Advanced Journal of Chemistry, Section A, 2024, 7(5), 522-539

Advanced Journal of Chemistry, Section A

journal homepage: www.ajchem-a.com

Original Research Article

QSAR Study on DYRK1A Inhibitors for Regenerative Therapy in Diabetes

Faezeh Khosravi [1], Roya Kiani-Anbouhi [1,*], Eslam Pourbasheer [2]

[1]*Department of Chemistry, Faculty of Science, Imam Khomeini International University, Qazvin, Iran*
[2]*Department of Chemistry, Faculty of Science, University of Mohaghegh Ardabili, Ardabil, Iran*

ARTICLEINFO

Article history

Submitted: 14 March 2024
Revised: 03 May 2024
Accepted: 19 May 2024

Manuscript ID: AJCA-2405-1530
Checked for Plagiarism: Yes
Language Editor Checked: Yes
DOI: 10.48309/ajca.2024.458482.1530

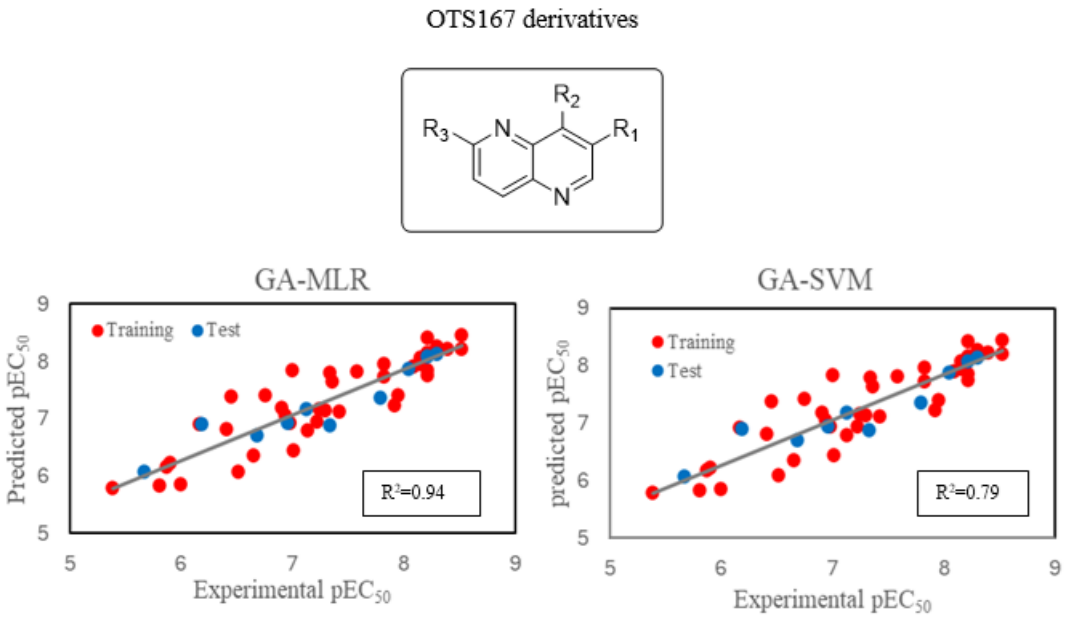
KEYWORDS

QSAR
GA-MLR
GA-SVM
DYRK1A inhibitors
Diabetes treatment

ABSTRACT

The QSAR models were developed for predicting DYRK1A biological activity (EC_{50}) with a series of 1,5-naphthyridines derivatives as highly potent DYRK1A-dependent inducers of human β-cell replication using multiple linear regressions (MLR) as a linear method and support vector machine (SVM) as a nonlinear method. The 49 chemicals in data set were randomly partitioned into training and test subsets. For the selection of molecular descriptors, the genetic algorithm (GA) feature selection approach was used, followed by MLR and SVM. Testing the prediction abilities of the obtained models were conducted using the tests of cross-validation, Y-randomization, and an external test set. By comparing the results of GA-MLR and GA-SVM models, it is clear that GA-SVM produced better results (R^2_{train}= 0.946, F_{train}= 78.641, RMSE $_{train}$= 0.203), although both models had adequate predictive quality. Using the predicted results of this study, new and potent DYRK1A inhibitors can be designed. In addition, this study provides insight into a new strategy to design diabetes drugs.

GRAPHICALABSTRACT



* Corresponding author: Kiani-Anbouhi, Roya
✉ E-mail: kiani@sci.ikiu.ac.ir

Introduction

High blood glucose levels cause diabetes, which is divided into two types: T1D (type 1 diabetes) and T2D (type 2 diabetes). Type 1 diabetes, also called insulin-dependent diabetes or adolescent diabetes, occurs when pancreatic beta-cells are destroyed by autoimmune factors, leading to decreased insulin production. A person with T1D cannot survive without insulin [1]. β-cell depletion is one of the main causes of T2D diabetes [2,3]. It is described as "insulin resistance" when individuals diagnosed with Type 2 Diabetes (T2D) exhibit an insufficient response to endogenous insulin. Importantly, the reduction in β-cell population leads to insufficient insulin production, which is a major cause of both T1D and T2D. The important point is that current approaches cannot address one of the main causes of T1D and T2D [3-5]. This unmet medical need provides an opportunity to discover the cure for T1D and T2D. Islet transplants, whether sourced from cadavers or stem cells, have the potential to become more prominent in future Type 1 Diabetes treatment. However, the associated expenses and intricacies hinder its widespread implementation. Therefore, the development of a safe regenerative drug for the expansion of residual β-cell mass can be transformative. As an alternative to these limitations, Quantitative structure-activity relationships (QSARs) are increasingly acknowledged as a valuable and effective tool in various fields, such as pharmaceutical research for drug development, in the last few decades [6-10]. QSAR models aim to establish a coherent relationship between the structures of the molecules under investigation and their associated activities. QSAR begins by calculating the theoretical parameters called descriptors, which describe each selected molecule's structure and shape using an algebraic value. For each molecule, many descriptors are calculated, but only a few have a decisive role in biological activity. Thus, it is necessary to use the variable selection tool to select effective descriptors in creating the method. Several methods are effective and widely used to select variables, such as stepwise (SW) [11,12], genetic algorithms (GAs) [13], and simulated annealing [14]. By obtaining the relevant descriptors, the model is constructed using various modeling methods such as multiple linear regression (MLR) [15,16], artificial neural network (ANN) [17], and support vector machine (SVM) [18].

In this study, QSAR models were constructed using the GA-MLR linear method and the GA-SVM nonlinear method and finally, a comparison was conducted between the results yielded by the two models. The primary objective of this study is to construct a robust QSAR model to consider the most important descriptors that affect β-cell proliferation stimulation.

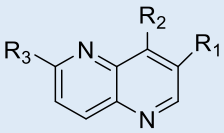
Experimental

Data set

The data set containing the values of EC_{50} of 49 compounds of OTS167 derivatives was collected from the literature [19]. In this study, the activity was interpreted as the half-maximal effective concentration (EC_{50}) of the compound. EC50 is the concentration of a drug compound that gives half-maximal response. The reported EC_{50} (nM) values were initially transformed into logarithmic scale as pEC_{50} (M) and subsequently used for QSAR analyses as the response variables.

