

Research Article

Linear and Nonlinear QSAR Study of N2 and O6 Substituted Guanine Derivatives as Cyclin-Dependent Kinase 2 Inhibitors

Nasser Goudarzi, M. Arab Chamjangali, and Payam Kalhor

Faculty of Chemistry, Shahrood University of Technology, P.O. Box 316, Shahrood 3619995161, Iran

Correspondence should be addressed to Nasser Goudarzi; goudarzi@shahroodut.ac.ir

Received 11 April 2013; Accepted 23 May 2013

Academic Editors: J. N. Latosinska and C. Y. Panicker

Copyright © 2013 Nasser Goudarzi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The inhibitory activities (pIC_{50}) of N2 and O6 substituted guanine derivatives as cyclin-dependent kinase 2 (CDK2) inhibitors have been successfully modeled using calculated molecular descriptors. Two linear (MLR) and nonlinear (ANN) methods were utilized for construction of models to predict the pIC_{50} activities of those compounds. The QSAR models were validated by cross-validation (leave-one-out) as well as application of the models for prediction of pIC_{50} of external set compounds. Also, the models were validated by calculation of statistical parameters and Y-randomization test. Two methods provided accurate predictions, although more accurate results were obtained by ANN model. The mean-squared errors (MSEs) for validation and test sets of MLR are 0.065, 0.069 and of ANN are 0.017 and 0.063, respectively.

1. Introduction

The cyclin-dependent kinases (CDKs) are a class of enzymes which play a fundamental role in cell cycle regulation [1, 2]. Particularly as their name suggests CDKs activation partially depends on the binding of another class of proteins named cyclins, for example, cyclins of the D family complex with CDK4 and CDK6 during G1 phase, cyclin E with CDK2 in late G1, cyclin A with CDK2 in S phase, and cyclin B with CDK1 (also known as cdc2) in late G2/M. Then, aberrant CDK control and consequent loss of cell cycle check point function have been directly linked to the molecular pathology of cancer [3]. It is well known that phosphorylation in a conserved threonine residue of the CDK subunit is required for its complete activation. This task is performed by the CDK activating kinase. These proteins properly regulate the cell cycle progress and DNA synthesis only as an active complex (T160pCDK/cyclin) [4]. Overall, the activity of the CDK/cyclin complex can be depleted by at least two different mechanisms that contain the phosphorylation of the CDK subunit at the inhibitory sites or the binding of the specialized natural inhibitors known as CDK inhibitors. In the first mechanism, the amino acid residue Y15 and to a lesser extend

T14 (in CDK2) are phosphorylated by human Wee 1 Hu [5]. This inhibitory phosphorylation is independent of previous cyclin binding [6]. The second mechanism involves the binding of natural CDK inhibitors. Four major mammalian CDK inhibitors have been discovered: P21 (CIP1/WAF190) and P27 (KIP1) inactive CDK2 and CDK4 cyclin complexes by binding to them. The two other inhibitors are P16_{INK4} and P15_{INK4B} that are specific for CDK4 and CDK6. They inhibit the formation of the active cyclin complexes by binding to the inactive CDK, and they can also bind to the active complex [2, 7]. However, it has been shown that natural CDK inhibitors are subexpressed in some carcinogenic cells, and medicinal chemists have put some of their effort in the search for new synthetic inhibitors to replace them [8–12]. Some of them have entered in clinical field; for instance, flavopiridol induces cell cycle arrest and tumor growth inhibition [13].

The search for more potent and selective CDK inhibitors is a daunting challenge due to the similarity of the ATP binding site along the different CDK subtypes [14]. According to this, the development and use of new strategies to overcome this problem are urgently needed. Nowadays, new and exciting strategies have emerged and become available to

find more potent and selective inhibitors, and they normally use quantitative structure-activity relationships (QSARs) derived from different computational calculation approaches [15–19].

Quantitative structure-activity/property relationship (QSAR/QSPR) was used for correlation of different activities and properties to characteristics of molecular structures. In recent years, several QSAR and QSPR models based on both linear and nonlinear methods that aimed to predict different activities and properties were used [20–30]. The reliable prediction of inhibition of CDK2 has an important role in medicinal researches. The ultimate role of the different formulations of the QSAR theory is to suggest mathematical models for estimating relevant endpoints of interest, especially when these cannot be experimentally determined for some reason. These studies simply rely on the assumption that the structure of a compound determines the related activity. The molecular structure is therefore translated into the so-called molecular descriptors through mathematical formulae obtained from several theories, such as chemical graph theory, information theory, and quantum mechanics [31, 32]. In this work, we carried out a QSAR study by predicting the inhibitory activity of a set of N2 and O6 substituted guanine derivatives by using multiple linear regression (MLR) and artificial neural network (ANN) dealing with linearity and nonlinearity, respectively.

