Journal of Physical and Theoretical Chemistry
of Islamic Azad University of Iran, 18 (2) 1-16: Summer 2021
(J. Phys. Theor. Chem. IAU Iran)
ISSN 1735-2126

Biological Activity and Lipophilicity Study of Several Doxazolidines as Cancer Cells Inhibitors

Maria Nikkar[1], Robabeh SayyadikordAbadi[*1], Asghar Alizadehdakhel[1], Ghasem Ghasemi[1]

[1] Department of Chemistry and Chemical Engineering, Rasht Branch, Islamic Azad University, Rasht, Iran

Received July 2021; Accepted November 2021

ABSTRACT

QSAR investigations of lipophilicity (XLOGP) and biological activity (IC50) values of some Doxazolidine derivatives were conducted using combinations of multiple linear regression (MLR) and artificial neural network (ANN) modeling methods and three different optimization techniques including simulated annealing (SA), genetic algorithm (GA) and Imperialist Competitive algorithm (ICA). In addition CORAL software was used to correlate the lipophilicity and biological activity to the structural parameters of the drugs. The obtained results were compared and GA-ANN and ICA-MLR combinations showed the best performance with regard to the correlation coefficient (R^2) and root-mean-square error (RMSE). The most effective descriptors extracted from lipophilicity and biological activity studies were presented and discussed. From GA-ANN method, the most important physico-chemical descriptors were found to be minimum value in atomic Sanderson electronegativities and maximum value in Squared Moriguchi Octanol-Water partition coeff. (log P $^{\wedge 2}$) descriptors. ICA-MLR method suggests the maximum value in polarizibility, electrotopological state and atom van der Walls volume as the most important physicochemical descriptors. It was concluded that QSAR study and Monte Carlo method can lead to a more comprehensive understanding of the relation between physico-chemical, structural or theoretical molecular descriptors of drugs to their biological activities and Lipophilicity.

Keywords: GA and ICA Algorithms, Doxazolidine derivatives, QSAR, Monte Carlo Method

1. INTRODUCTION

Doxazolidine reacts with DNA at 5′-NGC-3′ sites in the minor groove to form a virtual crosslink of the DNA strands, where N is an unspecified base [1]. Doxorubicin is an anthracycline antibiotic [2] that reacts with formaldehyde to form doxazolidine. Doxazolidine and doxoform inhibit human cancer cell line growth at 200-fold lower concentration than doxsaliform, a prodrug to an acyclic

[*]Corresponding author: sayyadi_04@yahoo.com, sayyadi@iaurasht.ac.ir

doxorubicin-formaldehyde conjugate [3].

QSAR study is modelling and optimization approaches that relates the descriptors (constitutional, geometrical, topological, quantum chemical, etc.) to the Maximal Inhibitory Concentration and Lipophilicity of drugs [4,5].

The partition coefficient (P) is a comparison of the solubilities of the solute in two liquids: water and hydrophobic solvent, such as 1-octanol [6,7]. The partition coefficient indicates the hydrophilicity (water-loving) or hydrophobicity (water-fearing) of drugs and is useful in estimating their distribution within the body [8, 9]. The logarithm of the partition coefficient between n-octanol and water, also referred to as logP, has been widely used in quantitative structure-activity relationship (QSAR) studies as a key parameter for characterizing lipophilicity. LogP was computed as a function of molecular physicochemical properties, such as molecular surfaces, volumes, dipoles, partial charges and HOMO/LUMO energies in the Property-based methods. XLOGP3 is a new method for the fast calculation of logP [10, 11]. XLOGP3 is an additive model and a total of 83 basic atom types are implemented in it to classify carbon, nitrogen, oxygen, sulfur, phosphorus and halogen atoms. This model also uses two correction factors. the first correction factor accounts for internal hydrogen bonds, which makes a molecule less hydrophilic than what is indicated by its chemical structure. The second correction factor is used for organic compounds with amino acid moieties [45]. The biological activity of a chemical compound is directly related to its physical properties. Lipophilicity is the most useful parameter in the assessment of biological activity of substances and in the predicting their toxic activity. Lipophilicity is described by partition processes taking place between two phases: non-polar (organic) and polar (typically water) [12].

Imperialist Competitive Algorithm (ICA) is a new population-based optimization algorithm that was proposed by Atashpaz-Gargari and Lucas in 2007 [13] and since then it was employed in solving a variety of optimization problems [14]. The algorithm, starts with an initial population. The individuals (countries) are two type: imperialists and colonies. The most powerful countries are selected as imperialists and the rest as the colonies of these imperialists. The total power of an empire depends on both power of the imperialist country and power of its colonies [15].

CORAL has been proposed as competent software for the QSAR studies. It uses Monte Carlo method to find the most important simplified molecular input-line entry system (SMILES)-based descriptors and calculate their correlation weights to predict an endpoint (e.g., pIC50). SMILES are lines of symbols, representing the molecular structure [16, 17].

In the current study, MLR and ANN modelling tools coupled with SA and GA [18-21] optimization techniques and Monte Carlo method were used to find the best set of descriptors for the prediction of inhibitory activity and lipophilicity of Doxazolidine derivatives, and the results were compared. In addition, formation of complexes of the chosen doxazolidine derivative with DNA at binding sites was investigated.

2. THEORY AND COMPUTATIONAL METHODS

2.1. Linear and Non-Linear Methods

Geometry optimizations of Doxazolidine compounds were carried out using the B3lyp/6-311g** at the Gaussian 03W [22] and then Dragon program was used for calculation of 3226 molecular

descriptors for each of the 25 compounds and then SPSS program [23] was used to reduce the number of descriptors [24] to 1309 and 1408 for the dependent variables of biological activity (pIC50) and lipophilicity (XLOGP), respectively. Then a stepwise multiple linear regression procedure, was employed to select the best descriptors of the 1309 and 1408 descriptors. Low standard deviation, least numbers of independent variables, high ability for prediction, high correlation coefficient (R^2) and RMSE are characteristics of an ideal model [25]. The RMSE and R^2 are defined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y_o)2}{n}} \qquad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_l)^2}{\sum_{i=1}^{n}(y_i - y_{avg})^2} \qquad (2)$$

Where, y_i, y_o, y_l and y_{avg} are the desired output, the predicted value by the model, values obtained using QSAR method and average value of all the values, respectively. Also, n is the number of molecules in the data set, respectively [12].

In QSAR methods which ANN was utilized as the modelling tool, 1309 and 1408 descriptors were fed into the input layer of the ANNs. All of the designed neural networks were three-layer and Levenberg-Marquart algorithm [19] was applied for training of the netwoks. Modelling and optimization calculations were carried out using Matlab. 7.12.

