Petroleum & Coal
ISSN 1337-7027

Available online at www.vurup.sk/pc
Petroleum & Coal 53 (2) 132-140, 2011

QUANTITATIVE STRUCTURE–PROPERTY RELATIONSHIP (QSPR) STUDY OF KOVATS RETENTION INDICES OF SOME OF ADAMANTANE DERIVATIVES BY THE GENETIC ALGORITHM AND MULTIPLE LINEAR REGRESSION (GA-MLR) METHOD

Z. Bayat*, M. Fakoor Yazdan Abad

Department of Chemistry,Islamic Azad University-Quchan Branch,Iran

Received December 17, 2010, Accepted May 15, 2011

Abstract

A quantitative structure–property relationship (QSPR) study was performed to develop models those relate the structures of 65 Kovats retention index (RI) of adamantane derivatives. Molecular descriptors derived solely from 3D structures of the molecular compounds. A genetic algorithm was also applied as a variable selection tool in QSPR analysis. The models were constructed using 52 molecules as training set, and predictive ability tested using 13 compounds. Modeling of RI of Adamantane derivatives as a function of the theoretically derived descriptors was established by multiple linear regression (MLR). The usefulness of the quantum chemical descriptors, calculated at the level of the DFT theories using 6-311+G** basis set for QSAR study of adamantane derivatives was examined. The use of descriptors calculated only from molecular structure eliminates the need to experimental determination of properties for use in the correlation and allows for the estimation of RI for molecules not yet synthesized. Application of the developed model to testing set of 13 drug organic compounds demonstrates that the model is reliable with goo predictive accuracy and simple formulation. The prediction results are in good agreement with the experimental value. A multi-parametric equation containing maximum Four descriptors at B3LYP/6-31+G** method with good statistical qualities (R^2_{train}=0.913, F_{train}=97.67, R^2_{test}=0.770, F_{test}=3.21, Q^2_{LOO}=0.895, R^2_{adj}=0.904, Q^2_{LGO}=0.844) was obtained by Multiple Linear Regression using stepwise method.

Keywords: Adamantane derivatives, Kovats retention indices(RI), genetic algorithm, MLR, QSPR.

1. Introduction

Diamondoids are classed with organic nanostructures; therefore, adamantane derivatives have become particularly popular with the development of nanotechnologies. The applications of adamantane derivatives are diverse: from antiviral drugs to nanorobots and molecular machines [1-3]. Particular attention is given to the chromatographic behavior of adamantane derivatives, because various chromatographic methods allow not only separation of multi component mixtures of isomers and structurally related framework hydrocarbons and their derivatives, but also qualitative and quantitative analysis of these mixtures [4]. Quantitative structure property relationships (QSPR), mathematical equations relating chemical properties such as acidity, electrochemistry, reactivity and chromatographic behavior to a wide variety of structural, topological and electronic features of the molecules [5], have been widely used in the field of chromatographic sciences [6–13]. Quantitative structure–retention relationships (QSSRs) represent statistical models which quantify the relation between the structure of the molecule and chromatographic retention indices of the compound, allowing the prediction of retention indices of novel compounds. QSPR on the RI have been reported for different types of organic compounds. Acevedo-Martinez et al. [14-18], developed linear and nonlinear models to study the Kovats retention indices of the immine family using topological, topographical and quantum chemical descriptors Correlations between the sorption and structural characteristics of adamantane derivatives were found based on the QSPR (Quantitative Structure Property Relationships) method. The success of a QSAR study depends on choosing robust statistical methods for producing the predictive model and also the relevant structural parameters for expressing the essential features within those chemical structures. Nowadays,

genetic algorithms (GA) are well known as interesting and widely used methods for variable selection [19]. In a QSAR study the model must be validated for its predictive value before it can be used to predict the response of additional chemicals. Validating QSPR with external data (i.e. data not used in the model development), although demanding, is the best method for validation [20–21]. In the present work, the data splitting was performed randomly and was confirmed by the factor spaces of the descriptors. Finally, the accuracy of the proposed model was illustrated using the following: leave one out, bootstrapping and external test set, cross-validations and Y-randomization techniques.

