Journal of Applied Chemistry

Molecular modeling studies of novel triazines as potent and selective phosphodiesterase 10A inhibitors using 2D quantitative structure-activity relationship

Zahra Rostami[1,*] , Negin Sarmasti[2], Ardeshir Khazaei[2], Mohammad Ali Zolfigol[2] and Jaber Yousefiseyf[3]

[1]*Department of Chemistry, Payame Noor University, I. R. of Iran*
[2]*Department of Chemistry, College of Science, Bu-Ali Sina University, Hamadan, Iran*
[3]*Department of Chemical Engineering, College of Technical and Engineering, Tehran University, Iran*

Article history:
Received: 10 April 2013
Received in revised form: 20 June 2013
Accepted: 24 August 2013

Abstract

The Two-dimensional Quantitative Structure-activity Relationship (2D-QSAR) of a series of 44 triazine derivatives with antischizophrenia and antihuntingdon's property has been studied. Multiple Linear Regression (MLR), Principal Component Regression (PCR) and Partial Least Squares (PLS) were used as regression analysis techniques with an attempt to derive a correlation between the biological activity as dependent variable and various descriptors as independent variables. The QSAR studies were performed using VLife MDS software. The models were validated for predictivity by both internal (q^2) and external (Pred_r^2) validation. Results indicated this is no significant statistical differences between calculated activities of these compounds with laboratory quantities thus, the obtained models allowed us to predict Antischizophrenia and Antihuntingdon activities of new Triazines derivatives.

Keywords: Antischizophrenia, Antihuntingdon, Triazines, Multiple Linear Regression (MLR), Principal Component Regression (PCR), Partial Least Squares (PLS)

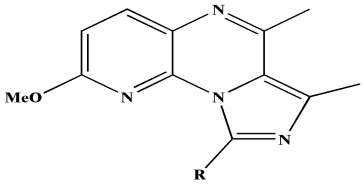
1. Introduction

Quantitative structure-activity relationship (QSAR) describes how a known biological activity can differ as a function of molecular descriptors derived from the chemical structure of a set of molecules. Many physiological activities of a molecule can be associated to their composition and structure. Molecular descriptors, which are numerical depictions of the molecular structures, are used for performing QSAR analysis [1]. Triazines represent the most important class of biologically active compounds as inhibitors of antischizophrenia and antihuntingdon's disease. Triazines and cyclic nucleotide phosphodiesterases (PDEs) are key regulators of cellular signal transduction by hydrolyzing the 30, 50 monophosphate bond of the intracellular second messengers adenosine 30, 50- monophosphate (cAMPa) and cyclic guanosine 30, 50-monophosphate (cGMP) [2] . PDEs are classified by their substrate specificity. Some selectively hydrolyze cAMP or cGMP, while others hydrolyze both substrates [3, 4]. PDE10A hydrolyzing both cAMP and cGMP, with a higher affinity for cAMP (Km = 0.05 lM) than for cGMP (Km = 3 lM).10 PDE10A is primarily a membrane bound enzyme containing a catalytic domain in the C-terminal portion of the protein. Key residues of the catalytic core form a well-defined hydrophobic clamp region that positions the planar rings of the nucleotide for interaction with an absolutely conserved glutamine residue (Gln716 in PDE10A) [5]. PDE10A mRNA is highly expressed only in brain and testes [6, 7]. In the brain, both PDE10A mRNA and protein are specifically enriched in the medium spiny neurons (MSNs) of the striatum [8]. PDE10A has been suggested to play a key role in regulating MSN activity and in turn striatal output as this region integrates dopaminergic and glutama tergic inputs from midbrain and cortical regions, respectively, to action motoric and cognitive function. Dysfunction in cortical–striatal neurotransmission has been implicated in the pathophysiology of schizophrenia and Huntingdon's disease, thus PDE10A inhibition has been suggested as a therapeutic strategy for this disease [9-12]. In recent disclosures, [13] Malamas et al have described the design of pyridyl- and phenyl-[2,3-e]pyrazines (1, 2; Figure 1) as potent inhibitors (IC50 ~ 50-500 pM) of PDE10A, with high selectivity (>500×) for the other

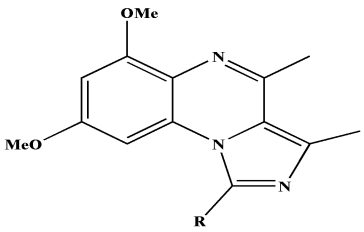
*. **Corresponding Author:** *E-mail address: Rostami@pnu.ac.ir*

known PDE isoforms, with the exception of PDE2, which are less selective (~50–200×) [14].

