Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 15 (2005) 3737–3743
Bioorganic & Medicinal Chemistry Letters

QSAR of adenosine A_3 receptor antagonist 1,2,4-triazolo[4,3-*a*]quinoxalin-1-one derivatives using chemometric tools

Prosenjit Bhattacharya and Kunal Roy*

Drug Theoretics and Cheminformatics Lab, Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700 032, India

Received 30 April 2005; revised 16 May 2005; accepted 17 May 2005
Available online 29 June 2005

Abstract—Considering the potential of selective adenosine A_3 receptor subtype ligands in the development of prospective therapeutic agents, an attempt has been made to explore physicochemical requirements of 1,2,4-triazolo[4,3-*a*]quinoxalin-1-one derivatives for A_3 receptor binding. In this study, lipophilicity ($\log P$), physicochemical substituent constants (π, MR, σ_p) of phenyl ring substituents, and Wang–Ford charges of common atoms of the quinoxaline nucleus (calculated from molecular electrostatic potential surface of energy-minimized geometry using AM1 technique) were used as independent variables along with suitable dummy parameters. The best multiple linear regression (MLR) equation obtained from factor analysis (FA-MLR) as the preprocessing step could explain and predict 72.6% and 65.3%, respectively, of the variance of the binding affinity. The same equation also emerged as the best equation in the population of 100 equations obtained from genetic function approximation (GFA-MLR). The results suggested that presence of an electron-withdrawing group at the para position of the phenyl ring would be favorable for the binding affinity. Again, the presence of a nitro group at position R_1 increases the binding affinity. When factor scores were used as predictor variables in the principal component regression analysis, the resultant model showed 78.6% explained variance and 63.1% predicted variance. The best equation derived from G/PLS could explain and predict 74.4% and 64.8%, respectively. The results have suggested the importance of Wang–Ford charges of atoms C_{15} and C_{19}, apart from positive contributions of electron-withdrawing para substituents of the variance of the phenyl ring and nitro group at the R_1 position.

1. Introduction

The principal aim of a medicinal chemist is to discover novel drugs with greater potency and reduced toxicity which may be achieved by molecular modification or tailoring of existing drugs, optimization of various lead compounds, isolation of active constituents from natural sources, or syntheses of new series of compounds.[1,2] A rational explanation of drug action is often restricted by our ability to correlate the observed physiological effects with a reasonable hypothesis or concept. Various structural, physicochemical, and biological parameters are used to correlate these with biological activity and the observed relations are used to predict the activity of a new compound and this information is exploited to develop newer molecules of optimum activity. Such a correlation may also help us in exploring the mechanistic features of biological activity.[3] Quantitative structure–activity relationship (QSAR) studies represent a non-experimental part of drug design encompassing the study of both structure–activity and structure–property relations in broad sense. This is an intellectual exercise of assembling, manipulating and examining data obtained from physical, chemical, and biological experiments, and correlating these to biological activity. Biological activity of a drug depends on the types and magnitude of interactions between the receptor and the drug molecule.[2] Various structural attributes of a drug molecule, such as electronic distribution, steric feature, etc., are the determining factors regulating the interactions.[4] All QSAR studies are based on the notion that biological activity (therapeutic or toxic) is a function of the chemical structure and/or property.[4,5] The

Keywords: QSAR; Adenosine A_3 receptor; Lipophilicity; Wang–Ford charges; 1,2,4-Triazolo[4,3-*a*]quinoxalin-1-one derivatives; Genetic function approximation; Factor analysis.
* Corresponding author. Tel: +913328670786; fax: +913328371078; e-mail: kunalroy_in@yahoo.com
URL: http://www.geocities.com/kunalroy_in

0960-894X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2005.05.051

goals of QSAR studies include a better understanding of the modes of actions, prediction of newer analogs with better activity, classification of active/inactive compounds, and optimization of the lead compound to reduce toxicity and increase selectivity.

Adenosine receptors represent a very important class of targets for new drug development; thus, adenosine receptor ligands are recently being well investigated for QSAR modeling. Adenosine, a prevalent nucleoside throughout the body, plays an important role in the function of normal nerve cells,[6] in controlling cell proliferation,[7] and as a signal of inflammation.[8] Inflamed tissues also release adenine nucleotides, which are converted to adenosine by Ecto-nucleotidases (CD39, CD73).[9,10] Cells that release these nucleotides include platelets, mast cells, nerves, and the endothelium.[11]

The four adenosine receptors which detect local changes in adenosine concentration are called A_1, A_{2A}, A_{2B}, and A_3.[12] They are 'seven-spanning' proteins that are coupled to various G-proteins.[13] A_2 receptors work on G_s,[14] but A_1 and A_3 interact with G_i and G_o. The interaction between the neuromodulator adenosine and adenosine receptors on the surface of neurons modifies the neuronal responses to neurotransmitters.[15]

