Iranian J. Math. Chem. 9 (3) September (2020) 179 − 199

Iranian Journal of Mathematical Chemistry

Journal homepage: ijmc.kashanu.ac.ir

Prediction of IC_{50} Values of 2–benzyloxy benzamide Derivatives using Multiple Linear Regression and Artificial Neural Network Methods

FARIBA MASOOMI SEFIDDASHTI, HEDAYAT HADDADI, SAEID ASADPOUR• AND SHIMA GHANAVATI NASAB

Department of Chemistry, Faculty of Sciences, Shahrekord University, P. O. Box 115, Shahrekord, Iran

ARTICLE INFO

Article History:

Received 9 February 2020
Accepted 25 March 2020
Published online 30 September 2020
Academic Editor: Boris Furtula

Keywords:

SMS2 inhibitor, benzyloxy benzamide derivatives, QSAR, Multiple Linear Regression (MLR), Artificial neural network (ANN).

ABSTRACT

In this study, six molecular descriptors were selected from a pool of variables using stepwise regression to built a QSAR model for a series of 2-benzyloxy benzamide derivatives as an SMS2 inhibitor to reduce atherosclerosis. Simple multiple linear regression (MLR) and a nonlinear method, artificial neural network (ANN), were used to modeling the bioactivities of the compounds. Modeling was carried out in total with 34 compounds of 2-benzyl oxybenzamide derivatives. PCA was used to divide the compounds into two groups of two training series and tests. The model was constructed with 27 combinations as training set, then the validity and predictive ability of the model were evaluated with the remaining 7 combinations. While the MLR provides an acceptable model for predictions, the ANN-based model significantly improves the predictive ability. In ANN model the average relative error (RE%) of prediction set is lower than 1% and square correlation coefficient (R^2) is 0.9912.

1 INTRODUCTION

Atherosclerosis is the name of a vein disease characterized by the deposition of low-density lipid and cholesterol on the inner wall of medium- and large-diameter arteries [1]. The result of this process is the formation of fibrous-fat plaques

•Corresponding author: (Email address: s.asadpour@sku.ac.ir)
DOI: 10.22052/ijmc.2020.217837.1483

(atheroma), which gradually increases with age and cause stenosis or other consequences. Arteriosclerosis is one of the leading causes of death in adults in advanced societies and countries with high levels of stress. SM or sphingomyelin is one of the most important circulating phospholipids [2, 3]. Sphingomyelin levels in human plasma have been shown to be an independent risk factor for atherosclerosis as well as a predictor for patients with acute coronary syndrome [4,5].

Inhibition of SPT, which is the first enzyme for SM biosynthesis, can dramatically reduce atherosclerosis by reducing SM. But there are many off-target side effects [6]. Therefore, inhibition of SMS2 is an alternative method of reducing SM. SMS is the last enzyme in the SM biosynthesis pathway that accelerates the conversion of Ceramide to SM. The SMS family consists of three members SMS1 and SMS2 and SMS related protein or SMSr. Studies have shown that SMS1 and SMS2 are two factors that affect sphingomyelin levels. Given the side effects of SMS1 depletion, SMS2 inhibition is an optimal strategy for reducing SM levels [6].

The process of discovering and developing new drugs based on trial and error is time consuming, difficult and costly. Another problem that plagues scientists is their lack of knowledge of the drug's activity before synthesizing and empirically investigating it, so one of the most important goals of chemists and pharmaceutical researchers is to evaluate the activity of drugs before they are manufactured. Therefore, the need for theoretical and computational methods to predict the properties or activity of pharmaceutical compounds without testing seems inevitable. The advent of chemistry has provided a solution to these problems [7−9].

One of the most successful approaches for predicting chemical properties that starts only with molecular structural information is quantitative structure-activity/properties (QSAR / QSPR) modeling. The notion that there is a close relationship between the bulk properties of the compounds and their molecular structure allows one to establish a complete relationship between the macroscopic and the microscopic properties of the material. The quantitative relationships between structure and property of mathematical equations are related to chemical structure with a wide range of physical, chemical, biological and technological properties. QSPR models, once created, can be used to predict the properties of compounds not yet measured or even unknown [10−13].

In this study, we attempted to establish a relationship between the structure of 2-benzyloxybenzamide derivatives and the pharmacological activity of these compounds as SMS2 enzyme inhibitors using (MLR) multiple linear regression [14] and (ANN) Artificial neural network [15, 16] methods. The descriptors were

selected using stepwise regression. Comparison of different linear and nonlinear methods in recent study has shown how different regression techniques affect the predictive ability of QSAR models [17]. To test the performance and stability of this model, we used different validation methods.

2. MATERIAL AND METHODS

2.1. DATASET

The dataset used in this study consisted of a series of 2-benzyloxy benzamide derivative that has been reported as SMS-2 inhibitors by Yali Li et al. [6]. 2D structure of 2-benzyloxy benzamide core is displayed in Figure 1. There were a total of 34 2-benzyloxy benzamide derivatives which are then split into a training set of 27 compounds for generating QSAR models and a test set of 7 compounds for validating the quality of the models. Remaining 1 compounds having IC_{50} value greater than 100 µM were removed. Compounds were selected randomly by Minitab with using PCA. All the structures and associated inhibitory activities are listed in Table 1.

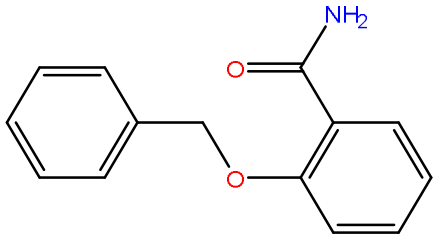


Figure 1. The 2D chemical structure of 2-benzyloxy benzamide core.

2.2. MOLECULAR MODELING

All the 2D & 3D structures were drawn and built by ChemDraw and Chem3D softwares respectively. Structures were optimized by MM2 algorithm in Chem3D.

