Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 13 (2005) 1159–1165
Bioorganic & Medicinal Chemistry

Exploring QSAR of thiazole and thiadiazole derivatives as potent and selective human adenosine A_3 receptor antagonists using FA and GFA techniques

Prosenjit Bhattacharya, J. Thomas Leonard 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 15 October 2004; revised 10 November 2004; accepted 11 November 2004
Available online 2 December 2004

Abstract—Binding affinity data of thiazole and thiadiazole derivatives ($n = 30$) for human adenosine A_3 receptor subtype have been subjected to Quantitative Structure–Activity Relationship (QSAR) analysis using quantum chemical and hydrophobicity parameters. Wang-Ford charges of the common atoms of the compounds [calculated from molecular electrostatic potential surface of energy minimized geometry using Austin Model 1 (AM1) technique] were used as independent variables apart from partition coefficient ($\log P$) and suitable dummy parameters. The variables for the multiple regression analyses were selected based on principal component factor analysis (FA), and generated equations were statistically validated using leave-one-out technique. The best equation thus obtained explained and predicted 74.4% and 68.9% respectively of the variance of the binding affinity. The results suggested importance of Wang-Ford charges of atoms C_2, C_5 and C_7. Furthermore, the A_3 binding affinity increases with decrease of lipophilicity of the compounds and in the presence of methyl or ethyl substituent at R position. Again, the binding affinity decreases in the presence of *tert*-butyloxy group at R position. When factor scores were used as predictor variables in principal component regression analysis, the resulted model showed 87.0% predicted variance and 89.5% explained variance. The data set was also modeled using genetic function approximation (GFA) technique. The best two equations derived from GFA show better predicted variance values (0.753 and 0.739) than that found in case of the best equation derived from FA. However, considerable intercorrelation was found between two predictor variables in case of GFA derived equations. GFA derived equations show importance of Wang-Ford charges of different atoms of the thiazole/thiadiazole nucleus and phenyl ring (S_9, X_8 and C_2, the effects of the first two being predominant) along with similar impact of lipophilicity and R group on the binding affinity as found in case of the FA derived relation.

1. Introduction

Adenosine is probably the most important neuromodulator in the central and peripheral nervous systems. The purine nucleoside adenosine modulates diverse physiological functions including induction of sedation, vasodilatation, suppression of cardiac rate and contractility, neurotransmitter release, inhibition of platelet aggregation and lipolysis.[1] Adenosine released from cells interacts with membrane receptors (adenosine receptors, ARs). At least four subtypes of adenosine receptor exist (A_1, A_{2A}, A_{2B} and A_3) which have been cloned from animal or human sources.[2]

The human adenosine A_1 receptor is chiefly linked to inhibition of adenylyl cyclase activity.[3] However, there is also good evidence for coupling (via G-proteins) to ion channels, and phospholipase C.[4] In the nervous system, the A_1 adenosine receptor is thought to mediate the inhibition of transmitter release and the reduction in neuronal activity. Blockade of this receptor in the heart leads to the accelerated, pronounced 'pounding' observed after drinking large amounts of strong coffee (due to the caffeine and theophylline).[5] A_1 AR protein is highly expressed in brain (especially cerebellum, hippocampus, thalamus, and cortex) and spinal cord and in part, modulates neurotransmitter release.[6] Other

Keywords: QSAR; Adenosine A_3 receptor; Lipophilicity; Wang-Ford charges; Thiazole and thiadiazole derivatives; Genetic function approximation.
* Corresponding author. Tel.: +91 33 2414 6676; fax: +91 33 2414 6677; e-mail: kunalroy_in@yahoo.com

0968-0896/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2004.11.022

tissues including kidney and testis also express A_1 AR. In white adipocytes A_1 AR inhibits lipolysis and stimulates glucose uptake.