One could not, however, confirm that the compounds designed would always possess good inhibitory activity to phosphodiesterase 10A, while experimental assessments of inhibitory activity of these compounds are time-consuming and expensive. Consequently, it is of interest to develop a prediction method for biological activities before the synthesis [15]. Computer-aided drug discovery techniques have tremendous effect in shortening the process of drug discovery investigations [13, 14]. Among different computational techniques, the quantitative structure-activity relationship (QSAR) methods are certainly the major factors in the contemporary drug design. Thus, it is quite clear why the industrial units are the prime users of the QSAR methods [16]. Using such an approach one could predict the activities of newly designed compounds before a decision is being made whether these compounds should be really synthesized and tested. The principal of the QSAR approach is to establish mathematical models that relate variations of factors with variations in responses. The factors are structural or physicochemical properties of the compounds, described using quantitative molecular descriptors. Responses are various quantifiable biological activities. In the present study, we reported 2D – QSAR studies on a series of phosphodiesterase 10A inhibitors to provide further insight into the key structural features required to design potential drug candidates of this class.



1: R=4-(3-Me-pyridyl)



2: R = 4-(3-Me-pyridyl)

Figure 1. PDE10A inhibitors.

2. Computational methods

Dataset, parameters and different statistical methods used to develop statistically significant QSAR models are described in the following sections.

2. 1. Chemical Data

A series of 44 molecules belonging to triazine derivatives as potent and selective phosphodiesterase 10A inhibitors were taken from the study by Malamas et al [14]. The 2D- QSAR models were generated using a training set of 11 molecules. The chemical structure and observed biological activities of the training and test set molecules are presented in Table 1. Predictive power of the resulting models was evaluated by a test set of 33 molecules with uniformly distributed biological activities. The observed selection of test set molecules was made by considering the fact that test set molecules represents a range of biological activity similar to the training set.

2.2. Biological Activities

The negative logarithm of the measured IC_{50} (μM) against phosphodiesterase 10A (PDE10A) as pIC_{50} [$pIC_{50} = -\log (IC_{50})$] was used as dependent variable, thus correlating the data linear to the free energy change. Since some compounds exhibited insignificant/no inhibition, such compounds were excluded from the present study. The pIC_{50} values of the molecules under study spanned a wide range from 5 to 10.

2.3. Molecular Descriptors

Various 2D descriptors (a total of 1038) like element counts, molecular weight, molecular refractivity, log P, topological index, Baumann alignment independent topological descriptors etc., were calculated using VLifeMDS QSAR plus module software. The preprocessing of the independent variables (i.e., descriptors) was done by removing invariable (constant column).

2.4. Selection of Training and Test Set

The dataset of 44 molecules was divided into training and test set by Sphere Exclusion Selection method for MLR, PCR and PCA models with pIC_{50} activity field as dependent variable and various 2D descriptors calculated for the molecules as independent variables.

2.5. Model Validation

This is done to test the internal stability and predictive ability of the QSAR models. Developed QSAR models were validated by the following procedure:

1) Internal Validation

Internal validation was carried out using leave-one-out (q^2_{LOO}) method. For calculating q^2, each molecule in the training set was eliminated once and the activity of the eliminated molecule was predicted by using the model developed by the remaining molecules. The q^2 was calculated using the equation which describes the internal stability of a model.

$$q^2 = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - y_{mean})^2}$$

Where y_i and $\hat{y}_i$ are the actual and predicted activity of the ith molecule in the training set, respectively, and y_{mean} is the average activity of all molecules in the training set.

Table 1. Structure, Experimental and Predicted Activity of triazines used in this study.