Stimulation of A_1 receptors inhibits nerve cells,[16] and these receptors also mediate the profound effects of adenosine on the heart.[17] Activation of either the A_1 adenosine receptor (A_1R) or A_3R elicits delayed cardioprotection against infarction, ischemia, and hypoxia.[18] By lowering heart rate, and especially, slowing down AV nodal conduction,[19] adenosine causes pharmacological cardioversion. In the basal forebrain, accumulation of adenosine (seen with prolonged wakefulness) is thought to inhibit cholinergic cells and induce sleep.[20] A_1 receptors also promote vasoconstriction.[21] A_1 receptors in the preglomerular vessels and tubules regulate renal fluid balance.[22] Antagonists to A_1 receptors cause diuresis and natriuresis without any major changes in glomerular filtration rate (GFR). A_1 antagonists decrease afferent arteriolar pressure.[23]

A_{2A} stimulation being anti-inflammatory, the receptors are used to sense excessive tissue inflammation.[24] These receptors also enhance neural communication,[25] promote coronary vasodilatation, and show anti-platelet effects. A_{2A} agonists cause profound vasodilatation, with a corresponding increase in plasma renin activity.[26] Studies demonstrate that agonists of A_{2A} adenosine receptors inhibit the release of the anti-angiogenic factor thrombospondin 1. Multiple cell types and all four adenosine receptors participate in these responses.[27]

A_{2B} is similar to A_{2A}, but not identical; these are, perhaps, the most poorly characterized of the adenosine receptors. A_{2B} receptor found on the human mast cell may be particularly relevant to the management of asthma[28] but A_{2B} receptors are widespread throughout the body. Like A_{2A} receptors, A_{2B} promote vasodilatation.[29]

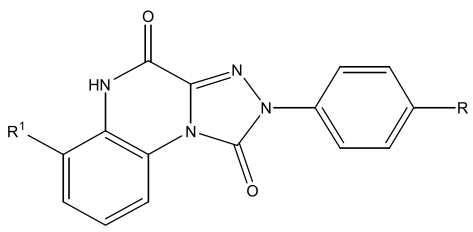
A variety of effects of A_3 receptor have been claimed, but other reports allege completely opposite effects. A_3 is a key receptor in both stimulation and inhibition of cell growth[30] (adenosine stimulates many normal cells in micromolar concentrations and induces apoptosis at higher concentrations in both normal and tumor cells). Lower concentrations may show antiproliferative effects on tumor cells, despite stimulating bone marrow cells.[31] However, others have claimed that adenosine may show many bad effects, promoting tumor growth and angiogenesis. A_3 receptor stimulation (at various concentrations and over various time-spans) may be harmful or beneficial to cerebral ischaemia.[32]

In continuation of our recent efforts to model adenosine receptor ligands,[33–36] the present paper attempts at quantitative structure–activity relationship (QSAR) modeling of A_3 subtype receptor binding data[37] of quinoxaline-1-one and quinoxaline-1,4-dione derivatives using quantum chemical, lipophilicity ($\log P$), physicochemical substituent constants, and suitable indicator parameters.

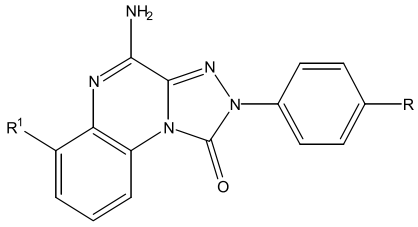
2. Materials and methods

Adenosine A_3 receptor binding affinity data[37] of quinoxaline-1-one and quinoxaline-1,4-dione derivatives (Table 1) have been used for the present QSAR study. The binding affinity data [K_i (nM)] were converted to logarithmic scale [pK_i (mM)] and then used for subsequent QSAR analyses as the response variable. For the present QSAR study, quantum chemical, lipophilicity ($\log P$), physicochemical substituent constants (π, MR, σ_p) of phenyl ring substituents, and appropriate indicator parameters (listed in Table 2) were used as predictor variables, as they were found to be appropriate for the development of models. The values of physicochemical substituent constants (Table 3) were taken from the literature.[38] Semi-empirical quantum chemical calculations were done according to the AM1 (Austin Model 1)[39–41] method using the *Chem 3D Pro*[42] package. The general structures of the compound (Fig. 1) were drawn in *Chem Draw Ultra ver 5.0*[42] and saved as a template structure. For every compound, the template structure was suitably changed considering its structural features, copied to *Chem 3D ver 5.0*[42] to create the 3-D model, and finally the model was 'cleaned up'. The non-hydrogen common atoms of the compounds were given a serial number so that they maintain the same serials in all the models (Fig. 1). Energy minimization was carried out under a Molecular Orbital Package (MOPAC) module [CS MOPAC Pro] of *Chem 3D Pro*[42] using restricted closed shell wave function[34,43] and setting a minimum root mean square (RMS) gradient at 0.100. The selected properties were heat of formation, gradient norm, and charges. The energy-minimized geometry was used for the calculation of Wang–Ford charges q_x (obtained from a molecular electrostatic potential surface) of different atoms (x). Henceforth, reference to a particular atom will be made using the numbering system given in Figure 1, for example, the common oxygen in the quinoxaline-1-one/quinoxaline-1,4-dione nucleus is indicated as O_{20}.