A_{2A} subtype appears to be almost exclusively coupled to stimulation of adenylyl cyclase activity.[7] Its distribution in the central nervous system (CNS) is very discrete, being heavily localised in the caudate and putamen bodies, and the nucleus accumbens and olfactory tubercle.[8] There is growing interest in this receptor as a means of influencing dopamine-mediated responses in these brain regions. A_{2A} has emerged as attractive target as novel therapeutic agents for Parkinson disease.[9] In the periphery, the A_{2A} receptor is present on platelets and is anti-aggregatory. Low affinity receptor A_{2A} stimulates the cyclase via G proteins. A_{2A} receptor is believed to cause vasorelaxation.

The function of the A_{2B} adenosine receptor is obscure, although it mediates the largest observed stimulation of cyclic AMP levels in human brain slices. A_{2B} receptor has been found to mediate vasodilatation in some vascular beds,[10] inhibit vascular smooth muscle growth and collagenase expression, stimulate cytokine synthesis, modulate intestinal functions and neurosecretions.[11,12] This receptor subtype also controls hepatic glucose balance, cell growth, gene expression,[13] mast cell degranulation.[14] Recently a growing interest for the discovery of selective A_{2B} receptor antagonist stemmed from their potential use as antidiabetic[15] and as anti asthmatic agent.[16,17]

The A_3 adenosine receptor is the most recently cloned of the adenosine receptors, and still awaits conclusive proof of its involvement in physiological processes, although a potential role in mast cells has been proposed. Activation of A_3 receptors has been shown to stimulate phospholipase C and to inhibit adenylate cyclase. A_3 agonist also stimulate phospholipase D, and release of inflammatory mediators, such as histamine from mast cells, which are responsible for inflammation and hypotension. For these reasons clinical use of A_3 adenosine receptor antagonists for the treatment of asthma and inflammatory disorders has been suggested. Recent studies also indicated a possible use of these derivatives as antitumor agent and hypoxia.[18,19] This subtype is the smallest receptor yet cloned (316 aa) and couples to inhibition of adenylyl cyclase activity.

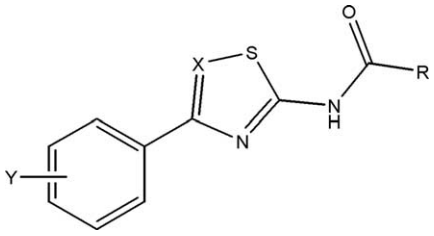
In continuation of our recent effort to model adenosine receptor ligands,[20–22] the present paper attempts Quantitative Structure–Activity Relationship (QSAR) modeling of A_3 subtype receptor binding data[23] of thiazole and thiadiazole derivatives using lipophilicity, quantum chemical and suitable indicator parameters.

2. Materials and methods

Adenosine A_3 receptor binding affinity data[23] of thiazole and thiadiazole 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 (μM)] and then used for subsequent QSAR analyses as the response variable. For the present QSAR study, lipophilicity, quantum chemical and appropriate indicator parameters (listed in Table 2) were used as the predictor variables as these were found appropriate for the development of models. Semi-empirical quantum chemical calculations were done according to AM1 (Austin Model 1)[24–26] method using Chem 3D Pro[27] package. The general structure of the compounds (Fig. 1) was drawn in Chem Draw Ultra ver 5.0[27] and it was saved as the template structure. For every compound, the template structure was suitably changed considering its structural features, copied to Chem 3D ver 5.0[27] 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 these maintain same serials in all the models (Fig. 1). Energy minimization was done under Molecular Orbital Package (MOPAC) module [CS MOPAC Pro] of Chem 3D Pro[27] using restricted closed shell wave function[21,22] and setting 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 calculation of Wang-Ford charges q_x (obtained from molecular electrostatic potential surface) of different atoms (x). Lipophilicity ($\log P$) values of the compounds (Table 1) were calculated according to Ghose and Crippen's fragmentation method[28] using Chem Draw Ultra ver 5.0[27] Henceforth, reference to a particular atom will be made using the numbering system appearing in Fig. 1, for example, sulphur in the thiazole/thiadiazole nucleus will be indicated as S_9.