No.	Substituent						pIC_{50}
	X	Y	R_1	R_2	R_3	R_4	
1	N	C	H	OMe	3-Me-4-pyridyl	Me	7.9586
2	C	C	OMe	OMe	3-Me-4-pyridyl	Me	9.6990
3	N	N	H	OMe	4-Me-3-pyridyl	H	6.4111
4	N	N	H	OMe	5-Me-3-pyridyl	H	7.2882
5	N	N	H	OMe	5-OMe-3-pyridyl	H	7.0395
6	N	N	H	OMe	2,4-Me, Me-thiazole	H	67.190
7	N	N	H	OMe	2-Me-Ph	H	7.3224
8	N	N	H	OMe	4-F, 2-Me-Ph	H	7.3382
9	N	N	H	OMe	2-Cl-Ph	H	7.8827
10	N	N	H	OMe	2,4-Cl, Cl-Ph	H	7.8239
11	N	N	H	OMe	2-Cl, 4-F-Ph	H	7.8827
12	N	N	H	OMe	2,5-Cl, Cl-Ph	H	7.2749
13	N	N	H	OMe	2-F, 5-OMe-Ph	H	8.0074
14	N	N	H	OMe	2-F, 5-OEt-Ph	H	6.9431
15	N	N	H	OMe	2-F, 5-O(CHMe2)-Ph	H	5.6144
16	N	N	OMe	H	4-Me-3-pyridyl	H	7.3726
17	N	N	OMe	H	3-Me-4-pyridyl	H	7.9957
18	N	N	OMe	H	2-Me-3-pyridyl	H	7.9066
19	C	N	OMe	H	2-Me-Ph	H	8.4815
20	C	N	OMe	H	2-Cl-Ph	H	8.6180
21	C	N	OMe	OMe	2-Cl-Ph	H	9.5528
22	C	N	OMe	OMe	2-Me-Ph	H	8.2899
23	C	N	OMe	OMe	2-CF3-Ph	H	8.4749
24	C	N	OMe	OMe	2-Me-3-pyridyl	H	8.8297
25	C	N	OMe	OMe	4-Me-3-pyridyl	H	8.3809
26	C	N	OMe	OMe	3-Me-4-pyridyl	H	8.6861
27	C	N	OMe	morpholine	2-Me-Ph	H	8.8153
28	C	N	OMe	morpholine	2-Cl-Ph	H	9.0706
29	C	N	OMe	morpholine	3-Me-4-pyridyl	H	8.3990
30	C	N	OMe	F	2-OMe-Ph	H	8.1851
31	C	N	OMe	F	3-OMe-Ph	H	8.0405
32	C	N	OMe	F	3-F, 2-Me-Ph	H	8.4498
33	C	N	OMe	F	4-Me-3-pyridyl	H	7.5952
34	C	N	OMe	F	3-Me-4-pyridyl	H	8.1427
35	C	N	OMe	F	2-Me-3-pyridyl	H	8.1057
36	C	N	OMe	F	3,5-Me,Me-N-Me-pyrazole	H	7.8827
37	C	N	OMe	F	3-Me-2-thienyl	H	8.7825
38	C	N	OMe	F	2,4-Me, Me-thiazole	H	7.9318
39	C	N	OMe	CF3	3-Me-4-pyridyl	H	8.8297
40	C	N	OMe	CF3	4-Me-3-pyridyl	H	8.2916
41	C	N	OMe	CF3	3-F, 2-Me-Ph	H	8.7328
42	C	N	OMe	OCF2	3-Me-4-pyridyl	H	8.1549
43	C	N	F	OMe	2-OMe-Ph	H	8.0793
44	C	N	F	OMe	3-Me-2-thienyl	H	8.0255

2) External Validation

For external validation, the activity of each molecule in the test set was predicted using the model developed by the training set. The pred_r^2 value is calculated as follows.

$$pred\_r^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - y_{mean})^2}$$

Where y_i and $\hat{y}_i$ are the actual and predicted activity of the ith molecule in the test set, respectively, and y_{mean} is the average activity of all molecules in the training set.

Both summations are over all molecules in the test set. Thus, the pred_r^2 value is indicative of the predictive power of the current model for external test set.

2.6. Randomization Test

To evaluate the statistical significance of the QSAR model for an actual dataset, one tail hypothesis testing was used [17, 18]. The robustness of the models for training sets was examined by comparing these models to those derived for random datasets. Random sets were generated by rearranging the activities of the molecules in the training set. The statistical model was derived using various randomly rearranged activities (random sets) with the selected descriptors and the corresponding q2 were calculated. The significance of the models hence obtained was derived based on a calculated Z score [17, 18].

A Z score value is calculated by the following formula:

$$Z_{score} = \frac{(h - \mu)}{\sigma}$$

Where h is the q^2 value calculated for the actual dataset, μ the average q^2, and σ is its standard deviation calculated for various iterations using models build by different random datasets.

The probability (α) of significance of randomization test is derived by comparing Z score value with Z score critical value as reported in reference [19], if Z score value is less than 4.0; otherwise it is calculated by the formula as given in the literature. For example, a Z score value greater than 3.10 indicates that there is a probability (α) of less than 0.001 that the QSAR model constructed for the real dataset is random. The randomization test suggests that all the developed models have a probability of less than 1% that the model is generated by chance.

2.7. QSAR by Multiple Linear Regression (MLR) Analysis

Multiple linear regression (MLR) analysis was used to develop QSAR models. Different combinations of parameters (factor loadings more than 0.7) belonging to different factors were tried to develop these models. Correlation analysis was carried out on these selected parameters. Intercorrelated parameters were eliminated. Statistical qualities of MLR equations were judged by parameters like correlation coefficient (R), adjusted R^2 (Ra^2), variance ratio (F), probability factor related to F-ratio (p) and standard error of estimate (s). The regression equation takes the form:

$$Y = b_1 \times x_1 + b_2 \times x_2 + b_3 \times x_3 + c$$

Where Y is the dependent variable, the 'b's are regression coefficients for corresponding 'x's (independent variable), 'c' is a regression constant or intercept.