2.3. 2D-QSAR METHODOLOGY

2.3.1. CALCULATION OF MOLECULAR DESCRIPTORS

Descriptors are the mathematics of a molecule that contains it Different sources of chemical information have been converted and encoded to counter chemicals, Biological and pharmaceutical problems. To develop QSAR 2D models, Different physicochemical descriptors are calculated for each of the compounds in the dataset using DRAGON software. Dragon converts the information of molecules

including bond energy, bond angle, bond type, molecular mass, electronic properties, and so on into numeric form and stores them in descriptive format.

2.3.2. FEATURE SELECTION

Feature Selection methods have been employed for selecting the best Descriptors among the Many Descriptors Containing Low Information for Model Construction or are correlated with other descriptors without incurring much loss of information. In this study, three methods were used to reduce descriptors. Initially, among the pair of descriptors with a correlation coefficient above 0.95, one was eliminated by the Dragon software. Dragoon reduced the number of 3224 calculated descriptors to 1419. Then descriptors that had constant or zero values that could not correlate the difference in structure to the difference in activity were removed. The number of remaining descriptors was thus reduced to 1152 descriptors.

The remaining descriptors, along with the activity function values, were entered into SPSS software. The important descriptors are selected under Stepwise approach. In the stepwise strategy, a multiple-linear equation was built step by step. First, an initial model was determined and then it was repeatedly changed by removing or adding a predictor variable based on stepping criteria for inclusion and exclusion. At every step, all variables were specified and evaluated to assign important descriptors. The software offered 10 models using stepwise regression Method. As can be seen from this table, as the number of descriptors increases, the validity of the model is improved and the statistical parameters corresponding to it are improved. But since in the QSAR studies the appropriate model is a model with the lowest number of descriptors to obtain the best fit, also with respect to the values of the RMSE, Q^2, R^2 and R^2adj statistical parameters finally, 6 descriptors was chosen as the final descriptors related to SMS-2 enzyme inhibition. The final descriptors given in Table 2. For the appointed models, the values of the RMSE, Q^2, R^2 and R^2adj parameters are calculated as shown in Table 2, since in MLR analysis the number of compounds in the samples should be at least 5 times the number of descriptors and the descriptors should be orthogonal values [18, 19].

According to the results shown in Table 2 and Due to the slope of this parameters changes, model 6 with 6 descriptors was selected as the top model and modeling was performed with 6 descriptors. The characteristics of the descriptors used in this study are presented in Table 3, as well as the values of these descriptors in Table 4. The selected descriptors should be independent of each other because in their high dependence only the descriptor with a higher correlation with the dependent variable is included in the model. Two-way correlation coefficient of descriptors was calculated by SPSS software and is

presented in Table 5. The results showed that the behavior of the selected descriptors was independent. As you can see, there is little connection between the descriptors.

3.1. DATA SELECTION

First, the data set that consisted 34 compounds were divided into a training set of 27 compounds and test set of 7 compounds with ratio 80% and 20% , respectively. Compound number 1, 2, 9, 18, 19, 27, 29 were selected as test set and rest of the 34 compounds as train set. In this study split of data set was done with Principal Component Analysis (PCA). The equation must use the minimum number of descriptors to obtain the best fit. To achieve this a Stepwise regression method is used to find out the best number of descriptors. Among the models given with the SPSS, after the sixth model no considerable improvement in regression coefficient (R^2) values has been observed.

3.2. MULTIPLE LINEAR REGRESSION (MLR)

MLR is one of the statistical methods that tries to establish a linear relationship between dependent variables and response variable. MLR method provides equation linking the structural features to the dependent variable (y) of the compounds:

$$\mathbf{y} = a_0 + a_1\boldsymbol{d}_1 + \cdots + a_n\boldsymbol{d}_n \quad (1)$$

where the intercept (a_0) and the regression coefficients of the descriptors (a_i) are determined by using the least-squares method. $\boldsymbol{d}_i$ has the common definition, variable or descriptor in this case, the elements of this vector are equivalent numerical values of a 3D structures of the molecules or structural descriptors.

In the present study, SPSS 23 was used to calculate MLR models. Stepwise-MLR method was used as multiple linear regression to select appropriate and important descriptors for training and test sets. This method has proven to be a very useful computational method in data analysis problems[20, 21].

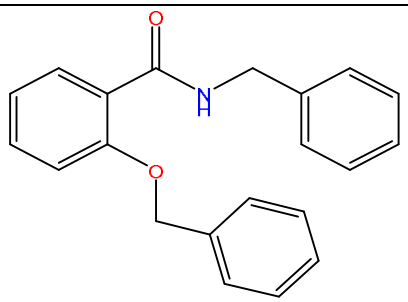
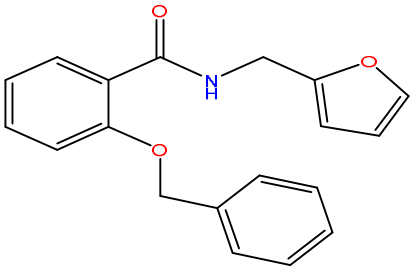
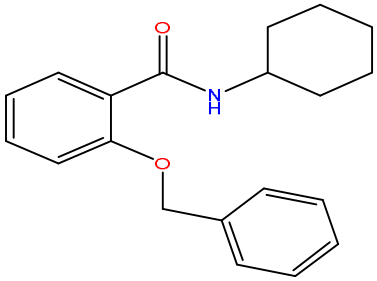
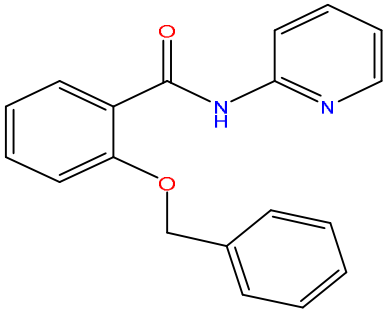
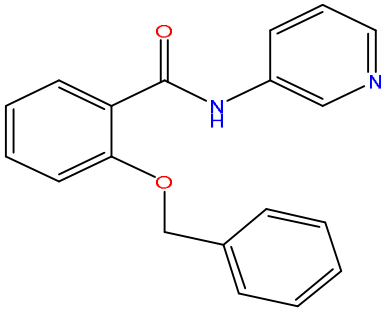
The best QSAR model built using multiple linear regression (MLR) method is represented by the following equation:

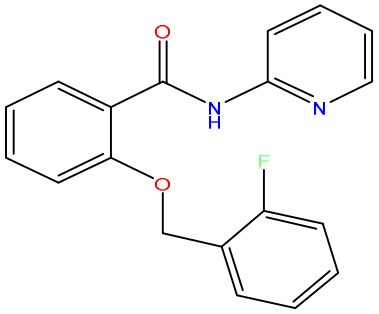
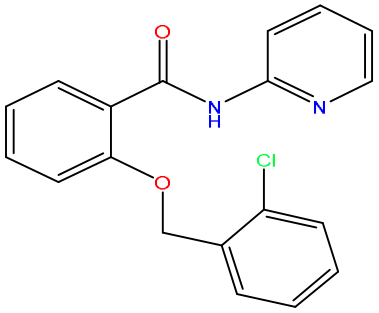
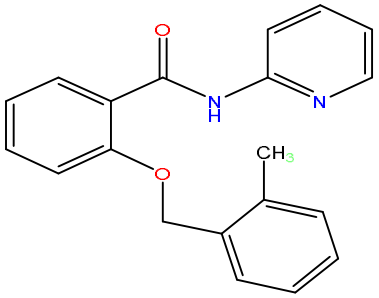
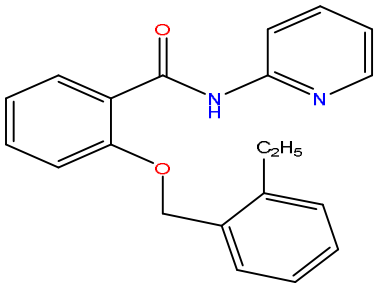
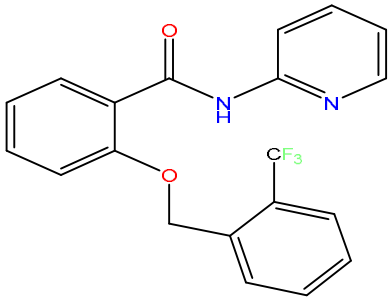
$$y = 8.2 + 9.2MATA8m + 9.4GATS8p + -6Mor30m + 3.7G2u + 6.9Mor06m + 3.98\ G3e \quad (2)$$

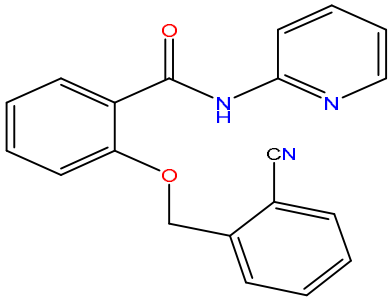
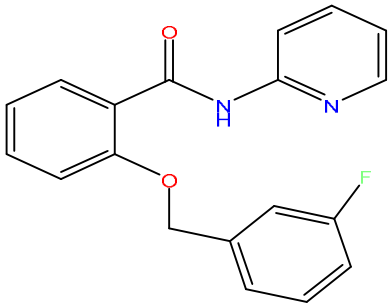
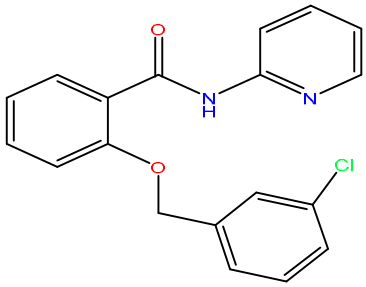
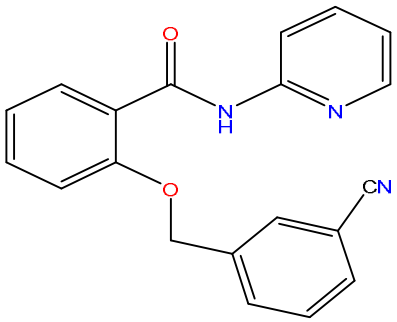
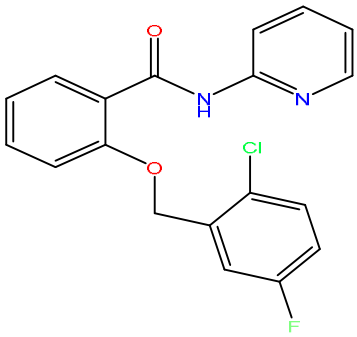
$$N=34; \quad R^2=0.767; \quad RMSE=7.46; \quad R^2_{CV}=0/766 \quad (3)$$

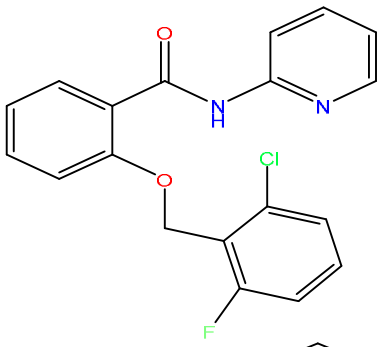
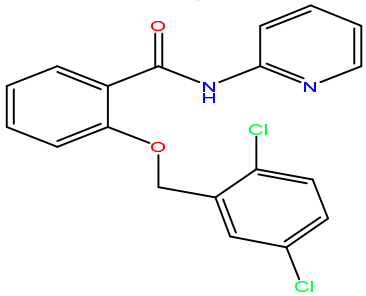
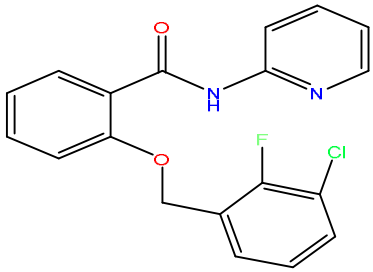
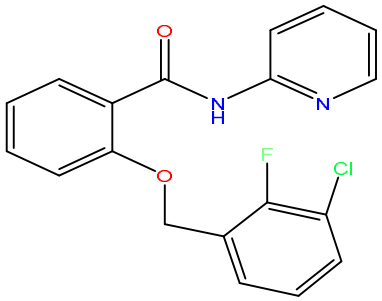
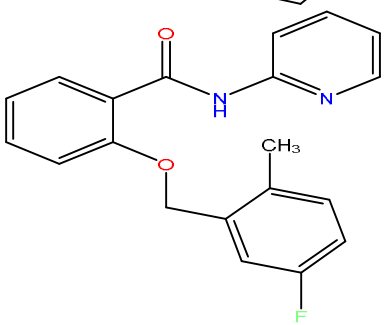
The values of the six descriptors and their corresponding correlation matrix are shown in tables 4 and 5.