Though classical approach of multiple regression technique was used as the final statistical tool for developing QSAR relations, factor analysis[29,30] 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 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 by principal component method and then rotated by VARIMAX rotation (a kind of rotation which is used in 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 largest possible number of factor loadings becomes zero. This is done to obtain 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 important in explaining variance of the activity. Further, variables with non-zero

Table 1. Structural features, lipophilicity (log *P*), observed, calculated and LOO predicted adenosine A_3 binding affinity data of thiazole and thiadiazole derivatives



Sl. No.	Structural features			log *P*	Adenosine A_3 receptor binding affinity [pK_i (μM)]						
	R	X	Y		Obs.[a]	Calc.[b]	Pred.[b]	Calc.[c]	Pred.[c]	Calc.[d]	Pred.[d]
1	CH_3	CH	H	2.76	4.738	5.003	5.063	5.015	5.045	5.029	5.079
2	$(CH3)_3CO$	CH	H	4.22	2.292	2.166	2.020	2.243	2.199	2.219	2.143
3	$NCCH_2$	CH	H	2.96	3.690	4.009	4.127	3.859	4.007	4.126	4.221
4	CH_3	CH	4-Cl	3.32	4.293	4.805	4.974	4.632	4.884	4.038	3.991
5	$C_6H_5CH_2$	CH	4-Cl	5.16	4.000	3.372	2.940	3.951	3.944	3.743	3.696
6	CH_3	CH	4-CH_3O	2.63	5.523	5.571	5.583	5.566	5.576	5.417	5.403
7	CH_3	CH	3-CH_3O	2.63	5.387	5.421	5.436	5.726	5.868	5.418	5.426
8	CH_3	CH	2-CH_3O	2.63	4.086	4.044	3.927	4.111	4.156	4.001	3.816
9	CF_3	CH	4-CH_3O	3.78	3.276	3.984	4.090	3.282	3.286	3.731	3.874
10	CH_3CH_2	CH	4-CH_3O	3.29	5.620	5.166	5.019	5.278	5.194	4.885	4.768
11	$CH_3CH_2CH_2$	CH	4-CH_3O	3.70	5.108	4.487	4.411	4.689	4.651	5.042	5.030
12	$(CH_3)_2CH$	CH	4-CH_3O	3.85	4.788	4.106	4.020	4.088	3.998	4.390	4.318
13	$NCCH_2$	CH	4-CH_3O	2.83	4.614	4.639	4.647	4.526	4.509	5.179	5.238
14	$(CH_3)_3C$	CH	4-CH_3O	4.56	4.496	4.206	4.188	4.339	4.332	4.196	4.174
15	$(CH_3)_3CO$	CH	4-CH_3O	4.09	2.487	2.613	2.759	2.324	2.152	2.560	2.636
16	C_6H_5	CH	4-CH_3O	4.53	4.542	4.484	4.479	4.578	4.582	4.294	4.266
17	$C_6H_5CH_2$	CH	4-CH_3O	4.48	4.848	4.501	4.470	4.690	4.677	4.428	4.394
18	$C_6H_5CH_2CH_2$	CH	4-CH_3O	4.89	4.536	4.128	4.096	4.291	4.275	4.080	4.049
19	*p*-$CH_3OC_6H_4CH_2$	CH	4-CH_3O	4.35	2.936	3.857	3.959	3.497	3.579	4.165	4.259
20	*p*-$CH_3OC_6H_4CH_2CH_2$	CH	4-CH_3O	4.77	4.544	3.749	3.652	4.092	4.049	3.933	3.876
21	$(C_6H_5)_2CH$	CH	4-CH_3O	5.88	3.279	4.078	4.307	3.706	3.822	3.969	4.157
22	$(C_6H_5)_2CHCH_2$	CH	4-CH_3O	6.41	3.398	3.646	3.768	3.540	3.638	3.414	3.424
23	2-Furan	CH	4-CH_3O	3.15	4.502	4.868	4.962	4.803	4.916	4.629	4.679
24	Thiophene-2-CH_2	CH	4-CH_3O	4.46	4.491	4.289	4.276	4.376	4.370	4.242	4.224
25	2-Thiophene	CH	4-CH_3O	4.51	4.159	4.483	4.514	4.392	4.409	4.284	4.298
26	CH_3	N	H	3.58	5.638	5.533	5.491	5.711	5.731	5.919	5.993
27	$C_6H_5CH_2$	N	H	5.42	4.102	4.309	4.399	4.275	4.344	4.445	4.512
28	CH_3	N	4-CH_3O	3.45	6.102	5.986	5.941	6.025	5.998	5.936	5.882
29	$C_6H_5CH_2$	N	4-CH_3O	5.30	4.623	4.737	4.770	4.613	4.610	4.515	4.496
30	CH_3CH_2	N	4-CH_3O	4.11	5.945	5.803	5.753	5.821	5.774	5.817	5.775