The 1309 and 1408 SPSS [26] screened descriptors were used as the feed to a MLR-ICA approach as the population matrix in order to find the best descriptors for the gas phase. The procedure begins from random points (matrix indices of descriptors) called the initial countries that are the counterpart of chromosomes in GA and it is a set of values of a candidate solution for the optimization problem. The empires are sub-populations of the countries. Assimilation, which can be considered as a primitive form of Particle Swarm Optimization [27-28, 16] moves through all non-best countries (called colonies) in an empire toward the best country (called imperialist) in the same empire to find the colonies with the lowest error (RMSE of MLR-predicted versus the empirical values). Different number of decision variables (nDes) and different number of empires (nEmp) were investigated to obtain the least RMSE and highest R^2. The number of decision variables (nDes)) and number of empires/ imperialists (nEmp) were considered 1up to 8 and 10, 20, 30, respectively.

2.2. Monte Carlo Method

CORAL [29] software was used for calculation of descriptor correlation weight (DCW) of the 25 Doxazolidine compounds with a hybrid optimization scheme including hydrogen-suppressed molecular graph (HSG), hydrogen-filled graphs (HFG) and SMILES representation of molecular structures. Modelling using CORAL software was carried out for thresholds of 1 up to 3 and 100 and 30 epochs with dependent variables biological activity (pIC50) and lipophilicity (XLOGP), respectively [30]. The SMILES-based and Graph -based optimal descriptors were obtained from our previous work [31]. The hybrid objective function for finding the optimal descriptors is defined as:

$$DCW(T, Nepoch)^{Hybrid} = DCW(T, Nepoch)^{SMILES} + DCW(T, Nepoch)^{Graph} \qquad (2)$$

2.3.Formation of Complexes with DNA

Active sites of the best Doxazolidine derivative were exposed to reaction with

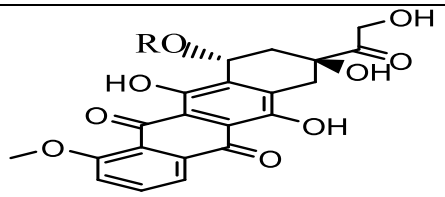
DNA at 5'-NGC-3' and geometrical optimization of the complexes were carried out by B3lyp/lanl 2dz at the Gaussian 03W [16]. DNA structure was obtained from Protein Data Bank (PDB) [27].

3. RESULTS AND DISCUSSION

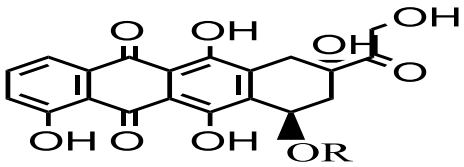
3.1. Linear and Non-Linear Combination Methods

All the studied Doxazolidine compounds [34-40] are presented in table 1. In MLR-PCR, MLR-PLS1 and MLR-MLR models, the best descriptors were selected using MLR procedure of SPSS software in three steps described in theory and computational methods section. Thereafter, the selected descriptors were employed as input in unscramble software and statistical parameters were calculated using PCR, PLS1and MLR models.

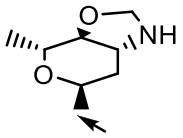
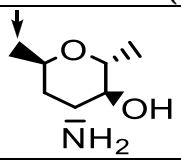
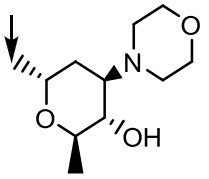
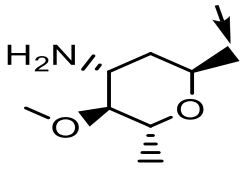
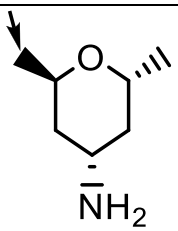
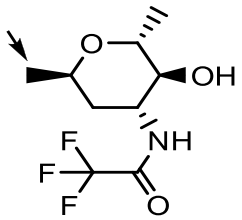
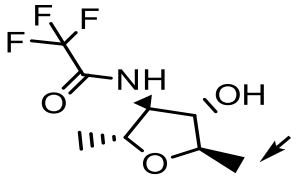
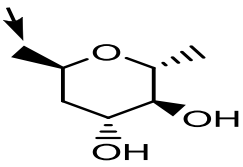
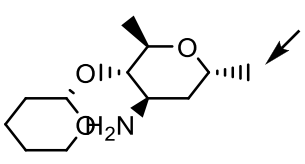
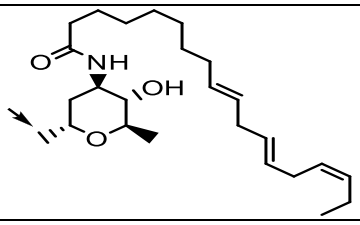
Table 1. Basic structure and corresponding compounds and branches of the Doxazolidine derivatives used to build QSAR models with B3lyp/6- 31g in gas phase

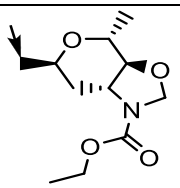
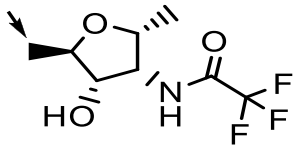
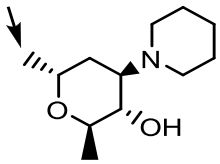
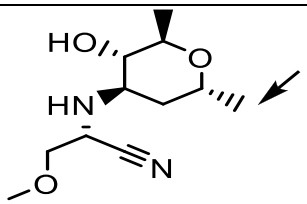
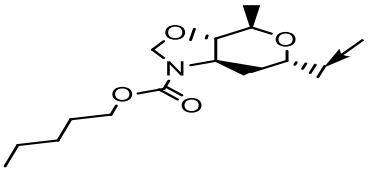
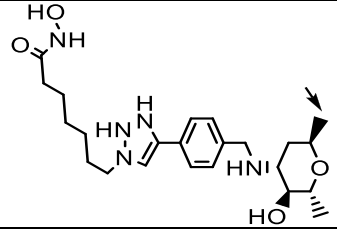
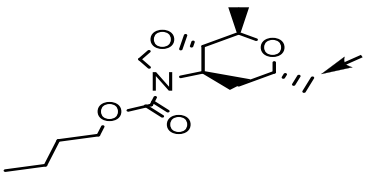
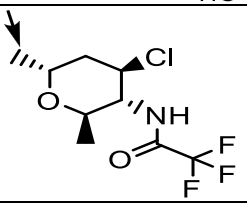
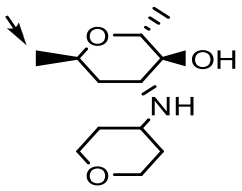
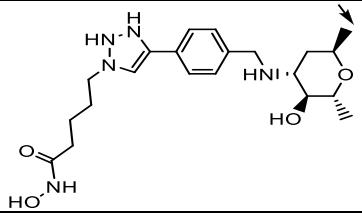
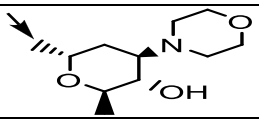
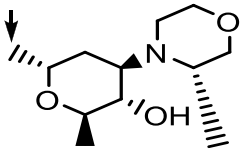
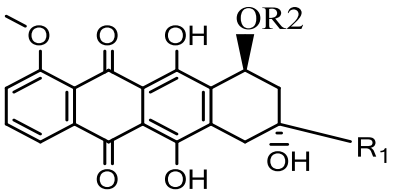
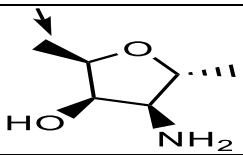
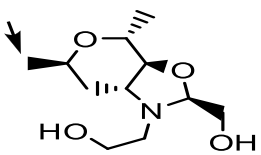
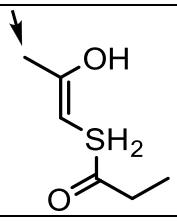
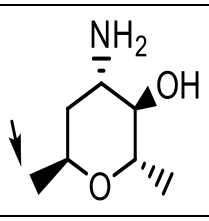