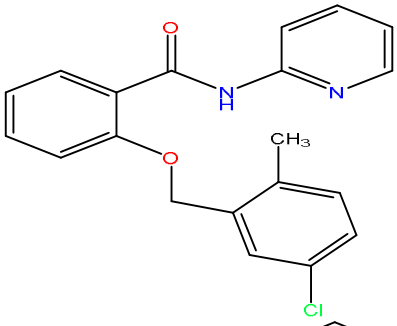
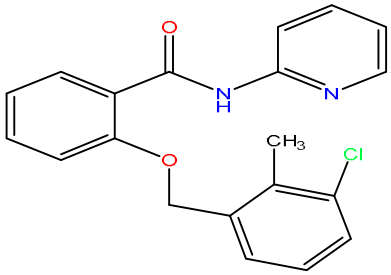
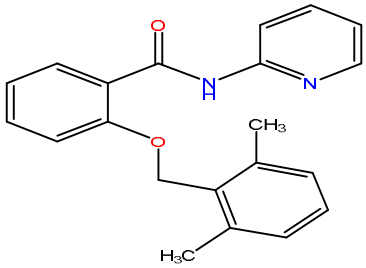
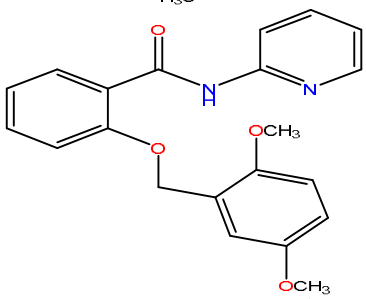
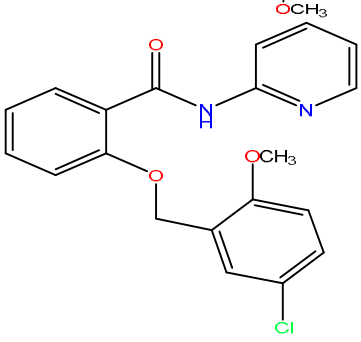
Table 1. Structural formulae of compounds and their IC_{50} values.

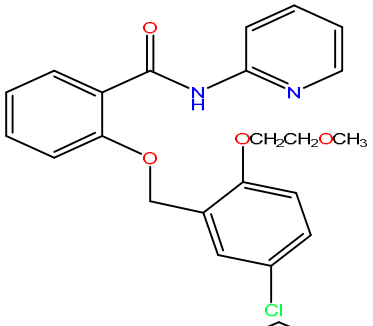
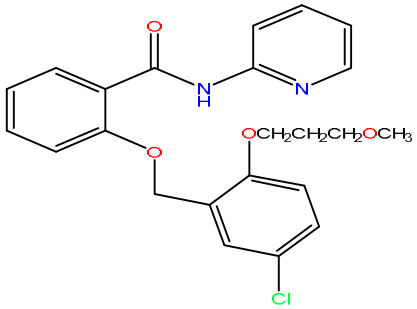
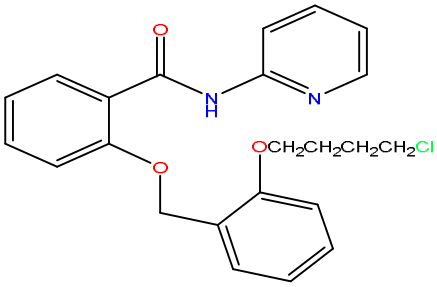
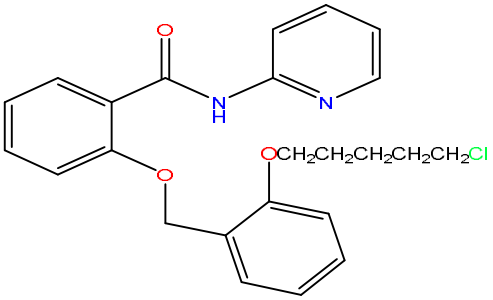
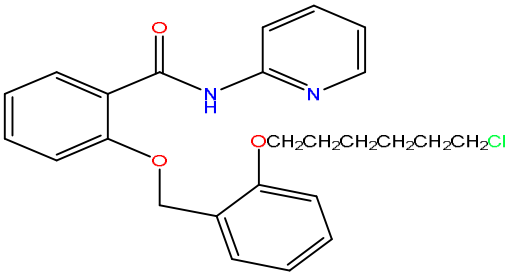
NUMBER	STRUCTURE	IC_{50}(μM)
1		34.2
2		49
3		43.4
4		11.7
5		3.2