[a] Ref. 23; Obs. = Observed; Calc. = Calculated; Pred. = Predicted (LOO).
[b] From Eq. 1.
[c] From Eq. 2.
[d] From Eq. 3.

Table 2. Definitions of variables

Variables	Definition
q_x	Wang-Ford charge of atom *x* (*x* may take values 1 to 14).
I_{OBu_t}	Indicator variable having value 1 if *tert*-butyloxy group is present at R position, value 0 otherwise.
I_X	Indicator variable having value 1 if *X* = N, value 0 otherwise.
I_{Me_Et}	Indicator variable having value 1 if R = methyl or ethyl, value 0 otherwise.

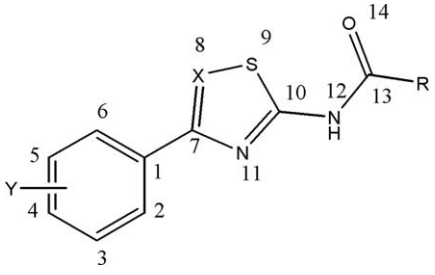


Figure 1. General structure of thiazole (X=CH) and thiadiazole (X=N) derivatives: the common atoms have been numbered 1 through 14.

loadings in different factors were combined in a multivariate equation. Attempt was also made to perform principal component regression analysis (PCRA)[29] taking factor scores as the predictor variables and adopting 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.[29] In PCRA, all descriptors are assumed to be important while the aim of factor analysis is to identify relevant descriptors. The factor analysis and multiple regression were performed using the statistical software SPSS.[31]

The data set was also modeled using genetic function approximation (GFA) technique[32,33] to generate a population of equations rather than one single equation for correlation between the binding affinity and descriptors. GFA involves the combination of multivariate adaptive regression splines (MARS) algorithm with genetic algorithm to evolve 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. This is done as follows: (i) an initial population of equations is generated by random choice of descriptors; (ii) pairs from the population of equations is chosen at random and 'crossovers' are performed and progeny equations are generated; (iii) it is better at discovering combinations of features that take advantage of correlations between multiple features; (iv) the fitness of each progeny equation is assessed by lack-of-fit (LOF) measure; (v) it can use a larger variety of equation term types in construction of its models; (vi) if the fitness of new progeny equation is better, then it is preserved. The models with proper balance of all statistical terms are used to explain variance in the biological activity. 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 done using the GFA module under QSAR + environment of Cerius2 software.[34] All default settings were used for the analysis (number of crossovers = 5000, linear terms, smoothness factor = 1, mutation probability for adding new term = 50%).

The statistical qualities of the equations[35] were judged by the parameters like 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. PRESS (leave-one-out)[36,37] and bootstrap 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}), standard deviation of error of prediction (SDEP) and bootstrap r^2 (bsr^2). Finally, 'leave-many-out' cross-validation was applied on the final equations by deleting 20% of the compounds in 5 cycles and predicting the binding affinities of the deleted compounds in each cycle from the corresponding equation derived from the reduced data set.

