consuming experiments for determining the refractivity indices of other organic compounds regarding its high predictive ability and simplicity.

Table 3. R^2 and Q^2 values after several *Y*-randomization tests.

Iteration	R^2	Q^2	Iteration	R^2	Q^2
1	0.212	0.130	**6**	0.205	0.324
2	0.054	0.085	**7**	0.182	0.081
3	0.332	0.121	**8**	0.152	0.023
4	0.072	0.123	**9**	0.178	0.049
5	0.121	0.135	**10**	0.311	0.126

CONCLUSIONS

A simple QSPR model was presented for prediction of the refractivity index as one of the prominent characteristics of organic liquids. This model is a simple and interpretable multivariate linear model, which has only four variables or molecular descriptors. These four molecular descriptors were selected using ES-SW technique. The calculation of the variables is mainly based on the chemical structure of the related molecules. The validation procedures by the aid of cross-validation, separation of data into independent training and validation sets as well as *Y*-randomization illustrated the accuracy and robustness of the produced QSPR model not only by calculating its fitness on sets of training data, but also by testing the predictive ability of the model. The QSPR model with simply calculated molecular descriptors could be employed to estimate the refractivity indices (RIs) of similar or new organic compounds.

ACKNOWLEDGEMENTS

The authors express their special thanks to the Office of Research Affairs of the Islamic Azad University of Shahrood for financial support and the time dedicated to this work. The authors declare that there is no conflict of interests.

REFERENCES

1. Dijksman J.A., Rietz F., Lorincz K.A., van Hecke M., Losert W., 2012. Invited Article: Refractive index matched scanning of dense granular materials. Rev Sci Instrum. 83(1), 011301.

2. Jorge P.A. S., Silva S.O., Gouveia C., Tafulo P., Coelho L., Caldas P., Viegas D., Rego G., Baptista J.M., Santos J.L., Frazao O., 2012. Fiber optic-based refractive index sensing at INESC Porto. Sensors. 12(6), 8371-8389.

3. Jaksic Z., Vukovic S., Matovic J., Tanaskovic D., 2011. Negative refractive index metasurfaces for enhanced biosensing. Materials. 4(1), 1-36.

4. Bandoz J., Le Liboux M., Nahoum R., Israel G., Raulin F., Torre J.P., 1992. A sensitive cryogenic refractometer. Application to the refractive index determination of pure or mixed liquid methane, ethane and nitrogen. Rev Sci Instrum. 63(5), 2967-2973.

5. Grange B.W., Stevenson W.H., Viskanta R., 1976. Refractive index of liquid solutions at low temperatures: an accurate measurement. Appl Opt. 15, 858-859.

6. Malitson I.H., 1965. Interspecimen comparison of the refractive index of fused silica. J Opt Soc Am. 55 (10), 1205-1208.

7. Moreels E., de Greef C., Finsy R., 1984. Laser light refractometer. Appl Opt. 23, 3010-3013.

8. Shenoy M., Sukhdev R., Thyagarajan K., 1990. Simple prism coupling technique to measure the refractive index of a liquid and its variation with temperature. Rev Sci Instrum. 61, 1010-1013.

9. Tilton L.W., Taylor J.K., 1938. Refractive index and dispersion of distilled water for visible radiation, and temperatures of 0 to 60°C. J Res Nat Bur Stand. 20, 419-477.

10. Wong L.Y., Anderson A., 1972. Refractive-index measurements of some low-temperature liquids. J Opt Soc Am. 62, 219-222.

11. Zaidi A.A., Makdisi Y., Bhatia K.S., Abutahun I., 1989. Accurate method for the determination of the refractive

index of liquids using a laser. Rev Sci Instrum. 60, 803-805.

12. Nemoto S., 1992. Measurement of the refractive index of liquid using laser beam displacement. Appl Opt. 31, 6690-6694.

13. Abbate G., Attanasio A., Bernini U., Ragozzino E., Somma F., 1976. The direct determination of the temperature dependence of the refractive index of liquids and solids. J Phys D: Appl Phys. 9 1945-1951.