2. Methodology

2.1 Data set

The property data used in this study are the Kovats retention index (RI) of of the set of 65 Adamantane derivatives [22]. The data set was randomly divided into two subsets: the training set containing 52 compounds (80%) and the test set containing 13 compounds (20%). The training set was used to build a regression model, and the test set was used to evaluate the predictive ability of the model obtained. The activity data for the complete set of compounds are presented in Table 1.To derive QSAR models, an appropriate representation of the chemical structure is necessary. For this purpose, descriptors of the structure are commonly used.

Table 1. Experimental values of RI for adamantane derivatives training set

Name	Exp.	Pred	Ref.
adamantane	1118	1131	22
1 3 dimethyl adamantine	1151	1198	22
1-fluoro adamantine	1159	1259	22
2-methylene adamantine	1160	1172	22
1,3,5 -trimethyl adamantine	1163	1226	22
2-methyl adamantine	1196	1219	22
1 2-dimethyl adamantine	1236	1231	22
1-ethyl adamantine	1260	1221	22
2 2-dimethyl adamantine	1269	1274	22
1-ethyl-3,5 di methyl adamantine	1279	1291	22
3-ethyl-1-adamantanol	1283	1348	22
3-methyl-1-adamantanol	1283	1290	22
3 5-dimethyl-1-adamantanol	1295	1290	22
1-chloroadamantane	1298	1295	22
3,5,7-trimethyl-1-adamantanol	1304	1318	22
2-adamantanon	1320	1322	22
2-chloro adamantine	1342	1342	22
1-propyl adamantine	1347	1298	22
2-methyl-2-adamantanol	1348	1397	22
2-isopropyl adamantine	1349	1391	22
2-propyl adamantine	1371	1391	22
1-bromo adamantine	1382	1376	22
1-hydroxy methyl adamantine	1402	1393	22
1-chloromethyladamantane	1404	1367	22
2-isobuthyl adamantine	1416	1383	22
3-ethyl-5,7-dimethyl -1-adamantanol	1421	1395	22
3-5 dimethyl 1 hydroxy methyl adamantine	1425	1440	22
5-7-dimethyl1-3 adamantandiol1.	1438	1458	22
1-buthyl adamantine	1443	1383	22
methyl-(1-adamanthyl) ketone	1443	1407	22
methyl-(2-adamanthyl)ketone	1445	1387	22
2-ethyl-2-adamantanol	1446	1445	22
2-buthyl adamantine	1465	1416	22
adamantane-2-carboxylic acid methyl ester	1467	1464	22

Name	Exp.	Pred	Ref.
methyl ester of 3,5 di methyl adamantane1-carboxilic acid	1467	1510	22
1-bromomethyl adamantine	1488	1494	22
2-methyl-1-hydroxy methyl adamantine	1490	1438	22
3-isopropyl-1-adamantanol	1506	1406	22
adamantane -1-carboxylic acid ethyl ester	1508	1459	22
Methyl esters of 2-methyl adamantane -1-carboxylic acid	1512	1520	22
Adamantane-2-carboxylic acid ethyl ester	1529	1557	22
ethyl-(1-adamanthyl)ketone	1529	1491	22
adamantane-1-carboxylic acid iso propyl ester	1532	1585	22
adamantane-1-carboxylic acid tert-buthyl ester	1556	1679	22
2-isobuthyl-2-adamantanol	1570	1547	22
Methyl ester of -3-ethyl adamantane -1-carboxylic asid	1579	1544	22
3-buthyy-1-adamantanol	1595	1528	22
esters of adamantane 1-carboxylic acid propyl ester	1603	1630	22
2-buthyl-2-adamantanol	1620	1596	22
adamantane -1-carboxylic acid sec-buthyl ester	1631	1632	22
adamantane -1-carboxylic acid iso buthyl ester	1658	1596	22
di methyl ester of 5,7-di methyl adamantane -1-3 di carboxylic acid	1769	1809	22

2.2 Molecular descriptor generation

To derive QSAR models, an appropriate representation of the chemical structure is necessary. For this purpose, descriptors of the structure are commonly used. These descriptors are generally understood as being any term, index or parameter conveying structure information. Commonly used descriptors in the QSAR analysis are presented in Table 2.