Basic structure A1 (Compounds 1, 5, 7-10, 13, 15-21, 24)



Basic structure A2 (Compounds 2, 3, 4, 6, 11, 12, 14, 22, 23, 25)

Compounds	Branch structure (R)	Compounds	Branch structure (R)
1		2	
5		3	
7		4	
8		6	
9		11	

10		12		
13		14		
15		22		
16		23		
17		25		
18				
19		 Basic structure A3 (Compound 20)		
21		Compounds	Branch structure (R1)	Branch structure (R2)
24		20		

Empirical values of pIC50 and lipophilicity (XLOGP) of these compounds were obtained from pubchem [28]. The RMSE and the correlation coefficient (R^2) in MLR–PCR, MLR–PLS1 and MLR–MLR for the predicted biological activity were found to be [0.3849 0.8661], [0.3776 0.8711] and [0.2892 0.9244], respectively. Furthermore, the calculated parameters indicated that MLR–MLR method were better than all other employed linear methods (MLR–PLS1 and MLR–PCR).

The 1309 and 1408 descriptors were fed to the MLR-SA, SA-ANN, GA-ANN and MLR-GA models and then the best descriptors were selected and the statistical parameters were calculated. Table 2 shows that RMSE and R^2 for predicted lipophilicity (XLOGP) in GA-ANN was found to be [0.2301 0.9748]. Therefore GA-ANN model were better than the other nonlinear models and hence, the selected descriptors using GA–ANN are discussed here. In addition, RMSE and R^2 of 0.0164 and 0.9854 were obtained from biological activity study, which confirms this result.

Table 2. Statistical parameters of different non-linear QSAR models

QSAR Models	Predicted(lipophilicity)	
	R^2	RMSE
MLR-SA	0.927	0.445
SA-ANN	0.9437	0.3270
MLR-GA	0.941	0.3290
GA-ANN	0.9748	0.2301

* The naming are according to Table 1.

In non-linear methods, 80%, 10% and 10% of data sets were randomly chosen as training, validation and test sets, respectively. Table 3 shows test and validation series of nonlinear models in Doxazolidine compounds.

RDF030e (Weighted by atomic Sanderson electronegativities), RDF155v (weighted by atomic van der Waals volumes) , RDF055m (Weighted by atomic masses) , RDF075e (Weighted by atomic Sanderson electronegativities) descriptors were selected using GA-ANN method. RDF descriptors are independent on the number of atom and the size of a molecule. It is unique regarding the three-dimensional arrangement of the atoms, and it is invariant against translational and rotational entire molecule [30]. Eigenvalues based indices descriptors (Eig1m: Leading eigenvalue from mass weighted distance matrix/in predicted biological activity) are computed from Weighted Distance Matrices of a Hydrogen-depleted Molecular Graph [33]. Mor01p (weighted by atomic polarizabilities) in predicted biological activity, Mor3e (3D-MoRSE- signal 03/ weighted by atomic Sanderson electronegativities) and Mor15u (3D-MoRSE- signal 15/ unweighted) in predicted lipophilicity (XLOGP) are 3D-MoRSE that were obtained through the molecular transformation employed in electron diffraction studies [34]. H6e (H autocorrelation of lag 6/ Weighted by atomic by atomic Sanderson electronegativities) in predicted biological

Table 3. Test and validation series in QSAR models

QSAR Models	Biological activity		lipophilicity	
	Test series	valid series	Test series	valid series
MLR-SA	4,19,22	11,21,24		
SA-ANN	1,2,20	4,7,13	5,6,24	1,15,16
MLR-GA	5,12,25	5,12,25	2,11,23	19,21,22
GA-ANN	7,14,17	3,6,8	7,14,17	3,6,8

The naming are according to Fig. 2.

activity, R4e (R autocorrelation of lag 4/ weighted by atomic Sanderson electronegativities), R8p (R autocorrelation of lag 8/ weighted by atomic Polarizabilities), and R1u+ (R maximal of lag 1/ unweighted) and H_{TP} (H total index/weighted by atomic polarizabilities) in predicted lipophilicity (XLOGP) are GETAWAY descriptors. GETAWAY (Geometry, Topology, and Atom-Weights Assembly) descriptor representations encode the geometrical information obtained from the molecular matrix, the topological information obtained from the molecular graph and the information obtained from atomic weights, which are specially designed with the aim of matching the 3D-molecular geometry [33]. Vv (V total size index/ Weighted by atomic van der Waals volumes) descriptor in predicted biological activity and G2v (2st component symmetry directional WHIM index/weighted by atomic van der Waals volumes) in predicted lipophilicity (XLOGP) are WHIM descriptors that were built in such a way to capture the relevant molecular 3D information regarding molecular size, shape, and symmetry and atom distribution with respect to invariant reference frames [33]. MLOGP2 (Squared Moriguchi Octanol-Water partition coeff.(log P $^{\wedge 2}$) is Molecular properties descriptor that is a method for the calculation of the n-octanol/water partition coefficient based on similarities in the structure or properties of chemical compounds [44].

The RMSE of the predicted set in GA–ANN model was 0.0164, which is acceptable in comparison to previous works [35-37].

A sensitivity analysis was carried out to investigate the effects of each single descriptor on the predicted lipophilicity (XLOGP) and the results are depicted in table 4. According to this table, for designing new drugs, R1u+, MLOGP2 and R8p descriptors in predicted lipophilicity are recommended to be at their maximum value, while, R4e, Mor3e, Mor15u,G2v,H_{Tp} descriptors are recommended to be at their minimum value.

GA-ANN method suggests the RDF030e, Eig1m, RDF155v, Mor01p, RDF075e, H6e, Vv descriptors as the most important physicochemical descriptors in predicted biological activity to be at their maximum value.

Table 4. Optimum value/range descriptors in GA-ANN method

Descriptor (lipophilicity)	range	Optimum value/range
R4e	1.9-2.2	minimum
Mor3e	-2.4- -10	minimum
Mor15u	-1- 1	minimum
G2v	0.135-0.165	minimum
R8p	0.32-0.38	maximum
R1u+	0-3	maximum
MLOGP2	16-28	maximum
H_{TP}	0.06-0.12	minimum

3.2. Molecular Descriptors Generation with MLR-ICA Approach

As a first trial, 1000 number of iterations were done to find the most powerful empires and, subsequently, the best descriptors. A plot of the best cost values versus the number of iterations was represented in Fig. 1. It implies that there is no variation in the best cost (MSE) after about 300 iterations. However, in order to ensure that the best descriptors are captured, the number of iterations for the rest of computations was set to 500 and 200 in the –logIC50 and lipophilicity (XLOGP) calculations, respectively.