14. Andreasson S.D.H., Gustafsson S.E., Halling N.O., 1971. Measurment of the refractive index of transparent solids and fluids. J Opt Soc Am. 61(5), 595-599.

15. Appleby R., James D.W., Bowie C.A., 1984. An interference refractometer to measure the pressure and temperature dependence of the refractive index of liquids. Spectrochim Acta Mol Biomol Spectros. 40(8), 785-787.

16. Dobbins H.M., Peck E.R., 1973. Change of refractive index of water as a function of temperature. J Opt Soc Am. 63, 318-320.

17. Gagnon R.E., Gammon P.H., Kiefte H., Clouter J.M., 1979. Determination of the refractive index of liquid carbon monoxide. Appl Opt. 18, 1237-1239.

18. Kerl K., Jeschek M., 1982. Equipment for precise measurements of the complex refractive index of gases as a function of wavenumber, temperature and pressure. J Phys E. 15, 955-960.

19. Khashan M.A., 1979. Application of the Fabry-Perot interferometer as a refractometer. Opt Acta. 26, 881-888.

20. Schellekens P., Wilkening G., Reinboth F., Downs M.J., Birch K.P., Spronck J., 1986. Measurements of the refractive index of air using interference refractormeters. Metrologia. 22, 279-287.

21. Subramanian G., 1980. Refractive index by image-moire technique. J Phys E. 13, 111.

22. Kim H., Paulson S.E., 2013. Real refractive indices and volatility of secondary organic aerosol generated from photooxidation and ozonolysis of limonene, alpha-pinene and toluene. Atmos Chem Phys. 13(15), 7711-7723.

23. Odhner H., Jacobs D. T., 2012. Refractive index of liquid D_2O for visible wavelengths. J Chem Eng Data. 57(1), 166-168.

24. Prabhakaran P., Jang K.K., Son Y., Yang D.Y., Lee K.S., 2013. Fabrication of microstructures containing high refractive index materials by two-photon lithography. Mol Cryst Liq Cryst. 578(1), 4-18.

25. Qu H., Skorobogatiy M., 2012. Resonant bio- and chemical sensors using low-refractive-index-contrast liquid-core Bragg fibers. Sensor Actuat B-Chem. 161(1), 261-268.

26. Wyngaard J.C., Seaman N., Kimmel S.J., Otte M., Di X., Gilbert K.E., 2001. Concepts, observations, and simulation of refractive index turbulence in the lower atmosphere. Radio Sci. 36 (4), 643-669.

27. Xie R.J., Hintzen H.T., 2013. Optical properties of (oxy)nitride materials: A review. J Am Ceram Soc. 96(3), 665-687.

28. Goudarzi N., Goodarzi M., Mohammadhosseini M.M., Nekooei M., 2009. QSPR models for prediction of half-wave potentials of some chlorinated organic compounds using SR-PLS and GA-PLS methods. Mol Phys. 107(17), 1739-1744.

29. Nekoei M., Salimi M., Dolatabadi M., Mohammadhosseini M., 2011. A quantitative structure-activity relationship study of tetrabutylphosphonium bromide analogs as muscarinic acetylcholine receptors agonists. J Serb Chem Soc. 76(8), 1117-1127.

30. Mohammadhosseini M., Deeb O., Alavi- Gharabagh A., Nekoei M., 2012. Exploring novel QSRRs for simulation of gas chromatographic retention indices of diverse sets of terpenoids in *Pistacia lentiscus* L. essential oil using stepwise and genetic algorithm multiple linear regressions. Anal Chem Lett. 2, 80-102.

31. Nekoei M., Mohammadhosseini M., 2014. Application of HS-SPME, SDME and cold-press coupled to GC/MS to analysis the essential oils of *Citrus sinensis* CV. Thomson Navel and QSRR study for prediction of retention indices by stepwise and genetic algorithm-multiple linear regression approaches. Anal Chem Lett. 4(2), 93-103.

32. Mohammadhosseini M., Zamani H.A., Akhlaghi H., Nekoei M., 2011. Hydrodistilled volatile oil constituents of the aerial parts of *Prangos serpentinica* (Rech.f., Aell. Esfand.) Hernnstadt and Heyn from Iran and quantitative

structure-retention relationship simulation. J Essent Oil-Bear Plants. 14(5), 559-573.