In the present study QSAR model was developed using multiple regression by stepwise selection method with pIC50 activity field as dependent variable and physico-chemical descriptors as independent variable having cross-correlation limit of 1. Selection of test and training set was done by sphere exclusion method.

2.8. QSAR by Principal Component Regression (PCR) Method

Principal components analysis rotates the data into a new set of axes such that the first few axes reflect most of the variations within the data. By plotting the data on these axes, we can spot major underlying structures automatically. The value of each point, when rotated to a given axis, is called the principal component value. Principal Components Analysis selects a new set of axes for the data. These are selected in decreasing order of variance within the data. They are also perpendicular to each other. Hence the principal components are uncorrelated. Some components may be constant, but these will be among the last selected. The problem noted with MLR was that correlated variables cause instability. So, how about calculating principal components, throwing away the ones which only appear to contribute noise (or constants), and using MLR on these? This process gives the modeling method known as Principal Components Regression. Rather than forming a single model, as with MLR, a model can be formed using 1, 2,… components and a decision can be made as to how many components are optimal. If the original variables contained collinearity, then some of the components will contribute only noise. So long as these are dropped, the models can be guarantee that our model will be stable. The QSAR model was developed using principal component regression by forward-backward variable selection method with pIC_{50} activity field as dependent variable. Selection of test and training set was done by Sphere exclusion Selection method.

2.9. QSAR by Partial Least Squares (PLS) Regression Method

PLS is an effective technique for finding the relationship between the properties of a molecule and

its structure. In mathematical terms, PLS relates a matrix Y of dependent variables to a matrix X of molecular structure descriptors, i.e., a latent variable approach to modeling the covariance structures in these two spaces. PLS have two objectives: to approximate the X and Y data matrices, and to maximize the correlation between them. Whereas the extraction of PLS components is performed stepwise and the importance of a single component is assessed independently, a regression equation relating each Y variable with the X matrix is created. PLS decomposes the matrix X into several latent variables that correlate best with the activity of the molecules. Despite its wide acceptance, a high value of q2 alone is an insufficient criterion for a QSAR model to be highly predictive. Use of greater number of descriptors particularly requires the model to be validated by external predictive power (r^2 predictive). Hence a set of 33 molecules covering different triazine derivatives was employed as test to evaluate the predictivity of training set.

3. Results and discussion

Rationalization of physicochemical characters for antischizophrenia and antiHuntingdon activity was performed using regression analysis techniques (partial least square analysis, multiple linear regressions, and principle component analysis) were applied to generate models. The Sphere Exclusion Selection method was adopted for division of the training and test sets. On the basis of k-mean cluster analysis, 11 compounds among 44 compounds were selected as the training set and remaining 33 compounds were selected as the test set. Unicolumn statics of methods shows that the max of the test is less than max of train set and the min of the test set is greater than of train set shown in Table 2 which is prerequisite analysis for further QSAR study.

Table 2. Unicolumn statistics of training and test set for sphere exclusion method

Unicolumn statistics	Average	Max	Min	Stand. Deviation
Training set	8.0524	9.6990	5.6144	0.7844
Test set	7.9866	8.8297	6.9431	0.7141

3.1. Generation of QSAR models

The dataset of 44 molecules were used for the present study. The common structure of triazine ring is shown in Figure 2.

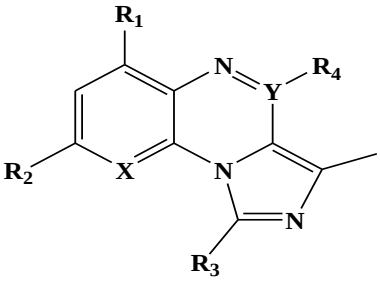


Figure 2. Structure of triazine analogs.

