Journal of Entomological Research

Volume 13, Issue 2, pages:35-53

Islamic Azad University, Arak Branch
ISSN 2008-4668
http://jer.iau-arak.ac.ir

Prediction of toxicity and octanol–water partition coefficient of Phenylcarbamate Derivativesas Insecticides Using Genetic Algorithm-Multiple Linear Regressions Method

A. sadat Navabi[1],T. Momeni Isfahani[2*], M. Ramazani[3], M. Alimoradi[4]

1- PHD student, Department of Chemistry, Arak Branch, Islamic Azad University, P.O. BOX 38135-567, Arak, Iran
2. Assistant Professor, Department of Chemistry, Arak Branch, Islamic Azad University, P.O. BOX 38135-567, Arak, Iran
3. Associate Professor, Department of Chemistry, Arak Branch, Islamic Azad University, P.O. BOX 38135-567, Arak, Iran
4. Assistant Professor, Department of Chemistry, Arak Branch, Islamic Azad University, P.O. BOX 38135-567, Arak, Iran

Abstract

A Quantitative Structure–Activity Relationship (QSAR) study based on Genetic Algorithm Multiple Linear Regressions (GA-MLR) were carried out for the prediction of the toxicity ($\log IC_{50}$) and the logarithm of octanol–water partition coefficient ($\log P_{ow}$) of some carbamate derivatives as insecticides. The optimized conformation of compounds were obtained at HF/6-31G* level with Gaussian 98 software. Dragon software is used to calculate molecular descriptors. A data set of these compounds was randomly divided into 2 groups: training and test sets. The QSAR models were optimized using multiple linear regressions (MLR).The most relevant molecular descriptors were collected by Genetic Algorithm (GA) and backward regression. The best GA-MLR models are obtained using statistical parameters, such as squared correlation coefficient (R^2), adjusted squared correlation coefficient (R^2adj), root mean square error (RMSE) values for training and test sets. The best QSAR models are obtained based on the statistical parameters Leave-one-out (LOO) cross-validation, external test set, external validation parameters (Q^2_{F1}, Q^2_{F2}, Q^2_{F3}) and the concordance correlation coefficient (CCC) were used to quantify the predictive ability of GA-MLR models. The results showed that GA-MLR models could be used to predict the activities of carbamate derivatives.

Keywords: carbamate derivatives, QSAR, $\log IC_{50}$, Octanol-water partition coefficients, insecticides

* Corresponding Author, E-mail*: t.momeni.@iau-arak.ac.ir*
Received: 23 May. 2021 – Accepted: 21 sep. 2021

Introduction

Quantitative structure activity relationship (QSAR) is the most widely used method in computational chemical biology and drug design (Sharma, 2011) .Various QSAR approaches have been developed for designing pharmaceuticals and agrochemicals (Muratov, 2020). Carbamate compounds are esters of carbamic acid that are commonly used as insecticides. These compounds are referred to as N-methylcarbamates. Apart from that, derivatives of carbamic acid, thiocarbamic acid, and dithiocarbamic acid are used as herbicides. When used properly, carbamate pesticides offer significant benefits to countries, as they protect agricultural yields, as well as human and animal health from insect-vector-mediated diseases. However, overexposure of humans and animals to these pesticides often results in poisonings. Also carbamate derivatives are synthesized for investigation of antimicrobial activity, biological activity and the pseudo-irreversible inhibitors (Chiou, 2009; Stepankova, 2008; Vorčáková, 2018).

As mentioned previously, carbomate derivatives represent a potential area in agricultural production, medicinal chemistry and modern drug discovery. Carbamate functional groups can be found in many approved drugs and prodrugs such as antiepileptic drug (Felbamate), nonopioid analgesic (Flupirtine), anthelmintic drug (Albendazole), inhibitor of HIV-1 and HIV-2 protease (Ritonavir) and inhibitor of reverse transcriptase (Efavirenz) (Ghosh, 2015). Biscarbamate with two carbamate groups in each molecule plays an important role in design and synthesis of antitumor compounds (Anderson, 1987; Anderson,1990) and Acetylcholinesterase (AChE) and Butyrylcholinesterase (BuChE) inhibitors (Yu, 1971; Bosak, 2013).

A QSAR study provides a theoretical basis for prediction of how active carbamate pesticides are by using quantitative reactivity index. The abuse of such pesticides has resulted in remaining large amount of carbamate in the ecological environment, posing a serious threat to the entire ecosystem (Jiahao, 2021).

The studies of mechanism of inhibitory effect and lipophilicity of a number of substituted N-phenylcarbamates have also been investigated using molecular modeling approach (Vorčáková, 2018).

Rivastigmine ((S)-N-ethyl-3-[(1-dimethylamino) ethyl] N-methylphenylcarbamate) has been used as Alzheimer's disease drug. It is a drug with a good selectivity for the brain enzyme and inhibitory activity against AChE and BChE (Anand, 2013; Yu,1999; Ahmadinejad,2018).

Besides, carbamate derivatives of indolines have been applied as antioxidants and Alzheimer's disease drug (Yanovsky, 2012).

The formations of non-covalent interaction and halogen bonding interactions have been proved between protein backbone of cholinesterase (ChE) enzymes and inhibitor (Hardegger, 2011; Zhou, 2012; Lu,2012; Xu,2014; Cavallo,2016).

The trifluoromethyl group ($-CF_3$) is one of the most important functional groups and producing them by anaesthetics (Halothane), antidepressants (Prozac, Paxil, Zoloft and Luvox), and antimalarial agent (Lariam) is a challenging task in medicinal chemistry, (Croft,2007; Beheshti,2016; Brodbeck, 1979).

Three dimensional quantitative structure-activity relationships (3D-QSARs) have been found for the fungicidal activity of N-phenyl-O-phenylthionocarbamate analogues against resistant and sensitive Botrytis cinerea (Allen, 1989).

QSAR analysis has been performed to predict toxicity of two groups of insecticides, namely the organophosphates and carbamates by using several types of descriptors including topological, spatial, thermodynamic, information content, lead likeness and E-state indices (Naik,2009).

The main aim of the present study is to develop QSAR models for predicting toxicity ($logIC_{50}$) and octanol-water partition coefficient ($logP_{ow}$) of new inhibitors of cholinesterases based on N-phenylcarbamates insecticides using Genetic Algorithm - Multiple Linear Regressions (GA-MLR) and molecular descriptors.

Materials and methods

Two groups of carbamate derivatives were studied: (i) N-phenylcarbamates with additional carbamate group (1–12) and (ii) N-phenylcarbamates with monosaccharide moiety (13–24). The inhibitory activity of carbamate derivatives against electrical acetylcholinesterase (eeAChE) is expressed as Median Inhibitory Concentration (IC_{50}). It is a quantitative measure that indicates how much of a particular inhibitory substance (e.g. drug) is needed to inhibit, in vitro, a given biological component (i.e. an enzyme, cell, cell receptor or microorganism, etc.) or biological process by 50%.

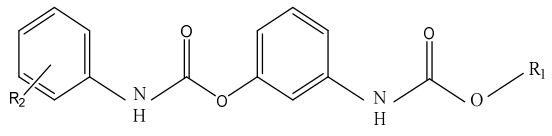
IC_{50} values are typically expressed as molar concentrations. In this study each of IC_{50} values were converted into a micro molar unit (μM) followed by transformation into a positive logarithmic scale ($logIC_{50}$).

The logarithm of octanol-water partition ratio or coefficient ($logP_{ow}$) of compounds is in fact a measure of hydrophobicity, and it has been widely used in many areas such as designing of drugs, specifying toxicology of compounds, modeling organic pollutants and determining the environmental effect of substances (Hansch.1995). $LogP_{ow}$ and $logIC_{50}$values were taken from the literature (Vorčáková, 2018).