3. Results and discussion

From the factor analysis on the data matrix consisting of adenosine A_3 binding affinity data, quantum chemical parameters, lipophilicity and indicator variables, it was observed that 8 factors could explain the data matrix to the extent of 97.0% (Table 3). The binding affinity is highly loaded in factor 8 (highly loaded with $I_{\mathrm{Me\_Et}}$), factor 7 (highly loaded with $I_{\mathrm{OBu\_t}}$) and factor 3 (loaded with q_7 and I_X), moderately in factor 4 (highly loaded with $\log P$ and $[\log P]^2$), factor 2 (highly loaded with q_{12}, q_{13} and q_{14}) and factor 6 (highly loaded with q_2) and poorly loaded in factor 1 (highly loaded with q_3, q_4, q_5 and q_6) and factor 5 (highly loaded with q_8, q_9 and q_{11}). Based on the factor analysis the following equation was derived with six predictor variables.

$$\begin{aligned} \mathrm{p}K_{\mathrm{i}} &= -3.951(\pm 2.415)q_2 - 2.623(\pm 2.136)q_5 \\ &\quad + 2.441(\pm 1.489)q_7 - 0.275(\pm 0.261)\log P \\ &\quad - 1.498(\pm 0.763)I_{\mathrm{OBu\_t}} \\ &\quad + 0.895(\pm 0.614)I_{\mathrm{Me\_Et}} + 4.030 \\ n &= 30, \quad R_a^2 = 0.744,\ R^2 = 0.797,\ R = 0.893, \\ F &= 15.0(\mathrm{df}6, 23),\ s = 0.483, \\ Q^2 &= 0.689,\ \mathrm{SDEP} = 0.523,\ S_{\mathrm{PRESS}} = 0.597, \\ \mathrm{PRESS} &= 8.2,\ \mathrm{bs}r^2(\pm \mathrm{sd}) = 0.798(\pm 0.007) \end{aligned} \tag{1}$$

The 95% confidence intervals of the regression coefficients are shown within parentheses. Eq. 1 could predict 68.9% and explain 74.4% of the variance of A_3 binding affinity. The bootstrap r^2 value of Eq. 1 is 0.798. The negative coefficient of $\log P$ in Eq. 1 indicates that hydrophobicity contributes negatively to the binding affinity. The negative coefficients q_2 and q_5 indicate that negative charges on C_2 and C_5 are conducive to the binding affinity (Fig. 1). The charges on C_2 and C_5 largely depend on type of phenyl ring substituent (Y). Again, the positive coefficient of q_7 indicates that the binding affinity increases as the positive charge on C_7 increases (Fig. 1). The charge on C_7 depends on the atom present at X position: nitrogen at X_8 increases positive charge on C_7 and thus increases the binding affinity. Hence, thiadiazole nucleus will be preferable over thiazole nucleus for the A_3 binding affinity. The negative coefficient of $I_{\mathrm{OBu\_t}}$ indicates that *tert*-butyloxy group at R position is not favorable for the binding affinity. Again, the positive coefficient of $I_{\mathrm{Me\_Et}}$ suggests that methyl or ethyl group at R position increases the binding affinity. Larger and more hydrophobic substituents at R position decrease the binding affinity, as further evidenced from the negative coefficient of $\log P$. This is further corroborated by the docking study of Jung et al.[23] which shows that there is a serine residue (S170) in close proximity of R group suggesting preference of a relatively hydrophilic group.