3.1.1 MODEL–1 (Multiple Linear Regression (MLR) Analysis)

After 2D QSAR study by Multiple Linear Regression method using forward-backward stepwise variable selection method, the final QSAR equation developed was as follows.pIC_{50} = + 0.1817(± 0.0094) [4pathClusterCount] + 0.1291(±0.0276) [SaaCHE-index]-1.0238(±0.2429) [SsssCHcount]-0.3706(±0.1296) [T_N_N_2] + 0.0018

Model–1 developed has a correlation coefficient (r^2) of 0.6658, significant cross validated coefficient of correlation of predicted data set (pred_r^2se) 0.3504 and degree of freedom 31.The model is validated by α_ran_r^2 = 0.00004, α_ran_q^2 = 0.00100, α_ran_pred_r^2 = 0.01, best_ran_r^2 = 0.33385, best_ran_q^2 = 0.14079, Z score_ran_r^2 =4.69114 and Z score_ran_q^2 = 3.72308.The randomization test suggests that the developed model have a probability of less than 1% that the model is generated by chance. Statistical data is shown in Table 3. The plot of observed vs. predicted activity is shown in Figure (3). From the plot it can be seen that MLR model is able to predict the activity of training set quite well (all points are close to regression line) as well as external.The descriptors which contribute for the pharmacological action are shown in Figure (4).The major group of descriptors involved sub groups like Path Cluster, Estate contributions, Estate Numbers and Alignment Independent (AI) descriptors help in understanding the effect of substituent at different position of triazines.

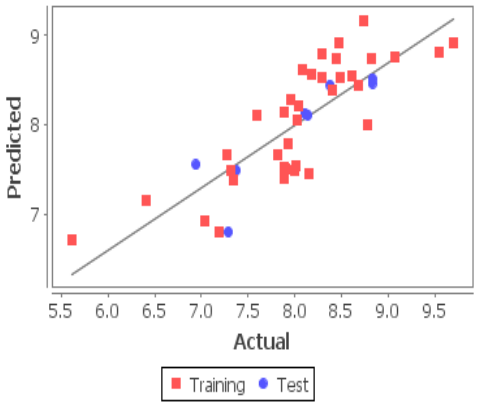


Figure 3. Graph of Actual vs. Predicted activities for training and test set molecules from the Multiple Linear Regression model. A) Training set (Red dots) B) Test Set (Blue dots).

The above study leads to the development of statistically significant QSAR model, which allows understanding of the molecular properties/features that play an important role in governing the variation in the activities.

In addition, this QSAR study allowed investigating influence of very simple and easy-to-compute descriptors in determining biological activities, which could shed light on the key factors that may aid in design of novel potent molecules.

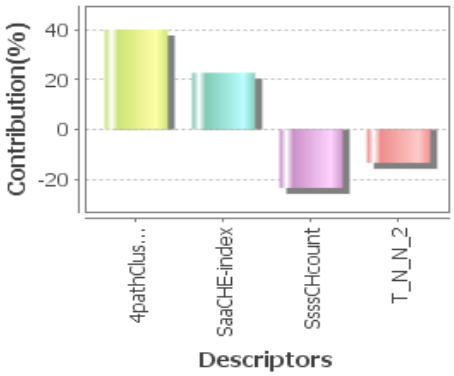


Figure 4. Plot of percentage contribution of each descriptor in developed MLR model explaining variation in the activity

The present QSAR model reveals that 4pathClusterCount descriptor in sub class Path Cluster has a major contribution in explaining variation in activities. The definition for the descriptors that were found to be dominating in the developed QSAR models is given below.

4pathClusterCount: This descriptor signifies total number of fragments of fourth order path cluster in a molecule.

SaaCHE-index: Electrotopological state indices for number of –CH group connected with two aromatic bonds.

SsssCHcount: This descriptor defines the total number of –CH group connected with three single bond.

T_N_N_2: This is the count of number of Nitrogen atoms (single double or triple bonded) separated from any other Nitrogen atom (single double or triple bonded) by 2 bonds in a molecule.

The careful examination of the descriptors in the model suggests that descriptor 4pathClusterCount is directly proportional to the activity and shows the role of the total number of fragments of fourth order path cluster in a molecule. It reveals that presence of fragments of fourth order path cluster over the triazines is favorable for the pharmacological activity and similar analogues.

The presence of descriptor SaaCHE-index (having positive MLR contribution) in the QSAR model reveals that the presence the number of –CH group connected with two aromatic bonds is favorable for the activity and structurally similar other analogs in the series.

The negative contribution of an estate number descriptor SsssCHcount which represents the total number of –CH group connected with three single bond reveals that –CH group group should not be directly attached with aniline ring for maximal activity.

An Alignment Independent (AI) descriptor T_N_N_2 which represents count of number of Nitrogen atoms is inversely proportional to the activity. It reveals that Nitrogen atoms should not be separated from any other Nitrogen atom for maximal activity.