Table 1 lists the chemical structures and their associated activity values of the data set. Based on 80% and 20% of the total data set, two training (39 compounds) and test sets (10 compounds) were divided randomly. Using the training set, the model was built and evaluated for its predictive power on the test set.

Table 1. Chemical structures of the 1,5-naphthyridines derivatives with experimental and predicted activities values (pEC_{50}) of DYRK1A inhibition potency



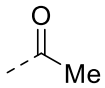
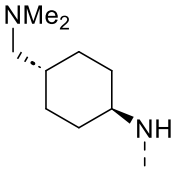
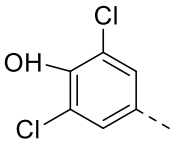
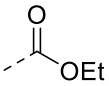
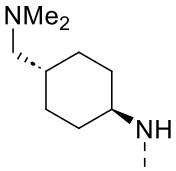
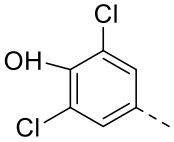
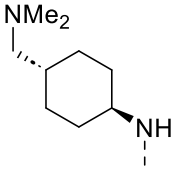
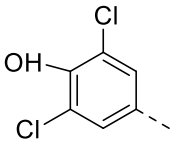
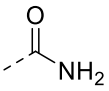
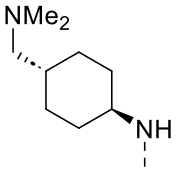
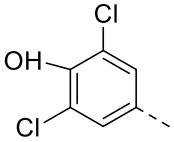
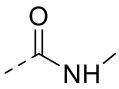
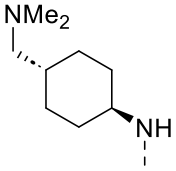
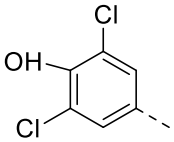
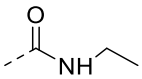
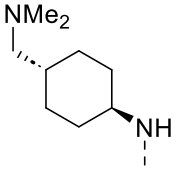
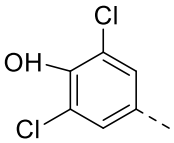
No.	R_1	R_2	R_3	pEC_{50} (Exp)	pEC_{50} GA-MLR	pEC_{50} GA-SVM
1[a]				8.301	8.127	7.986
2				8.154	7.935	8.010
3				7.346	7.785	7.411
4				7.823	7.955	7.948
5				8.221	7.743	8.080
6[a]				8.221	8.071	7.995

Table 1. Continued…

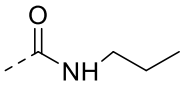
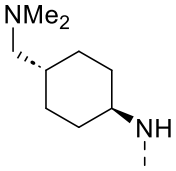
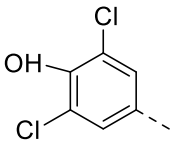
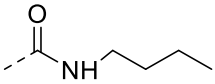
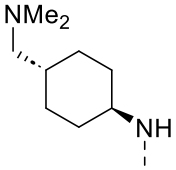
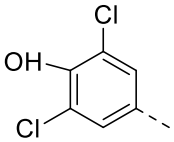
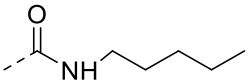
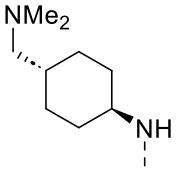
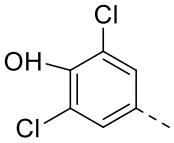
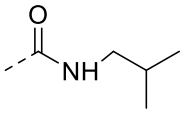
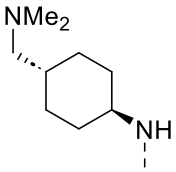
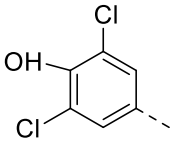
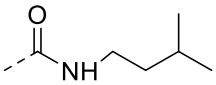
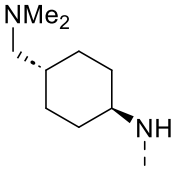
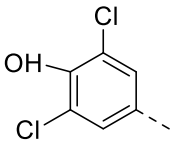
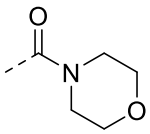
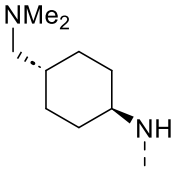
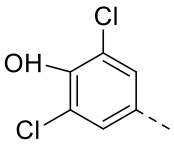
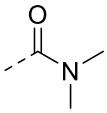
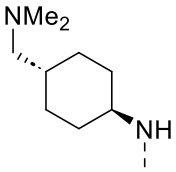
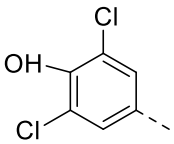
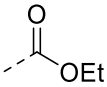
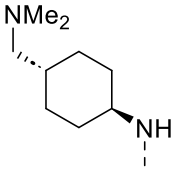
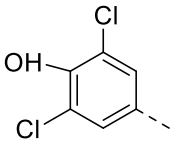
No.	R_1	R_2	R_3	pEC_{50} (Exp)	pEC_{50} GA-MLR	pEC_{50} GA-SVM
7				8.221	8.149	8.201
8				8.522	8.204	8.380
9				8.301	8.257	8.160
10				8.221	8.406	8.219
11				8.522	8.443	8.380
12[a]				6.193	6.899	6.998
13				6.450	7.368	6.590
14				7.585	7.804	7.725

Table 1. Continued…

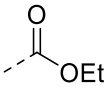
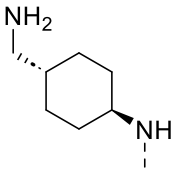
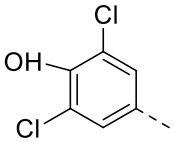
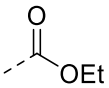
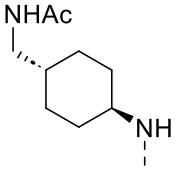
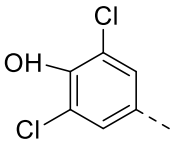
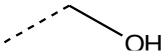
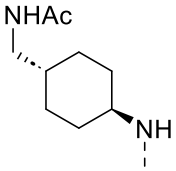
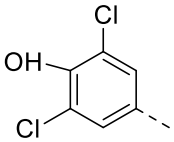
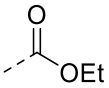
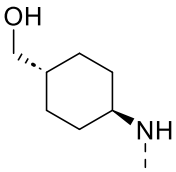
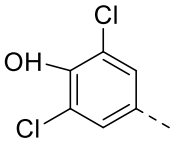
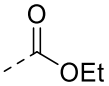
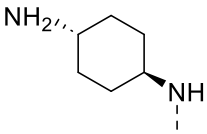
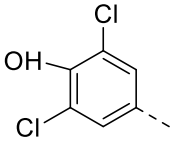
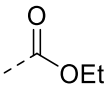
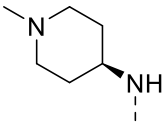
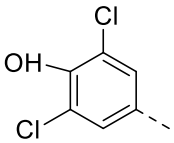
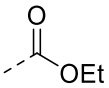
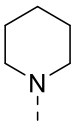
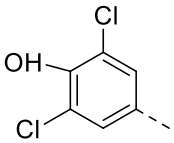
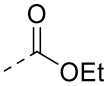
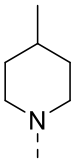
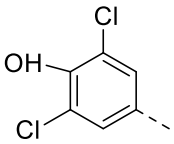
No.	R_1	R_2	R_3	pEC_{50} (Exp)	pEC_{50} GA-MLR	pEC_{50} GA-SVM
15				8.096	7.898	7.950
16				7.823	7.720	7.680
17[a]				7.795	7.355	6.857
18				7.000	7.836	7.879
19[a]				8.045	7.868	7.964
20				8.221	7.842	8.032
21[a]				6.692	6.694	6.901
22				6.910	7.180	7.050