Table 2. Experimental values of RI for adamantane derivatives test set

Name	EXP	Test	Ref.
1-methyladamantane	1137	1148	22
2-fluoro adamantine	1182	1281	22
1-adamantanol	1268	1374	22
2-ethyl adamantine	1284	1269	22
2-adamantanol	1329	1403	22
1-isopropyl adamantine	1358	1284	22
3 5-dimethyl -1-bromo adamantine	1401	1433	22
2-bromoadamantane	1426	1464	22
esters of adamantane1-carboxylic acid methyl ester	1449	1444	22
3-propyl-1-adamantanol	1495	1447	22
2-propyl-2-adamantanol	1526	1474	22
3-(1-adamanthyl)pentane	1559	1430	22
propyl-(1-adamanthyl) ketone	1609	1538	22

Some of the descriptors are obtained directly from the chemical structure, e. g. constitutional, geometrical, and topological descriptors. Other chemical and physicochemical properties were determined by the chemical structure (lipophilicity, hydrophilicity descriptors, electronic descriptors, energies of interaction). In this work, we used Gaussian 03 for ab initio calculations. DFT method at 6-31+G** were applied for optimization of adamantane derivatives and calculation of many of the descriptors. At first Adamantane derivatives were built by Hyperchem software and some o the descriptors such as surface area, hydration energy, and refractivity were calculated through it. The rest of the descriptors were obtained of Gaussian calculations. A large number of descriptors were calculated by Gaussian package and Hyperchem software. One way to avoid data redundancy is to exclude descriptors that are highly intercorrelated with each other before performing statistical analysis. Reduced multi co-linearity and redundancy in the data will facilitate selection of relevant variables and models for the investigated endpoint. Variable-selection for the QSAR modeling was carried out by stepwise linear regression method. A stepwise technique was employed that only one parameter at a time was added to a model and always in the order of most significant to least significant in terms of F-test

values. Statistical parameters were calculated subsequently for each step in the process, so the significance of the added parameter could be verified. The goodness of the correlation is tested by the regression coefficient (R^2), the F-test and the standard error of the estimate (SEE). The test and the level of significance, as well as the confidence limits of the regression coefficient, are also reported. The squared correlation coefficient, R^2, is a measure of the fit of the regression model. Correspondingly, it represents the part of the variation in the observed (experimental) data that is explained by the model.

2.3 Genetic algorithm

Genetic algorithms (GAs) are governed by biological evolution rules [23]. These are stochastic optimization methods that have been inspired by evolutionary principles. The distinctive aspect of a GA is that it investigates many possible solutions simultaneously, each of which explores a different region in the parameter of space [24]. To select the most relevant descriptors, the evolution of the population was simulated [25-27]. The first generation population was randomly selected; each individual member in the population was defined by a chromosome of binary values and represented a subset of descriptors. The number of the genes at each chromosome was equal to the number of the descriptors. A gene was given the value of 1, if its corresponding descriptor was included in the subset; otherwise, it was given the value of zero. The number of the genes with the value of 1 was kept relatively low to have a small subset of descriptors [28]. As a result, the probability of generating 0 for a gene was set greater (at least 60 %) than the value of 1. The operators used here were the crossover and mutation operators. The application probability of these operators was varied linearly with a generation renewal (0–0.1 % for mutation and 60–90 % for crossover). The population size was varied between 50 and 250 for the different GA runs. For a typical run, the evolution of the generation was stopped when 90% of the generations took the same fitness. The fitness function used here was the leave-one-out cross-validated correlation coefficient, Q^2_{LOO}. The GA program was written in Matlab 6.5 [29].