2. Basic Theory

2.1. Multiple Linear Regression (MLR). The general goal of multiple linear regression (MLR) is to model the relationship between some independent variables and a dependent variable by fitting a linear equation to observed data. Generally, the multiple linear regression model is as the following equation:

$$y = b_1x_1 + b_2x_2 + b_3x_3 \cdots b_nx_n + \varepsilon, \quad (1)$$

where n is the number of independent variables, b_1, \dots, b_n are the regression coefficients, and y is the dependent variable [33].

2.2. Artificial Neural Network (ANN). ANNs have large numbers of computational units connected in a vast parallel construction. Neural networks do not need an obvious formulation of the handled problem. They act as a means to introduce scaled data to the network. The data from the input layer (input neurons) propagate through the network via interconnections. Scalar weights are specialized to each connection. A remarkable aspect of the neural networks is their learning step. In this step, the value of weights and biases would be optimized based on a set of measured numerical values (training set). More details about neural networks are given in [34, 35].

3. Materials and Methods

3.1. Data. The experimental data used in the present study to model IC₅₀ were taken from [36]. The whole data set included

56 compounds, whose biological activities (pIC₅₀ values) were determined for inhibition of CDK2. It is worthy to say that pIC₅₀ values span a broad range from 4.11 to 7.22 M. In this work, the structure-activity model is generated by ANN and MLR. The names of these compounds, their experimental and calculated pIC₅₀ values by ANN and MLR methods, and also their values using leave-one-out are shown in Table 1. As can be seen, this set contains 56 inhibitory activity (pIC₅₀) data of CDK2s. The data set was split into training, validation, and test sets to increase the network's generalization. The training set of 34 compounds, with pIC₅₀ values ranging from 4.11 to 7.22, was used to construct the model. The validation set of 11 compounds, with pIC₅₀ values ranging from 4.19 to 6.62, was used to prevent overtraining/over fitting of the ANN model. The test set of 11 compounds, with pIC₅₀ values ranging from 4.19 to 6.96, was used as an external set to evaluate the predictive ability of the model.

3.2. Descriptor Generation and Screening. The inhibitory activation of compounds is related to some of their structural, electronic, and geometric properties. The value of these properties can be encoded quantitatively by numerical values named molecular descriptors. These molecular parameters are to be used to search for the best QSAR model of the inhibitory activation. The 2D structures of the molecules were drawn using Hyperchem8 software [37]. The molecular structures were optimized using the Polak-Ribiere algorithm until the root-mean-square gradient was 0.001 kcal mol⁻¹. The resulting geometry was transferred into the Dragon program package, and 1481 descriptors were produced [38]. Then, these descriptors were given to SPSS 17 for statistical work [39]. It is worth mentioning that in the first preselected analysis, some descriptors were removed because many of them included zero or other constant/near-constant values and did not have enough information of structure. On the other hand, to decrease the redundancy existing in the descriptor data matrix, the correlation coefficient R of the descriptors with each other (Pearson's correlation) was examined, and the collinear descriptors (with $R > 0.9$) were removed. By doing so, 238 descriptors were remained. Then by using the stepwise mode for regression, 14 models were given. With considering some parameters such as F , t , R , and standard error (SE), model number 14 containing 10 descriptors was used as MLR model to predict of pIC₅₀. These descriptors are 3D-MoRSE-signal 20/unweighted (Mor20u), Geary autocorrelation-lag 2/weighted by atomic Sanderson electronegativities (GATS2e), R autocorrelation of lag 5/weighted by atomic polarizabilities (R5p), Moran autocorrelation-lag 2/weighted by atomic Sanderson electronegativities (MATS2e), mean topological charge index of order 6 (JGI6), mean topological charge index of order 6 (JGI4), R autocorrelation of lag 5/weighted by atomic masses (Rlm), leverage-weighted autocorrelation of lag 0/weighted by atomic masses (HATS0m), 3D-MoRSE-signal 20/weighted by atomic van der Waals volumes (Mor06v), 1st component accessibility directional WHIM index/weighted by atomic electrotopological states (E1s). The name, class, and meaning of these descriptors are shown in Table 2.

TABLE 1: The names of compounds, their experimental and calculated pIC_{50} values by ANN and MLR methods, and also their values were calculated using leave-one-out.