Table 1. Continued…

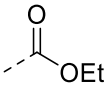
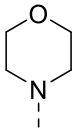
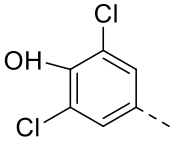
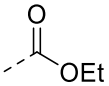
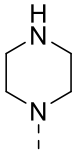
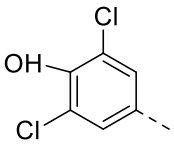
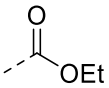
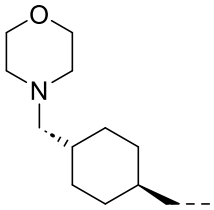
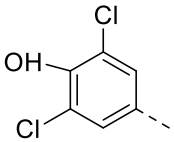
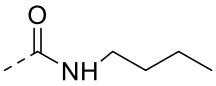
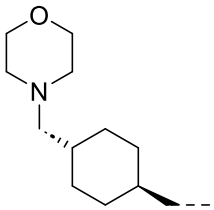
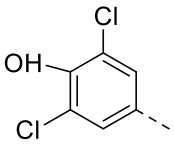
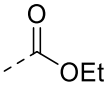
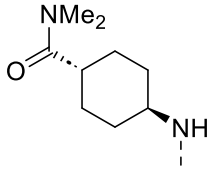
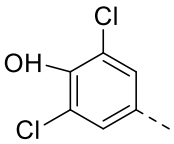
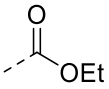
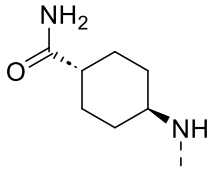
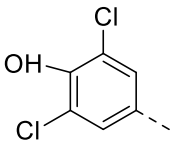
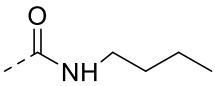
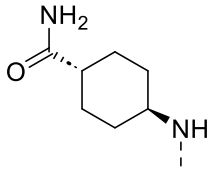
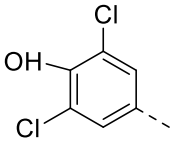
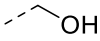
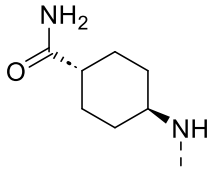
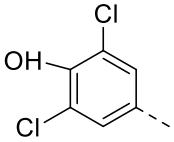
No.	R_1	R_2	R_3	pEC_{50} (Exp)	pEC_{50} GA-MLR	pEC_{50} GA-SVM
23				6.978	6.925	7.039
24				7.136	6.781	6.990
25				8.154	8.063	8.010
26				8.397	8.217	8.250
27				5.902	6.226	6.042
28				7.301	7.146	7.252
29				6.756	7.405	6.896
30				6.168	6.904	6.308

Table 1. Continued…

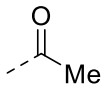
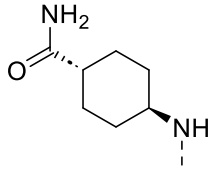
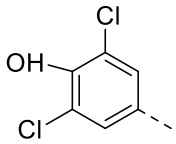
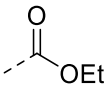
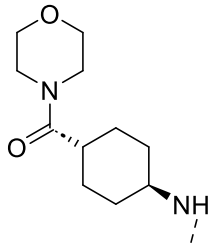
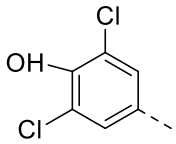
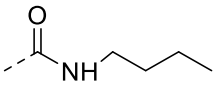
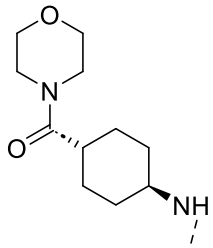
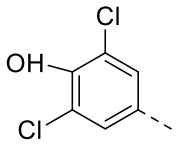
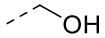
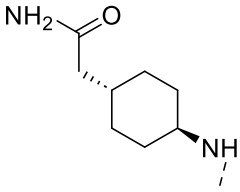
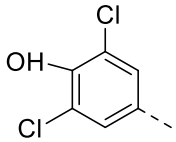
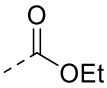
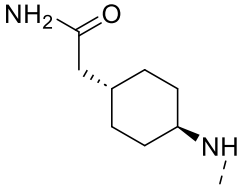
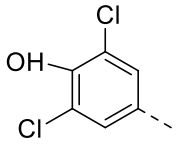
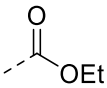
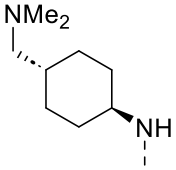
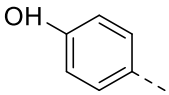
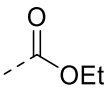
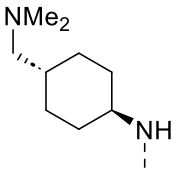
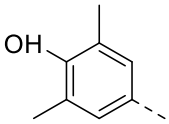
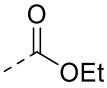
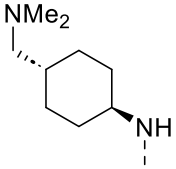
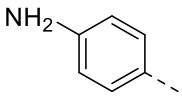
No.	R_1	R_2	R_3	pEC_{50} (Exp)	pEC_{50} GA-MLR	pEC_{50} GA-SVM
31				7.958	7.390	7.810
32				7.013	6.436	6.870
33				7.251	7.159	7.110
34[a]				7.130	7.167	6.964
35				7.366	7.642	7.506
36				6.522	6.077	6.252
37				7.920	7.223	7.700
38[a]				5.678	6.067	6.245