3. Results and discussion

In a QSAR study, generally, the quality of a model is expressed by its fitting ability and prediction ability, and of these the prediction ability is the more important. In order to build and test the model, a data set of 65 compounds was separated into a training set of 52 compounds, which were used to build the model and a test set of 13 compounds, which were applied to test the built model. With the selected descriptors, we have built a linear model using the training set data, and the following equation was obtained:

$$RI = -4.45754\ (\pm 1.161528)\ \sigma_9\ -80.1305\ (\pm 11.72555)\ \Delta G_{CYCLO+}\ 5.768715\ (\pm 0.292762)\ M - 121.607\ (\pm 42.44063)\ MC_{9+}\ 0.072961\ (\pm 0.015957)\ HF + 177.4361\ (\pm 112.5648)$$

(B3LYP/6-31+G**)

R^2_{train}=0.914 F_{train}=97.674 R^2_{test}=0.770 F_{test} = 3.214 R^2_{adj}=0.904
Q^2_{LOO}=0.895 Q^2_{LGO}= 0.84451 N_{train}= 52, N_{test} = 13

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 leave one out, bootstrapping and external test set respectively, RMSE is the root mean square error and F is the Fisher F statistic. The figures in parentheses are the standard deviations. The built model was used to predict the test set data and the prediction results are given in Table 1. and the test results are given in Table 3. As can be seen from Table 1, the calculated values for the RI are in good agreement with those of the experimental values. The predicted values for RI for the compounds in the training and test sets using equation RI were plotted against the experimental RI values in Figure 1.and the comparison between Retention Index using prediction and the experimental .A plot of the residual for the predicted values of RI for both the training and test sets against the experimental RI values are shown in Figure 2. As can be seen the model did not show any proportional and systematic error, because the propagation of the residuals on both sides of zero are random. The real usefulness of QSAR models is not just their ability to reproduce known data, verified by their fitting power (R^2), but is mainly their potential for predictive application. For this reason the model calculations were performed by maximizing

the explained variance in prediction, verified by the leave-one-out cross-validated correlation coefficient,Q^2_{LOO} maximizing the explained variance in prediction, verified by the leave-one-out cross-validated correlation coefficient, Q^2LOO. To avoid the danger of over fitting and the possibility of overestimating the model predictivity by using Q^2LOO procedure, as is strongly recommended for QSAR modeling. The Q^2 LOO and Q^2 GLO for the MLR model are shown in Equation RI. This indicates that the obtained regression model has a good internal and external predictive power.

Table 3. The calculated descriptors used in this study.

Descriptors	Symbol	Abbreviation	Descriptors	Symbol	Abbreviation
Quantum chemical descriptors	Molecular Dipole Moment	MDP	Quantum chemical descriptors	difference between LUMO and HOMO	E_{GAP}
	Molecular Polarizability	MP		Hardness [η=1/2 (HOMO+LUMO)]	H
	Natural Population Analysis	NPA		Softness (S=1/ η)	S
	Electrostatic Potential0	EP		Electro negativity [χ= -1/2 (HOMO–LUMO)]	*X*
	Highest Occupied Molecular Orbital	HOMO		El Electro philicity (ω=$χ^2$ /2 η)	Ω
	Lowest Unoccupied Molecular Orbital	LUMO		Mullikenl Chargeg	MC
Chemical properties	Partition Coefficient	Log P	Chemical properties	Molecule surface area	SA
	Mass	M		Hydration Energy	HE
	Molecule volume	V		Refractivity	REF