6		3.5
7		1.1
8		1.5
9		0.88
10		2.1

11		31.8
12		1.6
13		1.4
14		60.7
15		0.74

16		1.5
17		0.99
18		0.69
19		3.1
20		0.74

21		1.1
22		3.8
23		0.89
24		2.8
25		0.67

26		11.6
27		4.1
28		1.3
29		0.67
30		0.47

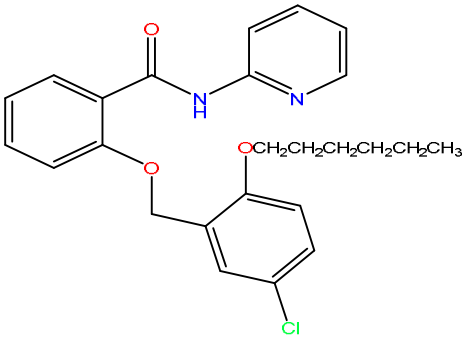
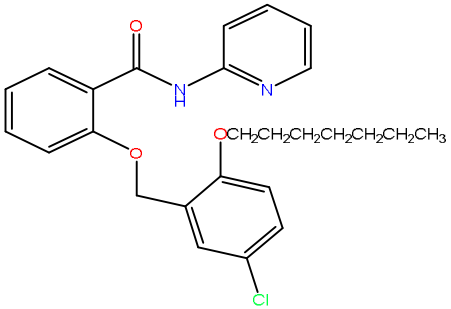
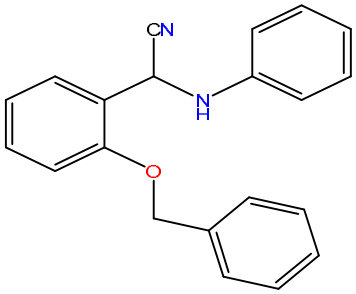
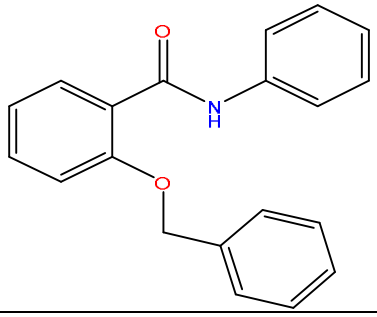
31		0.52
32		0.43
33		13.5
34		13.4

Table 2. R^2, RMSE, Q^2, R^2adj values for models with different number of descriptors.

# Descriptors	R^2	RMSE	Q^2	R^2adj
1	0.324	13.2	0.278	0.302
2	0.587	10.14	0.569	0.56
3	0.647	9.38	0.632	0.611
4	0.705	8.77	0.677	0.664
5	0.76	8.53	0.695	0.717
6	0.81	7.46	0.766	0.767
7	0.853	6.6	0.816	0.814
8	0.882	6.17	0.998	0.844

Table 3. Descriptors used in the 2D-QSAR study.

Descriptor Types	Descriptor blocks type	Descriptor description
MATS8m	2D autocorrelation	Moran autocorrelation-lag8 / weighted by atomic masses
GATS8p	2Dautocorrelation	Geary autocorrelation – lag8 / weighted by atomic polarizabilities
Mor30m	3DMoRSE	3D-MORSE – signal 30 / weighted by atomic masses
G2u	WHIM descriptors	2st component symmetry directional WHIM index / unweighted
Mor06m	3DMoRSE	3D-MORSE – signal 06 / weighted by atomic masses
G3e	WHIM descriptors	3st component symmetry directional WHIM index weighted by atomic Sanderson electronegativities

Table 4. Values of the obtained parameters of the studied derivatives of 2-benzyloxy benzamide.

Number	MATS8m	GATS8p	Mor30m	G2u	Mor06m	G3e
1	-0.577	0.165	-0.976	2.938	-0.923	0.205
2	3.228	-0.701	-0.504	-1.575	-0.980	1.184
3	1.642	1.478	0.248	0.176	0.029	1.184
4	-0.758	0.537	-0.854	-0.160	-0.362	0.874
5	-0.453	0.141	-0.574	0.985	-0.515	-0.826
6	-1.949	1.395	-2.559	-0.901	-1.720	-0.053
7	-1.070	0.660	-0.180	0.917	-1.513	-0.208
8	-1.086	0.802	-1.440	-0.632	-1.338	-0.414
9	-0.462	-0.622	-0.163	-0.632	0.031	0.874
10	-0.199	-0.488	-1.553	-0.363	-1.570	-0.517
11	1.092	1.589	0.449	0.176	-0.598	0.874
12	1.535	0.335	1.997	-0.632	-0.315	-0.414
13	0.007	0.873	0.755	-0.632	0.712	-1.135
14	1.001	2.622	0.257	0.176	0.867	1.184
15	-0.281	-1.014	-0.469	1.052	0.925	0.514
16	-1.210	1.111	0.117	0.648	-0.805	-0.053
17	-0.371	0.050	0.580	-1.306	0.635	0.565

Table 4 (Continued).

Number	MATS8m	GATS8p	Mor30m	G2u	Mor06m	G3e
18	-0.371	0.050	0.108	-0.565	1.115	-0.620
19	-1.119	1.221	0.554	-0.093	-0.159	-0.259
20	1.503	-0.531	1.744	0.042	-0.402	0.102
21	-0.133	-1.247	1.123	0.513	1.047	1.029
22	0.623	-0.436	-1.055	-1.508	0.892	-1.547
23	-0.174	-1.635	0.257	-0.497	0.961	-0.774
24	0.015	-0.974	1.140	0.783	1.145	-1.650
25	0.475	-1.140	0.782	-1.238	0.913	0.668
26	0.861	-1.219	1.088	0.715	1.546	-0.980
27	-0.117	-1.208	0.974	0.715	1.618	-1.393
28	-0.084	-1.196	0.904	-0.834	1.887	-1.135
29	-0.256	0.137	0.458	1.254	-0.794	1.699
30	-0.938	0.588	-1.186	-0.767	-0.881	-1.187
31	-0.125	-0.203	-0.040	-0.565	0.323	1.596
32	-0.289	-0.254	-1.212	1.928	-0.833	-1.135
33	0.163	-0.685	-0.653	0.648	-0.291	-0.053
34	-0.125	-0.203	-0.119	-0.767	-0.646	1.802

Table 5. Correlation matrix between different obtained descriptors.

	MATS8M	GATS8P	MOR30M	G2U	MOR06M	G3E
MATS8M	1					
GATS8P	-0.13286	1				
MOR30M	0.429756	-0.23036	1			
G2U	-0.17285	0.00051	0.104558	1		
MOR06M	0.198226	-0.45511	0.578589	-0.03966	1	
G3E	0.217833	0.235266	0.031318	-0.08778	-0.22842	1

3. RESULTS AND DISCUSSION

The predicted values of IC_{50} of the train data set, using this model is plotted against experimental values and is shown in Figure 1. The above linear model was used to predict the 7external test dataset which was never used in descriptor selection or model building. The result show an R^2= 0.767, Q^2=0.642 and RMSE =7.46. The predicted values of IC_{50} of the train and test set using the MLR equation is given in the Table 6.

Table 6. Observed values and calculated values of IC_{50} according to MLR methods.