No.	Name	pIC_{50}	ANN	MLR	ANN (LOO)	MLR (LOO)
1	6-Propoxy-9H-purin-2-amine	4.17	4.22	4.0983	4.2054	4.4244
2	6-Butoxy-9H-purin-2-amine	4.32	4.26	4.3	4.3073	4.3991
3	6-Pentyloxy-9H-purin-2-amine	4.31	4.31	4.27	4.4568	4.3289
4	6-Isopropoxy-9H-purin-2-amine	4.12	4.25	4.5953	4.3075	4.0767
5	6-Sec-butoxy-9H-purin-2-amine	4.6	4.38	4.31	4.4773	4.4061
6	6-Isobutoxy-9H-purin-2-amine	4.38	4.54	4.1	4.6696	4.6664
7	6-(2-Methylbutoxy)-9H-purin-2-amin	4.82	4.69	4.93	4.7491	4.7295
8	6-(Isopenthyloxy)-9H-purin-2-amin	4.59	4.49	4.505	4.5325	4.4889
9	6-(Hex-5-enyloxy)-9H-purin-2-amine	4.33	4.28	4.4022	4.964	4.4241
10	6-((E)-Hex-3-enyloxy)-9H-purin-2-amin	4.16	4.09	4.4578	4.0104	4.0491
11	6-(Allyloxy)-9H-purin-2-amine	4.11	4.22	4.1781	4.3261	3.7617
12	6-(2-Methylallyloxy)-9H-purin-2-amin	4.46	4.43	4.33	4.4437	4.4786
13	6-(2-Methylenebutoxy)-9H-purin-2-amin	4.68	4.56	3.5157	4.842	4.8973
14	6-(3-Methyl-2-ethylenebutoxy)-9H-purin-amin	4.8	4.86	4.67	5.0957	4.8305
15	6-(Cyclopentyl methoxy)-9H-purin-2-amin	4.68	4.69	5.0611	4.9247	4.7728
16	6-(Cyclopentenyl methoxy)-9H-purin-2-amin	4.51	4.65	4.42	4.7164	4.6733
17	6-(Cyclohexylmethoxy)-9H-purin-2-amin	4.66	4.49	4.6965	4.4498	4.4222
18	6-((Cyclohex-3-enyl)methoxy)-9H-purin-2-amin	4.8	4.71	5.0266	4.773	4.8678
19	6-(2-Cyclohexylethoxy)-9H-purin-2-amin	4.36	4.55	4.48	4.4313	4.482
20	6-(Benzyloxy)-9H-purin-2-amine	4.46	4.5	4.5365	4.512	4.4843
21	6-(Phenethyloxy)-9H-purin-2-amine	4.19	4.13	4.28	4.2465	4.1987
22	6-(2,2-Diethoxypropoxy)-9H-purin-2-amin	4.7	4.64	4.1429	5.0631	4.7004
23	6-((2-Isopropyl-1,3-dioxolan-2-yl)methoxy)-9H-purin-2-amine	4.19	4.26	4.04	4.5904	4.259
24	6-(Cyclohexylmethoxy)9H-purin-2-amin	4.77	4.92	4.7561	4.4964	4.5035
25	6-(Cyclohexylmethoxy)-N-phenyl-9H-purin-2-amin	6.01	6.19	5.826	6.0775	5.9886
26	6-(Cyclohexylmethoxy)-N-methyl-9H-purin-2-amin	5.3	5.18	4.92	5.2791	5.2741
27	6-(Cyclohexylmethoxy)-N-ethyl-9H-purin-2-amin	5.55	5.78	5.6238	5.2692	5.26
28	N-(3-Chlorophenyl)-6-(cyclohexylmethoxy)-9H-purin-2-amine	5.64	5.41	5.76	5.4811	5.7077
29	N-(3-Bromophenyl)-6-(cyclohexylmethoxy)9H-purin-2-amine	5.17	5.34	4.947	5.4014	5.3896
30	(3-(6-(Cyclomethylmethoxy)9H-purin-2-ylamino)phenyl)methanol	6.4	6.48	6.4982	6.2763	6.317
31	6-(Cyclohexylmethoxy)-N-(3-methoxyphenyl)-9H-purin-2-amine	5.74	6.29	6.08	6.0901	6.0515
32	6-(Cyclohexylmethoxy)-N-(3-(methylthio)phenyl)-9H-purin-2-amine	5.77	5.96	5.9264	5.9645	5.9312
33	6-(Cyclohexylmethoxy)-N-(4-methoxyphenyl)-9H-purin-2-amine	6.19	6.37	5.99	5.7645	5.9589
34	6-(Cyclomethylmethoxy)N-(4-(methylsulfonyl)phenyl)-N,N-dimethyl-9H-purin-2-amine	6.19	6.37	5.99	5.7645	5.9589
35	6-(Cyclomethylmethoxy)-N-(4-(methylsulfonyl)phenyl)-9H-purin-2-amine	7.22	7.24	6.7582	6.773	6.9168
36	6-(Cyclohexylmethoxy)-N-(4-(methylsulfonyl)phenyl)-9H-purin-2-amine	7.00	7.07	6.8456	6.5432	6.6683
37	6-(Cyclohexylmethoxy)-N-(4-(ethyl sulfonyl)phenyl)-9H-purin-2-amine	6.68	6.82	6.67	6.4289	6.912
38	4-(6-(Cyclohexylmethoxy)-9H-purin-2-ylamino)benzamide	6.19	6.4	6.74	6.3105	6.5367
39	4-(6-(Cyclohexylmethoxy)-9H-purin-2-ylamino)-N-methylbenzamide	6.7	6.91	6.7797	6.4012	6.5837
40	4-(6-(Cyclohexylmethoxy)-9H-purin-2-ylamino)-N,N-dimethylbenzamide	6.7	6.5	6.4905	6.5163	6.6349
41	6-(Cyclohexylmethoxy)-N-(4-(prop-1-en-2-yl)phenyl)-9H-purin-2-amine	6.52	6.44	7.0214	6.416	6.5527
42	1-(4-(6-(Cyclohexylmethoxy)-9H-purin-2-ylamino)phenyl)ethanol	6.1	6.5	6.68	6.4374	6.4769
43	2-(4-(6-(Cyclohexylmethoxy)-9H-purin-2-ylamino)phenyl)acetonitril	6.52	6.36	6.3441	6.5529	6.453
44	N-(4-(2-(4-Methylpiperazin-1-yl)ethylsulfonyl)phenyl)-6-(cyclohexylmethoxy)-9H-purin-2-amine	6.92	7.05	6.8082	6.9098	6.8732
45	N-(4-(2-(Piperidin-1-yl)ethylsulfonyl)phenyl)-6-(cyclohexylmethoxy)-9H-purin-2-amine	6.47	6.61	6.66	6.6328	6.652
46	N-(4-(2-Thiomorpholinoethylsulfonyl)phenyl)-6-(cyclohexylmethoxy)-9H-purin-2-amine	6.27	6.41	7.347	5.983	6.1249
47	N-(4-(2-(Diethylamino)ethylsulfonyl)phenyl)-6-(cylohexylmethoxy)-9H-purin-2-amine	6.35	6.38	6.3132	6.7114	6.613