Table 1. Structural features, observed, calculated, and LOO predicted adenosine A_3 binding affinity data of quinoxaline-1,4-dione (**1–9**) and quinoxaline-1-one (**10–17**) derivatives



1-9



10-17

Sl. No.	Structural features		Adenosine A_3 receptor binding affinity [pK_i (mM)]						
	R	R^1	Obs.[a]	Calc.[b]	Pred.[b]	Calc.[c]	Pred.[c]	Calc.[d]	Pred.[d]
1	OCH_3	H	4.796	4.137	4.006	4.021	3.848	3.874	3.726
2	NO_2	H	6.222	6.146	6.043	6.366	7.056	6.470	5.647
3	NH_2	H	2.444	3.391	3.551	2.840	3.154	3.351	3.587
4	$N(CH_3)_2$	H	3.368	3.066	2.980	3.446	3.500	3.345	3.703
5	OC_2H_5	H	3.757	4.195	4.228	3.880	3.931	3.980	4.028
6	OH	H	4.328	3.946	3.914	3.772	3.679	3.979	3.954
7	$OCOCH_3$	H	4.951	5.247	5.334	4.661	4.507	5.122	5.043
8	OCH_3	NO_2	5.328	4.821	4.562	5.215	5.129	4.939	4.642
9	OCH_3	NH_2	4.081	4.137	4.142	4.497	4.740	4.172	4.201
10	OCH_3	H	4.347	4.137	4.121	4.339	4.335	4.350	4.353
11	H	NO_2	5.323	5.338	5.347	5.105	4.169	5.308	4.752
12	NH_2	H	3.475	3.391	3.376	3.486	3.489	3.379	3.430
13	OC_2H_5	H	3.693	4.194	4.239	3.946	4.072	3.946	4.014
14	OH	H	4.135	3.946	3.930	3.843	3.793	4.020	4.016
15	OCH_3	NO_2	4.328	4.821	5.071	4.702	4.996	4.487	4.640
16	OCH_3	NH_2	4.658	4.137	4.097	4.799	4.816	4.289	4.241
17	OH	NH_2	3.754	3.946	3.962	4.093	4.170	3.991	4.080

[a] Ref. 37; Obs., Observed; Calc., Calculated; Pred., Predicted (LOO).
[b] From Eq. 1 [FA-MLR and GFA-MLR derived eq.].
[c] From Eq. 2 [PCRA derived eq.].
[d] From Eq. 3 [G/PLS derived eq.].

Table 2. Definitions of variables

Variables	Definition
q_x	Wang–Ford charge of atom x (x may take values 1–20)
I_{NO_2}	Indicator variable having value 1 if nitro group is present at the R^1 position, value 0 otherwise
I_{NH_2}	Indicator variable having value 1 if amino group is present at the R^1 position, value 0 otherwise
I'_{NH_2}	Indicator variable having value 1 if amino group present at position 8 (Fig. 1), value 0 otherwise

Table 3. Physicochemical substituent constants[a] of phenyl ring substituents

Substituents	Substituent constants		
	π	MR[b]	σ_p
OCH_3	−0.02	0.787	−0.27
NO_2	−0.28	0.736	0.78
NH_2	−1.23	0.54	−0.66
$N(CH_3)_2$	0.18	1.555	−0.83
OC_2H_5	0.38	1.247	−0.24
OH	−0.67	0.285	−0.37
$OCOCH_3$	−0.64	1.247	0.31
H	0.00	0.103	0.00

[a] Ref. 38.
[b] MR values are scaled to a factor of 0.1 as usual.

Though a classical approach of multiple linear regression (MLR) technique was used as the final statistical tool for developing QSAR relations, factor analysis (FA)[44,45] was used as the data-preprocessing step to identify the important predictor variables contributing to the response variable and to avoid collinearities among them. The principal objectives of factor analysis are to display multidimensional data in a space of lower dimensionality, with minimum loss of information (explaining >95% of the variance of the data matrix) and to extract the basic features behind the data with the ultimate goal of interpretation and/or prediction. Factor analysis was performed on the data set(s) containing biological activity and all descriptor variables, which were to be considered. The factors were extracted

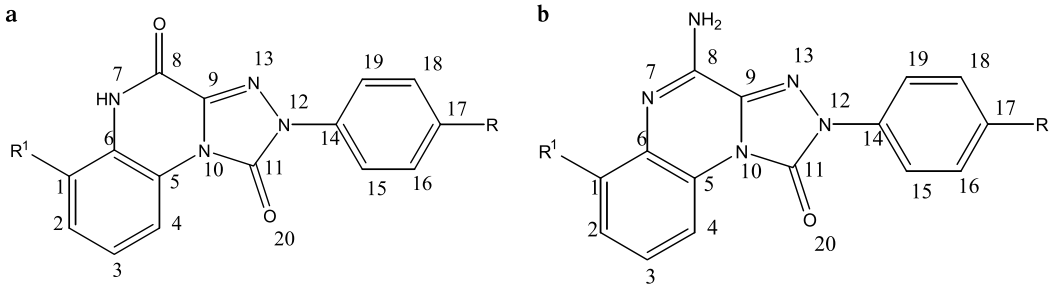


Figure 1. General structures of (a) quinoxaline-1,4-dione and (b) quinoxaline-1-one derivatives: the common atoms have been arbitrarily numbered 1–20 (it has no relation to the chemical nomenclature system).