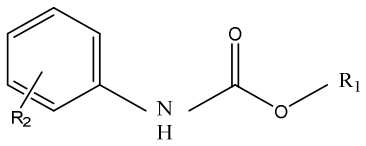
Studied carbamate drugs derivatives are listed in Table 1. The data set of 24 carbamate drugs derivatives were randomly split into a training set of 18 compounds (75%) that was applied to build the model and a testing set of 6 compounds (25%) used to assess model performance.

In order to calculate the molecular descriptors, Dragon software package Version 5.4-2006 was used. For this propose the GaussView program of Gaussian software was used to draw the chemical structure of molecules. These chemical structures then were optimized applying the Gaussian 98 software package based on Hartree-Fock (HF) method and a 6-31G* basis set. Finally, the output of Gaussian software for each compound was fed into Dragon program and the descriptors were calculated.

Table 1.The name and template structure of two groups of carbamate drugs derivatives used in present study.



A



B

No.	Main backbone form	R_1	R_2	Name
1	A	Methoxy	H	O-3-[(Methoxycarbonyl)amino]phenyl-N-phenyl carbamate
2	A	Ethoxy	H	O-3-[(Ethoxycarbonyl)amino]phenyl-N-phenyl carbamate
3	A	n-Butoxy	H	O-3-[(Butoxycarbonyl)amino]phenyl-N-phenyl carbamate

4	A	Ethoxy	3-Methyl	O-3-[(Ethoxycarbonyl)amino]phenyl-N-(3-methylphenyl) carbamate
5	A	n-Butoxy	3-Methyl	O-3-[(Butoxycarbonyl)amino]phenyl-N-(3-methylphenyl) carbamate
6	A	Ethoxy	4-Methoxy	O-3-[(Ethoxycarbonyl)amino]phenyl-N-(4-methoxyphenyl) carbamate-
7	A	n-Butoxy	4-Methoxy	O-3-[(Butoxycarbonyl)amino]phenyl-N-(4-methoxyphenyl) carbamate-
8	A	Ethoxy	Cl	O-3-[(Ethoxycarbonyl)amino]phenyl-N-(4 chlorophenyl)carbamate-
9	A	n-Butoxy	Cl	O-3-[(Butoxycarbonyl)amino]phenyl-N-(4 chlorophenyl)carbamate-
10	A	Methoxy	di(Cl)	O-3-[(Methoxycarbonyl)amino]phenyl-N-(3,4 dichlorophenyl)carbamate-
11	A	Ethtoxy	di(Cl)	O-3-[(Ethoxycarbonyl)amino]phenyl-N-(3,4 dichlorophenyl)carbamate
12	A	n-Butoxy	di(Cl)	O-3-[(Butoxycarbonyl)amino]phenyl-N-(3,4 dichlorophenyl)carbamate-
13	B	galactopyranose	di(Cl)	O-(1,2:3,4-Di-O-isopropylidene-a-D-galactopyranose-6-yl)- N–(3,4-dichlorophenyl)carbamate–
14	B	galactopyranose	CF3	O-(1,2:3,4-Di-O-isopropylidene-a-D-galactopyranose-6-yl)- N–(3-(trifluoromethyl)phenyl)carbamate
15	B	glucofuranose	di(Cl)	O-(1,2:5,6-Di-O-isopropylidene-a-D-glucofuranose-3-yl)–N(3,4-dichlorophenyl)carbamate–
16	B	glucofuranose	CF3	O-(1,2:5,6-Di-O-isopropylidene-a-D-glucofuranose-3-yl)–N (3-(trifluoromethyl)phenyl)carbamate–
17	B	fructopyranose	di(Cl)	O-(2,3:4,5-Di-O-isopropylidene-b-D-fructopyranose-1-yl)- N–(3,4-dichlorophenyl)carbamate
18	B	fructopyranose	CF_3	O-(2,3:4,5-Di-O-isopropylidene-b-D-fructopyranose-1-yl) -N–(3-(trifluoromethyl)phenyl)carbamate
19	B	Galactopyranose	di(Cl)	O-(a/b-D-Galactopyranose-6-yl)–N–(3,4-dichlorophenyl) carbamate
20	B	Galactopyranose	CF_3	O-(a/b-D-Galactopyranose-6-yl)–N–(3-(trifluoromethyl phenyl)carbamate(
21	B	Glucopyranose	di(Cl)	O-(a/b- -3-yl)–N–(3,4-dichlorophenyl)carbamate
22	B	Glucopyranose	3-CF_3	O-(a/b-D-Glucopyranose-3-yl)–N–(3-(trifluoromethyl phenyl)carbamate(
23	B	Fructofuranose	di(Cl)	O-(a/b-D-Fructofuranose-1-yl)–N–(3,4-dichlorophenyl)carbamate
24	B	Fructofuranose	CF_3	O-(a/b-D-Fructofuranose-1-yl)–N–(3-(trifluoromethyl phenyl)carbamate(

These descriptors are obtained using various techniques, especially by mathematical equations, and are divided into 20 logical blocks. Some of these blocks entail Topological, RDF, 3D-MoRSE, Information, 2D autocorrelations and GETAWAY descriptors (Moriwaki, 2018; Xu, 2011; Bahadori, 2017; Todeschini, 2000).

As a result, a total set of 1143 and 1532 dragon molecular descriptors were calculated for each compound in the data sets of log IC50 and logPow respectively. These descriptors were discarded based on the following condition

1) Descriptors that are constant have been eliminated (350 descriptors for logIC50 and 423 descriptors for logPow).

2) To decrease the redundancy among descriptors, the correlation of descriptors with each other and with activities (log IC50 and logPow) of the molecules were examined, and collinear descriptors (R> 0.9) were detected. Among the collinear descriptors, the one having the highest correlation with activity was retained, and the others were removed from the data matrix (545descriptors for $\log IC_{50}$ and 668 descriptors for $\log P_{ow}$).

Finally, for the selection of the most significant descriptors from the pool of remaining ones, genetic algorithm (GA) with variable subset selection method (Xu, 2011) was used. These descriptors were then used as inputs of backward stepwise MLR method.

Genetic algorithm (GA): Genetic algorithm (GA) is a stochastic method to solve optimization problems based on a fitness function, chromosome representation terms, and biological-inspired operators (Shen, 2004; Baumann, 2002). The fitness function is used to assign a value for all the chromosomes in the population. Chromosomes are considered as points in the solution space. Each chromosome- or individual- in the population represents a model with a subset of variables. The biological-inspired operators are called Selection, Crossover, and Mutation (Katoch, 2020). In the present work, the GA program was written in Matlab 6.5 (Mathworks, 2002). The number of genes at each chromosome was equal to the number of remaining descriptors. The probability of the application of crossover and mutation was varied linearly with generation renewal (0–0.1% for mutation and 60–90% for crossover). The population size was varied between 50 and 250 for different GA runs. In order to cope with stochastic noise, each GA run was repeated 50 times and results were averaged.

The Backward MLR model was then built using the Statistical Package for the Social Science (SPSS) statistics version 20 to determine the relationships between activities ($\log IC_{50}$, $\log P_{ow}$) and the chemical structural of the carbamate drugs derivatives.

Regression parameters

The backward stepwise MLR regression method was used as a linear technique for the construction of QSAR models in this work.

The toxicity (log IC_{50}) and $\log P_{ow}$ values were considered as dependent variables while independent variables were dragon molecular descriptors. Statistics parameters were used to determine the quality of the models. These parameters include the coefficient of determination (R^2) the root mean square error (RMSE), adjusted correlation coefficient (R^2_{adj}), Durbin-Watson statistic (DW) and fisher index of quality (F). Each of mentioned statistical parameters can be described as the followings

The coefficient of determination (R^2) (Chatterjee, 2013) is a ratio of the explained sum of squares to the total sum of squares, and it is given by the following equation:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i}^{n}(y_i - \bar{y}_i)^2} \qquad (1)$$

Where y_i is the observed activity, $\hat{y}_i$ is the activity calculated by the model, and $\bar{y}_i$ is the average activity.