The effects of number of selected descriptors on the chosen descriptors and the prediction quality (according to R^2 and RMSE) was investigated and as it is expected, the model accuracy regarding to R^2 and RMS increased by increasing the number of model parameters (descriptors in this case). In order to choose the most suitable number of empires, the model was

run using different number of empires and the results are demonstrated in table 5. The optimum number of empires was chosen as 30 and 50 in the pIC50 and lipophilicity (XLOGP) calculations, respectively.

Plot of the predicted versus empirical values of –logIC50 and XLOGP are depicted in fig.2. The figure implies that the developed model possesses a high correlation coefficient, indicating that the experimental and predicted values are well correlated.

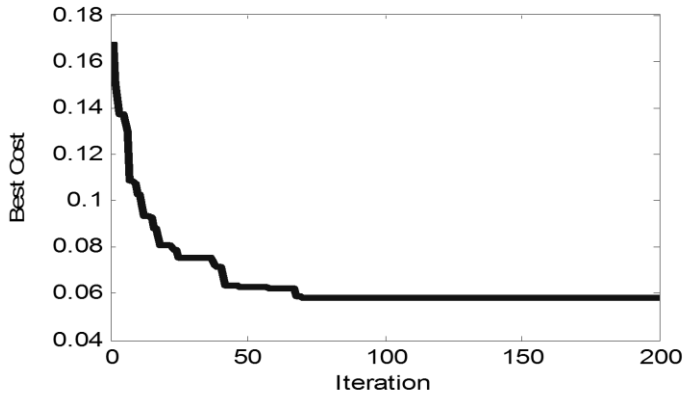


Fig. 1. Plot between Best Cost values versus the variation of Iteration in lipophilicity (XLOGP) predicted.

Table 5. Statistical parameters of ICA-MLR models with different nEmp (Max.It=1000)

nVar- nEmp	Predicted (lipophilicity)	
	R^2	RMSE
8-10	0.970	0.1913
8-20	0.9773	0.1989
8-30	0.9786	0.1931
8-40	0.9777	0.1969
8-50	0.9787	0.1907
8-60	0.9777	0.1973
8-70	0.970	
8-80	0.9773	
8-90	0.9786	
8-100	0.9777	

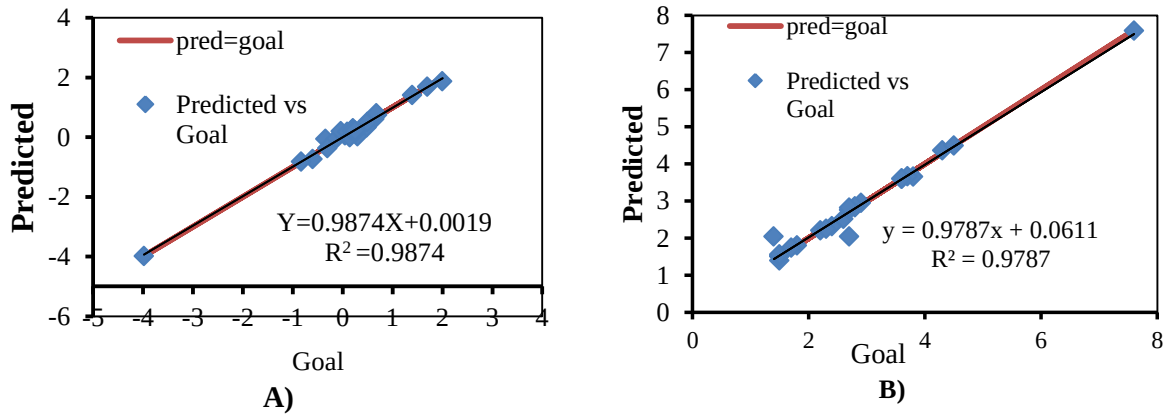


Fig. 2. A) Plot between predicted values versus Goal with nVar=8 and nEmp=30, in predicted biological activity B) Plot between predicted values versus Goal with nVar=8 and nEmp=50 in predicted lipophilicity (XLOGP).

The best selected descriptors using MLR-ICA Method with nDes=8 and nEmp=30 are F10[C-C] (Frequency of C-C at topological distance 10), Mor15m (3D-MoRSE-signal 15/weighted by atomic masses), F09[C-C] (Frequency of C-C at topological distance 09), SP04 (Shape profile no.4), MATS8v (2D autocorrelation-lag 8/weighted by atomic Van der Waals), Mor14u (2D autocorrelation-lag 8/weighted by atomic Van der Waals), Mor31p (3D-MoRSE-signal 31/weighted by atomic polarizabilities) and L3e (3rd Component size directional WHIM index/weighted by atomic Sanderson electronegativities) with biological activity (pIC50) as the dependent variable.

From ICA-MLR method, the most important physico-chemical descriptors including F10[C-C], Mor15m, RDF155v, Mor14u, RDF075e, Mor31p and L3e should be kept at their minimum value. Adversely, SP04, F09[C-C] descriptors should have the least value, and MATS8v in the range of -0.05 to 0.1.

The best selected descriptors using MLR-ICA Method in lipophilicity (XLOGP) calculations with nDes=8 and nEmp=50 are RDF145p (Radial Distribution Function-1.5/ weighted by atomic Polarizabilities/ RDF descriptors), Ts (T total size index/weighted by atomic electrotopological states/ WHIM descriptor), G1v (st component symmetry directional WHIM index/weighted by atomic van der Waals volumes/WHIM descriptors), Av (A total size index/weighted by atomic van der Waals volumes/WHIM descriptors), L1u (1st component size directional WHIM index/unweighted/WHIM descriptors), VEZ2 (Average Randic-type eigenvector-based index from distance matrix/ eigenvalue-based indices), Mor30u (3D-MoRSE-signal 30/unweighted/ 3D-MoRSE descriptor) and BLTA96 (Verhaar model of Algae based-line toxicity from MLOGP (mmol/l)/ Molecular properties).

The graphs of RDF125p, Ts, G1v, Av, L1u, VEZ2, Mor07u and BLTA96 descriptors versus lipophilicity (XLOGP) were plotted in Fig. 3.

The charts showed that the response value increases by increasing RDF15p, Ts, G1v, VEZ2, Mor07u, while the effects of Av and BLA96 descriptors are inverse. Also, the L1u descriptor in the range of 0 to 60 shows no effect on the response.

3.3. Result of the Monte Carlo Method

The statistical parameters of the models obtained using molecular graphs (HSG) and SMILES are shown in Table 6. Performance of the models were compared with each other by the criterion of the predictability in test set (R^2m) which should be larger than 0.5 [38], correlation coefficient (R^2) in each set, cross-validated correlation coefficient (Q^2) and standard error of estimation (s). The difference between R^2m and R'^2m values (ΔRmTEST) was used as another criterion in this issue. The results showed that for all of the three splits, threshold of 3 gives the best results. The results with threshold of 3 for the probe 3 including the function of lipophilicity (XLOGP) in terms of DCW and the model fit results are presented in Table 6.