Number	IC_{50} Observed	IC_{50} predicted	Residual	Relative error%
1t	34.20	15.65	18.55	54.24
2t	49.00	26.45	22.55	46.03
3a	43.40	41.33	2.07	4.78
4a	11.70	11.79	-0.09	-0.79
5a	3.20	5.61	-2.41	-75.31
6a	3.50	3.32	0.18	5.26
7a	1.10	-2.22	3.32	302.03
8a	1.50	1.16	0.34	22.71
9t	0.88	0.42	0.46	52.05
10a	2.10	-3.15	5.25	249.76
11a	31.80	30.53	1.27	3.98
12a	1.60	7.34	-5.74	-358.99
13a	1.40	10.00	-8.60	-614.26
14a	60.70	51.92	8.78	14.46
15a	0.74	11.22	-10.48	-1416.00
16a	1.50	3.46	-1.96	-130.35
17a	0.99	3.56	-2.57	-259.73
18t	0.69	7.74	-7.05	-1021.71
19t	3.10	3.60	-0.50	-16.16
20a	0.74	4.36	-3.62	-488.72
21a	1.10	1.73	-0.63	-56.90
22a	3.80	10.55	-6.75	-177.72
23a	0.89	-8.64	9.53	1070.34
24a	2.80	-3.44	6.24	222.99
25a	0.67	1.51	-0.84	-126.03
26a	11.60	7.54	4.06	34.99
27t	4.10	-1.81	5.91	144.25
28a	1.30	-3.84	5.14	395.75
29t	0.67	10.33	-9.66	-1441.55
30a	0.47	-1.44	1.91	405.87
31a	0.52	11.88	-11.36	-2184.04
32a	0.43	7.30	-6.87	-1598.26
33a	13.50	7.34	6.16	45.61
34a	13.40	5.73	7.67	57.26

[t] test set
[a] training set

3.2. ARTIFICIAL NEURAL NETWORKS (ANN)

Artificial Neural Networks (ANN) is a computer-based system that originates from a simple brain model, the performance of networks is very different from the performance of multi-linear regression. In ANN, there is no final regression equation showing the relationship between independent and dependent variables. The fitting model is determined by learning, unsupervised or unsupervised, such as the re-propagation algorithm. In general, a neural network architecture consists of

multiple layers, an input layer, an output layer, and a number of hidden layers of each neuron.

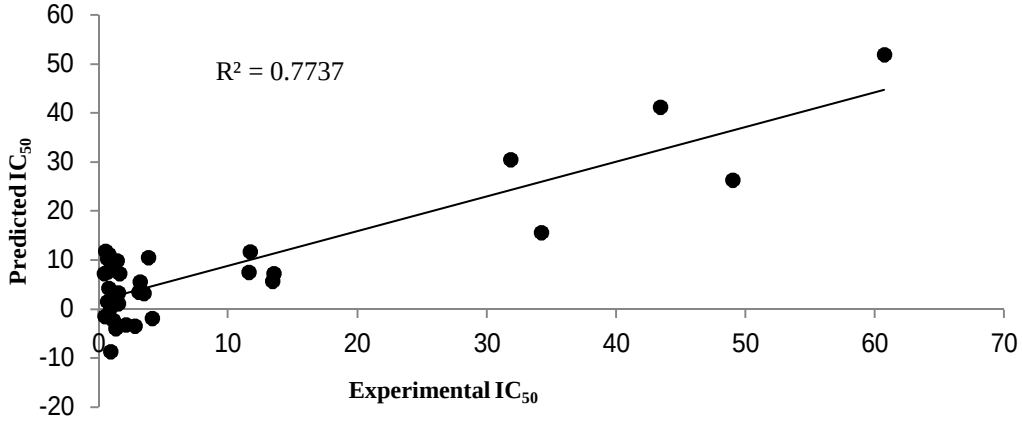


Figure 2. Predicted IC_{50} activities by MLR in comparison with experimental.

In the network, it connects to each node of the connection layers and is influenced by the amount of weights affected by the units connected to it. During the random weight training and initial random crash, adjustments are made to find the minimum difference between the output value and the target value. After a sufficient number of training iterations, the ANN learns to recognize patterns in the data, so it can be used for predicting new input values [22-26]. The networks used in this study consisted of three layers consisting of an input layer, a hidden layer, and an output layer. The input nodes contain five parameters in the regression equation and one constant. The output neuron refers to the retention index. Before entering the neural network, input data were stored at a ratio of 0 to 1. IC_{50} values were also used with this rule. Sigmoid transfer functions were applied in all layers. The weights were adjusted through a back propagation algorithm to correct the model behavior. This computer program is designed to generate the desired number of neurons in the hidden layer. In order to select the optimal model, different topological networks with different hidden units were performed. On the other hand, the values of learning factor, coefficient of movement, and core values of weight and bias were tested to find the best performance and fastest convergence.

$N=34;\ R_{train}=1;\ R_{test}=1;\ R_{validation}=0.995;\ R_{all}=0.99999;\ R^2_{CV}=0.99998$ (4)
RMSE=15/4076;

The predicted values of IC_{50} of the train and test set using the ANN Model is given in the Table 7.

Table 7. Observed values and calculated values of IC_{50} according to ANN methods.

Number	IC_{50} Observed	IC_{50} predicted	Residual	Relative error%
1a	34.2	34.16	0.04	0.13
2a	49	49.0	0.00	0.00
3a	43.4	43.39	0.07	0.02
4a	11.7	11.69	0.01	0.08
5a	3.2	3.20	0.00	0.00
6a	3.5	3.50	0.00	0.00
7t	1.1	1.14	-0.04	-3.36
8v	1.5	1.17	0.32	21.73
9a	0.88	0.88	0.01	0.10
10t	2.1	2.09	0.01	0.33
11t	31.8	31.80	-0.01	-0.02
12t	1.6	1.68	-0.08	-5.27
13a	1.4	1.40	-0.01	-0.11
14t	60.7	60.70	0.00	0.00
15a	0.74	0.74	0.00	-0.01
16a	1.5	1.50	0.00	0.02
17a	0.99	0.99	-0.01	-0.10
18v	0.69	0.68	0.01	1.66
19v	3.1	2.95	0.15	4.90
20a	0.74	0.74	0.00	-0.02
21a	1.1	1.10	0.01	0.04
22v	3.8	3.76	0.03	0.91
23a	0.89	0.89	-0.01	-0.12
24a	2.8	2.80	-0.01	-0.02
25a	0.67	0.67	0.00	0.02
26a	11.6	11.59	0.01	0.03
27a	4.1	4.10	0.00	-0.01
28a	1.3	1.30	-0.01	-0.04
29a	0.67	0.67	0.00	0.01
30v	0.47	0.48	-0.01	-1.45
31a	0.52	0.52	0.00	0.00
32a	0.43	0.43	0.00	0.00
33a	13.5	13.50	0.00	0.00
34a	13.4	13.44	-0.04	-0.27