by the principal component method and then rotated by VARIMAX rotation (a kind of rotation, which is used in the principal component analysis so that the axes are rotated to a position in which the sum of the variances of the loadings is the maximum possible) to obtain Thurston's simple structure. The simple structure is characterized by the property that as many variables as possible fall on the coordinate axes when presented in common factor space, so that the largest possible number of factor loadings becomes zero. This is done with a view obtaining a numerically comprehensive picture of the relatedness of the variables. Only variables with non-zero loadings in such factors where biological activity also has non-zero loading were considered to be important for explaining the variance of activity. Further, variables with non-zero loadings in different factors were combined in a multivariate equation. Attempt was also made to perform the principal component regression analysis (PCRA),[41] taking factor scores as the predictor variables and adopting a backward stepwise regression method. In this case, the principal components serve as latent variables. PCRA has an advantage that collinearities among X variables are not a disturbing factor and that the number of variables included in the analysis may exceed the number of observations.[45] The factor analysis and multiple regression were performed using the statistical software SPSS.[46]

The data set was also modeled using the genetic function approximation (GFA) technique,[47,48] to generate a population of equations rather than one single equation for correlation between the binding affinity and descriptors. GFA involves a combination of multivariate adaptive regression spline (MARS) algorithm with a genetic algorithm to evolve the population of equations that best fit the training set data. It provides an error measure, called the lack of fit (LOF) score that automatically penalizes models with too many features. It also inspires the use of splines as a powerful tool for non-linear modeling. GFA is carried out as follows: (i) an initial population of equations is generated by a random choice of descriptors; (ii) pairs from the population of equations are chosen at random, 'crossovers' are performed, and progeny equations are generated; (iii) it is better at discovering combinations of features that take advantage of the correlations between multiple features; (iv) the fitness of each progeny equation is assessed by a lack-of-fit (LOF) measure; (v) it can use a larger variety of equation term types in the construction of its models; and (vi) if the fitness of a new progeny equation is better, then it is preserved. A distinctive feature of GFA is that it produces a population of models (e.g., 100), instead of generating a single model, as do most other statistical methods. The range of variations in this population gives added information on the quality fit and importance of the descriptors. The GFA study was conducted using a GFA module under QSAR+ environment of Cerius2 software.[49] All default settings were used for the analysis (linear terms, smoothness factor = 1, mutation probability for adding a new term = 50%) and the number of crossovers was set to 100,000.

The genetic partial least-squares (G/PLS) algorithm may be used as an alternative to a GFA calculation. G/PLS[50] is derived from two QSAR calculation methods: GFA and partial least-squares (PLS). The G/PLS algorithm uses GFA to select appropriate basis functions to be used in a model of the data and PLS regression as the fitting technique to weigh the basis functions' relative contributions in the final model. PLS is a generalization of regression, which can handle data with strongly correlated and/or noisy or numerous X variables.[51] The linear PLS model finds 'new variables' (latent variables or X scores), which are linear combinations of the original variables. To avoid overfitting, a strict test for the significance of each consecutive PLS component is necessary and then stopping when the components are non-significant. Cross-validation is a practical and reliable method of testing this significance.[51] Application of G/PLS thus allows the construction of larger QSAR equations, while still avoiding overfitting and eliminating most variables.[50] The G/PLS study was conducted using the G/PLS option under QSAR+ environment of Cerius2 software.[49] The settings used were as follows: initial number of variables being 4, number of PLS components being 1 (optimized by cross-validation), number of iterations being 5000, and no fixed length being set for the final equation.

The statistical qualities of the MLR equations[52] were judged by parameters, such as explained variance (R_a^2), correlation coefficient (R), standard error of estimate (s), and variance ratio (F), at specified degrees of freedom (df). All accepted equations have regression coefficients and F ratios significant at 95% and 99% levels, respectively, if not stated otherwise. For PLS equation R_a^2, R^2 and least square error (LSE) were taken as statis-

tical measure while lack-of-fit (LOF) was noted for the GFA derived equations. PRESS (leave-one-out)[53,54] statistics were calculated and the reported parameters are cross-validation R^2 (Q^2), predicted residual sum of squares (PRESS), standard deviation based on PRESS (S_{PRESS}), and standard deviation of error of prediction (SDEP).