The variation of correlation coefficient (test set) with respect to threshold and the number of epochs are plotted in figure 4. This figure confirms that [3 23] are the most appropriate values for threshold and number of epochs including the functions of lipophilicity (XLOGP), respectively.

The experimental and calculated activities (pIC50) and lipophilicity (XLOGP) for the sequence of compounds are plotted against each other in figure 5. A good correlation between the calculated and empirical values of pIC50 and lipophilicity (XLOGP) can be observed in

this figures that approves the appropriateness of the developed model.

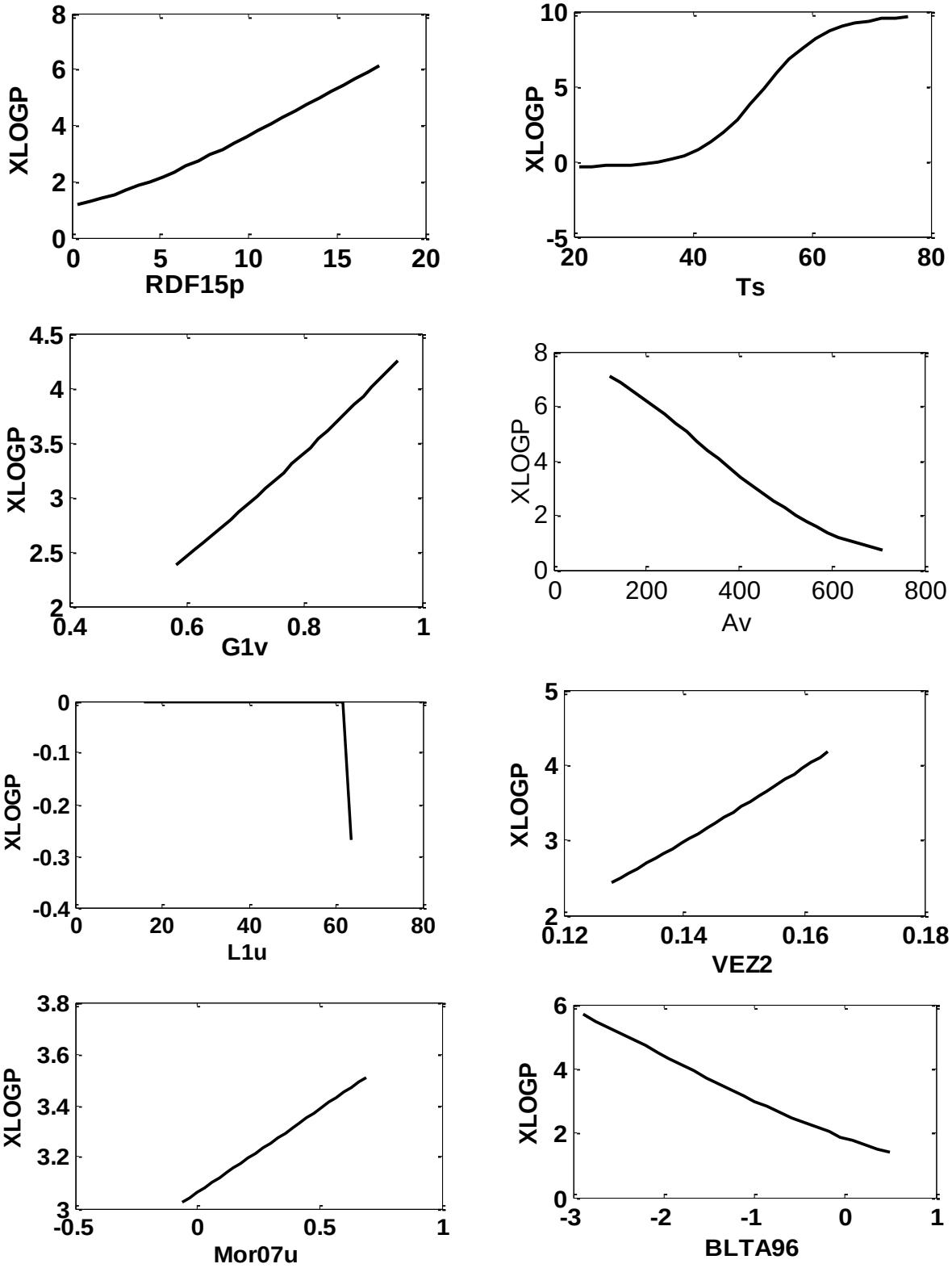
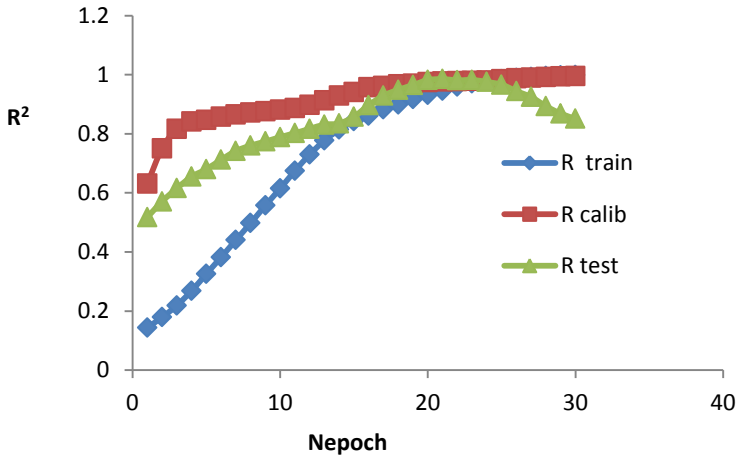


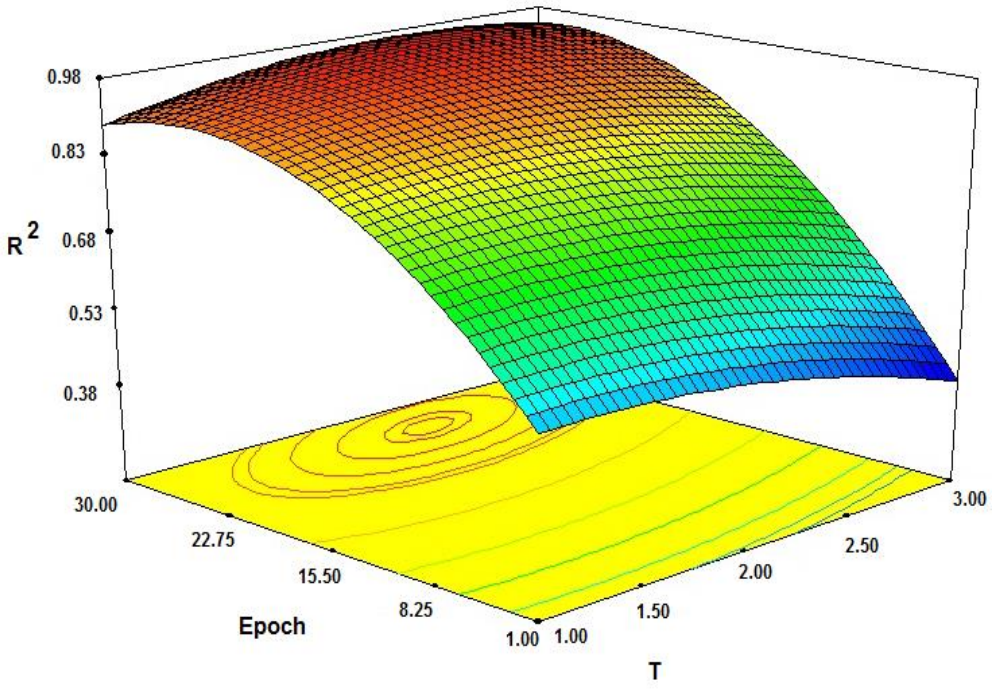
Fig. 3. Plot between lipophilicity (XLOGP) versus of the RDF15p, Ts, G1v, Av, L1u,VEZ2, Mor07u and BLTA96 descriptors.