33. Mohammadhosseini M., 2013. Novel PSO-MLR algorithm to predict the chromatographic retention behaviors of natural compounds. Anal Chem Lett. 3(4), 226-248.

34. Mohammadhosseini M., 2014. Prediction of the GC-MS retention indices for a diverse set of terpenes as constituent components of *Camu-camu* (*Myrciaria dubia* (HBK) McVaugh) volatile oil, using particle swarm optimization-multiple linear regression (PSO-MLR). Journal of Chemical Health Risks. 4(1), 75-95.

35. Mohammadhosseini M., 2012. 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. Asian J Chem. 24(9), 3814-3820.

36. Mohammadhosseini M., Nekoei M., 2013. Quantitative structure-electrochemistry relationship (QSER) study for prediction of half-wave reduction potentials of some chlorinated organic compounds by GA-MLR. Asian J Chem. 25, 349-352.

37. Nekoei M., Goudarzi N., Nekoei S., Mohammadhosseini M., 2014. QSAR study of arylsulfonylpiperazine inhibitors of 11β-HSD1 by GA-MLR, GA-PLS and GA-ANN. Anal Chem Lett. 4(1), 14-28.

38. Nekoei M., Mohammadhosseini M., Pourbasheer E., 2015. QSAR study of VEGFR-2 inhibitors by using genetic algorithm-multiple linear regressions (GA-MLR) and genetic algorithm-support vector machine (GA-SVM): A comparative approach. Med Chem Res. 24(7), 3037-3046.

39. Acevedo-Martinez J., Escalona-Arranz J. C., Villar-Rojas A., Tellez-Palmero F., Perez-Roses R., Gonzalez L., Carrasco-Velar R., 2006. Quantitative study of the structure-retention index relationship in the imine family. J Chromatogr A. 1102(1-2), 238-244.

40. Garkani-Nejad Z., Karlovits M., Demuth W., Stimpfl T., Vycudilik W., Jalali-Heravi M., Varmuza K., 2004. Prediction of gas chromatographic retention indices of a diverse set of toxicologically relevant compounds. J Chromatogr A. 1028(2), 287-295.

41. Héberger K., Kowalska T., 1999. Quantitative structure–retention relationships: VI. Thermodynamics of Kováts retention index–boiling point correlations for alkylbenzenes in gas chromatography Chemometr Intell Lab Syst. 47, 205-217.

42. Jalali-Heravi M., Fatemi M.H., 2001. Artificial neural network modeling of Kovats retention indices for noncyclic and monocyclic terpenes. J Chromatogr A. 915(1-2), 177-183.

43. Tulasamma P., Reddy K. S., 2006. Quantitative structure and retention relationships for gas chromatographic data: Application to alkyl pyridines on apolar and polar phases. J Mol Graphics Model. 25(4), 507-513.

44. Heberger K., Gorgeny M., 1999. Principal component analysis of Kovats indices for carbonyl compounds in capillary gas chromatography. J Chromatogr A. 845, 21-31.

45. Heberger K., Milczewska K., Voelkel A., 2005. Principal component analysis of polymer-solvent and filler-solvent interactions by inverse gas chromatography. Colloids and Surfaces A: Physicochem Eng Aspects. 260(1-3), 29-37.

46. Dean J. 1992. Lange's Handbook of Chemistry. McGraw-Hill. 14th Edn, New York.

47. Todeschini R., Consonni V., 2000. Handbook of Molecular Descriptors. Wiley-VCH, Mannhold, R., Kubinyi, H., Timmerman, H. Weinheim, Germany.

48. Efron B., 1983. Estimating the error rate of a prediction rule: Improvement on cross-validation. J Amer Statist Assoc. 78, 316-331.

49. Tropsha A., Gramatica P., Gombar V. K., 2003. The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. QSAR Comb Sci. 22(1), 69-77.

50. Wold S., Eriksson L. 1995. Statistical Validation of QSAR Results. In Chemometrics Methods in Molecular Design. VCH Weinheim. Germany.