TABLE 1: Continued.

No.	Name	pIC ₅₀	ANN	MLR	ANN (LOO)	MLR (LOO)
48	N-(4-(2-(4-Isopropylpiperazin-1-yl)ethylsulfonyl)phenyl)-6-(cyclohexylmethoxy)-9H-purin-2-amine	6.47	6.4	6.8005	6.8036	6.6025
49	2-(4-(2-(4-(6-(Cyclohexylmethoxy)-9H-purin-2-ylamino)phenylsulfonyl)ethyl)piperazin-1-yl)ethanol	6.59	6.63	6.59	6.7072	6.8705
50	2-(1-(2-(4-(6-(Cyclohexylmethoxy)-9H-purin-2-ylamino)phenylsulfonyl)ethyl)piperidin-4-yl)ethanol	6.62	6.53	6.36	6.6902	6.6265
51	2-(2-(4-(2-(4-(6-(Cyclohexylmethoxy)-9H-purin-2-ylamino)phenylsulfonyl)ethyl)piperazin-1-yl)ethoxy)ethanol	6.51	6.44	6.5378	6.4673	6.636
52	N-(4-(2-(4-(Ethylpiperazin-1-yl)ethylsulfonyl)phenyl)-6-(cyclohexylmethoxy)-9H-purin-2-amine	6.59	6.72	6.6713	6.7344	6.4404
53	1-(4-(2-(4-(6-(Cyclohexylmethoxy)-9H-purin-2-yl amino)phenylsulfonyl)ethyl)piperazin-1-yl)ethanone	6.64	6.67	6.2815	6.6566	6.6462
54	N-(4-(2-(4-(2-(Methoxyethyl)piperazin-1-yl)ethylsulfonyl)phenyl)-6-(cyclohexylmethoxy)-9H-purin-2-amine	6.54	6.47	6.8	6.0738	6.6812
55	N-(4-(2-(Pyrrolidin-1-yl)ethylsulfonyl)phenyl)-6-(cyclohexylmethoxy)-9H-purin-2-amine	6.74	6.89	6.8736	6.5724	6.6375
56	6-(Cyclohexylethoxy)-9H-purin	6.96	6.56	6.96	6.8179	6.5037

TABLE 2: Descriptors were selected for construction of model.

No.	Symbol	Class	Meaning
1	Mor20u	3D-MoRSE	3D-MoRSE-signal 20/unweighted
2	GATS2e	2D autocorrelation	Geary autocorrelation-lag 2/weighted by atomic Sanderson electronegativities
3	R5p	GETAWAY	R autocorrelation of lag 5/weighted by atomic polarizabilities
4	MATS2e	2D autocorrelation	Moran autocorrelation-lag 2/weighted by atomic Sanderson electronegativities
5	JGI6	Galvez topol. Charge	Mean topological charge index of order 6
6	JGI4	Galvez topol. Charge	Mean topological charge index of order 4
7	R1m	GETAWAY	R autocorrelation of lag 5/weighted by atomic masses
8	HATS0m	GETAWAY	Leverage-weighted autocorrelation of lag 0/weighted by atomic masses
9	Mor06v	3D-MoRSE	3D-MoRSE-signal 20/weighted by atomic van der Waals volumes
10	EIs	WHIM	1st component accessibility directional WHIM index/weighted by atomic electrotopological states

TABLE 3: Correlation matrix for the selected descriptors.