Table 1. Continued…

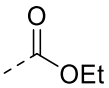
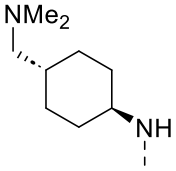
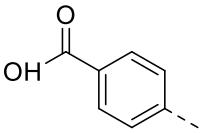
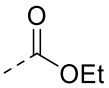
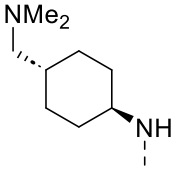
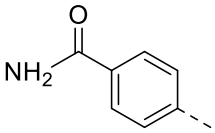
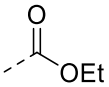
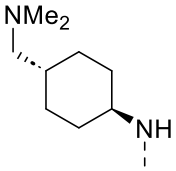
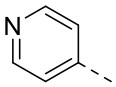
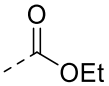
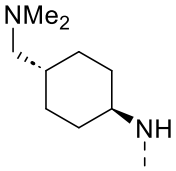
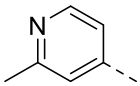
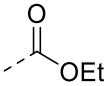
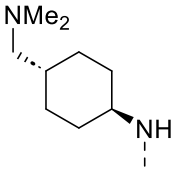
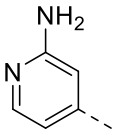
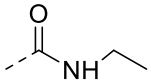
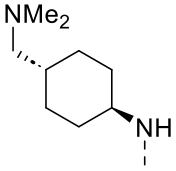
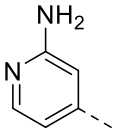
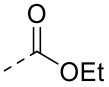
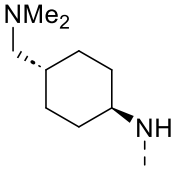
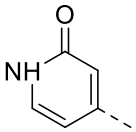
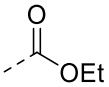
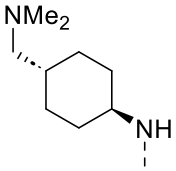
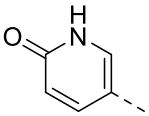
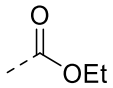
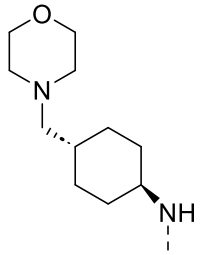
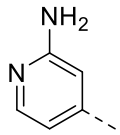
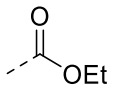
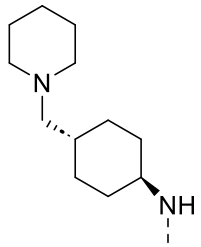
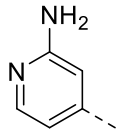
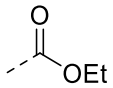
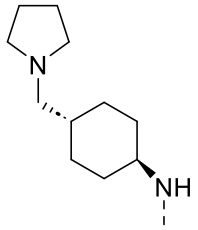
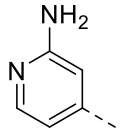
No.	R_1	R_2	R_3	pEC_{50} (Exp)	pEC_{50} GA-MLR	pEC_{50} GA-SVM
39				6.004	5.843	5.860
40				5.392	5.778	5.737
41				6.651	6.345	6.522
42				5.815	5.834	5.955
43				7.229	6.937	7.080
44[a]				7.337	6.872	6.821
45				6.416	6.800	6.556
46				5.874	6.162	6.065

Table 1. Continued…

No.	R_1	R_2	R_3	pEC_{50} (Exp)	pEC_{50} GA-MLR	pEC_{50} GA-SVM
47				7.431	7.118	7.290
48				6.943	7.045	7.083
49[a]				6.966	6.930	7.107

[a]: test

Descriptors calculation

Hyperchem 7.5 software was used to draw the 2D chemical structures of the 49 molecules, and then molecular mechanics force field (MM+) and semi-empirical method (AM1) were used for pre-optimization and optimization, respectively [20]. The molecular configurations were fine-tuned to achieve a root mean square gradient of 0.01 kal/mol. After that, DRAGON v2.2, which utilizes minimum-energy molecular geometries, was employed for the acquisition of molecular descriptors.

For every molecule within the dataset, a total of 1481 descriptors were computed consisting of 0D descriptors (constitutional), 1D descriptors (atom-centered fragments, functional group counts), 2D descriptors (such as topological, walk, and path counts, Burden eigenvalues, and topological charge indices descriptors), 3D descriptors (such as RDF, WHIM, GETAWAY, and 3D-MoRSE descriptors) and the others [21,22].

Following an analysis for constant or near constant variables, several constant or nearly constant values descriptors were eliminated from the computed descriptors.

Likewise, only the descriptor with the highest correlation with the pEC_{50} will be used in further development of the QSAR models among those with correlation coefficients over 0.90. Subsequently, the remaining molecular descriptors (481) were organized in the n x m data matrix, with n and m denoting the compounds and descriptors quantity, respectively.

Variable selection

Based on the objective function, genetic algorithms were employed for the purpose of identifying the most pertinent descriptors

[23,24]. The initial step in performing genetic algorithms involves generating a considerable amount of randomly selected variables in the context of chromosomes for the genetic algorithm [13]. Subsets of variables selected for this analysis are then tested by their fitness for forecasting inhibitory activity levels. The fitness function utilized in the genetic algorithm was defined as the cross-validation correlation coefficient of the leave-one-out method (Q^2_{LOO} derived using MLR) [25]. After excluding the worst subsets, the remaining subsets will be bred. The mutation is finally taking place. The genesis of the genetic algorithm can be attributed to Leardi *et al.* [24] and has become one of the most efficient methods for the selection of variables in recent years.

The implementation of the genetic algorithm method was conducted in the Matlab 6.5 program [26] to serve as a selection tool in this project. To correlate among the chosen descriptors, using the genetic algorithm, with biological response, MLR, and SVM methods were employed. Matlab 6.5 program implements both MLR and SVM methods [26].

Results and Discussion

A total of 49 compounds were partitioned into two groups with 80% and 20% ratios, respectively. There were 39 compounds used in the training set and 10 compounds used in the test set. Despite the fact that a random split was performed on the data set, the distribution of structural diversity and biochemical data was one of the objectives when choosing the compounds in the test set. After building the model with the training set, the predictability of the model was tested with some series of compounds.

GA-MLR method

After the selection of suitable descriptors by the genetic algorithm, multiple linear regression method was performed on the training data and the outcomes were assessed through the test data.

Six descriptors were selected using genetic algorithms: EEig03r, GGI6, GGI7, RDF145m, Mor13m, and HATS6p in which contribute to the EC_{50}. To guarantee the independent nature of the chosen descriptors, a correlation matrix (Table 2) involving their correlation coefficients is needed. Based on Table 2, these variables behave independently in the models due to their low correlation coefficients. In this case, the maximum numerical correlation coefficient observed between two descriptors is 0.511.