3.1.2. MODEL–2 (Principal Component Regression (PCR) Analysis)

Model-2 is having following QSAR equation which forward-backward stepwise variable selection method was used to data reduction procedure.

pIC_{50}= -0.3176 [T_N_N_4] + 0.5732 [T_N_O_3]-0.3078 [T_N_O_7]-0.0869 [T_2_F_5]+8.5933

The model-2 gave correlation coefficient (r^2) of 0.6676, cross validated correlation coefficient (q^2) of 0. 5390, F test of 21.421, r^2 for external test set (pred_r^2) 0.7681, coefficient of correlation of predicted data set (pred_r^2se) 0.3455 and degree of freedom 32. The model is validated by α_ran_r^2 = 0.00000, α_ran_q^2 = 0.00000, α_ran_pred_r^2 = 0.01, best_ran_r^2 = 0.3983, best_ran_q^2 = 0.1843, Z score_ran_r^2 = 8.5720 and Z score_ran_q^2 = 7.8550.The randomization test suggests that the developed model have a probability of less than 1% that the model is generated by chance. Statistical data is shown in Table 3. The plot of observed vs predicted activity is shown in Figure (5).

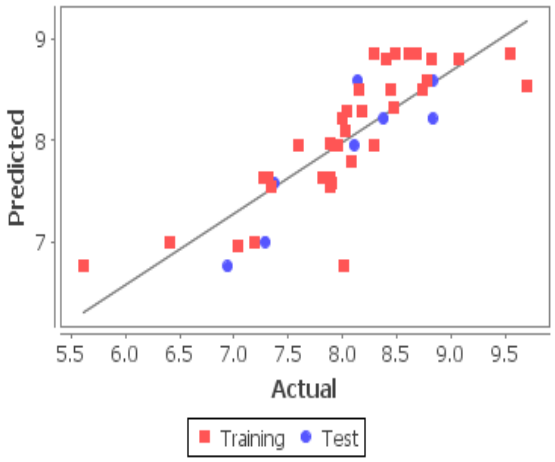


Figure 5. Graph of Actual vs. Predicted activities for training and test set molecules by Principal Component Regression model. A) Training set (Red dots) B) Test Set (Blue dots).

The descriptors which contribute for the pharmacological action are shown in Figure (6).

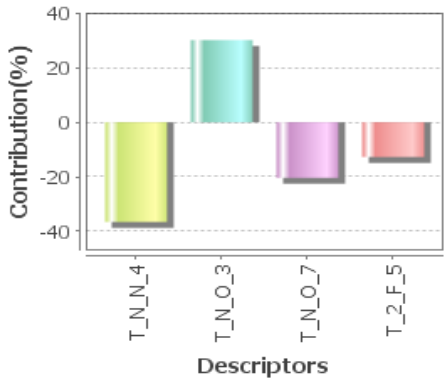


Figure 6. Plot of percentage contribution of each descriptor in developed PCR model explaining variation in the activity.

The definition of the descriptors that were found to be dominating in the developed QSAR models is given below.

1. T_N_N_4: This is the count of number of Nitrogen atoms (single double or triple bonded) separated from any other Nitrogen atom (single double or triple bonded) by 4 bonds in a molecule.

T_N_O_3: This is the count of number of Nitrogen atoms (single double or triple bonded) separated from any other Oxygen atom (single double or triple bonded) by 3 bonds in a molecule.

T_N_O_7: This is the count of number of Nitrogen atoms (single double or triple bonded) separated from any other Oxygen atom (single double or triple bonded) by 7 bonds in a molecule.

T_2_F_5: This is the count of number of double bounded atoms (i.e. any double bonded atom, T_2) separated from Florin atom by 5 bonds. The presence of descriptor T_N_O_3 which having positive PCR coefficient model reveals that the count of number of Nitrogen atoms separated from any other Oxygen atom by 3 bonds in a molecule of triazines is essential for Antischizophrenia, Antihuntingdon activities.

3.1.3. MODEL– 3 (Partial Least Squares (PLS) Regression Method)

2D QSAR study by Partial Least Squares method using Simulated Annealing variable selection method, the final QSAR equation developed was as follows.

pIC_{50} =-0.2090 [T_2_N_3] + 0.5310 [T_N_O_3] - 0 .3302 [T_O_F_4] + 0.1037 [T_2_Cl_3] + 11.4473

The model-3 gave 67% variance of prediction with correlation coefficient (r^2) of 0.6460, cross validated correlation coefficient (q^2) of 0.5449, F test of 19.467 , r^2 for external test set (pred_r^2) 0.7460 , coefficient of correlation of predicted data set (pred_r^2se) 0.3616 and degree of freedom 32.

Table 3.statistical parameters of MLR, PCR and PLS in sphere exclusion method.