Attempt was also made to use factor scores as the predictor variables to avoid loss of information on selection of relevant molecular descriptors from the set of descrip-

Table 3. 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]	Factor 8[a]	Communalities[b]
pK_i	−0.113	0.293	0.436	−0.386	0.100	−0.250	−0.444	0.481	0.939
q_1	0.610	−0.299	−0.657	0.059	−0.023	−0.263	−0.012	−0.116	0.981
q_2	−0.297	0.136	0.197	−0.168	−0.069	**0.887**	0.033	0.091	0.975
q_3	**0.830**	0.021	0.031	−0.106	0.119	−0.407	0.056	0.116	0.898
q_4	**−0.945**	0.221	−0.007	0.148	0.014	−0.031	0.007	−0.046	0.967
q_5	**0.970**	−0.184	−0.002	−0.060	−0.008	−0.029	−0.010	0.028	0.981
q_6	**−0.845**	0.217	0.398	0.009	−0.024	0.066	−0.047	0.001	0.926
q_7	−0.171	0.132	**0.938**	0.047	−0.203	0.097	−0.040	0.115	0.994
q_8	0.188	−0.147	−0.592	−0.345	**0.673**	−0.051	0.077	−0.019	0.989
q_9	−0.044	0.002	0.012	0.483	**−0.798**	0.102	−0.012	−0.269	0.955
q_{10}	0.211	−0.750	0.236	0.264	−0.491	−0.067	0.016	−0.082	0.985
q_{11}	−0.037	0.074	−0.660	−0.199	**0.683**	−0.017	0.082	−0.083	0.962
q_{12}	−0.160	**0.945**	0.044	−0.060	−0.124	−0.004	−0.152	0.124	0.977
q_{13}	0.166	**−0.921**	−0.010	−0.144	0.231	−0.063	0.179	0.037	0.987
q_{14}	−0.215	**0.889**	0.235	0.003	0.226	0.085	−0.135	−0.041	0.970
$\log P$	−0.129	−0.020	0.104	**0.944**	−0.223	−0.089	0.026	−0.134	0.995
$[\log P]^2$	−0.096	−0.003	0.090	**0.959**	−0.213	−0.066	0.007	−0.095	0.995
I_{OBu_t}	0.015	−0.447	−0.102	0.011	0.082	0.003	**0.873**	−0.072	0.986
I_X	0.028	−0.089	**0.976**	0.050	−0.029	0.045	−0.072	0.048	0.974
I_{Me_Et}	0.261	0.042	0.263	−0.449	0.188	0.215	−0.095	**0.735**	0.971
Variance	0.201	0.184	0.183	0.136	0.106	0.059	0.053	0.048	0.970

[a] Factors 1–8 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 sum of squares of its loadings in different factors.

tors and significant increase in statistical qualities was obtained.

$$\begin{aligned} pK_i &= 0.280(\pm0.118)f2 + 0.415(\pm0.118)f3 \\ &\quad - 0.368(\pm0.118)f4 - 0.239(\pm0.118)f6 \\ &\quad - 0.424(\pm0.118)f7 + 0.459(\pm0.118)f8 + 4.401 \\ n &= 30, \quad R_a^2 = 0.895,\ R^2 = 0.916,\ R = 0.957, \\ F &= 42.0(\mathrm{df}6, 23),\ s = 0.310, \\ Q^2 &= 0.870,\ \mathrm{SDEP} = 0.338,\ S_{\mathrm{PRESS}} = 0.386, \\ \mathrm{PRESS} &= 3.4,\ \mathrm{bs}r^2(\pm\mathrm{sd}) = 0.917(\pm0.001) \end{aligned} \quad (2)$$

The descriptors used in Eq. 2 represent factor scores for the compounds. Eq. 2 could predict and explain 87.0% and 89.5%, respectively, of the variance of the binding affinity. The factor scores as mentioned in Eq. 2 signify the importance of different variables as shown in bold face in Table 3.