[t] test set
[v] validation set
[a] training set

Finally, the main results of these two models are shown in Table 8. We evaluated the best QSAR equations created in this study. Based on these results, comparing the quality of the MLR model shows that the ANN model is significantly more predictive because the ANN approach has better results than the MLR. As can be seen from this table, ANN establishes a favorable relationship between the molecular descriptors and the activity of the compounds studied.

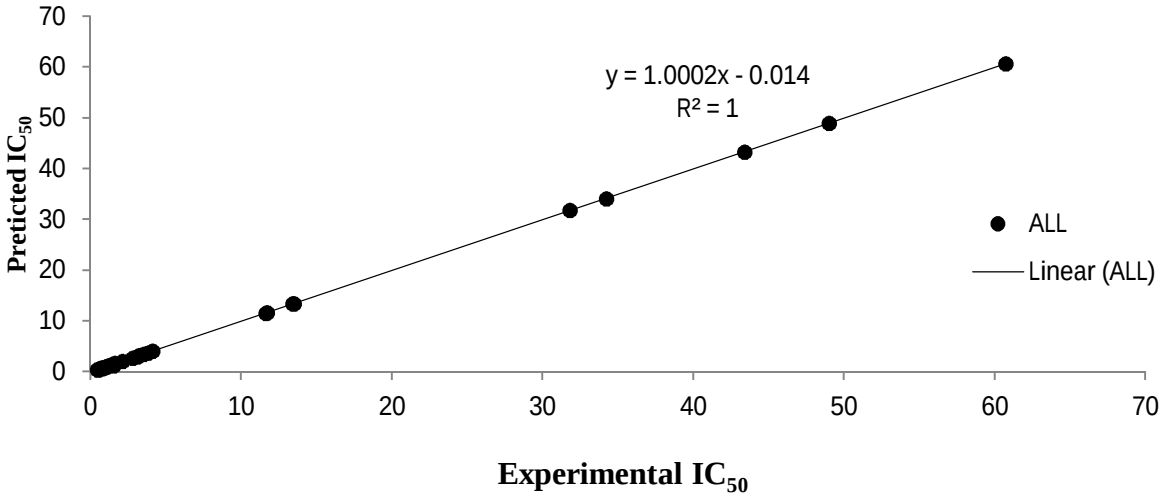


Figure 3. Predicted IC_{50} activities by ANN in comparison with experimental.

4. CONCLUSION

In the present study, quantitative analysis of the structure-activity relationship (QSAR) was performed on 34 molecules. The QSAR model was developed using multiple linear regression (MLR) and neural network (ANN) paradigms. Stepwise regression method was used to select the most significant descriptors. Six types of descriptors were used to construct the MLR model and the neural network for 2-benzyloxy benzamide derivatives, which include MATS8m, GATS8p, MOR30m, G2u, MOR06m and G3e. While the MLR provides an acceptable model for predictions, the ANN-based model significantly improves the predictive. It provides the best results among those we have tested. To compare the accuracy and predictability of the models presented, key statistical indicators such as R^2 and RMSE are presented in different models using different statistical tools and descriptions. To compare the results of the two models, Table 7, is presented.

ACKNOWLEDGMENT. The authors appreciate faculty of sciences, department of Chemistry, Shahrekord University.

Table 8. Comparing values of IC_{50} experimental and predicted results using MLR & ANN methods.

Number	$IC_{50\ observed}$	$MLR_{\ predicted}$	Relative error%	$ANN_{\ predicted}$	Relative error%
1	34.2	15.65	54.24	34.16	0.13
2	49	26.45	46.03	49.00	-0.01
3	43.4	41.33	4.78	43.39	0.02
4	11.7	11.79	-0.79	11.69	0.08
5	3.2	5.61	-75.31	3.199	0.04
6	3.5	3.32	5.26	3.50	-0.02
7	1.1	-2.22	302.03	1.137	-3.36
8	1.5	1.16	22.71	1.17	21.73
9	0.88	0.42	52.05	0.88	0.10
10	2.1	-3.14	249.76	2.09	0.33
11	31.8	30.53	3.98	31.80	-0.02
12	1.6	7.34	-358.99	1.68	-5.27
13	1.4	10.00	-614.26	1.40	-0.11
14	60.7	51.92	14.46	60.70	0.00
15	0.74	11.22	-1416.00	0.74	-0.01
16	1.5	3.45	-130.35	1.50	0.02
17	0.99	3.56	-259.73	0.99	-0.10
18	0.69	7.74	-1021.71	0.68	1.66
19	3.1	3.60	-16.16	2.95	4.90
20	0.74	4.38	-488.72	0.74	-0.02
21	1.1	1.73	-56.90	1.10	0.04
22	3.8	10.55	-177.72	3.76	0.91
23	0.89	-8.64	1070.34	0.89	-0.12
24	2.8	-3.44	222.99	2.80	-0.02
25	0.67	1.52	-126.03	0.67	0.02
26	11.6	7.54	34.99	11.59	0.03
27	4.1	-1.81	144.25	4.10	-0.01
28	1.3	-3.84	395.75	1.30	-0.04
29	0.67	10.33	-1441.55	0.67	0.01
30	0.47	-1.44	405.87	0.48	-1.45
31	0.52	11.88	-2184.04	0.52	-0.70
32	0.43	7.30	-1598.26	0.43	-0.86
33	13.5	7.34	45.61	13.50	-0.01
34	13.4	5.73	57.26	13.44	-0.27

REFERENCES

[1] J. W. Gofman, F. Lindgren, H. Elliott, W. Mantz, J. Hewitt, B. Strisower, V. Herring, T. P. Lyon, The role of lipids and lipoproteins in atherosclerosis, *Science* **111** (1950) 166−186.