The root mean square error (RMSE) for the training or test sets is calculated as follows:

$$RMSE = \sqrt{\frac{\sum(y_{obs}-y_{pred})^2}{n}} \qquad (2)$$

The adjusted correlation coefficient (R^2_{adj}) (Chatterjee, 2015), is obtained by the following formula, where *p* is the total number of regressors in the model, n is the sample size, and R^2 is the coefficient of determination.

$$R^2_{adj} = R^2 - \frac{p(1-R^2)}{(n-p-1)} \qquad (3)$$

Known as one of the most well-known statistical tests, Fisher index of quality (F) or fisher's ratio is a measure for (linear) discriminating power of some variables.

The F-ratio test is defined as the ratio of the model sum of squares (MSS) over the residual sum of squares (RSS) by the following formula (Allen, 2018):

$$F = \frac{MSS(n-p-1)}{RSS.p} \qquad (4)$$

In the above expression, p is the number of predictor variables used in the model development. The computed F value of a model should be significant at $P < 0.05$.

MSS is defined as the sum of the square differences between the estimated response and the average observed response:

$$MSS = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \qquad (5)$$

RSS is defined as the sum of square differences between the observed (y) and estimated response ($\hat{y}$) over all the sample objects.

$$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (6)$$

The Durbin Watson (DW) statistic is a test for autocorrelation in the residuals from a statistical regression analysis and also it is a test for first-order serial correlation. The DW test statistic can be computed using equation (7).

$$DW = \frac{\sum_{t=2}^{T}(E_t - E_{t-1})^2}{\sum_{t=1}^{T}{E_t}^2} \qquad (7)$$

Where E_t is residual from an ordinary least squares regression. This test reports a test statistic with a value from 0 to 4; if there is no autocorrelation, the DW statistic will be around 2. The DW statistic will fall below 2 if there is positive autocorrelation while it will lie somewhere between 2 and 4 if there is negative autocorrelation, (Hisamatsu, 1994; Hatekar, 2010).

Results and Discussion

QSAR models for the toxicity (Log Ic_{50})

The GA–MLR analysis led to the derivation of five different models for the toxicity, with 3-7 descriptors. Table **2** shows the regression coefficients and statistical parameters of models for log IC_{50}. The statistical parameters of the five models are almost similar; thus, model 5, which has the lowest number of descriptors, has been selected. Equation (8) shows that the suitable model for the toxicity includes three molecular descriptors: Gs, AMW and Mor04m). These descriptors have been classified into topological, constitutional and 3D-MoRSE indices.

Log IC_{50}= 4.419 + 0.113 Mor04m - 0.194 AMW - 4.554 Gs (8)

R= 0.990, R^2=0.980, R^2_{adj}=0.976, RMSE=0.041, F=231.030, Sig= 0.000

Table 2. Statistical parameters of the models calculated with the SPSS software for the log IC_{50}.

models	Indepdndent Variables	R	R^2	R^2_{adj}	RMSE	F	Sig
1	**Gs** (G total symmetry index/weighted by atomic electrotopological states **),Mor08u** (signal 08 / unweighted) **,AMW** (average molecular weight) **,Mor16u** (signal 16 / unweighted) **,Mor04m** (signal 04 / weighted by mass) **, HATS6m** (leverage-weighted autocorrelation of lag 6 / weighted by mass)**, ,PipC04**(molecular multiple path count of order 4)	0.996	0.992	0.987	0.041	188.439	0.000
2	Gs ,Mor08u ,AMW ,Mor16u ,Mor04m ,HATS6m	0.995	0.991	0.985	0.041	192.211	0.000
3	Gs , AMW ,Mor16u ,Mor04m ,HATS6m	0.995	0.990	0.986	0.041	233.435	0.000
4	Gs , AMW ,Mor16u ,Mor04m	0.995	0.989	0.986	0.041	293.858	0.000
5	Gs , AMW,Mor04m	0.990	0.980	0.976	0.041	231.030	0.000

QSAR models for the logarithm of octanol- water partition coefficient (logP_{ow})

Several linear QSAR models with between 3 to 7 descriptors are established to predict logP_{ow}. Table **3** shows the statistical parameters of different models for the logP of carbamate drugs derivatives.

values of R^2, RMSE and adjusted correlation coefficient (R^2_{adj}), and also involvement of fewer descriptors in the model are good measures of the quality of QSAR models. The best linear model for the logP includes three molecular descriptors BLI, RDF020u and EEig02d.

The statistical parameters of the best model are presented in equation (**9**). These descriptors have been classified into topological, RDF and eigenvalues indices.

Log P_{OW} = -38.031 +6.225 EEig02d – 0.217 RDF020u + 24.311 BLI (9)

R= 0.959, R^2=0.919, R^2_{adj}=0.901, RMSE=0.398, F=52.778, Sig= 0.000

Table 3. Statistical parameters of the models calculated with the SPSS software for the log P_{OW}.

models	Indepdndent Variables	R	R^2	R^2_{adj}	RMSE	F	Sig
1	**BLI**(Kier benzene-likeliness index) **, GGI8**(topological charge index of order 8) **, J3D** (3D matrix-based descriptors)**, RDF020u**(Radial Distribution Function - 020 / unweighted) **, EEig02d**(Eig02_AEA(dm)) **,Mor13u**(signal 13 / unweighted)**, BEHe4**(SpMax4_Bh(e))**,BEHm1**(SpMax1_Bh(m))	0.969	0.940	0.886	0.398	17.472	0.000
2	**BLI, GGI8, J3D, RDF020u, EEig02d, Mor13u ,BEHm1**	0.969	0.938	0.895	0.398	21.792	0.000
3	**BLI, GGI8, J3D, RDF020u, EEig02d, BEHm1**	0.965	0.932	0.895	0.398	25.139	0.000
4	**BLI, GGI8, RDF020u, EEig02d, BEHm1**	0.964	0.930	0.900	0.398	31.711	0.000
5	**BLI, RDF020u, EEig02d, BEHm1**	0.960	0.923	0.899	0.398	38.704	0.000
6	**BLI, RDF020u, EEig02d**	0.959	0.919	0.901	0.398	52.778	0.000

Multicollinearity

, to specify the best model for predicting the mentioned activities,
Multicollinearity criterion is used

In regression analysis methods multicollinearity should be checked. In statistics, multicollinearity (also collinearity) is a state of very high inter-correlations or inter-associations among the independent variables and it generally occurs when there is a high correlation between two or more predictor variables; in other words, when one predictor variable can be used in order to predict the other one (Srivastava, 2012; Pourbasheer, 2017). This creates redundant information, skewing the results in a regression model. An easy way to detect multicollinearity is to calculate correlation coefficients for all pairs of predictor variables; the variance inflation factor (VIF) and Pearson correlation coefficient (PCC) values were elicited for the selected descriptors using SPSS. The VIF is calculated as presented in equation (10)

$$\mathrm{VIF} = \frac{1}{1-R^2} \qquad (10)$$

Where R^2 is the correlation coefficient of the multiple regression between the variables within the model. If the VIF value lies between 1 and 10, there would be no multicollinearity and the related model is acceptable. On the contrary, if VIF is larger than 10, the related model is unstable and a recheck is necessary (Roy,2016; Tropsha, 2010). The corresponding VIF values of the three descriptors for both models 8, 9 are presented in Tables **4** and **5**. As can be seen from these Tables, all the variables have VIF values lower than 10 indicating that the selected descriptors are not highly correlated and hence the developed model has high statistical significance with reasonably orthogonal descriptors (Srivastava, 2012). What is more, in these tables all the correlation coefficient value for each pair of descriptors was less than 0.49 that meant the selected descriptors were independent.