Method	MLR			PLS			PCR		
Parameters	Step-wise	genetic algorithm	Simulated annealing	Step-wise	genetic algorithm	Simulated annealing	Step-wise	genetic algorithm	Simulated annealing
N	36	36	36	36	36	36	36	36	36
Deg. of freedom	31	31	32	33	32	32	32	32	32
r^2	0.6658	0.5694	0.6381	0.6105	0.5750	0.6460	0.6676	0.5162	0.6244
q^2	0.5486	0.4198	0.4871	0.4685	0.4259	0.5449	0.5390	0.4086	0.4493
F test	15.438	10.249	18.806	25.865	14.430	19.467	21.421	11.382	17.731
r^2 se	0.4818	0.5469	0.4935	0.5041	0.5348	0.4881	0.4730	0.5706	0.5028
q^2 se	0.5599	0.6349	0.5875	0.5889	0.6216	0.5534	0.5570	0.6308	0.6088
pred_r^2	0.7615	0.5259	0.7268	0.7894	0.5306	0.7460	0.7681	0.5803	0.8334
pred_r^2se	0.3504	0.4941	0.3750	0.3293	0.4916	0.3616	0.3455	0.4649	0.2928
best_ran_r^2	0.3339	0.3705	0.1637	0.2814	0.3723	0.3537	0.3983	0.2173	0.3199
best_ran_q^2	0.1407	0.1284	-0.1141	0.0312	0.0852	0.1249	0.1843	0.0258	0.1390
Z score_ran_r^2	4.6911	5.9259	11.520	6.8929	6.1476	6.8500	8.5720	6.0522	8.0634
Z score_ran_q^2	3.7230	5.0268	7.8631	6.4875	5.9746	7.0257	7.8550	5.2363	6.1846
α_ran_r^2	0.00004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
α_ran_q^2	0.0010	0.00001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
α_ran_pred_r^2	0.0100	0.0500	0.0100	0.0010	0.0500	0.0100	0.0100	0.0500	0.0100

In this table: MLR = Multiple Linear Regression, PCR = Principal Component Regression, PLS = Partial Least Squares, n = number of molecules of training set, df =degree of freedom, r^2 = coefficient of determination, q^2 = cross validated r^2, pred_r^2 = r^2 for external test set, pred_r^2se = coefficient of correlation of predicted data set.

The model is validated by α_ran_r^2 = 0.00000, α_ran_q^2 = 0.00000, α_ran_pred_r^2 = 0.0100, best_ran_r^2 = 0.3537, best_ran_q^2 = 0.1249, Z score_ran_r^2 = 6.8500 and Z score_ran_q^2 =7.0257.The randomization test suggests that the developed model have a probability of less than 1% that the model is generated by chance. Statistical data is shown in table 3. The plot of observed vs. predicted activity is shown in Figure (7).

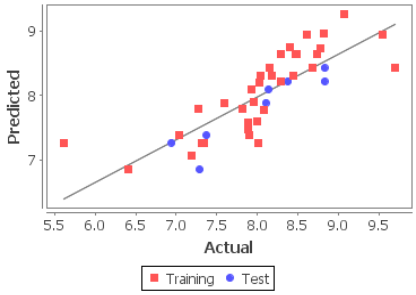


Figure 7. Graph of Actual vs. Predicted activities for training and test set molecules by Partial Least Squares model. A) Training set (Red dots) B) Test Set (Blue dots).

The descriptors which contribute for the pharmacological action are shown in Figure (8).The descriptor T_N_O_3 is common between PCR and PLS; only differs from each other in their percentage of contribution. The definition of the remaining descriptors that were found to be dominating in the developed QSAR models is given below.

T_2_N_3: This is the count of number of double bounded atoms (i.e. any double bonded atom, T_2) separated from Nitrogen atom by 3 bonds.

T_O_F_4: This is the count of number of Oxygen atoms (single double or triple bonded) separated from Florinatom by 4 bond distance in a molecule.

T_2_Cl_3: This is the count of number of double bounded atoms (i.e. any double bonded atom, T_2) separated from Chlorine atom by 3 bonds.

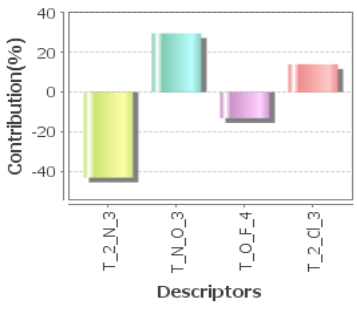


Figure 8. Plot of percentage contribution of each descriptor in developed PLS model explaining variation in the activity

The descriptors which contribute for the pharmacological action are shown in Figure 8. The QSAR model by PLS reveals that Alignment Independent (AI) descriptors (T_N_O_3) and T_2_Cl_3 have positive contributions in explaining variation in activities and the remaining Alignment Independent (AI) descriptors that were found to be dominating in the developed QSAR models are inversely proportional to the activity. It reveals that presence the Nitrogen atoms which separated from any other Oxygen atom and the count of number of double bounded atoms are favorable for the pharmacological activity. The numerical values of loading descriptors of Different models are listed in Table 4.