	Mor20u	GATS2e	R5p	MATS2e	JGI6	JGI4	R1m	HATS0m	Mor06v	EIs
Mor20u	1									
GATS2e	0.12	1								
R5p	0.121	0.163	1							
MATS2e	-0.58	-0.599	0.088	1						
JGI6	0.614	0.117	-0.22	-0.528	1					
JGI4	-0.459	0.304	0.106	0.199	-0.102	1				
R1m	0.226	-0.225	0.238	0.097	0.248	-0.043	1			
HATS0m	-0.452	-0.079	0.157	0.333	0.126	0.148	0.419	1		
Mor06v	0.601	-0.045	-0.141	-0.512	0.558	-0.462	0.332	-0.138	1	
EIs	0.354	-0.195	-0.017	-0.24	0.352	-0.205	0.373	-0.112	0.431	1

The correlation matrix for the selected 10 descriptors presented in the model is shown in Table 3. These results show there is not any correlation between the selected descriptors.

4. Results and Discussion

The prediction ability of QSAR/QSPR models is affected by two factors: one is the descriptors, which should carry enough

TABLE 4: The training settings for the ANN model.

No. of descriptors	Training fcn.	Transfer fcn.	No. of nodes in hidden layers	No. of epochs	MSE
10	Levenberg-	Log sigmoid	4	5	0.0171
10	Bayesian	Log sigmoid	6	5	0.0186
6	Levenberg-	Tan sigmoid	7	3	0.0285
10	Bayesian	Tan sigmoid	3	4	0.0218

TABLE 5: Comparison of the statistical parameters obtained from ANN and MLR models.

Parameter	Test set ($N = 11$)		Validation set ($N = 11$)	
	MLR	ANN	MLR	ANN
MAE (%)	13.43	13.29	13.82	10.405
MSE	0.069	0.063	0.065	0.0171
PRESS	0.1394	0.1333	0.1349	0.0693
SEP	0.964	0.956	0.936	0.982
R^2	0.8122	0.7422	0.5518	0.4181
RMAE	3.607	3.433	3.68	2.221

information of molecular structure for the interpretation of the activity/property. The other is the modeling method employed [20]. The number of descriptors available for QSAR/QSPR studies is often so large that it is difficult to obtain a model including all of them. Therefore, identifying important descriptors certainly plays an important role in QSAR/QSPR.

Descriptors should represent the maximum information in activity variations, and collinearity among them must be kept to a minimum. As can be seen from the correlation matrix (Table 3), there is no significant correlation between the selected descriptors. In the present work, these descriptors were used for construction of both linear and nonlinear models. The following linear model was obtained by the training set compounds and 10 selected molecular descriptors:

$$\text{pIC}_{50} = 11.746 + 0.684 \text{ Mor20u} - 2.228 \text{ GATS2e} - 18.175 \text{ R5p} - 5.114 \text{ MATS2e} + 65.07 \text{ JGI6} - 40.405 \text{ JGI4} + 2.155 \text{ R1m} - 4.434 \text{ HATS om} - 0.149 \text{ Mor06v} - 0.715 \text{ EIs}.$$

This model was then used to predict the validation and test sets of data. Then, artificial neural network (ANN) was used to make a nonlinear model to calculate the inhibitory activities (pIC_{50}) of the compounds. To do so, a 3-layer feedforward network with backpropagation pattern was used in which mean squared error (MSE) was applied as the performance function. The MLR selected descriptors were used as the input layer of the network. To have a strong network, 5 parameters were optimized. The optimized parameters are (1) the number of descriptors (between 2 and 10), (2) the number of nodes in the hidden layer, (3) the transfer function (including *log sigmoid* and *tan sigmoid*), (4) training function (including *Bayesian regulation (trainbr)* and *Levenberg-Marquardt (trainlm)*), and finally (5) number of epochs. Table 4 shows the training settings of the optimized network. It should be noted that the training of the network for the prediction of pIC_{50} was interrupted when the MSE of the validation set started to increase, to avoid overfitting.

According to Table 4, a network with a Levenberg-Marquardt training function and log-sigmoid transfer function with 10 descriptors (the same MLR descriptors) has the

least MSE value (0.0171). In order to evaluate the predictive ability of the linear and nonlinear models and to compare them, we employed the percentage of mean absolute error (MAE), mean squared error (MSE), predictive residual sum of squares (PRESSs), standard error of prediction (SEP), determination coefficient (R^2), percentage of relative error prediction (REP (%)), and relative mean absolute error (RMAE). These statistical parameters for MLR and ANN are listed in Table 5.

As can be seen from Table 5, all the error parameters' values of ANN for both test and validation sets are smaller than those of MLR. This is believed to be due to the nonlinear capabilities of the ANN model.

The used statistical parameters are defined as:

$$\begin{aligned}
 R^2 &= 1 - \frac{\sum_{i=1}^n (y_{\text{pred}} - y_{\text{obs}})^2}{\sum_{i=1}^n (y_{\text{obs}} - y_{\text{meas}})^2}, \\
 \text{RMSEP} &= \sqrt{\frac{\sum_{i=1}^n (y_{\text{pred}} - y_{\text{obs}})^2}{n}}, \\
 \text{RSEP} (\%) &= 100 \times \sqrt{\frac{\sum_{i=1}^n (y_{\text{pred}} - y_{\text{obs}})^2}{\sum_{i=1}^n (y_{\text{obs}})^2}}, \\
 \text{MAE} (\%) &= \frac{100}{n} \times \sqrt{\sum_{i=1}^n |(y_{\text{pred}} - y_{\text{obs}})|}, \\
 \text{MSE} &= \frac{\sum (y_{\text{obs}} - y_{\text{pred}})}{n}, \\
 \text{RMSE} &= \sqrt{\frac{(y_{\text{obs}} - y_{\text{pred}})^2}{n}}, \\
 F &= \frac{\sum ((y_{\text{pred}} - y)^2 / p)}{\sum ((y_{\text{pred}} - y)^2 / n - p - 1)},
 \end{aligned}
 \tag{2}$$

TABLE 6: R^2 values of the test set after several Y-randomization tests.