Table 4. Internal correlation matrix between molecular descriptors (Equation 8).

Descriptor	Gs	AMW	Mor04m	VIF
Gs	1	0.084	0.279	6.320
AMW		1	-0.488	2.594
Mor04m			1	3.719

Table 5. Internal correlation matrix between molecular descriptors (Equation 9).

Descriptor	BLI	RDF020u	EEig02d	VIF
BLI	1	0.227	0.251	1.212
RDF020u		1	-0.268	1.197
EEig02d			1	1.185

QSAR model validation

The capability of the QSAR equations to predict $\log P_{ow}$ and $\log IC_{50}$ of new compounds were determined using internal validation by squared cross- validation coefficient for leave –one- out (Q^2_{LOO}) (Richardson, 2017) and also using external validation. Cross-validated squared correlation coefficient is calculated by equation (11).

$$Q^2 = 1 - \frac{\sum(y_{pred}-y_{obs})^2}{\sum(y_{pred}-\bar{y})^2} \qquad Q^2 \leq 1 \qquad (11)$$

In Equation (11), y_{pred} and y_{obs} indicate predicted and observed activity values respectively and $\bar{y}$ indicates mean activity value. The value of Q^2LOO should be greater than 0.5 for an acceptable model. For LOO cross-validation, a data point is removed from the training set of compounds, and the model is recalculated. This will be repeated until each data point is omitted once. The Q^2LOO values of the toxicity (log IC_{50}) and $logP_{ow}$ were calculated 0.858 and 0.842 respectively. These results illustrated the quality of the obtained GA-MLR models.

In order to create and test the models, 24carbamate drugs derivatives randomly separated into two groups, a training set of 18 compounds (75%) was applied to test the fabricated model and a prediction set of 6 compounds was considered as an external validation set (25%).

Statistical parameters including correlation coefficient (R), coefficient of multiple determination (R^2), adjusted correlation coefficient (R^2adj), Fisher ratio (F) and root mean square error (RMSE), for training and external validation sets are listed in Table **6 pertained to both** of Equations **8** and **9**. From the table, it can be seen that the statistical coefficients are very satisfactory and there are desirable linear correlations between activities (logIC_{50}, $logP_{ow}$) of carbamate drugs derivatives and the selected descriptors.

Table 6. Statistical parameters obtained by the GA- MLR model for the training and test sets (Equations 8and 9).

Activity	Data set	N	R	R^2	R^2_{adj}	RMSE	DW	F	Q^2_{LOO}
log IC_{50}	training	18	0.990	0.980	0.976	0.041	2.038	231.030	0.858
	test	6	0.997	0.994	0.972	0.034	1.992	79.853	-
$logP_{ow}$	testing	18	0.959	0.919	0.901	0.398	2.026	52.778	0.892
	test	6	0.968	0.937	0.842	0.429	1.987	23.146	-

To verify predictive performance of QSAR models, different validation criteria were proposed using external validation parameters Q^2_{F1} (Shi,2001), Q^2_{F2} is based on prediction of the test set compounds (Schüürmann, 2008), Q^2_{F3} has a metric with a threshold value of 0.5 required for validation of a QSAR model (Consonni, 2009). the concordance correlation coefficient (CCC) parameter can be calculated in order to check the model reliability (Chirico, 2011) and the predictive parameter R^2_m (Pratim, 2009). Q^2_{F1}, Q^2_{F2} and Q^2_{F3} can be computed using equations (12)-(14).

$$Q^2_{F1} = 1 - \frac{\sum_{i=1}^{nEXT}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{nEXT}(y_i - \bar{y}_{TR})^2} \quad (12)$$

$$Q^2_{F2} = 1 - \frac{\sum_{i=1}^{nEXT}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{nEXT}(y_i - \bar{y}_{EXT})^2} \quad (13)$$

$$Q^2_{F3} = 1 - \frac{[\sum_{i=1}^{nEXT}(\hat{y}_i - y_i)^2]/n_{EXT}}{[\sum_{i=1}^{nTR}(y_i - \bar{y}_{TR})^2]n_{TR}} \quad (14)$$

In equations (12)-(14), y_i is the experimental response, $\hat{y}_i$ is the predicted value of the response calculated excluding the i^{th} element, $\bar{y}$ is the mean value of the dependent variable, $\bar{y}_{TR}$is the average of the training, $\bar{y}_{EXT}$is the average of the external data set (Shi, 2001).

Fitting of statistical results of the BW-MLR models are reported in Tables 6 while cross-validation of results is presented in table 7. External predictability of both the activities is evident from high values of external validation parameters e.g. R^2_{ext} and Q^2-Fnindicates the statistical robustness of the models. Closer values of RMSE of the training and test sets were observed for the logIC_{50} and $logP_{ow}$, and variance explained in external prediction (Q^2_{F1}, Q^2_{F2}, and Q^2_{F3} > 0.93, and concordance correlation coefficient (CCC > 0.96) (see Table).

Table 7. The QSAR models predictivity results.

Activity	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	$RMSE_{ext}$	CCC_{ext}	R^2_{ext}	R^2_{adj}
log IC_{50}	0.998	0.991	0.999	0.034	0.992	0.994	0.972
$logP_{ow}$	0.937	0.936	0.960	0.429	0.960	0.937	0.842

Because all of the validation techniques prove that the obtained models 7and 8 are valid ones, these models can be used to predict logIC_{50} and logP_{ow} of carbamate drugs derivatives.

The experimental and predicted values based on the GA-MLR model are shown in Table 8. Also, Figs. 1 and 2 show the predicted versus experimental values of logIC_{50} and logP_{ow} for the training and test sets with GA-MLR method.

As can be seen, the predicted values for the logIC_{50} and logP_{ow} are in good agreement with experimental data.

Table 8. The experimental, predicted and residual values of log (IC_{50}) and logP_{ow} of carbamate drugs derivatives. (The inhibitory activity of carbamate drugs derivatives against electric eel acetylcholinesterase (eeAChE) expressed as IC_{50}(μM)).

No.	**Experimental IC_{50} (μM) eeAChE**	**Experimental log IC_{50}**	**Prediction log IC_{50}**	**Residual log IC_{50}**	**Experimental logP**	**Prediction log P**	**Residual log P**
1	20.60	1.314	1.290	0.023	3.300	3.541	0.516
2	12.90	1.111	1.164	-0.054	3.900	3.928	0.048
3*	15.70	1.196	1.152	0.044	5.100	5.238	-0.138
4	18.30	1.262	1.193	0.069	4.600	4.291	0.034
5	20.90	1.320	1.283	0.037	5.300	5.289	0.068
6	12.40	1.179	1.172	0.007	3.500	3.301	0.057
7*	14.10	1.093	1.135	-0.042	5.000	4.181	0.819
8	15.00	1.149	1.119	0.030	4.800	4.567	0.049
9	15.00	1.176	1.199	-0.023	5.400	5.568	0.090
10	7.700	0.886	0.923	-0.037	4.900	5.184	0.079
11*	8.600	0.934	0.941	-0.006	5.300	5.031	0.269
12	9.900	0.996	1.062	-0.066	6.300	6.110	0.157
13	46.90	1.671	1.658	0.013	6.200	5.938	0.109
14	61.40	1.788	1.779	0.009	5.000	5.152	0.105
15*	69.40	1.841	1.846	-0.005	7.300	7.551	-0.251
16	70.00	1.845	1.950	-0.105	5.900	5.322	0.125
17	76.30	1.883	1.807	0.075	4.900	5.861	0.165
18	76.50	1.884	1.935	-0.051	6.700	6.176	0.384
19*	94.30	1.975	2.029	-0.055	2.000	2.217	-0.217
20	99.20	1.997	2.014	-0.018	2.000	1.564	0.337
21	89.10	1.950	1.879	0.071	3.100	2.975	0.299
22	103.4	2.015	2.083	-0.068	1.900	2.199	0.194
23*	98.00	1.991	1.927	0.064	3.000	3.482	-0.482
24	103.50	2.015	1.928	0.087	3.000	3.732	0.182