Table 6. The split models in Monte Carlo Method

Split 1: (T=3, prob = 3 in lipophilicity)
XLOGP = -8.9340157 (± 0.0323538) + 0.0815940 (± 0.0002382) * DCW(2,30)
n = 12, R^2 = 0.9989, Q^2= 0.9985, s = 0.028 (training set)
n = 8, R^2= 0.9954, Q^2 = 0.9927, s = 0.587 (calibration set)
n = 5, R^2 = 0.8510, Q^2 = 0.8065, s = 0.460 (test set), R^2m TEST = 0. 0.5987

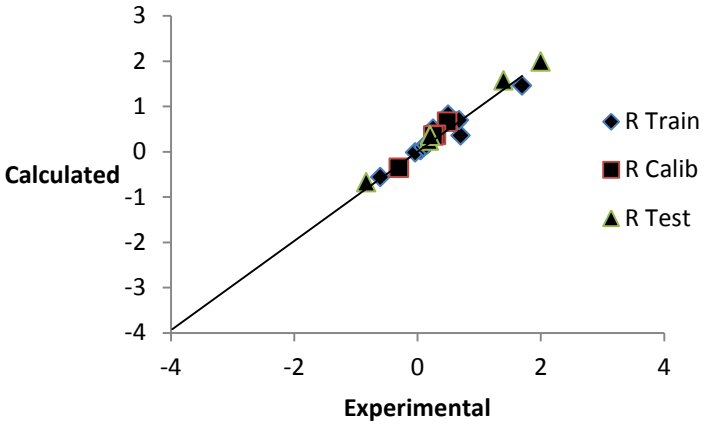


A)

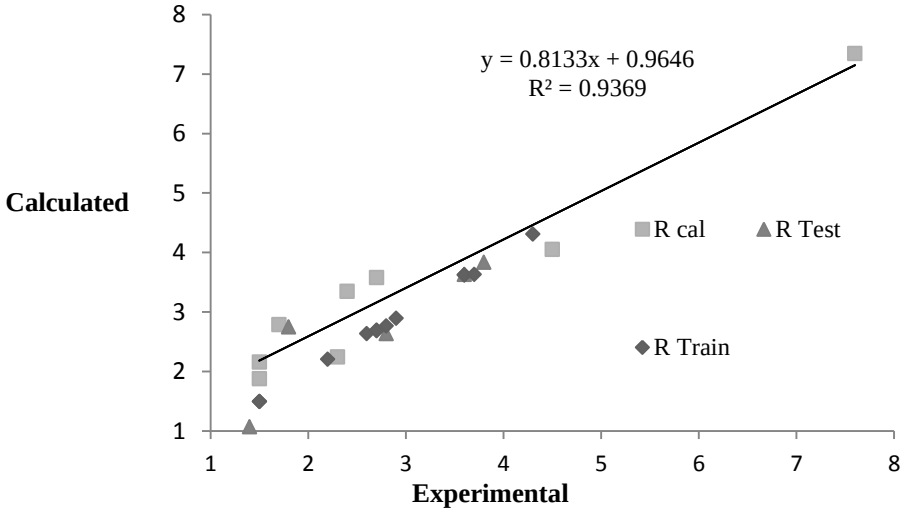


B)

Fig. 4. The variation of correlation coefficient for test set by threshold and number of epochs. A) Effects of the number of epochs in predicted lipophilicity. B) 3D surface plot of R^2 according to the threshold and the number of epochs.



A)



B)

Fig. 5. Correlation between experimental and predicted (A) –logIC50 and (B) lipophilicity calculated using Eq.3.

Molecular features are sorted according to their correlation weights and are given in table 7.

The higher the correlation weigh of a molecular feature, the lower the value of IC50, and therefore, the features are more significant. Definitions of the molecular features are given by Kumar and Chauhan [39]. The Presence of cyclic rings, Presence of sp2 is power carbon connected to ring, presence of oxygen, nitrogen, but absence of sulphur and phosphorus, absence of halogens, presence of double bond in the cyclic rings, presence of oxygen connected to cyclic rings are the most important molecular features in biological activity (pIC50) that might be considered in designing new drugs .

According to table 7 with the dependent variables lipophilicity (XLOGP), presence of cyclic rings, presence of double bond and cyclic ring with branching, presence of double bond, presence of oxygen connected to double bond and ring, number of carbon in sp^2, presence of N atom connected to ring, presence of carbon in sp^2, presence of sp^2 Carbon connected to ring, absence of halogen, presence of cyclic ring with branching, presence Oxygen five-member

cycles, presence of 5, 6 six-mmber cycles with aromaticity and H atom in cycle, are the most important molecular features that might be considered in designing new drugs.

3.4. Result of the Doxazolidine-DNA

DNA reacts with Doxazolidine in the 3-amino group of each Doxorubicin to the 2-amino group of a Guanin-base via a methylene originating from formaldehyde as an impurity in the crystallization solvent [41]. Compound № 1 of Doxazolidine derivatives has the highest pIC50 value [40] and therefore was chosen as the best potential drug for the investigation using Complex formation study with DNA. The Wang crystal structure also showed hydrogen bonding from the 9-OH group of the Doxorubicin to the amino group and nitrogen at the 1-position of the Guanin-base on the opposing strand [42].

The structure-optimized Doxazolidine-DNA complexes is shown in fig.6 and the covalent bond and hydrogen bond were found to be 1.4 and 1.9, respectively. It can be observed that the predicted values are in good accordance with their experimental values [43,44]. The computed complex structure formation energy was obtained to be -4417.2877 Hartree.