[2] J. Liu, H. Zhang, Z. Li, T. K. Hailemariam, M. Chakraborty, K. Jiang, D. Qiu, H. H. Bui, D. A. Peake, M. S. Kuo, Sphingomyelin synthase 2 is one of the determinants for

plasma and liver sphingomyelin levels in mice, *Arterioscler. Thromb. Vasc. Biol.* **29** (2009) 850−856.
[3] S. G. Nasab, A. Semnani, F. Marini, A. Biancolillo, Prediction of viscosity index and pour point in ester lubricants using quantitative structure-property relationship (QSPR), *Chemom. Intell. Lab. Syst.* **183** (2018) 59−78.
[4] Å. Nilsson, R. D. Duan, Absorption and lipoprotein transport of sphingomyelin, *J. Lipid Res.* **47** (2006) 154−171.
[5] A. Schlitt, S. Blankenberg, D. Yan, H. von Gizycki, M. Buerke, K. Werdan, C. Bickel, K. J. Lackner, J. Meyer, H. J. Rupprecht, Further evaluation of plasma sphingomyelin levels as a risk factor for coronary artery disease, *Nutr. Metab.* **3** (2006) 5.
[6] Y. Li, T. Huang, B. Lou, D. Ye, X. Qi, X. Li, S. Hu, T. Ding, Y. Chen, Y. Cao, Discovery, synthesis and anti-atherosclerotic activities of a novel selective sphingomyelin synthase 2 inhibitor, *Eur. J. Med. Chem.* **163** (2019) 864−882.
[7] A. Habibi-Yangjeh, E. Pourbasheer, M. Danandeh-Jenagharad, Application of principal component-genetic algorithm-artificial neural network for prediction acidity constant of various nitrogen-containing compounds in water, *Monatsh. Chem.* **140** (2009) 15−27.
[8] A. Beheshti, E. Pourbasheer, M. Nekoei, A. Banaei, Quantitative structure-activity relationship study of amino acid derivatives as histone deacetylase inhibitors using the genetic algorithm–Multiple linear regression, *Anal. Chem. Lett.* **2** (2012) 33−43.
[9] M. Adimi, M. Salimi, M. Nekoei, E. Pourbasheer, A. Beheshti, A quantitative structure-activity relationship study on histamine receptor antagonists using the genetic algorithm-multi-parameter linear regression method, *J. Serb. Chem. Soc.* **77** (2012) 639−650.
[10] M. Ravi, A. J. Hopfinger, R. E. Hormann, L. Dinan, 4D-QSAR analysis of a set of ecdysteroids and a comparison to CoMFA modeling, *J. Chem. Inf. Comput. Sci.* **41** (2001) 1587−1604.
[11] B. Luke, Comparison of three different QSAR/QSPR generation techniques, *J. Mol. Struct. THEOCHEM* **468** (1999) 13−20.
[12] P. Bruneau, Search for predictive generic model of aqueous solubility using Bayesian neural nets, *J. Chem. Inf. Comput. Sci.* **41** (2001) 1605−1616.
[13] A. R. Katritzky, R. Petrukhin, D. Tatham, S. Basak, E. Benfenati, M. Karelson, U. Maran, Interpretation of quantitative structure− property and− activity relationships, *J. Chem. Inf. Comput. Sci.* **41** (2001) 679−685.
[14] P. C. Jurs, Computer software applications in chemistry, John Wiley & Sons, 1996.
[15] W. S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, *Bull. Math. Biol.* **5** (1943) 115−133.
[16] S. Asadpour, S. Jazayeri Farsani, S. Ghanavati Nasab, A. Semnani, QSAR Modeling of Some Derivatives of Thiazolidinedione With Antimalarial Properties, *Frontiers in Chemical Research*, **1** (2019) 17−24.
[17] S. Wold, L. Eriksson, S. Clementi, Statistical validation of QSAR results, Chemometric methods in molecular design, Weinheim 1995.

[18] P. Zhou, H. Mei, F. Tian, J. Wang, S. Wu, Z. Li, A new two-dimensional approach to quantitative prediction for collision cross-section of more than 110 singly protonated peptides by a novel moecular electronegativity-interaction vector through quantitative structure-spectrometry relationship studies, *Front. Chem. China*, **2** (2007) 55−63.
[19] M. Randic, Resolution of ambiguities in structure-property studies by use of orthogonal descriptors, *J. Chem. Inf. Comput. Sci.* **31** (1991) 311−320.
[20] M. Zakariazadeh, A. Barzegar, S. Soltani, H. Aryapour, Developing 2D-QSAR models for naphthyridine derivatives against HIV-1 integrase activity, *Med. Chem. Res.* **24** (2015) 2485−2504.
[21] A. Afantitis, G. Melagraki, H. Sarimveis, P. A. Koutentis, O. Igglessi-Markopoulou, G. Kollias, A combined LS-SVM & MLR QSAR workflow for predicting the inhibition of CXCR3 receptor by quinazolinone analogs, *Mol. Diversity* **14** (2010) 225−235.
[22] J. Gasteiger, J. Zupan, Neural networks in chemistry, *Angew. Chem. Int. Ed. Engl.* **32** (1993) 503-527.
[23] D. W. Salt, N. Yildiz, D. J. Livingstone, C. J. Tinsley, The use of artificial neural networks in QSAR, *Pestic. Sci.* **36** (1992) 161−170.
[24] T. Aoyama, Y. Suzuki, H. Ichikawa, Neural networks applied to pharmaceutical problems. III. Neural networks applied to quantitative structure-activity relationship (QSAR) analysis, *J. Med. Chem.* **33** (1990) 2583−2590.
[25] W. Guo, P. Zhu, H. Brodowsky, The study for optimization of chromatographic condition by means of artificial neural networks, *Talanta* **44** (1997) 1995−2001.
[26] W. Guo, Y. Lu, X.M. Zheng, The predicting study for chromatographic retention index of saturated alcohols by MLR and ANN, *Talanta* **51** (2000) 479−488.