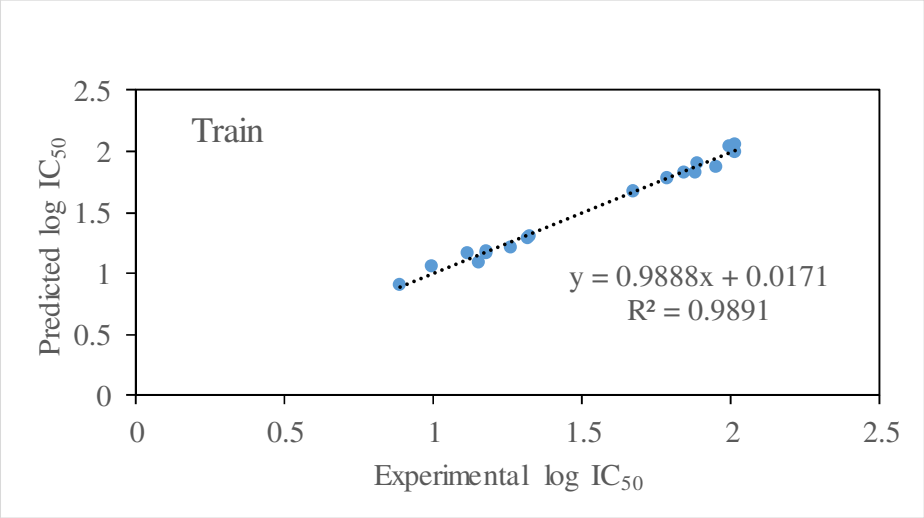


Figure 1: Plot of the calculated values of $logIC_{50}$ from the GA-MLR model versus the experimental values of it for training set.

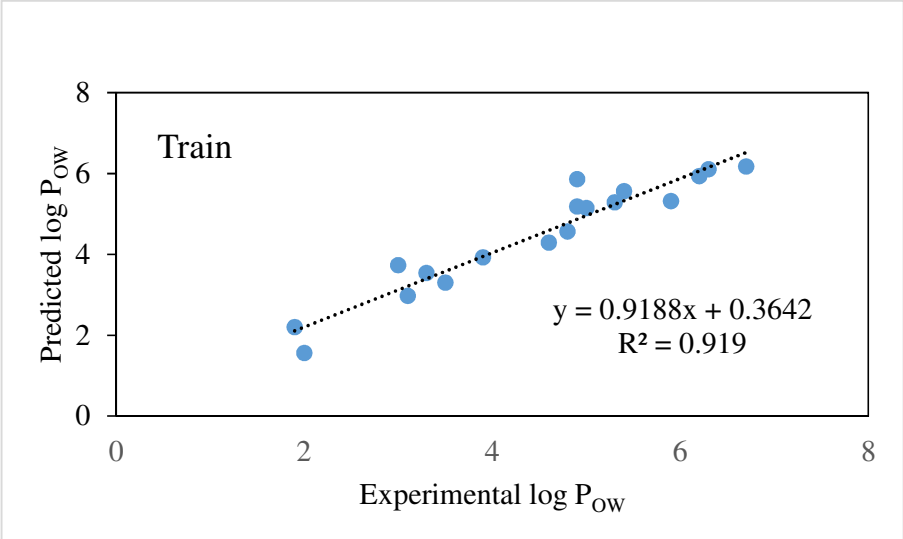


Figure 2: Plot of the calculated values of logPow from the GA-MLR model versus the experimental values of it for training set.

Residual

Residuals, in the context of regression models, are the difference between the observed value of the target variable (y) and the predicted value (ŷ), i.e. the error of the prediction. The residuals plot shows the residuals on the vertical axis and the dependent variable on the horizontal axis. The residual plots can be used to assess the quality of the regression and the underlying statistical assumptions about residuals such as constant variance, independence of variables and normality of the distribution. For these assumptions to hold true for a particular regression model, the residuals would have to be randomly distributed around zero.

As can be seen in figures 3 and 4, the points in residual plots are randomly scattered around the horizontal axis therefore a linear model provides a decent fit to the data.

The residual values of the $logIC_{50}$ and $logP_{ow}$, expressed by equations **8** and **9, are** shown in Table **8**.

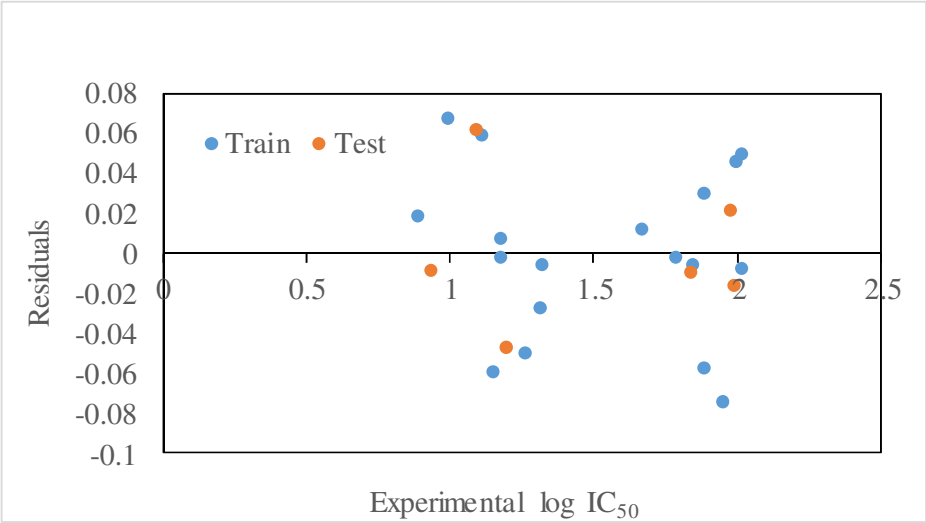


Figure 3 : Residuals plotted against the observed toxicity for training and test sets of carbamate drugs derivatives.

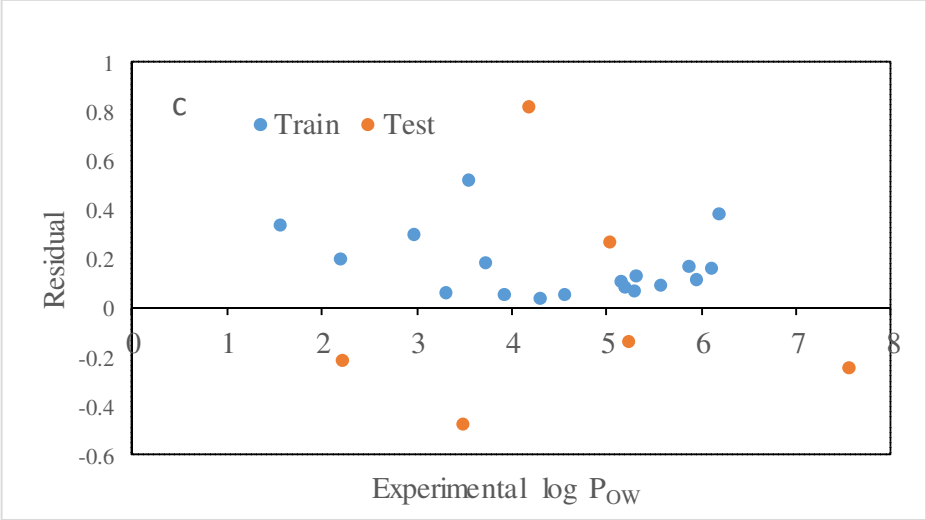


Figure 4: Residuals plotted against the observed logP_{ow} for training and test sets of carbamate drugs derivatives.