3. Results and discussion

From the factor analysis on the data matrix consisting of adenosine A_3 binding affinity data, quantum chemical parameters, lipophilicity ($\log P$), physicochemical substituent constants of phenyl ring substituents and indicator variables, it was observed that seven factors could explain the data matrix to the extent of 95.1% (Table 4). The loading matrix shows that the binding affinity is highly loaded with factor 4 (highly loaded in $\log P$ and $[\log P]^2$) and factor 6 (highly loaded in π), moderately with factor 5 (loaded in q_{11} and q_{20}), factor 3 (highly loaded in q_{15} and q_{16}), and factor 7 (highly loaded in q_{19}), and poorly with factor 2 (loaded in q_1 and q_2) and factor 1 (loaded in q_5, q_6, q_7, q_8, q_9, and q_{10}). Based on the factor analysis, the best equation derived was the following one (Eq. 1), which also emerged as the best equation in the population of 100 equations obtained from GFA.

$$\begin{aligned} &pK_i = 1.913\sigma_p + 0.684I_{NO_2} + 4.654 \\ &n = 17, \quad R_a^2 = 0.726, \quad R^2 = 0.760, \quad R = 0.872, \\ &F = 22.2(\mathrm{df}\ 2, 14), \quad s = 0.465, \\ &Q^2 = 0.653, \quad \mathrm{SDEP} = 0.507, \quad S_{PRESS} = 0.559, \\ &\mathrm{PRESS} = 4.4 \end{aligned} \tag{1}$$

The regression coefficients in Eq. 1 are significant at the 95% level. Eq. 1 could predict and explain 65.3% and 72.6%, respectively, of the variance of the binding affinity. The positive coefficient of σ_p signifies that the binding affinity is directly proportional to the value of σ_p, which means that presence of an electron-withdrawing group at the *para* position of the phenyl ring will be favorable for binding affinity. Furthermore, in the presence of the term σ_p, no quantum chemical term was found to be important. Thus, electronic substituent constant (σ_p) was found to be the best descriptor for modeling the data set. The positive coefficient of I_{NO_2} suggests that the presence of nitro group at position R_1 increases the binding affinity.

An attempt was made to use factor scores as predictor variables to avoid any loss of information on the selection of relevant molecular descriptors from a set of descriptors and significant increase in statistical qualities was obtained.

Table 4. Factor loadings of the variables after VARIMAX rotation

Descriptors	Factor 1[a]	Factor 2[a]	Factor 3[a]	Factor 4[a]	Factor 5[a]	Factor 6[a]	Factor 7[a]	Communalities[b]
pK_i	0.070	0.111	0.360	0.583	0.373	0.446	−0.211	0.870
q_1	0.007	**−0.989**	0.043	−0.041	0.006	−0.006	−0.089	0.989
q_2	0.047	**0.969**	−0.087	0.171	−0.090	0.065	0.020	0.992
q_3	0.482	−0.826	0.005	−0.134	0.080	−0.051	0.143	0.962
q_4	−0.725	0.629	0.099	0.172	0.156	−0.008	−0.030	0.986
q_5	**0.985**	0.067	−0.080	−0.008	−0.102	0.073	−0.027	0.999
q_6	**−0.923**	0.351	0.038	0.099	0.012	−0.017	0.105	0.998
q_7	**0.980**	−0.131	−0.071	−0.014	−0.029	0.031	−0.095	0.994
q_8	**−0.987**	0.038	0.029	−0.064	−0.012	−0.006	−0.008	0.980
q_9	**0.978**	0.100	0.134	0.016	−0.015	0.049	−0.006	0.988
q_{10}	**−0.929**	−0.322	0.006	−0.006	0.138	−0.102	0.038	0.998
q_{11}	0.321	0.361	0.127	−0.153	**−0.802**	0.162	−0.076	0.948
q_{12}	0.790	0.263	−0.115	0.116	0.447	0.045	−0.273	0.996
q_{13}	0.131	0.011	−0.647	0.180	−0.558	0.142	0.281	0.880
q_{14}	−0.590	−0.306	0.608	−0.079	−0.110	0.001	0.377	0.971
q_{15}	0.174	0.276	**−0.883**	0.125	0.028	0.037	0.082	0.910
q_{16}	0.061	0.013	**0.954**	0.186	0.143	0.106	−0.057	0.984
q_{17}	−0.018	−0.142	−0.819	−0.450	−0.129	−0.142	−0.195	0.968
q_{18}	0.043	0.250	0.529	0.470	0.231	0.209	0.501	0.913
q_{19}	0.358	−0.049	0.073	−0.060	0.046	−0.183	**−0.901**	0.987
q_{20}	0.075	0.115	0.430	0.306	**0.827**	−0.004	0.082	0.988
π	0.038	−0.008	0.151	−0.041	−0.139	**0.903**	0.275	0.936
MR	0.324	0.040	−0.180	−0.550	0.127	0.601	−0.044	0.821
σ_p	0.154	0.093	0.348	0.395	0.564	0.380	−0.201	0.812
I_{NO_2}	−0.114	0.334	−0.153	0.777	−0.123	0.192	0.260	0.872
I_{NH_2}	−0.239	**−0.942**	0.021	0.124	−0.058	0.039	−0.061	0.968
$\log P$	−0.030	−0.079	−0.166	**−0.921**	−0.211	0.060	−0.029	0.932
$[\log P]^2$	−0.026	0.152	−0.040	**−0.901**	−0.214	0.273	0.066	0.962
I'_{NH_2}	**−0.984**	0.069	0.032	0.048	−0.035	−0.021	0.114	0.991
Variance	0.314	0.211	0.169	0.093	0.078	0.053	0.034	0.951

[a] Factors 1–7 represent factor loadings and have the character of correlation coefficients between the common factors and the variables.