Variation of inflation factors (VIF) [27] is another important parameter for evaluating molecular descriptors, which helps determine if each descriptor has multi-collinearity. The VIF is described as Equation 1:

$$VIF = \frac{1}{1-r2} \quad (1)$$

The correlation coefficient 'r' is used to express the correlation coefficients between each variable and the others in the QSAR model. VIF values between 1 and 5 are considered acceptable and predictive for models. A value of 1 indicates no inter-correlation. In the case of a VIF value over 10.0, the model becomes unstable and unacceptable. In Table 2, we show correlation coefficients and VIF values based on GA-MLR for selected descriptors. VIF values under 2 are shown in Table 2, confirming the predictability of suggested models based on these descriptors. A predictive QSAR model was developed with six descriptors using GA-MLR analysis, represented as Equation 2:

$$pEC_{50} = 25.341(+3.669) - 6.486\ (+1.010)\ EEig03r - 3.004\ (+0.980)\ GGI6 + 8.094\ (+1.003)\ GGI7 + 0.216\ (+0.055)\ RDF145m + 0.5337(+0.181)\ Mor13m + 21.175(+5.856)\ HATS6p \quad (2)$$

N_{train}=39, R^2_{train}=0.792, R^2_{test}= 0.871, R^2_{adj}= 0.753, F_{train}=20.39, F_{test}=1.706, Q^2_{Loo}=0.680, Q^2_{LGO}=0.588

Table 2. The correlation coefficient between chosen descriptors and their respective VIF values as determined through GA-MLR

	EEig03r	GGI6	GGI7	RDF145m	Mor13m	HATS6p	VIF
EEig03r	1	0	0	0	0	0	1.254
GGI6	-0.002	1	0	0	0	0	1.382
GGI7	0.321	0.474	1	0	0	0	1.805
RDF145m	0.329	0.313	0.511	1	0	0	1.722
Mor13m	-0.054	-0.082	-0.028	-0.020	1	0	1.012
HATS6p	-0.013	-0.279	-0.454	-0.5455	0.176	1	1.550

Table 3. Statistical results of different QSAR models

	Training			Test		
	R2	**RMSE**	**F**	**R2**	**RMSE**	**F**
GA-MLR	0.792	0.388	20.399	0.871	0.338	1.706
GA-SVM	0.946	0.203	78.641	0.688	0.486	0.682

The symbol N denotes the number of molecules present in the training dataset, and Q^2_{Loo} and Q^2_{LGO} represent the cross-validation coefficients for leaving one out and leaving a group out (usually, 20% of molecules are excluded), respectively. The built model exhibits remarkable reliability based on Q^2_{Loo}'s value (0.680). R^2_{adj}, R^2 and F are a squared correlation coefficient, adjusted correlation coefficient and a Fisher F statistics, respectively. A statistical model for GA-MLR is presented in Table 3. Based on the calculated R^2 values for both sets, the test set clearly showed better results. Model predictive capability is demonstrated by low root mean square errors ($RMSE_{train}$ = 0.388 and $RMSE_{test}$ = 0.338) and high R^2 and F values. According to GA-MLR model, Table 1 presents predicted inhibitory activities for whole molecules. Figures 1 and 2 demonstrate the prediction and residual plots, respectively. According to Figure 2, the GA-MLR method does not produce systematic errors.

A Y-randomization test was executed to assess the robustness of the constructed model. In this approach, values of pEC_{50} are shuffled, and a novel model was constructed utilizing randomized data. The validation of the effectiveness of the primary derived model necessitates the new models to exhibit lower R^2 and Q^2_{Loo} values. According to Table 4, the values below 0.32 indicate that it is impossible to attribute the goodness of the built model to chance. An examination for potential outliers within this dataset is crucial; we visualized the domain of applicability with the William plot. Williams plot is depicted in Figure 3. The metric known as the warning leverage (h*) is defined by Equation 3:

$$h^* = 3p/n \quad (3)$$

Table 4. Y-randomisation tests

No.	Q^2	R^2
1	0.030	0.298
2	0.068	0.105
3	0.059	0.084
4	0.089	0.316
5	0.009	0.199
6	0.044	0.116
7	0.067	0.094
8	0.012	0.117
9	0.000	0.204
10	0.024	0.232

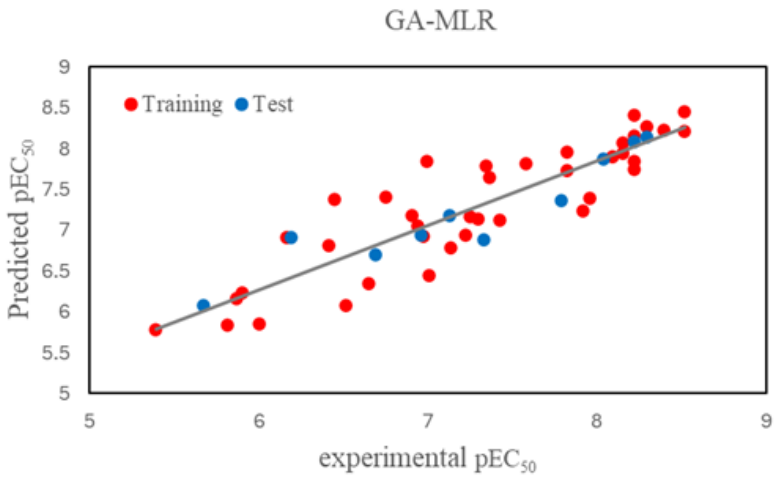


Figure 1. The plot of predicted versus experimental pEC_{50} values by the GA-MLR model.

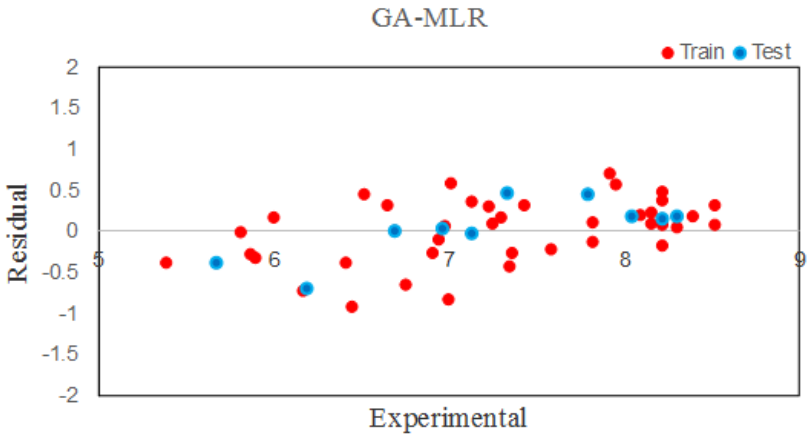


Figure 2. The plot of residual vs. the experimental pEC_{50} values (GA-MLR model).

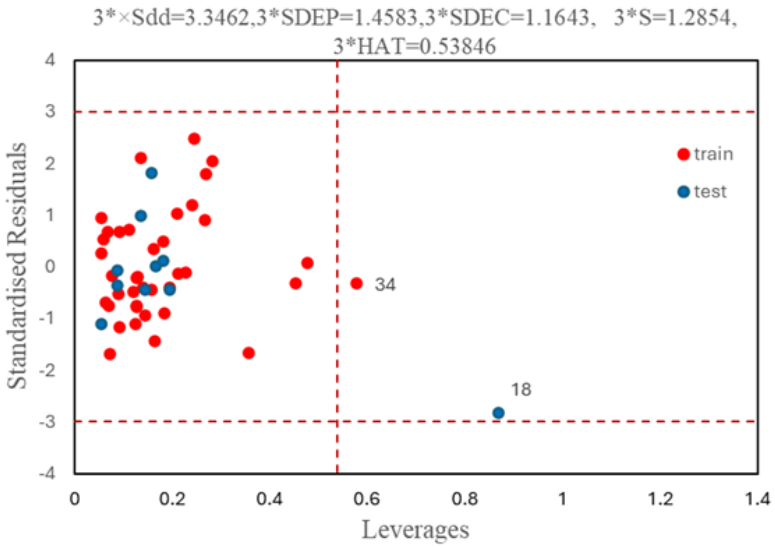


Figure 3. The Williams plot of GA-MLR model for the training and test sets.