Table 7. SMILES attributes with positive correlation weights for split 1

SMILES attributes (in lipophilicity)	CWs	SMILES attributes (in lpophilicity)	CWs
1.....	2.79805	C...2.....	4.52826
2.....	4.00058	N...1.....	4.05179
3.....	2.59214	HALO00000000	3.50178
4.....	2.36306	=...4.....	6.90227
5.....	3.46578	5...(.....	2.50147
=...(.....	2.24612	C5.....0...	3.40285
=.....	3.22375	C6....H.5...	2.79721
BOND10000000	4.34849	C6-H6	2.79866
O...=.....	3.42557	O...4.....	3.86564
NNC-C...110	7.00077	3...(.....	3.80289

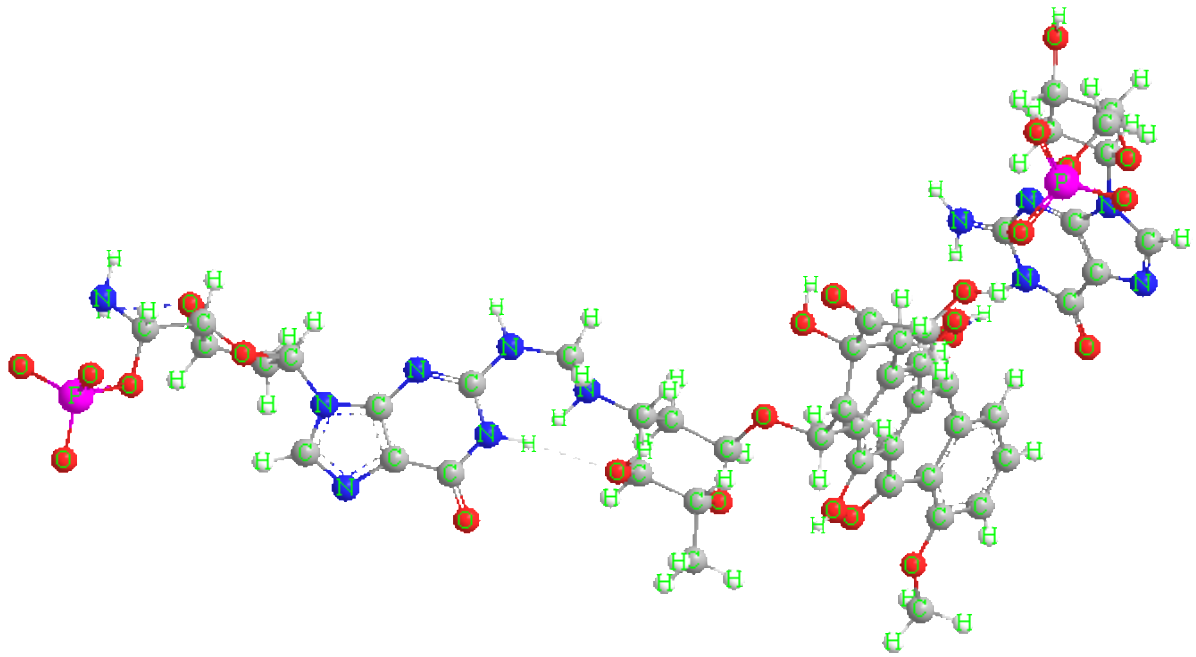


Fig. 6. Structure of Doxazolidine-DNA.

4. CONCLUSIONS

The GA-ANN and ICA-MLR show the best performance among the considered approaches. In GA-ANN method with the dependent variable of biological activity (pIC50**),** the RDF030e, Eig1m, RDF155v and Mor01p physico-chemical descriptors and in MLR-ICA combination F09[C-C] descriptor were found to have the most important role on the drug activity. According to the Monte Carlo studies, the structural descriptors including present of cyclic rings and absence of halogens are the most important factors. Precise setting of these physico-chemical and structural descriptors can reduce the half maximal inhibitory concentration (IC50). Noting that the aforementioned descriptors are the most effective descriptors amongst the eight ones in GA-ANN and MLR-ICA methods, van der Waals volumes, Sanderson electronegativities, frequency of C-C and dipole moment should be maximized in designing new drugs. In ICA-MLR method, the lipophilicity (XLOGP) value grows via increasing RDF15p, Ts, G1v, VEZ2, BLTA96 descriptors. The polarizibility, electrotopological state, atom van der Walls volume, average Randic-type eigenvector-based index from distance matrix should be maximized.

From Monte Carlo method, the structural descriptors including number of carbon in sp^2, presence of double bond and cyclic ring with branching are important. In GA-ANN method atomic Sanderson electronegativities descriptors are recommended to be at their minimum value. Adversely, Squared Moriguchi Octanol-Water partition coeff.(log P $^{\wedge 2}$) descriptor should be maximized. It was concluded that the simulation use of Monte Carlo gives deeper and more comprehensive knowledge of the effect of structural descriptors including presence of cyclic rings, Presence of branching and Doxazolidine reacts with 5'-NGC-3' sites in DNA.

5. ACKNOWLEDGEMENT

The authors gratefully acknowledge the support provided by the Islamic Azad University of Rasht.

6. REFERENCES

[1] GC. Post, BL. Barthel, DJ. Burkhart, JR. Hagadorn, TH. Koch., J. Med. Chem. 48 (2005) 7648–7657.
[2] A. Di Marco, M. Gaetani, B. Scarpinato. Cancer. Chemother. Rep. (Part I), 53 (1969) 33–37.
[3] G. C. Post, B. L. Barthel, D. J. Burkhart, JR. Hagadorn, TH. Koch, J. Med. Chem. 48 (2005) 7648-57.
[4] R. SayyadikordAbadi, A. Alizadehdakhel. Rev. Roum. Chim., 63 (2018) 171-180.
[5] D. B. Horvath, B. Mao, QSAR Comb Sci. 22 (2003**)** 498.
[6] Y. Kwon, Partition and Distribution Coefficients. Handbook of Essential Pharmacokinetics, Pharmacodynamics and Drug Metabolism for Industrial Scientists. New York: Kluwer Academic/Plenum Publishers. 2001.
[7] A. Leo, C. Hansch, D. Elkins. Chem. Rev. 71 (1971) 525–616.
[8] J. Sangster. Octanol–Water Partition Coefficients: Fundamentals and Physical Chemistry. Wiley Series in Solution Chemistry. Chichester: John Wiley & Sons Ltd. 1997.
[9] L. Shargel, WP. Susanna, AB. Yu., Applied Biopharmaceutics & Pharmacokinetics (6th ed.). New York: McGraw-Hill Medical. 2012.
[10] R. Wang, Y. Fu, L. Lai, J. Chem. Inf. Comput. Sci. 37 (1997) 615-621.
[11] R. Wang, Y. Gao, L. Lai, Drug Discov. Des. 19 (2000) 47-66.
[12] J. Guziałowska-Tic1., E3S Web of Conferences 19 (2017), 02034 EEMS.