Interpretation of the best descriptors

As it was already shown, three descriptors: Gs, AMW and Mor04m, can be successfully used for modeling and predicting the toxicity (logIC_{50}) of carbamate drugs derivatives. Gs descriptor is symmetrical index which is weighted by atomic electrotopological states. Gs has been classified as a topological index.

The topological indices have been derived based on descriptors including the molecular graph adjacency matrix, the distance matrix, the distance vector, the number of edges and valence electrons in molecular graph (Hu, 2016). These descriptors were successfully used to represent the characterization of structures and for anticipation of wide spectrum of properties and activities (Consonni, 2002; Balaban, 1992).

AMW is a descriptor of the average molecular weight that has been classified as a constitutional index. These descriptors (0D-descriptors) are the most simple and commonly used descriptors, reflecting the composition of a molecule without any information about molecular structure and atom connectivity (Hu, 2016; Consonni, 2002; Balaban, 1992).

The numbers of atoms, rings, bonds and sum or average of the atomic properties are examples of constitutional descriptors. These descriptors show the ability to prognosticate a wide range of thermodynamic properties and biological activities (Ahmadinejad, 2018; Redžepović, 2020; Estrada, 2000).

Mor04m descriptor is a signal 04, weighted by mass has been classified into 3D-MoRSE (3Dmolecular representation of structure based on electron diffraction) indices.

On the other hand, three other descriptors: BLI, RDF020u and EEig02d can be used successfully to predict the $logP_{ow}$ of carbamate drugs derivatives. These descriptors belong to topological, Radial Distribution Function (RDF) and Eigenvalues descriptors respectively. RDF is a tridimensional descriptor and it has been applied for modeling and predicting the physicochemical, biological and potential toxicological of organic compounds (Consonni, 2002; González, 2006).

The eigenvalue descriptor or eigenvalue-based topological molecular indices (EI) has been classified into several groups owing to the nature of graph matrices (e.g. adjacency matrix) which are used in their definitions (Redžepović, 2020; Estrada, 2000; González, 2006). These indices have been successfully used to studying the folding in biomolecules.

To examine the relative importance and contribution of each descriptor in the final model, the value of the mean effect (MF) was calculated for each descriptor (Hayat,2019; Pourbasheer, 2009) using equation (15).

$$MF_j = \frac{\beta_j \sum_{i=1}^{i=n} d_{ij}}{\sum_j^m \beta_j \sum_i^n d_{ij}} \qquad (15)$$

In equation **15**, MF_j represents the mean effect of descriptor j, βj is the coefficient of descriptor j, dij stands for the value of descriptor j in the data matrix, m is the number of descriptors included in the model and n is the number of molecules in the training set (Golmohammadi, 2010; Oluwaseye, 2018; Habibi-Yangjeh, 2009).

The selected descriptors were used in the GA-MLR models and their corresponding MF values are shown in Table **9**.

The value and sign of MF value shows the relative contribution and direction of influence of each descriptor on the partition coefficient.

From Table 9, it can be seen that the most influential descriptor based on its mean effect is Gs descriptor for $logIC_{50}$ and is BLI descriptor for $logP_{ow}$. These descriptors had the highest contribution to the studied compounds.

Gs descriptor has a positive sign for its mean effect which reveals that as it increases, the values of $logIC_{50}$ increases too. BLI has negative sign for its mean effect, which means when it increases, the values of $logP_{ow}$ decreases.

Table 9. The meaning of the descriptors used in the GA-MLR models and their MF, VIF and coefficient values**.**

The best descriptors for $logIC_{50}$	**Notation**	**Natation Block**	**Coefficient**	**MF[a]**	**VIF[b]**
G total symmetry index/weighted by atomic electrotopological states	Gs	Topological indices	0.113	0.955	6.320
average molecular weight	AMW	Constitutional indices	-0.194	0.041	2.594
signal 04 / weighted by mass	Mor04m	3D-MoRSE descriptors	-4.554	-0.024	3.719

The best descriptors for $logP_{ow}$		Natation	Coefficient	MF[a]	VIF[b]
Kier benzene-likeliness index	BLI	Topological indices	24.311	-0.802	1.212
Radial Distribution Function - 020 / unweighted	RDF020u	RDF descriptors	-0.217	-0.007	1.197
Eig02_AEA(dm)	EEig02d	Eigenvalues	6.225	0.205	1.185

Conclusion

The pesticides have attracted more and more attentions not only for their benefits in improving the crop yield but also for the potential of causing environmental danger simply becasue almost all pesticides are detrimental, to all living things**.** In present study, QSAR models have been used to predict the logarithm of octanol-water partition coefficient ($logP_{ow}$) and the toxicity ($logIC_{50}$) of phenylcarbamate derivatives by the genetic algorithm -multiple linear regressions (GA-MLR). Recognizing the modeling descriptors of the obtained models, LogP and logIC50 pertaining descriptors were taken into account in every model. the major significance of descriptors on the pesticide toxicity was proved. Three descriptors Gs, AMW and Mor04m are more important for describing $logIC_{50}$, while $logP_{ow}$ can be better modeled using three descriptors BLI, RDF020u and EEig02d.The predictive efficiency of constructed QSAR models was examined by different validation criteria.

The desirable results of validation parameters and high statistical quality of GA-MLR models indicate that generated models can be accurately applied to predict the $logIC_{50}$ and $logP_{ow}$ of the carbamate derivatives. Also these developed QSAR models can be used to design some new carbamate drugs derivatives with promising characteristics.

References

Abbasitabar, F., and Zare-Shahabadi, V. 2017. In silico prediction of toxicity of phenols to Tetrahymena pyriformis by using genetic algorithm and decision tree-based modeling approach. Chemosphere, 172: 249-259.

Ahmadinejad, N., Shafiei, F., and Isfahani, T. M. 2018. Quantitative Structure-Property Relationship (QSPR) Investigation of Camptothecin Drugs Derivatives. Combinatorial chemistry & high throughput screening, 21(7): 533-542.

Allen, K. N., and Abeles, R. H. 1989. Inhibition kinetics of acetylcholinesterase with fluoromethyl ketones. Biochemistry, 28(21): 8466-8473.

Allen, R. 2018. Fisher's F-ratio illustrated graphically. The Mathematical Gazette, 102(553): 50-62.

Anand, P., and Singh, B. 2013. A review on cholinesterase inhibitors for Alzheimer's disease. Archives of pharmacal research, 36(4): 375-399.

Anderson, W. K., and Mach, R. H. 1990. Synthesis and chemical reactivity of c-5 substituted 6, 7-bis-(hydroxymethyl)-1H-pyrrolizine biscarbamate tumor inhibitors. Journal of heterocyclic chemistry, 27(4): 1025-1030.

Anderson, W. K., and Milowsky, A. S. 1987. 3-Pyrroline N-oxide bis (carbamate) tumor inhibitors as analogs of indicine N-oxide. Journal of medicinal chemistry, 30(11): 2144-2147.

Bahadori, B., and Atabati, M. 2017. Harmony search as a powerful tool for feature selection in QSPR study of the drugs lipophilicity. Combinatorial chemistry & high throughput screening, 20(4): 321-327.

Balaban, A., and Balaban, T. 1992. Correlations using topological indices based on real graph invariants. Journal de chimie physique, 89(7-8): 1735-1745.