Iteration	R^2 test set
1	0.29
2	0.02
3	0.00
4	0.09
5	0.00
6	0.00
7	0.46
8	0.01
9	0.13
10	0.01

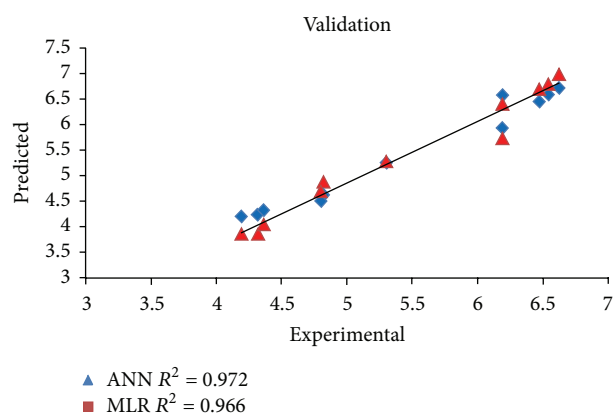


FIGURE 1: Plot of the predicted values versus the experimental ones for the validation set.

where y_i is the experimental value, \hat{y}_i is the predicted value, \bar{y} is the mean value, and n is the number of compounds.

To avoid chance correlation and to guarantee the network's predictability power, Y-randomization test was carried out. The results of several repetition of this test are shown in Table 6. The low values of R^2 show that there is no chance correlation in the developed model.

Figures 1 and 2 show plots of the predicted values versus experimental ones of ANN and MLR models for validation and test sets. The obtained results show the superiority of ANN model than MLR to predict of pIC_{50} of these compounds. The ANN and MLR residuals of leave-one-out are plotted against the experimental values in Figure 3. The symmetric distribution of residuals at both sides of the zero line indicates that no systematic error exists in the development of the MLR and ANN models.

5. Conclusion

From the analysis of the obtained results, we can conclude that (1) the proposed models can sufficiently represent structure-activity relationship of the compounds. (2) By comparison of results from the MLR and ANN, the performance of the ANN model is clearly better than that of MLR, which indicates that nonlinear model can simulate the relationship between the structures of the compounds and their activities

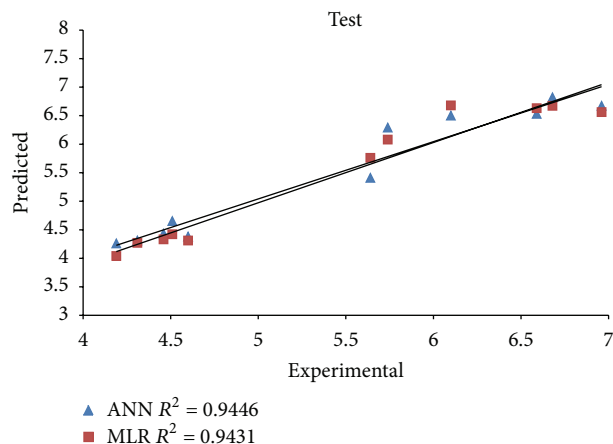


FIGURE 2: Plot of the predicted values versus the experimental ones for the test set.

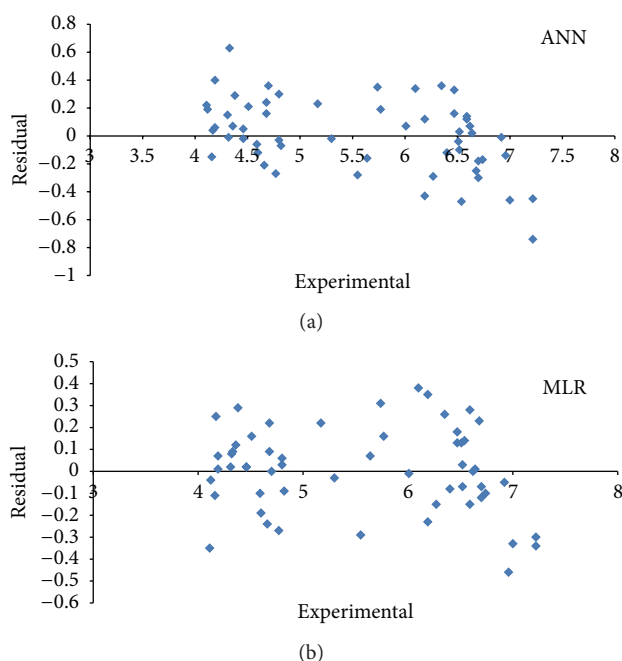


FIGURE 3: Plot of the ANN and MLR residuals of leave-one-out versus experimental values.

more accurately. (3) The calculated statistical parameters of these models reveal the superiority of ANN over MLR model.