[b] Representing the proportion of the data variance, which can be assigned to common factor space, the communality of a variable is the sum of squares of its loadings in different factors.

$$\begin{aligned} pK_i &= 0.320f3 + 0.518f4 + 0.332f5 \\ &\quad + 0.396f6 - 0.187f7 + 4.294 \\ n &= 17, \quad R_a^2 = 0.786, \quad R^2 = 0.853, \quad R = 0.923, \\ F &= 12.7(\text{df } 5, 11), \quad s = 0.411, \\ Q^2 &= 0.631, \quad \text{SDEP} = 0.524, \quad S_{\text{PRESS}} = 0.651, \\ \text{PRESS} &= 4.6 \end{aligned} \tag{2}$$

The descriptors used in Eq. 2 represent factor scores for the compounds. Eq. 2 could predict and explain 63.1% and 78.6%, respectively, of the variance of the binding affinity. Though the explained variance value of Eq. 2 is better than that of Eq. 1, the predicted variance of the former is slightly lower than that of the latter. The factor scores as mentioned in Eq. 2 signify the importance of different variables, as shown in boldface in Table 4.

Applying the G/PLS technique (number of components = 1, number of iterations = 5000), the following equation was obtained as the best one among the population of equations.

$$\begin{aligned} pK_i &= 1.847\sigma_p + 0.792I_{NO_2} - 2.375q_{15} + 2.507q_{19} + 4.452 \\ n &= 17, \quad R_a^2 = 0.744, \quad R^2 = 0.808, \quad R = 0.899, \\ Q^2 &= 0.648, \quad \text{LSE} = 0.143, \quad \text{SDEP} = 0.511, \\ S_{\text{PRESS}} &= 0.608, \quad \text{PRESS} = 4.4 \end{aligned} \tag{3}$$

Eq. 3 can predict 64.8% and explain 74.4% of the variance of the binding affinity. The positive coefficients of σ_p indicate that electron-withdrawing substituents at the *para* position of the phenyl ring would be favorable for the binding affinity. Again, positive coefficient of I_{NO_2} indicates that presence of nitro group at R_1 increases the binding affinity. The negative coefficient of q_{15} indicates that negative charge on C_{15} is conducive to binding affinity. The positive coefficient of q_{19} signifies that positive charge on C_{19} is favorable for binding affinity. Both q_{15} and q_{19} are mainly dependent on the type of R substituent.

The calculated and leave-one-out predicted A_3 binding affinity values according to Eqs. 1–3 are given in Table 1. A comparison among equation statistics and prediction statistics of different models is given in Table 5. Appearance of the two parameters σ_p and I_{NO_2} in all the models (in the form of factor scores in Eq. 2) suggests the importance of these terms in modeling the data set. The LOO statistic of Eq. 1 is better than those of the other two equations, while R^2 value of Eq. 2 is higher than those of the other two equations. The scatter plots of observed versus LOO predicted binding affinity values according to Eqs. 1–3 are given in Figure 2.

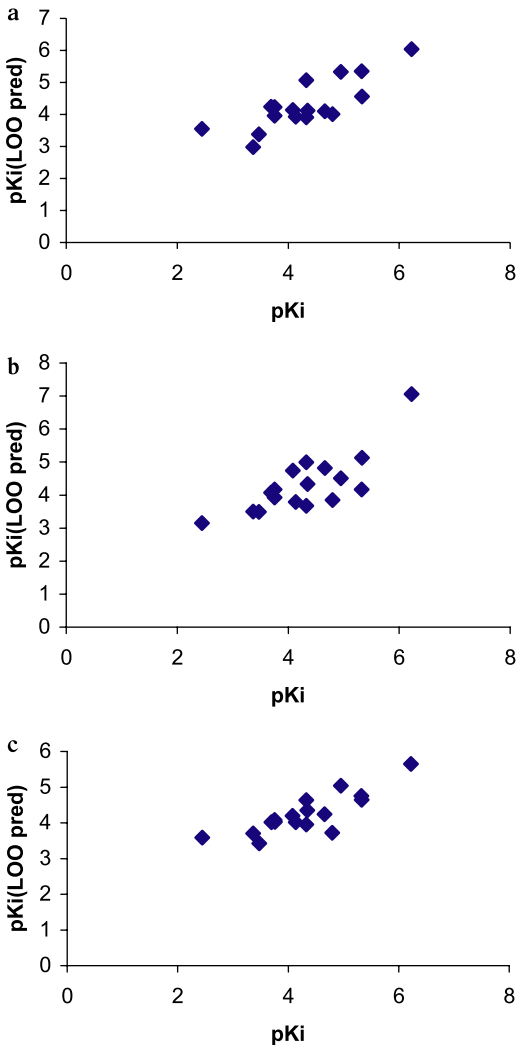


Figure 2. Scatter plots of observed versus leave-one-out predicted binding affinity values according to (a) Eq. 1; (b) Eq. 2; (c) Eq. 3.