Baumann, K., Albert, H., and Von Korff, M. 2002. A systematic evaluation of the benefits and hazards of variable selection in latent variable regression. Part I. Search algorithm, theory and simulations. Journal of Chemometrics: A Journal of the Chemometrics Society, 16(7): 339-350.

Beheshti, A., Pourbasheer, E., Nekoei, M., and Vahdani, S. 2016. QSAR modeling of antimalarial activity of urea derivatives using genetic algorithm–multiple linear regressions. Journal of Saudi Chemical Society, 20(3): 282-290.

Bosak, A., Smilović, I. G., Štimac, A., Vinković, V., Šinko, G., and Kovarik, Z. 2013. Peripheral site and acyl pocket define selective inhibition of mouse butyrylcholinesterase by two biscarbamates. Archives of biochemistry and biophysics, 529(2): 140-145.

Brodbeck, U., Schweikert, K., Gentinetta, R., and Rottenberg, M. 1979. Fluorinated aldehydes and ketones acting as quasi-substrate inhibitors of acetylcholinesterase. Biochimica et Biophysica Acta (BBA)-Enzymology, 567(2): 357-369.

Cavallo, G., Metrangolo, P., Milani, R., Pilati, T., Priimagi, A., Resnati, G., and Terraneo, G. 2016. The halogen bond. Chemical reviews, 116(4): 2478-2601.

Chatterjee, S., and Hadi, A. S. 2015. Regression analysis by example: John Wiley & Sons.

Chatterjee, S., and Simonoff, J. S. 2013. Handbook of regression analysis: Wiley Online Library.

Chiou, S. Y., Huang, C. F., Hwang, M. T., and Lin, G. 2009. Comparison of active sites of butyrylcholinesterase and acetylcholinesterase based on inhibition by geometric isomers of benzene-di-N-substituted carbamates. Journal of biochemical and molecular toxicology, 23(5): 303-308.

Chirico, N., and Gramatica, P. 2011. Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. Journal of chemical information and modeling, 51(9): 2320-2335.

Consonni, V., Ballabio, D., and Todeschini, R. 2009. Comments on the definition of the Q 2 parameter for QSAR validation. Journal of chemical information and modeling, 49(7): 1669-1678.

Consonni, V., Todeschini, R., Pavan, M., and Gramatica, P. 2002. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 2. Application of the novel 3D molecular descriptors to QSAR/QSPR studies. Journal of chemical information and computer sciences, 42(3): 693-705.

Croft, A. M. 2007. A lesson learnt: the rise and fall of Lariam and Halfan. Journal of the Royal Society of Medicine, 100(4): 170-174.

Estrada, E. 2000. Characterization of 3D molecular structure. Chemical Physics Letters, 319(5-6): 713-718.

Ghosh, A. K., and Brindisi, M. 2015. Organic carbamates in drug design and medicinal chemistry. Journal of medicinal chemistry, 58(7): 2895-2940.

Golmohammadi, H., and Dashtbozorgi, Z. 2010. Quantitative structure–property relationship studies of gas-to-wet butyl acetate partition coefficient of some organic compounds using genetic algorithm and artificial neural network. Structural Chemistry, 21(6): 1241-1252.

González, M. P., Terán, C., Teijeira, M., and Helguera, A. M. 2006. Radial distribution function descriptors: an alternative for predicting A2 A adenosine receptors agonists. European journal of medicinal chemistry, 41(1): 56-62.

Habibi-Yangjeh, A., and Danandeh-Jenagharad, M. 2009. Application of a genetic algorithm and an artificial neural network for global prediction of the toxicity of phenols to Tetrahymena pyriformis. Monatshefte Für Chemie-Chemical Monthly, 140(11): 1279-1288.

Hansch, C., Leo, A., Hoekman, D., and Livingstone, D. 1995. Exploring QSAR: hydrophobic, electronic, and steric constants: American Chemical Society Washington, DC.

Hardegger, L. A., Kuhn, B., Spinnler, B., Anselm, L., Ecabert, R., Stihle, M., Gsell, B., Thoma, R., Diez, J., & Benz, J. 2011. Systematic investigation of halogen bonding in protein–ligand interactions. Angewandte Chemie International Edition, 50(1): 314-318.

Hatekar, N. R. 2010. *Principles of Econometrics: An Introduction (using R)*: SAGE Publications India.

Hayat, S., Imran, M., and Liu, J. B. 2019. Correlation between the Estrada index and π-electronic energies for benzenoid hydrocarbons with applications to boron nanotubes. International Journal of Quantum Chemistry, 119(23): e26016.

Hisamatsu, H., and Maekawa, K. 1994. The distribution of the Durbin-Watson statistic in integrated and near-integrated models. Journal of Econometric*s*, 61(2): 367-382.

Hu, B., Kuang, Z.-K., Feng, S.-Y., Wang, D., He, S.-B., and Kong, D.-X. 2016. Three-dimensional biologically relevant spectrum (BRS-3D): shape similarity profile based on PDB ligands as molecular descriptors. Molecules, 21(11): 1554.

Jiahao, T., Liangliang, W., Li, P., Junjie, D., and Xiaoqin, D. 2021. Quantitative structure-activity relationship of carbamate pesticides based on quantitative reactivity index. Chinese Journal of Pesticide Science, 23(4): 688-693.

Katoch, S., Chauhan, S. S., and Kumar, V. 2020. A review on genetic algorithm: past, present, and future. Multimedia Tools and Applications: 1-36.

Lu, Y., Liu, Y., Xu, Z., Li, H., Liu, H., and Zhu, W. 2012. Halogen bonding for rational drug design and new drug discovery. Expert opinion on drug discover*y*, 7(5): 375-383.

Moriwaki, H., Tian, Y.-S., Kawashita, N., and Takagi, T. 2018. Mordred: a molecular descriptor calculator. Journal of cheminformatics, 10(1): 1-14.

Muratov, E. N., Bajorath, J., Sheridan, R. P., Tetko, I. V., Filimonov, D., Poroikov, V., Oprea, T. I., Baskin, I. I., Varnek, A., and Roitberg, A. 2020. Correction: QSAR without borders. Chemical Society Reviews, 49(11): 3716-3716.

Naik, P., Sindhura, Singh, T., and Singh, H. 2009. Quantitative structure–activity relationship (QSAR) for insecticides: development of predictive in vivo insecticide activity models. SAR and QSAR in Environmental Research, 20(5-6): 551-566

Oluwaseye, A., Uzairu, A., Shallangwa, G., and Abechi, S. 2018. Qsar studies on derivatives of Quinazoline-4 (3h)-Ones with anticonvulsant activities. The Journal of Engineering and Exact Sciences, 4(2): 0255-0264.

Pourbasheer, E., Ahmadpour, S., Zare-Dorabei, R., and Nekoei, M. 2017. Quantitative structure activity relationship study of p38α MAP kinase inhibitors. Arabian Journal o*f* Chemistry, 10(1): 33-40.

Pourbasheer, E., Riahi, S., Ganjali, M. R., and Norouzi, P. 2009. Application of genetic algorithm-support vector machine (GA-SVM) for prediction of BK-channels activity. European journal of medicinal chemistry, 44(12): 5023-5028.

Pratim Roy, P., Paul, S., Mitra, I., and Roy, K. 2009. On two novel parameters for validation of predictive QSAR models. Molecules, 14(5): 1660-1701.

Redžepović, I., and Furtula, B. 2020. Predictive potential of eigenvalue-based topological molecular descriptors. Journal of Computer-Aided Molecular Design, 34(9): 975-982.

Richardson, R. J., and Johnson, D. E. 2017. Computational Systems Pharmacology and Toxicology: Royal Society of Chemistry.