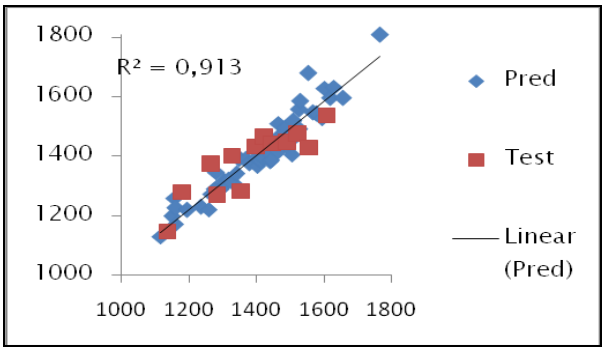

Fig.1. The predicted versus the experimental RI by MLR

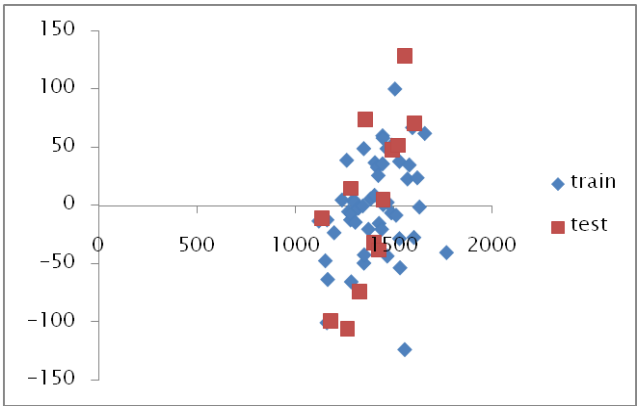

Fig. 2. The residual versus the experimental LogP by GA-MLR. (See colour version of this figure online at www.informahealthcare.com/enz)

Also, in order to assess the robustness of the model, the Y-randomization test was applied in this study. The dependent variable vector (RI) was randomly shuffled and The new QSAR models (after several repetitions) would be expected to have low R^2 and Q^2_{LOO} values (Table 4). If the opposite happens then an acceptable QSAR model cannot be obtained for the specific modeling method and data.

Table 4. The R^2_{train} and Q^2_{LOO} values after several Y-randomization tests

NO	R^2_{train}	Q^2_{LOO}
1	0.1045	0.0312
2	0.000022	0.0976
3	0.0939	0.0614
4	0.0042	0.1282
5	0.0457	0.0570
6	0.0340	0.1927
7	0.0060	0.1442
8	0.2991	0.0125
9	0.0175	0.1700
10	0.0251	0.0608

The MLR analysis was employed to derive the QSAR models for different adamantane derivatives. MLR and correlation analyses were carried out by the statistics software SPSS (Table 5).

Table 5. The correlation coefficient existing between the variables used in MLR with B3lyp/6-31+G** method.

	HF	MC_9	M	ΔG_{CYCLO}	σ_9
HF	1	0	0	0	0
MC_9	0.048869	1	0	0	0
M	0.39506	0.245901	1	0	0
ΔG_{CYCLO}	0.099875	0.22142	0.226936	1	0
σ_9	0.17506	0.485565	0.070032	0.04617	1

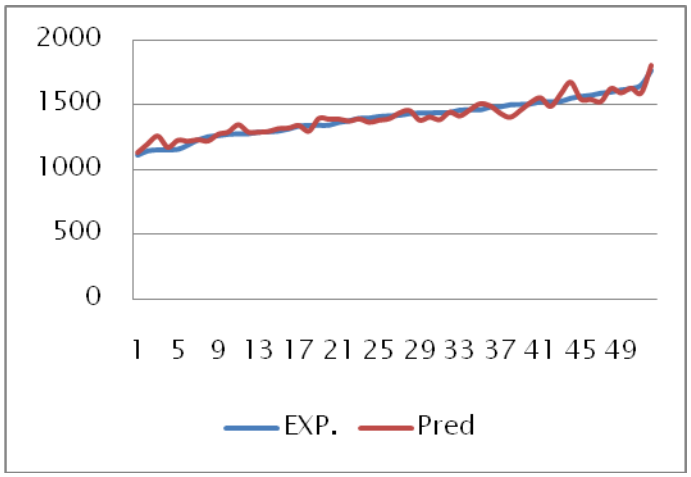

Fig. 3. The comparison between properties (RI) using experimental and prediction
Series 1: Values of log P were obtained by using prediction. Series 2: Values of log P were obtained by using experimental methods.

3.1. Interpretation of descriptors

The QSPR developed indicated that Nuclear magnetic Resonance (σ_9), free energy solvation (ΔG_{CYCLO}), mulliken charge (MC), Hartee-fuck energy (HF) compound Kovats retention index. Positive values in the regression coefficients indicate that the indicated descriptor contributes positively to the value of RI, whereas negative values indicate that the greater the value of the descriptor the lower the value of RI. In other words, increasing the σ_9,ΔG_{CYCLO} and MC will decrease RI and increasing the HF and M increases extent of RI of the adamantane derivatives. The standardized regression coefficient reveals the significance of an individual descriptor presented in the regression model.