4. Conclusion

The study suggests the importance of R and R_1 substituents for A_3 binding affinity. The binding affinity increases with the presence of electron-withdrawing substituents at the R position. Furthermore, Wang–Ford charges of atoms C_{15} and C_{19} (Fig. 1) also play an important role for binding affinity. Again, the presence of a nitro group at R_1 is also favorable for binding affinity. However, more data points covering

Table 5. A comparison of different models

Equation No.	Chemometric tool	Variables appearing in the equation	Equation statistics			Prediction statistics		
			R_a^2	R^2	R	Q^2	PRESS	SDEP
1	FA-MLR and GFA-MLR	σ_p, I_{NO_2}	0.726	0.760	0.872	0.653	4.4	0.507
2	PCRA	Factor scores	0.786	0.853	0.923	0.631	4.7	0.524
3	G/PLS	σ_p, I_{NO_2}, q_{15}, q_{19}	0.744	0.808	0.899	0.648	4.4	0.511

wider features of substitution pattern need to be considered to reach a conclusion.

Acknowledgment

One of the authors (K.R.) thanks the All India Council for Technical Education (AICTE), New Delhi, for a financial grant under the Career Award for Young Teachers (CAYT) scheme.

References and notes

1. Roy, K.; Sengupta, C.; De, A. U. *J. Sci. Ind. Res.* **2001**, *60*, 699.
2. Franke, R. *Theoretical Drug Design Methods*, Vol. 7, Elsevier, Amsterdam 1984.
3. Jambhekar, S. S., 4th ed In *Pinciples of Medicinal Chemistry*; Foye, W. O., Lemk, T. L., Williams, D. A., Eds.; BI Waverly Pvt Ltd: New Delhi, 1995; Vol. 12.
4. Hansch, C.; Leo, A.; Hoekman, D. H. In *Exploring QSAR*; American Chemical Society: Washington, DC, 1995.
5. Hansch, C.; Fujita, T. *J. Am. Chem. Soc.* **1964**, *86*, 1616.
6. Ramirez, S. H.; Fan, S.; Maguire, C. A.; Perry, S.; Hardiek, K.; Ramkumar, V.; Gelbard, H. A.; Dewhurst, S.; Maggirwar, S. B. *J. Neurosci. Res.* **2004**, *77*, 258.
7. Merighi, S.; Benini, A.; Mirandola, P.; Gessi, S.; Varani, K.; Leung, E.; Maclennan, S.; Borea, P. A. *J. Biol. Chem.* **2005**, *280*, 19526.
8. Cadieux, J. S.; Leclerc, P.; StOnge, M.; Dussault, A. A.; Laflamme, C.; Picard, S.; Ledent, C.; Borgeat, P.; Pouliot, M. *J. Cell Sci.* **2005**, *118*, 1437.
9. Fleisher, L. N.; Ferrell, J. B.; McGahan, M. C. *Exp. Eye Res.* **1995**, *60*, 165.
10. <http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=601752=601752>.
11. Teran, G. H.; Conteno, B. N.; Pastor, M.; Sanz, F. *Proteins: Struct. Funct. Bioinformatics* **2003**, *54*, 705.
12. <http://www.anaesthetist.com/anaes/drugs/atp.htm/>.
13. <http://www.anaesthetist.com/icu/manage/drugs/adenosine/index.htm/>.
14. <http://pharmrev.aspetjournals.org/cgi/content/full/49/4/381#B147/>.
15. Dohrman, D. P.; Diamond, I.; Gordon, A. S. *Alcohol Health Res. World* **1997**, *21*, 136.
16. Kuwahata, T.; Kurume *Med. J.* **2004**, *51*, 109.
17. Reichelt, M. E.; Willems, L.; Molina, J. G.; Sun, C. X.; Noble, J. C.; Ashton, K. J.; Schnermann, J.; Blackburn, M. R.; Headrick, J. P. *Circ. Res.* **2005**, *96*, 363.
18. Shneyvays, V.; Leshem, D.; Zinman, T.; Mamedova, L. K.; Jacobson, K. A.; Shainberg, A.; Gonda. *Am. J. Physiol. Heart. Circ. Physiol.* **2005**, 288, H2792.
19. Wu, L.; Belardinelli, L.; Zablocki, J. A.; Palle, V.; Shryock, J. C. *Am. J. Physiol. Heart Circ. Physiol.* **2001**, *280*, H334.
20. Basheer, R.; Strecker, R. E.; Thakkar, M. M.; McCarley, R. W. *Prog. Neurobiol.* **2004**, *73*, 379.