Where, n signifies the calibration compounds quantity, while p represents the model variables quantity plus one. If a compound possesses a leverage (h) exceeding the warning leverage (h*), it indicates that the compound holds significant influence. A cut-off value of three standardized residuals is commonly used to accept predictions since it covers about 99% of normally distributed data. Figure 3 shows that two compounds, 34 and 18, exhibit leverage (h) values greater than the warning h* value of 0.538. Therefore, they are structural outliers.

GA - SVM method

A nonlinear model was also established using the SVM technique with the same chosen descriptors and was compared to the GA-MLR model. The results of both methods were summarized in Table 3. Within SVM regression, various factors are taken into account, such as the type of kernel function, the capacity parameter, ε -insensitive loss function, and its related parameters [28]. Sample distribution in space is determined by the Karnel function type. Thus, it is necessary to declare a Karnel function type. Due to its good performance, the radial basis function (RBF) was applied [29]. The RBF is defined by the mathematical expression denoted as Equation 4:

$$\exp(-\gamma * |u-\upsilon|^2) \quad (4)$$

In this particular formula, γ represents a kernel parameter while u and υ are considered as independent variables. The parameter γ plays a crucial role in regulating the Radial Basis Function (RBF) and holds direct influence over the performance of Support Vector Machines (SVM) as well as the duration required for training. To enhance the γ parameter, a method involving cross-validation utilizing leave-one-out technique was implemented on the initial training dataset to execute a thorough grid search.

To determine the optimal value of γ, incremental steps of 0.1 were taken from 0.1 to 10. Cross-validation RMSEs were additionally ascertained. Figure 4 presents a plot of gamma (γ) parameter values against RMSE of cross-validation, showing that gamma (γ) has an optimal value of 6.3.

Due to the presence of the ε-insensitive parameter, the entirety of the training set may not satisfy boundary constraints, thus allowing for sparsity within the dual formulation's resolution. The optimal values of this parameter vary depending on the noise type found in the data.

Based on the different values of ε, the cross-validation RMSE varies from 0.01 to 1.0 in increments of 0.01. The ε -insensitive values are depicted as a function of the achieved RMSE of cross-validation in Figure 5. According to this figure, the optimal value for this parameter is 0.14.

Another crucial parameter in SVM modeling is the C parameter, which governs the balance between maximizing margins and minimizing training inaccuracies. From 1 to 100, parameter C was incrementally increased by 1 until it reached an optimal value, as shown in Figure 6. The findings derived from the analysis presented in Figure 6 indicate that 97 is the optimal capacity parameter. Figure 7 and Table 1 show the results of predicting the pEC_{50} using GA-SVM. Using the above analysis, the optimum values for constructing a SVM model were determined as follows: C = 97, ε = 0.14, γ = 6.3. A statistical analysis of the optimal model for the training set (R^2 =0.946, F=78.641, RMSE=0.203) and test set (R^2 =0.688, F=0.682, RMSE=0.486) indicates a good predictive capability. The training set compounds performed better in prediction compared to GA-MLR (Table 3). The GA-SVM model exhibits better performance than the GA-MLR model for the training set, showcasing lower RMSE alongside higher F and R^2 values, whereas GA-MLR gave remarkable results for

test set compared to GA-SVM. Likewise, SVM-based genetic algorithms can be applied to predict inhibitory activity of DYRK1A inhibitors using the GA-SVM method developed.