4. Conclusion

In this article, a QSAR study of 65 adamantane derivatives was performed based on the theoretical molecular descriptors calculated by the GAUSSIAN software and selected. The built model was assessed comprehensively (internal and external validation) and all the validations indicated that the QSPR model built was robust and satisfactory, and that the selected descriptors could account for the structural features responsible for the adamantane derivatives property of the compounds. The QSPR model developed in this study can provide a useful tool to predict the RI of new compounds and also to design new compounds with high RI.

References

[1] Janssens S, Beyaert R.: Functional diversity and regulation of different interleukin-1 receptor-associated kinase (IRAK) fami members, Mol. Cell. 2003;11:293–302.

[2] Li S, Strelow A, Fontana EJ, Wesche H. IRAK-4: A novel member of the IRAK family with the properties of an IRAK-kinase, Proc. Natl. Acad. Sci. USA 2002;99:5567–5572.

[3] Medvedev AE, Lentschat A, Kuhns DB, Blanco JC, Salkowski C, Zhang S, Arditi M, Gallin JI, Vogel SN.: Distinct mutations in IRAK-4 confer hyporesponsiveness to lipopolysaccharide and interleukin-1 in a patient with recurrent bacterial infections, J. Exp. Med. 2003;198:521–531.

[4] Picard C, Puel A, Bonnet M, Ku CL, Bustamante J, Yang K, Soudais C, Dupuis S, Feinberg J, Fieschi C, Elbim C, Hitchcock R, Lammas D, Davies G, Al- Ghonaium A, Al-Rayes H, Al-Jumaah S, Al-Hajjar S, Al-Mohsen IZ, Frayha HH, Rucker R, Hawn TR, Aderem A, Tufenkeji H, BHaraguchi S, Day NK, Good RA, Gougerot-Pocidalo MA, Cassanova JL.: Pyogenic bacterial infections in humans with IRAK-4 deficiency, Science 2003;299:2076–2079.

[5] Li X.: IRAK4 in TLR/IL-1R signaling: possible clinical applications, Eur. J. Immunol. 2008;38:614–618.

[6] Buckley GM, Ceska TA, Fraser JL, Gowers L, Groom CR, Higueruelo AP, Jenkins K, Mack SR, Morgan T, Parry DM, Pitt WR, Rausch O, Richard MD, Sabin V.: IRAK-4 inhibitors. Part II: A structure-based assessment of imidazo[1,2-a]pyridine binding. Bioorg. Med. Chem. Lett., 2008;18:3291–3295.

[7] Buckley GM, Fosbeary R, Fraser JL, Gowers L, Higueruelo AP, James LA, Jenkins K, Mack SR, Morgan T, Parry DM, Pitt WR, Rausch O, Richard MD, Sabin V.: IRAK-4 inhibitors. Part III: A series of imidazo[1,2-a]pyridines, Bioorg. Med. Chem. Lett. 2008;18:3656–3660.

[8] Buckley GM, Gowers L, Higueruelo AP, Jenkins K, Mack SR, Morgan T, Parry DM, Pitt WR, Rausch O, Richard MD, Sabin V, Fraser JL.: IRAK-4 inhibitors. Part 1: a series of amides, Bioorg. Med. Chem. Lett., 2008;18:3211–3214.

[9] Sammes PG, Taylor JB.: Comprehensive Medicinal Chemistry. Oxford, Pergamon Press, 1990:766.

[10] Riahi S, Pourbasheer E, Dinarvand R, Ganjali MR, Norouzi P.: Quantitative structure-activity relationship study on the anti-HIV-1 activity of novel 6-naphthylthio HEPT analogs, Chem. Biol. Drug. Des., 2008;74:165–172.