[13] E. Atashpaz-Gargari, C.Lucas, Imperialist competitive algorithm: An algorithm for optimization inspired by imperialistic competition. in *IEEE* Congress on Evolutionary Computation. Singapore. (2007) 4661–4667.
[14] S. Hosseini, A. Al Khaled, 2014, Appl. Soft Comput, 24 (2014) 1078-1094.
[15] Z. Aliniya, S. A. Mirroshandel, Expert Syst Appl, 117 (2019) 243-266.
[16] AP. Toropova, A. A. Toropov, E. Benfenati, G. Gini, D. Leszczynska, J. Leszczynski, J. Comput. Chem, 32 (2011) 2727.
[17] A. A. Toropov, A. P. Toropova, SE. Martyanov, E. Benfenati, G. Gini, D. Leszczynska, J. Leszczynski, Chemom. Intell. Lab., 109 (2011) 94.
[18] E. Tanış, N. Çankaya, S. Yalçın, Russ. J. Phys. Chem. B, 13 (2019**)** 49.
[19] K. Levenberg, Quarterly of Applied Mathematics, 2 (1944**)** 164-168.
[20] S. H. Sadat Hayatshahi, P. Abdolmaleki, M. Ghiasi, S. Safarian**,** FEBS Lett. 581 (2007) 506-14.
[21] R. Sayyadi kord Abadi, A. Alizadehdakhel, S. Dorani Shiraz, Russ. J. Phys. Chem. B, 11(2017) 307.
[22] M. J. Frisch, et al. Gaussian 09 (Gaussian, Inc., Wallingford CT, 2009). https://gaussian.com/glossary/g09/.
[23] SPSS, Version 19, available at http://www.spssscience.com (2010).
[24] O. G. Maksimova, A. V. Maksimov, Russ. J. Phys. Chem. B, 11 (2017) 15.
[25] Yu. M. Sivergin, S. M. Kireeva, S. M. Usmanov, Russ. J. Phys. Chem. B, 9 (2015) 952.
[26] S. Putta, J. Eksterowicz, C. Lemmen, R. Stanton, J. Chem. Inf. Comput. Sci. 43 (2003) 1623-1635.
[27] E. A. Gargari, F. Hashemzadeh, R. Rajabioun, C. Lucas, Inter. J. Intellig.Comput. Cybernetics, 1 (2008): 337-355.
[28] J.L. Lin, Yu-H Tsai, C.-Y. Yu; M-S. Li, Algorithms, 5 (2012) 433-448.
[29] (http://www.insilico.eu/coral)
[30] AM. Veselinović, JB. Milosavljević, A.A. Toropov, GM. Nikolić, Eur. J. Pharm. Sci. 48(2013) 532-41.
[31] J. Veselinović, A. Veselinović, A. Toropov, A. Toropova, I. Damnjanović, G, Nikolić, Scientific Journal of the Faculty of Medicine in Niš, 31 (2014) 95-103.
[32] http://www.rcsb.org/.
[33] R. Todeschini, V. Consonni, Hand Book of Molecular Descriptors. Wiley-VCH Verlag GmbH, Weinheim, Germany, 2008**.**
[34] J. H. Schurz, P. Selzer, J. Gasteiger, J. Chem. Inform. Comput. Sci., 36 (1996) 334.
[35] R. Sayyadikord Abadi, A. Alizadehdakhel, S. Tajadodi Paskiabei, J. Korean Chem. Soci., 60 (2016) 225.
[36] R. Sayyadikord Abadi, A. Alizadehdakhel, G. R. Najafi, M. Masomparast. Rev. Roum. Chim. 63 (2018) 217-226.
[37] R. Sayyadi kord Abadi, A. Alizadehdakhel, F. Moosaour, Indian j.chem. 56 B (2017) 677-686.
[38] A. Golbraikh, A. Tropsha, J. Mol. Graph. Model. 20 (2002) 269-276.
[39] A. Kumar, S. Chauhan. SAR and QSAR in Environmental Research, 28 (2017**)** 179.
[40] https://pubchem.ncbi.nlm.nih.gov/
[41] AH-J. Wang, YG. Gao, Y. C. Liaw, Y. K. Li, Biochemistry, 30 (1991**)** 3812.
[42] H. Zhang, Y. G. Gao, G. A. van der Marel, J. H. van Boom , A. H. Wang .J. Biol. Chem. 268 (1993) 10095.
[43] B. T. Kalet, M. B. McBryde, J. M. Espinosa, T. H. Koch, J. Med. Chem. 50 (2007) 4493–4500.

[44] I. Moriguchi, S. Hirono, Q. Liu, I. Nagakome, Y. Matushita, Chem. Pharm. Bull. 42 (1994), 976.

[45] T. Cheng, Y. Zhao, X. Li, F. Lin, Y. Xu, X. Zhang, Y. Li, R. Wang, J. Chem. Inf. Model. 47 (2007) 2140.

مجله شیمی فیزیک و شیمی نظری
دانشگاه آزاد اسلامی واحد علوم و تحقیقات
جلد ۱۸، شماره ۲، تابستان ۱۴۰۰
ISSN ۱۷۳۵-۲۱۲۶

مطالعه فعالیت بیولوژیکی و چربی دوستی دوکسازولیدین به عنوان مهار کننده سلول‌های سرطانی

ماریا نیک کار[۱]، **ربابه صیادی کردابادی**[۱*]، اصغر علیزاده داخل[۱]، قاسم قاسمی[۱]

[۱] گروه شیمی و مهندسی شیمی، واحد رشت، دانشگاه آزاد اسلامی، رشت، ایران

چکیده

تحقیق QSAR در مورد چربی دوستی (XLOGP) و فعالیت بیولوژیکی (IC50) برخی از مشتقات دوکسازولیدین با استفاده از روشهای مدل سازی رگرسیون خطی چندگانه (MLR) و شبکه عصبی مصنوعی (ANN) و سه تکنیک بهینه سازی مختلف از جمله الگوریتم شبیه سازی (SA)، الگوریتم ژنتیک (GA) و الگوریتم رقابت استعماری (ICA) انجام شد. علاوه بر این از نرم افزار CORAL برای ارتباط چربی دوستی و فعالیت بیولوژیکی با پارامترهای ساختاری داروها استفاده شد. نتایج بدست آمده مقایسه شد و روش‌های ترکیبی GA-ANN و ICA-MLR با توجه به ضریب همبستگی (R^2) و میانگین خطای مجموع مربعات (RMSE) بهترین عملکرد را نشان دادند. موثرترین توصیف کننده های استخراج شده از مطالعات چربی دوستی و فعالیت بیولوژیکی ارائه و مورد بحث قرار گرفت. از روش GA-ANN مشخص شد که مهمترین توصیف کننده‌های فیزیکی – شیمیایی دارای حداقل مقدار در الکترونگاتیویته ساندرسون (Sanderson) و حداکثر مقدار در ضریب توزیع اکتانول- آب موریگوچی ((Moriguchi, (log P ^2)) هستند. در روش ICA-MLR حداکثر مقدار در قطبش‌پذیری، حالت الکتروتوپولوژیکی و **حجم وندر والس** اتم ها به عنوان مهمترین توصیف کننده‌های فیزیکی- شیمیایی پیشنهاد شد. نتیجه گیری شد که مطالعه QSAR و روش مونت کارلو می تواند به درک جامع‌تری از رابطه بین توصیف کننده های فیزیکی – شیمیایی، ساختاری یا تئوری داروها با فعالیت بیولوژیکی و چربی دوستی منجر شود.

کلید واژه‌ها: الگوریتم GA و CA ؛ مشتقات دکسازولیدین؛ QSAR؛ روش مونت کارلو

[*] مسئول مکاتبات: sayyadi@iaurasht.ac.ir ,sayyadi_04@yahoo.com