References

- [1] C. Norbury and P. Nurse, "Animal cell cycles and their control," *Annual Review of Biochemistry*, vol. 61, pp. 441–470, 1992.
- [2] D. O. Morgan, "Principles of CDK regulation," *Nature*, vol. 374, no. 6518, pp. 131–134, 1995.
- [3] M. Hall and G. Peters, "Genetic alterations of cyclins, cyclin-dependent kinases, and Cdk inhibitors in human cancer," *Advances in Cancer Research*, vol. 68, pp. 67–108, 1996.
- [4] T. M. Sielecki, J. F. Boylan, P. A. Benfield, and G. L. Trainor, "Cyclin-dependent kinase inhibitors: useful targets in cell cycle

- regulation," *Journal of Medicinal Chemistry*, vol. 43, no. 1, pp. 1–18, 2000.
- [5] N. Watanabe, M. Broome, and T. Hunter, "Regulation of the human WEE1Hu CDK tyrosine 15-kinase during the cell cycle," *EMBO Journal*, vol. 14, no. 9, pp. 1878–1891, 1995.
- [6] K. Coulonval, L. Bockstaele, S. Paternot, and P. P. Roger, "Phosphorylations of cyclin-dependent kinase 2 revisited using two-dimensional gel electrophoresis," *Journal of Biological Chemistry*, vol. 278, no. 52, pp. 52052–52060, 2003.
- [7] N. P. Pavletich, "Mechanisms of cyclin-dependent kinase regulation: structures of Cdks, their cyclin activators, and Cip and INK4 inhibitors," *Journal of Molecular Biology*, vol. 287, pp. 821–828, 1999.
- [8] M. D. Losiewicz, B. A. Carlson, G. Kaur, E. A. Sausville, and P. J. Worland, "Potent inhibition of Cdc2 kinase activity by the flavonoid L86-8275," *Biochemical and Biophysical Research Communications*, vol. 201, no. 2, pp. 589–595, 1994.
- [9] A. M. Senderowicz and E. A. Sausville, "Preclinical and clinical development of cyclin-dependent kinase modulators," *Journal of the National Cancer Institute*, vol. 92, no. 5, pp. 376–387, 2000.
- [10] I. R. Hardcastle, B. T. Golding, and R. J. Griffin, "Designing inhibitors of cyclin-dependent kinases," *Annual Review of Pharmacology and Toxicology*, vol. 42, pp. 325–348, 2002.
- [11] M. Knockaert, P. Greengard, and L. Meijer, "Pharmacological inhibitors of cyclin-dependent kinases," *Trends in Pharmacological Sciences*, vol. 23, no. 9, pp. 417–425, 2002.
- [12] P. L. Toogood, "Progress toward the development of agents to modulate the cell cycle," *Current Opinion in Chemical Biology*, vol. 6, pp. 472–478, 2002.
- [13] G. I. Shapiro, "Preclinical and clinical development of the cyclin-dependent kinase inhibitor flavopiridol," *Clinical Cancer Research*, vol. 10, no. 12, pp. 4270s–4275s, 2004.
- [14] M. Vieth, R. E. Higgs, D. H. Robertson, M. Shapiro, E. A. Gragg, and H. Hemmerle, "Kinomics—structural biology and chemogenomics of kinase inhibitors and targets," *Biochimica et Biophysica Acta*, vol. 1697, no. 1–2, pp. 243–257, 2004.
- [15] M. Fernandez, A. Tundidor-Camba, and J. Caballero, "Modeling of cyclin-dependent kinase inhibition by 1H-pyrazolo[3,4-d]pyrimidine derivatives using artificial neural network ensembles," *Journal of Chemical Information and Modeling*, vol. 45, no. 6, pp. 1884–1895, 2005.
- [16] M. P. Gonzalez, J. Caballero, A. M. Helguera, M. Garriga, G. Gonzalez, and M. Fernandez, "2D autocorrelation modelling of the inhibitory activity of cytokinin-derived cyclin-dependent kinase inhibitors," *Bulletin of Mathematical Biology*, vol. 68, no. 4, pp. 735–751, 2006.
- [17] H. Dureja and A. K. Madan, "Topochemical models for prediction of cyclin-dependent kinase 2 inhibitory activity of indole-2-ones," *Journal of Molecular Modeling*, vol. 11, no. 6, pp. 525–531, 2005.
- [18] J. Z. Li, H. X. Liu, X. J. Yao, M. C. Liu, Z. D. Hu, and B. T. Fan, "Structure-activity relationship study of oxindole-based inhibitors of cyclin-dependent kinases based on least-squares support vector machines," *Analytica Chimica Acta*, vol. 581, pp. 333–342, 2007.
- [19] S. Samanta, B. Debnath, A. Basu, S. Gayen, K. Srikanth, and T. Jha, "Exploring QSAR on 3-aminopyrazoles as antitumor agents for their inhibitory activity of CDK2/cyclin A," *European Journal of Medicinal Chemistry*, vol. 41, no. 10, pp. 1190–1195, 2006.