4. Conclusion

In this QSAR study, the proposed QSAR model, due to the high predictive ability, can therefore act as a useful aid to the costly and time consuming experiments for determining the maximal Antischizophrenia and Antihuntingdon activities. We first tried to identify descriptors trends which lead to biological activities based on the proposed QSAR equation. We have obtained three optimized mathematical models. All models (MLR, PCR and PLS) have shared four Path Cluster , Estate contributions, Estate Numbers and Alignment Independent (AI) descriptors class and as mentioned, 4pathClusterCount and T_N_O_3 descriptors are the most important variables for predicting Antischizophrenia and Antihuntingdon activities. Since these molecular descriptors are the main factors which influence the biological activities of Triazines, it is necessary to explore such descriptors. Meanwhile, studying their applicability could lead to a vital improvement in QSAR studies.

Therefore, obtained data by adequate designed QSAR studies allow observing aspects and essential molecular characteristics to have an increased biological activity, suggesting certain structural requirements for an increased Antischizophrenia and Antihuntingdon potential. Our results open very interesting perspectives regarding triazine derivatives. Finally the QSAR model could be helpful to predict the biological activities of compounds by calculating the descriptors involved in the QSAR equation. Hence the model proposed in this work is useful in describing QSAR of novel triazines as potent and selective phosphodiesterase 10A inhibitors and can be employed to design new derivatives of triazines with specific inhibitory activity.

5. Acknowledgment

The author is grateful to the Bu-ali Sina and Payame Noor Universities, Hamedan, for providing facilities to conduct this study.

References

[1] G. Yapısal, O. Benzeyen, B. Yarı and M. İ. Konsantrasyonlarının, *Türk Biyokimya Dergisi [Turkish Journal of Biochemistry–Turk J Biochem]* **35** (2010) 287.

[2] D. T Manallack, R. A.Hughes and P. E. Thompson, *J. Med. Chem.* 48 (2005) 3449.

[3] J. O. Odingo, *Exp. Opin. Ther. Patents* **15** (2005) 773.
[4] D. Rotella, *Nat. Rev. Drug Discov.* **1** (2002) 674.
[5] K. Zhang, G. Card, Y. Suzuki, R. Artis, D. Fong, S. Gillette, D. Hsieh, J. Neiman, B. West, C. Zhang, M. Milburn and S. Kim, J. *Bollag, G. Mol. Cell* **15** (2004) 279.
[6] K. Fujishige, J. Kotera and K. Omori, *Eur. J. Biochem.* **266** (1999) 1118.
[7] K. Fujishige, J.Kotera, H. Michibata, K.Yuasa and K. Omori, *J. Biol. Chem.* **274** (1999) 18438.
[8] T. F. Seeger, B. Bartlett, T. M.Coskran, A. H.Varghese, and R. D. Williams *Brain Res.* **985** (2003) 113.
[9] H. Kaiya, *Prostaglandins Leukot. Essent. Fatty Acids* **46** (1992) 33.
[10] E. C. Muly, Y. Smith, P.Allen, and P. Greengard, *Psychopharmacol. Bull* **36** (2002) 92.
[11] D. L.Garver, C.Carl Johnson and D. R. Kanter, *Life Sci.* **1982** (1987) 31.
[12] C. Giampa, D. Laurenti, S. Anzilotti, G. Bernardi, F. Menniti and F. Fusco, *PLoS One* **5** (2010) 13417.
[13] N. Hoefgen, H. Stange, R. Rudolf Schindler, H. Lankau, C. Grunwald, B. Langen and N. Brandon, *Med. Chem*. **53** (2010) 4399.
[14] M. S. Malamas , H. Stange, R. Schindler, H-J. Lankau, C. Grunwald, B. Langen, U. Egerland, T. Hage, Y. Ni, J. Erdei, K. Fan, K. Parris, and N.Hoefgen, *Bioorganic & Medicinal Chemistry Letters* **22** (2012) 5876.
[15] M. N. Noolvi, H. Patel and V. Bhardwaj, *Digest Journal of Nanomaterials and Biostructures* **5** (2010) 387.
[16] M. S. Malamas, Y.Ni, J. Erdei, H. Stange, R. Schindler, T. Hage, M. Pangalos, N. Brandon and N. Hofgen, *J. Med. Chem*. **54** (2011) 7621.
[17] W. Zheng and A. Tropsha, *J. Chem. Inf. Comput. Sci.* **40** (2000) 185.
[18] N. Gilbert, W.B. Statistics, Co Saunders, P.A. Philadelphia (1976).