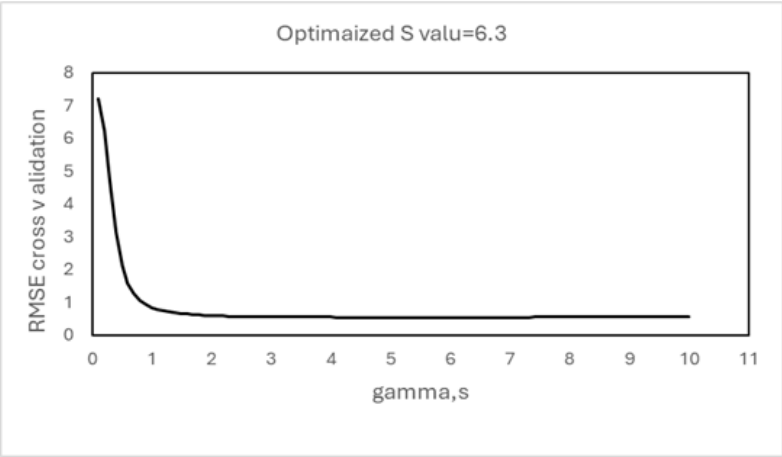


Figure 4. The gamma(γ) vs. RMSE for the training set

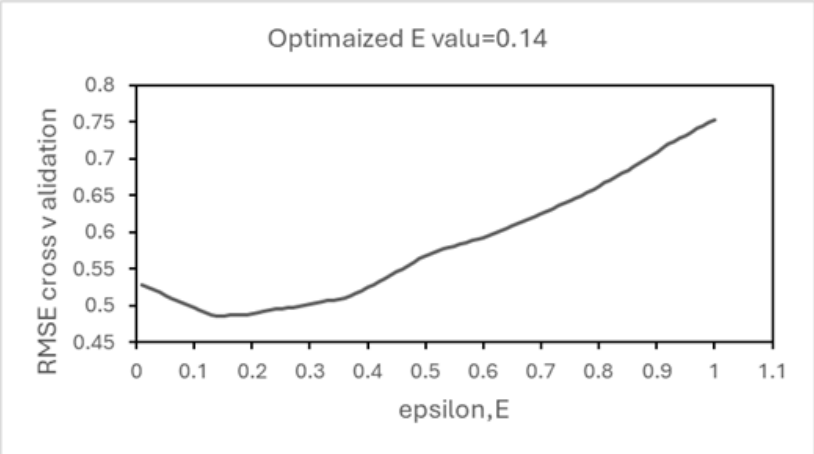


Figure 5. The epsilon (ε) vs. RMSE for the training set

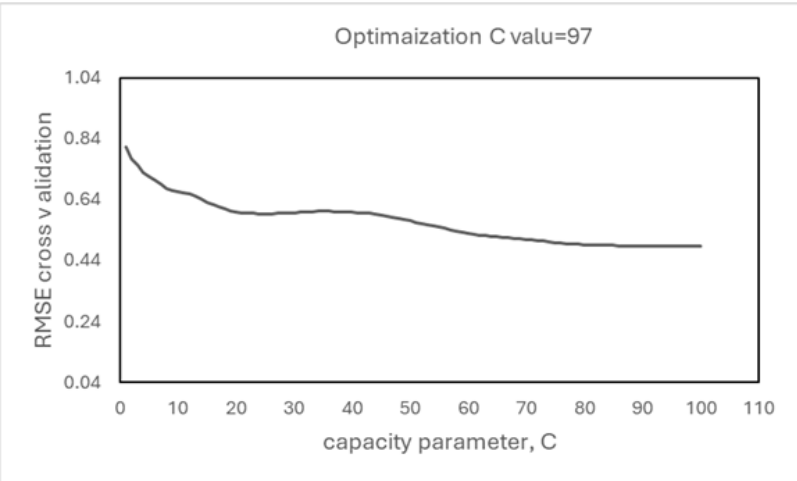


Figure 6. The capacity parameter(C) vs. RMSE for the training set

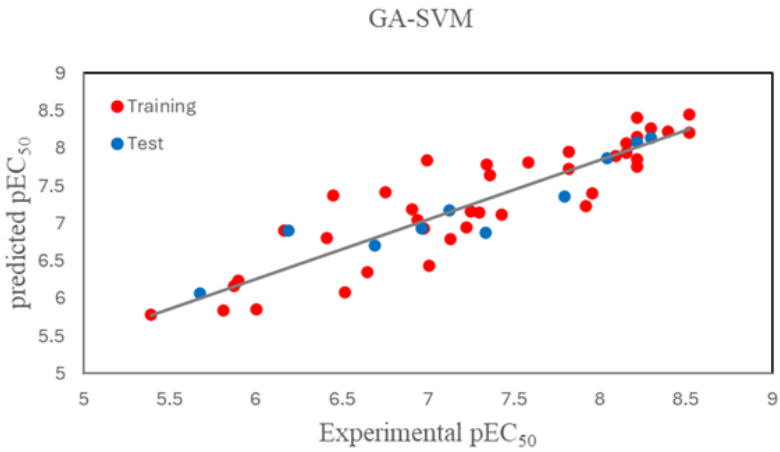


Figure 7. The predicted versus experimental pEC_{50} plot by GA-SVM

Molecular descriptors of the proposed models: discussion

Analyzing the mechanism of inhibition and developing new drugs with higher inhibitory activities can be achieved through an analysis of the selected descriptors and their respective effects on inhibitory activity. The proposed models consisted of six descriptors: EEig03r, GGI6, GGI7, RDF145m, Mor13m, and HATS6p.

The first descriptor of the established model is EEig03r which represents Eigenvalue 03 from edge adj. matrix weighted by resonance integrals and belonging to the Edge adjacency indices. According to Equation 2, EEig03r descriptor with negative sign indicates that pEC_{50} is inversely related to it. The second and third descriptors in the model are GGI6 and GGI7 which describe topological charge index of order 6 and topological charge index of order 7, respectively. There is an inverse relationship between GGI6 and the dependent variable (pEC_{50}) when its sign is negative and the positive sign of GGI7 indicates that the pEC_{50} is directly related to this descriptor.

The next descriptor in the proposed model is RDF145m which represents the Radial distribution function – 14.5 / weighted by atomic masses. The RDF indicates the requirements for compound 3D structures [30]. Descriptors of this type are independent of atom number, for example, at the size of a molecule. Moreover, RDF descriptors can be used to show specific information in a particular 3D structure space based on specific atom types or distance ranges. In RDF descriptors, distance distributions are used as a basis for the descriptors. This descriptor describes the weighting schemes based on atomic masses. It is evident from Equation 2 that a positive value for this descriptor is directly related to the pEC_{50} value, and an increase in inhibitory activity can be achieved by increasing the mass and distribution of a specific group of atoms. The fifth descriptor chosen, Mor13m, signifies a 3D-Morse descriptor weighted by atomic masses. Positive signs also accompany this descriptor. By examining the distance distribution in the geometric depiction of molecules, the 3D-MoRSE descriptors play a role in creating the radial distribution function code and are evaluated based on the sum of atomic weights during divergent angular scattering [31].

The final descriptor is HATS6p (leverage-weighted autocorrelation of lag 6/weighted by polarizability) which is among the GETAWAY descriptors. These descriptors can provide significant information regarding substituents and fragments within molecules [32,33]. HATS6p has a positive sign, indicating an increase in its value would increase pEC_{50}.

Conclusion

This study utilized support vector machine and multiple linear regression techniques to analyze QSAR for a series of compounds that acts as highly potent DYRK1A-dependent replicators. To select the most relevant descriptors, the algorithm genetic method was applied. Based on the results of validation methods including cross-validation and Y- randomization, the built models appear to be accurate and strong. When compared to GA-MLR, the GA-SVM approach offers more precise predictions for compounds within the training set. This study demonstrates that utilizing QSAR models can aid in forecasting the activity of novel compounds acting as DYRK1A inhibitors, and also provide insight into how to develop more potent inhibitors for diabetes treatment.

Conflict of interest

No potential conflicts of interest were disclosed.

Orcid

Faezeh Khosravi: 0009-0005-0018-0851
Roya Kiani-Anbouhi: 0009-0007-2344-9010
Eslam Pourbasheer: 0000-0001-8969-4341

References

[1] S. Mustofa, V. Anisya, Type 1 and 2 diabetes mellitus: A review on current treatment approach and gene therapy as potential intervention, *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, **2020**, *4*, 12-17. [Crossref], [Google Scholar], [Publisher]

[2] A.E. Butler, J. Janson, S. Bonner-Weir, R. Ritzel, R.A. Rizza, P.C. Butler, *β*-Cell deficit and increased *β*-cell apoptosis in humans with type 2 diabetes, *Diabetes*, **2003**, *52*, 102-110. [Crossref], [Google Scholar], [Publisher]

[3] P. Wang, N.M. Fiaschi-Taesch, R.C. Vasavada, D.K. Scott, A. Garcia-Ocana, A.F. Stewart, Diabetes mellitus-advances and challenges in human *β*-cell proliferation, *Nature Reviews Endocrinology*, **2015**, *11*, 201-212. [Crossref], [Google Scholar], [Publisher]

[4] E. Ferrannini, The stunned *β* cell: A brief history, *Cell metabolism*, **2010**, *11*, 349-352. [Google Scholar]

[5] M. Campbell-Thompson, A. Fu, J.S. Kaddis, C. Wasserfall, D.A. Schatz, A. Pugliese, M.A. Atkinson, Insulitis and *β*-cell mass in the natural history of type 1 diabetes, *Diabetes*, **2016**, *65*, 719-731. [Crossref], [Google Scholar], [Publisher]

[6] L. El Mchichi, A. El Aissouq, R. Kasmi, A. Belhassan, R. El-Mernissi, A. Ouammou, T. Lakhlifi, M. Bouachrine, In silico design of novel pyrazole derivatives containing thiourea skeleton as anti-cancer agents using: 3D QSAR, drug-likeness studies, ADMET prediction and molecular docking, *Materials Today: Proceedings*, **2021**, *45*, 7661-7674. [Google Scholar]

[7] A. Abdolmaleki, J. B Ghasemi, F. Ghasemi, Computer aided drug design for multi-target drug design: SAR/QSAR, molecular docking and pharmacophore methods, *Current Drug Targets*, **2017**, *18*, 556-575. [Google Scholar], [Publisher]

[8] A. Beheshti, E. Pourbasheer, M. Nekoei, S. Vahdani, QSAR modeling of antimalarial activity of urea derivatives using genetic algorithm–multiple linear regressions, *Journal of Saudi Chemical Society*, **2016**, *20*, 282-290. [Crossref], [Google Scholar], [Publisher]