- [20] M. Goodarzi and M. P. Freitas, "Predicting boiling points of aliphatic alcohols through multivariate image analysis applied to quantitative structure-property relationships," *Journal of Physical Chemistry A*, vol. 112, no. 44, pp. 11263–11265, 2008.
- [21] M. Goodarzi and M. P. Freitas, "Augmented three-mode MIA-QSAR modeling for a series of anti-HIV-1 compounds," *QSAR and Combinatorial Science*, vol. 27, no. 9, pp. 1092–1097, 2008.
- [22] M. Goodarzi, T. Goodarzi, and N. Ghasemi, "Spectrophotometric simultaneous determination of manganese(II) and iron(II) in pharmaceutical by orthogonal signal correction-partial least squares," *Annali di Chimica*, vol. 97, no. 5–6, pp. 303–312, 2007.
- [23] N. Goudarzi, M. H. Fatemi, and A. Samadi-Maybodi, "Quantitative structure-properties relationship study of the ^{29}Si -NMR chemical shifts of some silicate species," *Spectroscopy Letters*, vol. 42, no. 4, pp. 186–193, 2009.
- [24] N. Goudarzi and M. Goodarzi, "Prediction of the logarithmic of partition coefficients ($\log P$) of some organic compounds by least square-support vector machine (LS-SVM)," *Molecular Physics*, vol. 106, pp. 2525–2535, 2008.
- [25] N. Goudarzi and M. Goodarzi, "Prediction of the acidic dissociation constant (pK_a) of some organic compounds using linear and nonlinear QSPR methods," *Molecular Physics*, vol. 107, no. 14, pp. 1495–1503, 2009.
- [26] N. Goudarzi and M. Goodarzi, "Prediction of the vapor pressure of some halogenated methyl-phenyl ether (anisole) compounds using linear and nonlinear QSPR methods," *Molecular Physics*, vol. 107, no. 15, pp. 1615–1620, 2009.
- [27] N. Goudarzi and M. Goodarzi, "QSPR models for prediction of half wave potentials of some chlorinated organic compounds using SR-PLS and GA-PLS methods," *Molecular Physics*, vol. 107, pp. 1739–1744, 2009.
- [28] Z. Elmi, K. Faez, M. Goodarzi, and N. Goudarzi, "Feature selection method based on fuzzy entropy for regression in QSAR studies," *Molecular Physics*, vol. 107, no. 17, pp. 1787–1798, 2009.
- [29] N. Goudarzi, M. Goodarzi, M. C. U. Araujo, and R. K. H. Galvao, "QSPR modeling of soil sorption coefficients (KOC) of pesticides using SPA-ANN and SPA-MLR," *Journal of Agricultural and Food Chemistry*, vol. 57, pp. 7153–7158, 2009.
- [30] N. Goudarzi, M. Goodarzi, and M. Arab Chamjangali, "Prediction of inhibition effect of some aliphatic and aromatic organic compounds using QSAR method," *Journal of Environmental Chemistry and Ecotoxicology*, vol. 2, pp. 47–50, 2010.
- [31] N. Trinajstić, *Chemical Graph Theory*, CRC Press, Boca Raton, Fla, USA, 1992.
- [32] A. R. Katritzky, V. S. Lobanov, and M. Karelson, "QSPR: the correlation and quantitative prediction of chemical and physical properties from structure," *Chemical Society Reviews*, vol. 24, no. 4, pp. 279–287, 1995.
- [33] N. R. Draper and H. Smith, *Applied Regression Analysis*, Wiley Series in Probability and Statistics, New York, NY, USA, 1998.
- [34] J. Zupan and J. Gasteiger, *Neural Networks in Chemistry and Drug Design*, Wiley-VCH, Weinheim, Germany, 1999.
- [35] N. K. Bose and P. Liang, *Neural Networks, Fundamentals*, McGraw-Hill, New York, NY, USA, 1996.
- [36] J. H. Alzate-Morales, J. Caballero, A. Vergara Jague, and F. D. Gonzalez, "Insights into the structural basis of N2 and O6 substituted guanine derivatives as cyclin-dependent kinase 2 (CDK2) inhibitors: prediction of the binding modes and potency of the inhibitors by docking and ONIOM calculations," *Journal of Chemical Information and Modeling*, vol. 49, no. 4, pp. 886–899, 2009.

- [37] HyperChem Release 7, HyperCube, Inc., <http://www.hyper.com>.
- [38] R. Todeschini, Milano Chemometrics and QSPR Group, <http://michem.disat.unimib.it/chm/staff/staff.htm>.
- [39] SPSS for windows Statistical package for IBM PC, SPSS Inc., <http://www.spss.com>.