[11] Riahi S, Pourbasheer E, Ganjali MR, Norouzi P.: Investigation of different linear and nonlinear chemometric methods for modeling of retention index of essential oil components: Concerns to support vector machine, J. Hazard. Mater. 2009;166:853–859.

[12] Riahi S, Pourbasheer E, Ganjali MR, Norouzi P.: Support vector machinebased quantitative structure-activity relationship study of cholesteryl ester transfer protein inhibitors, Chem. Biol. Drug. Des., 2009;73:558–571.

[13] Depczynski U, Frost VJ, Molt K.: Genetic algorithms applied to the selection of factors in principal component regression, Anal. Chim. Acta, 2000;420:217.

[14] Alsberg BK, Marchand-Geneste N, King RD.: A new 3D molecular structure representation using quantum topology with application to structureproperty relationships, Chemometr. Intel. Lab., 2000;54:75–91.

[15] Jouanrimbaud D, Massart DL, Leardi R, Denoord OE.: Genetic algorithms as a tool for wavelength selection in multivariate calibration, Anal. Chem., 1995;67:4295–4301.

[16] Riahi S, Ganjali MR, E Pourbasheer, Divsar F, Norouzi P, Chaloosi M.: Development and validation of a rapid chemometrics-assisted spectrophotometry and liquid chromatography methods for the simultaneous determination of the phenylalanine, tryptophan and tyrosine in the pharmaceutical products, Curr. Pharm. Anal., 2008;4:231–237.
[17] Riahi S, Ganjali MR, Pourbasheer E, Norouzi P.: QSRR study of GC retention indices of essential-oil compounds by multiple linear regression with a genetic algorithm, Chromatographia, 2008;67:917–922.
[18] Riahi S, Pourbasheer E, Ganjali MR, Norouzi P, Zeraatkar Moghaddam A.: QSPR study of the distribution coefficient property for hydantoin and 5-arylidene derivatives. A genetic algorithm application for the variable selection in the MLR and PLS methods, J. Chin. Chem. Soc., 2008;55:1086–1093.
[19] Riahi S, Ganjali MR, Moghaddam AB, Pourbasheer E, Norouzi P.: Development of a new combined chemometrics method, applied in the simultaneous voltammetric determination of cinnamic acid and 3, 4-dihydroxy benzoic acid, Curr. Anal. Chem., 2009;5:42–47.
[20] Tropsha A, Gramatica P, Gombar VK.: The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models, QSAR Comb. Sci., 2003;22:69–77.
[21] Riahi S, Ganjali MR, Norouzi P, Jafari F.: Application of GA-MLR, GA-PLS and the DFT quantum mechanical (QM) calculations for the prediction of the selectivity coefficients of a histamine-selective electrode, Sens. Actuators, B 2008;132:13–19.
[22] Burkhard, J., Vais, J., Vodicka L. and Landa S.: Adamantane and its derivatives. XVI. The gas chromatographic characterization of adamantane derivatives, Journal of Chromatography. Chrom. 4057.
[23] Holland H.: Adaption in Natural and Artificial Systems. Ann Arbor, MI: The University of Michigan,1975;342–375.
[24] Cartwright HM.: Applications of Artificial Intelligence in Chemistry. Oxford: Oxford University, 1993;760–765.
[25] Hunger J, Huttner G.: Optimization and analysis of force field parameters by combination of genetic algorithms and neural networks, J. Comput. Chem. 1999;20:455–471.
[26] Ahmad S, Gromiha MM.: Design and training of a neural network for predicting the solvent accessibility of proteins, J. Comput. Chem. 2003; 24:1313–1320.
[27] Waller CL, Bradley MP.: Development and validation of a novelb variable selection technique with application to multidimensional quantitative structure-activity relationship studies, J. Chem. Inf. Comput. Sci. 1999;39:345–355.
[28] Aires-de-Sousa J, Hemmer MC, Casteiger J.: Prediction of H-1 NMR chemical shifts using neural networks, Anal. Chem. 2002;74:80–90.
[29] The Mathworks. Genetic Algorithm and Direct Search Toolbox Users Guide. Massachusetts: MathWorks, 2002;50–65.