Roy, K., Das, R. N., Ambure, P., and Aher, R. B. 2016. Be aware of error measures. Further studies on validation of predictive QSAR models. Chemometrics and Intelligent Laboratory Systems, 152: 18-33.

Schüürmann, G., Ebert, R.-U., Chen, J., Wang, B., and Kühne, R. 2008. External validation and prediction employing the predictive squared correlation coefficient Test set activity mean vs training set activity mean. Journal of Chemical Information and Modeling, 48(11): 2140-2145.

Sharma, O., Saini, N., Gupta, V., Sachdeva, K., and Arya, H. 2011. Evolutionary history of QSAR: a review. J. Natur. Cons, 1(4): 266-272.

Shen, Q., Jiang, J.-H., Jiao, C.-X., Huan, S.-Y., Shen, G.-l., and Yu, R.-Q. 2004. Optimized partition of minimum spanning tree for piecewise modeling by particle swarm algorithm. QSAR studies of antagonism of angiotensin II antagonists. Journal of chemical information and computer sciences, 44(6): 2027-2031.

Shi, L. M., Fang, H., Tong, W., Wu, J., Perkins, R., Blair, R. M., Branham, W. S., Dial, S. L., Moland, C. L., and Sheehan, D. M. 2001. QSAR models using a large diverse set of estrogens. Journal of Chemical Information and Computer Sciences, 41(1): 186-195.

Srivastava, A. K., and Shukla, N. 2012. QSAR based modeling on a series of lactam fused chroman derivatives as selective 5-HT transporters. Journal of Saudi Chemical Society, 16(4): 405-412.

Stepankova, S., and Komers, K. 2008. Cholinesterases and cholinesterase inhibitors. Current Enzyme Inhibition, 4(4): 160-171.

Sung, N.-D., Park, K.-H., Jang, S.-C., and Soung, M.-K. 2007. 3D-QSAR Analysis on the Fungicidal Activity of N-phenyl-O-phenylthionocarbamate Analogues against Gray Mold (Botrytis cinerea). The Korean Journal of Pesticide Science, 11(2): 59-66.

Todeschini, R., and Consonni, V. 2000. Handbook of Molecular Properties: Wiley-VCH Verlag: Weinheim.

Tropsha, A. 2010. Best practices for QSAR model development, validation, and exploitation. Molecular informatics, 29(6-7): 476-488.

Vorčáková, K., Májeková, M., Horáková, E., Drabina, P., Sedlák, M., and Štěpánková, Š. 2018. Synthesis and characterization of new inhibitors of cholinesterases based on N-phenylcarbamates: In vitro study of inhibitory effect, type of inhibition, lipophilicity and molecular docking. Bioorganic chemistry, 78: 280-289.

Xu, J., Wang, L., Wang, L., Shen, X., and Xu, W. 2011. QSPR study of Setschenow constants of organic compounds using MLR, ANN, and SVM analyses. Journal of computational chemistry, 32(15): 3241-3252.

Xu, Z., Yang, Z., Liu, Y., Lu, Y., Chen, K., and Zhu, W. 2014. Halogen bond: its role beyond drug–target binding affinity for drug discovery and development. Journal of chemical information and modeling, 54(1): 69-78

Yanovsky, I., Finkin-Groner, E., Zaikin, A., Lerman, L., Shalom, H., Zeeli, S., Weill, T., Ginsburg, I., Nudelman, A., and Weinstock, M. 2012. Carbamate derivatives of indolines as cholinesterase inhibitors and antioxidants for the treatment of Alzheimer's disease. Journal of medicinal chemistry, 55(23): 10700-10715.

Yu, C.-C., Kearns, C. W., Metcalf, R. L., and Davies, J. 1971. Acetylcholinesterase inhibition by bis-carbamates. Pesticide Biochemistry and Physiology, 1(2): 241-247.

Yu, Q.-s., Holloway, H. W., Utsuki, T., Brossi, A., and Greig, N. H. 1999. Synthesis of novel phenserine-based-selective inhibitors of butyrylcholinesterase for Alzheimer's disease. Journal of medicinal chemistry, 42(10): 1855-1861.

Zhou, P., Huang, J., and Tian, F. 2012. Specific noncovalent interactions at protein-ligand interface: implications for rational drug design. Current medicinal chemistry, 19(2): 226-238.

دانشگاه آزاد اسلامی، واحد اراک
شاپا ٤٦٦٨-٢٠٠٨
http://jer.iau-arak.ac.ir

فصلنامه تخصصی تحقیقات حشره‌شناسی
(علمی- پژوهشی)
جلد ۱۳، شماره۲، سال ۱٤۰۰، (۵۳-۳۵)

پیش‌بینی سمیت و ضریب تقسیم اکتانول-آب آفت کش ها فنیل کاربامات کاربامات با استفاده از روش الگوریتم ژنتیک- رگرسیون چندگانه خطی

عاطفه سادات نوابی[۱] ،طاهره مومنی اصفهانی[۲] ، مجید رمضانی[۳]، محمد علیمرادی[٤]*

۱- دانشجوی دکتری، گروه شیمی، واحد اراک، دانشگاه آزاد اسلامی اراک، اراک، ایران
۲-استادیار،گروه شیمی، واحد اراک، دانشگاه آزاد اسلامی اراک، اراک، ایران
۳-دانشیار،گروه شیمی، واحد اراک، دانشگاه آزاد اسلامی اراک، اراک، ایران
٤-استادیار،گروه شیمی، واحد اراک، دانشگاه آزاد اسلامی اراک، اراک، ایران

چکیده

مطالعه ارتباط کمی ساختار-فعالیت (QSAR) مبتنی بر الگوریتم ژنتیک رگرسیون خطی چندگانه (GA-MLR) برای پیشگویی سمیت ($logIC_{50}$) و لگاریتم ضریب توزیع اکتانول-آب ($logP_{ow}$) برخی مشتقات کاربامات به عنوان آفت کش انجام شد. ساختار ترکیبات شیمیایی با نرم افزار گوسین ۹۸ و روش هارتری‌فاک و سری پایه *٦-٣١G (HF/6-31G*) بهینه شدند. توصیفگرهای مولکولی با نرم‌افزار دراگون محاسبه شد. مجموعه داده‌ها به‌طور تصادفی به دو دسته آموزش و آزمون تقسیم گردیدند. مناسب‌ترین توصیف‌گرهای با استفاده از روش الگوریتم ژنتیک وبرگشتی تعیین شدند. بهترین مدل GA-MLR با استفاده از پارامترهای آماری مانند مجذور ضریب همبستگی (R^2)، ضریب همبستگی تنظیم شده (R^2_{adj})، ریشه مربعات میانگین خطا (RMSE) برای دودسته آموزش و آزمون انتخاب گردید. بهترین مدل QSAR مبتنی بر پارامترهای آماری اعتبارسنجی تقاطعی آزمون خارجی (LOO)، پارمترهای اعتبارسنجی خارجی (Q^2_{F1}, Q^2_{F2}, Q^2_{F3}) و ضریب همبستگی تطابق (CCC) برای کیفیت توانایی پیشگویی مدل GA-MLR بررسی گردید. این نتایج نشان می دهد که مدل‌های GA-MLR می‌تواند برای پیشگویی فعالیت مشتقات کاربامات مورد استفاده قرار گیرد.

واژه‌های کلیدی:"مشتقات کاربامات"، "QSAR"؛"توصیف‌گرهای مولکولی"؛"کولین استراز"؛"$logIC_{50}$ "؛" ضریب توزیع اکتانول-آب "؛" آفت‌کش ".

* نویسنده رابط، پست الکترونیکی: *t.momeni.@iau-arak.ac.ir*

تاریخ دریافت مقاله: ۱٤۰۰/۳/۲ - تاریخ پذیرش مقاله: ۱٤۰۰/٦/۳۰