21. Donoso, M.V.; Lopez, R.; Miranda, R.; Briones, R.; HuidobroToro, J. P. *Am. J. Physiol. Heart Circ. Physiol.* **2005**, 288 (in press).
22. Welch, W. J. *Curr. Opin. Pharmacol.* **2002**, *2*, 165.
23. Randhawa, M. P., Jr.; Lasley, R. D.; Mentzer, R. M., Jr. *J. Card. Surg.* **1993**, *8*, 332.
24. <http://www.cahr.acrv.ca/english/resources/abstracts_2002/toc004.html/>
25. Feoktistov, I.; Goldstein, A.; Sheler, J. R.; Schwartz, L. B.; Biaggioni, I. *Drug Dev. Res.* **2003**, *58*, 461.
26. Pernille, B.; Hansen; Schnermann, J. *Am. J. Physiol. Renal Physiol.* **2003**, *0363*, 6127.
27. Desai *Mol. Pharmacol.* **2005**, *67*, 1406.
28. <http://www.anaesthetist.com/icu/manage/drugs/adenosine/asthma/>
29. Rubino, A.; Ralevic, V.; Burnstock, G. *Br. J. Pharmacol.* **1995**, *115*, 648.
30. Madi, L.; Ochaion, A.; Rath-Wolfson, L.; Bar-Yehuda, S.; Erlanger, A.; Ohana, G.; Harish, A.; Merimski, O.; Barer, F.; Fishman, P. *Clin. Cancer Res.* **2004**, *10*, 4472.
31. Kiryat-Matalon, P. T. *Clin. Cancer Res.* **2004**, *10*, 472.
32. VonLubitz, D. K. *Eur. J. Pharmacol.* **1999**, *365*, 9.
33. Roy, K. *Indian J. Chem. B* **2003**, *42*, 1485.
34. Roy, K. *QSAR Comb. Sci.* **2003**, *22*, 614.
35. Bhattacharya, P.; Leonard, J. T.; Roy, K. *Bioorg. Med. Chem.* **2005**, *13*, 1159.
36. Bhattacharya, P.; Leonard, J. T.; Roy, K. *J. Mol. Model.* **2005**, 11 (in press) http://dx.doi.org/10.1007/s00894-005-0273-6
37. Colotta, V.; Catarzi, D.; Varano, F.; Calabri, F. R.; Lenzi, O.; Filacchioni, G.; Martini, C.; Trincavelli, L.; Deflorian, F.; Moro, S. *J. Med. Chem.* **2004**, *47*, 3580.
38. Kubinyi, H. In *Burger's Medicinal Chemistry and Drug Discovery*; 5th ed. Wolff, M. E., Ed.; Wiley: New York, 1995; Vol. 1, pp 507–509.
39. Dewar, M. J. S.; Zoebisch, E. G.; Healey, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
40. Dewar, M. J. S.; Hwang, C. H.; Kuhn, D. R. *J. Am. Chem. Soc.* **1991**, *113*, 735.
41. Civcir, P. U. *J. Mol. Struct.: THEOCHEM* **2001**, *535*, 121.
42. *Chem 3D Pro version 5.0* and *Chem Draw Ultra version 5.0* are programs of CambridgeSoft Corporation, USA.
43. Roy, K.; Leonard, J. T.; Sengupta, C. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3705.
44. Franke, R. *Theoretical Drug Design Methods*; Elsevier: Amsterdam, 1984, pp 184–195.
45. Franke, R.; Gruska, A. In *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; VCH: Weinheim, 1995, pp 113–163.
46. SPSS is statistical software of SPSS. Inc, II USA.
47. Rogers, D.; Hopfinger, A. J. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854.
48. Fan, Y.; Shi, L. M.; Kohn, K. W.; Pommier, Y.; Weinstein, J. N. *J. Med. Chem.* **2001**, *44*, 3254.
49. Cerius^2 version 4.8 is a product of Accelyrs Inc, San Diego, USA.
50. <http://mmlin.1.pha.unc.edu/~jin/QSAR/GA_PLS/gapls.html/>.
51. Wold, S. In *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; VCH: Weinheim, 1995, pp 195–218.
52. Snedecor, G. W.; Cochran, W. H. *Statistical Methods*; Oxford and IBH Publishing Co Pvt Ltd: New Delhi, 1967, pp 381–418.
53. Wold, S.; Eriksson, L. In *Chemometric Methods in Molecular Design*; van de Waterbeemed, H., Ed.; VCH: Weinheim, 1995, pp 312–317.
54. Debnath, A. K. In *Combinatorial Library Design and Evaluation*; Ghose, A. K., Viswandhan, V. N., Eds.; Marcel Dekker Inc: New York, 2001, pp 73–129.