[9] E. Pourbasheer, R. Aalizadeh, 3D-QSAR and molecular docking study of LRRK2 kinase inhibitors by CoMFA and CoMSIA methods, *SAR and QSAR in Environmental Research*, **2016**, *27*, 385-407. [Crossref], [Google Scholar], [Publisher]

[10] M. Zivkovic, M. Zlatanovic, N. Zlatanovic, M. Golubović, A.M. Veselinović, The application of the combination of Monte Carlo optimization method based QSAR modeling and molecular

docking in drug design and development, *Mini Reviews in Medicinal Chemistry*, **2020**, *20*, 1389-1402. [Crossref], [Google Scholar], [Publisher]

[11] N.R. Draper, H. Smith, Applied regression analysis, *John Wiley & Sons*, **1998**. [Google Scholar], [Publisher]

[12] R.R. Hocking, A biometrics invited paper. The analysis and selection of variables in linear regression, *Biometrics*, **1976**, 1-49. [Crossref], [Google Scholar], [Publisher]

[13] A. Arbor, J. Holland, Adaptation in natural and artificial systems, *The University of Michigan Press: Ann Arbor, MI, USA*, **1975**. [Google Scholar], [Publisher]

[14] Q. Shen, Q.Z. Lü, J.H. Jiang, G.L. Shen, R.Q. Yu, Quantitative structure–activity relationships (QSAR): Studies of inhibitors of tyrosine kinase, *European Journal of Pharmaceutical Sciences*, **2003**, *20*, 63-71. [Crossref], [Google Scholar], [Publisher]

[15]. T. Yang, Z. Yang, F. Pan, Y. Jia, S. Cai, L. Zhao, L. Zhao, O. Wang, C. Wang, Construction of an MLR-QSAR model based on dietary flavonoids and screening of natural *α*-glucosidase inhibitors, *Foods*, **2022**, *11*, 4046. [Crossref], [Google Scholar], [Publisher]

[16] R. Kiani-Anbouhi, M.R. Ganjali, P. Norouzi, Prediction of the complexation stabilities of La 3+ ion with ionophores applied in lanthanoid sensors, *Journal of Inclusion Phenomena and Macrocyclic Chemistry*, **2014**, *78*, 325-336. [Crossref], [Google Scholar], [Publisher]

[17] T.W. Quadri, L.O. Olasunkanmi, O.E. Fayemi, E.D. Akpan, C. Verma, E.-S.M. Sherif, K.F. Khaled, E.E. Ebenso, Quantitative structure activity relationship and artificial neural network as vital tools in predicting coordination capabilities of organic compounds with metal surface: A review, *Coordination Chemistry Reviews*, **2021**, *446*, 214101. [Crossref], [Google Scholar], [Publisher]

[18] E. Pourbasheer, R. Aalizadeh, M.R. Ganjali, QSAR study of CK2 inhibitors by GA-MLR and GA-SVM methods, *Arabian Journal of Chemistry*, **2019**, *12*, 2141-2149. [Crossref], [Google Scholar], [Publisher]

[19] P.A. Allegretti, T.M. Horton, Y. Abdolazimi, H.P. Moeller, B. Yeh, M. Caffet, G. Michel, M. Smith, J.P. Annes, Generation of highly potent DYRK1A-dependent inducers of human *β*-cell replication via multi-dimensional compound optimization, *Bioorganic & Medicinal Chemistry*, **2020**, *28*, 115193. [Crossref], [Google Scholar], [Publisher]

[20] D. Laxmi, S. Priyadarshy, HyperChem 6.03. *Biotech Software & Internet Report: The Computer Software Journal for Scientists*, **2002,** *3*, 5-9. [Crossref], [Google Scholar], [Publisher]

[21] R.R. Todeschini, V. Consonni, Handbook of molecular descriptors, *John Wiley & Sons*, **2008**. [Google Scholar]

[22] R. Todeschini, V. Consonni, A. Mauri, M. Pavan, DRAGON for Windows (Software for Molecular Descriptor Calculations), *Milano Chemometrics and QSAR Research Group: Milano*, **2005**, *5*. [Google Scholar]

[23] C.L. Waller, M.P. Bradley, Development and validation of a novel variable selection technique with application to multidimensional quantitative structure–activity relationship studies, *Journal of Chemical Information and Computer Sciences*, **1999**, *39*, 345-355. [Crossref], [Google Scholar], [Publisher]

[24] J. Aires-de-Sousa, M.C. Hemmer, J. Gasteiger, Prediction of ^{1}H NMR chemical shifts using neural networks, *Analytical Chemistry*, **2002**, *74*, 80-90. [Crossref], [Google Scholar], [Publisher]

[25] R. Leardi, R. Boggia, M. Terrile, Genetic algorithms as a strategy for feature selection, *Journal of Chemometrics*, **1992**, *6*, 267-281. [Crossref], [Google Scholar], [Publisher]

[26] G.A. Mathworks, Direct search toolbox users guide, *The Mathworks Inc., Natick, MA, USA*, **2005**. [Google Scholar],

[27] V. Agrawal, P. Khadikar, QSAR prediction of toxicity of nitrobenzenes, *Bioorganic & Medicinal Chemistry*, **2001**, *9*, 3035-3040. [Crossref], [Google Scholar], [Publisher]

[28] V.N. Vapnik, V. Vapnik, Statistical learning theory, **1998**. [Google Scholar]

[29] E. Pourbasheer, R. Aalizadeh, M.R. Ganjali, P. Norouzi, QSAR study of IKKβ inhibitors by the genetic algorithm: Multiple linear regressions, *Medicinal Chemistry Research*, **2014**, *23*, 57-66. [Crossref], [Google Scholar], [Publisher]

[30] R. Todeschini, V. Consonni, Handbook of molecular descriptors, John Wiley & Sons, **2008**. [Google Scholar], [Publisher]

[31] S. Gosav, M. Praisler, D. Dorohoi, ANN expert system screening for illicit amphetamines using molecular descriptors, *Journal of Molecular Structure*, **2007**, *834*, 188-194. [Crossref], [Google Scholar], [Publisher]

[32] V. Consonni, R. Todeschini, M. Pavan, Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors, *Journal of Chemical Information and Computer Sciences*, **2002**, *42*, 682-692. [Crossref], [Google Scholar], [Publisher]

[33] M.P. Gonzalez, C. Teran, L. Saiz-Urra, M. Teijeira, Variable selection methods in QSAR: An overview, *Current Topics in Medicinal Chemistry*, **2008**, *8*, 1606-1627. [Crossref], [Google Scholar], [Publisher]

HOW TO CITE THIS ARTICLE

F. Khosravi, R. Kiani-Anbouhi, E. Pourbasheer. QSAR Study on DYRK1A Inhibitors for Regenerative Therapy in Diabetes. *Adv. J. Chem. A*, 2024, 7(5), 522-539.

DOI: 10.48309/ajca.2024.458482.1530
URL: https://www.ajchem-a.com/article_196569.html