Among the population of models obtained from GFA, the best two models are the followings:

$$\begin{aligned} pK_i &= -1.940(\pm2.072)q_2 - 11.413(\pm3.878)q_8 \\ &\quad - 13.611(\pm5.005)q_9 - 0.038(\pm0.029)[\log P]^2 \\ &\quad - 1.556(\pm0.722)I_{OBu\_t} + 3.083 \\ n &= 30, \quad R_a^2 = 0.775,\ R^2 = 0.814,\ R = 0.902, \\ F &= 21.0(\mathrm{df}5, 24),\ s = 0.452, \\ Q^2 &= 0.753,\ \mathrm{SDEP} = 0.466,\ S_{\mathrm{PRESS}} = 0.521, \\ \mathrm{PRESS} &= 6.5,\ \mathrm{bs}r^2(\pm\mathrm{sd}) = 0.814(\pm0.007) \end{aligned} \quad (3)$$

$$\begin{aligned} pK_i &= -1.932(\pm2.150)q_2 - 11.413(\pm4.012)q_8 \\ &\quad - 13.741(\pm5.189)q_9 - 0.311(\pm0.260)\log P \\ &\quad - 1.518(\pm0.747)I_{OBu\_t} + 3.686 \\ n &= 30, \quad R_a^2 = 0.762,\ R^2 = 0.803,\ R = 0.896, \\ F &= 19.6(\mathrm{df}5, 24),\ s = 0.465, \\ Q^2 &= 0.739,\ \mathrm{SDEP} = 0.479,\ S_{\mathrm{PRESS}} = 0.536, \\ \mathrm{PRESS} &= 6.9,\ \mathrm{bs}r^2(\pm\mathrm{sd}) = 0.804(\pm0.007) \end{aligned} \quad (4)$$

The coefficients of Q_2 in Eqs. 3 and 4 are significant at 90% level. Eq. 3 could predict and explain 75.3% and 77.5% respectively of the variance while in case of Eq. 4 predicted and explained variance values are 73.9% and 76.2% respectively. The negative coefficients of $[\log P]^2$ and $\log P$ in Eqs. 3 and 4, respectively, indicate the negative contribution of hydrophobicity to the binding affinity as was found in case of the factor analysis derived equation (Eq. 1). The squared term of $\log P$ in Eq. 3 indicates that the binding affinity decreases geometrically (in second order) as the value of $\log P$ increases arithmetically. As the statistical quality of Eq. 3 is better than that of Eq. 4 as evident from Q^2 and PRESS values, and as both equations contain the same terms except the hydrophobicity one, the dependence of the binding affinity on $\log P$ is clearly evident. The negative coefficients of q_8 and q_9 indicate that negative charge on X_8 and less positive charge on S_9 are favorable for the binding affinity (Fig. 1). This reconfirms that nitrogen at X_8 increases the binding affinity in comparison to the thiazole congener presumably due to a hydrogen bonding interaction with any donor site present in the receptor (Fig. 1). This is further corroborated by the docking study of Jung et al.[23] which shows that the thiadiazole nitrogen (N^8 ; Fig. 1) is involved in

Table 4. Intercorrelation (r) matrix

	q_5	q_7	q_8	q_9	$\log P$	$[\log P]^2$	I_{OBu_t}	I_{Me_Et}
q_2	−0.330	0.351	−0.232	0.059	−0.173	−0.161	−0.065	0.275
q_5	1.000	−0.193	0.229	−0.086	−0.172	−0.152	0.096	0.279
q_7		1.000	−0.771	0.182	0.180	0.171	−0.220	0.258
q_8			1.000	−0.724	−0.550	−0.540	0.251	0.141
q_9				1.000	0.669	0.653	−0.055	−0.540
$\log P$					1.000	0.991	0.027	−0.599
$[\log P]^2$						1.000	−0.005	−0.562
I_{OBu_t}							1.000	−0.175

Table 5. Results of leave-20%-out cross-validation Model equation, $pK_i = \sum \beta_i x_i + \alpha$

Key eq. no.	Number of cycles	Average regression coefficients (standard deviations of mean)	Statistics	
			Q^2	PRESS
			(Average Pres)	(SDEP)
(1)	5[a]	pK_i = −4.083 (0.697) q_2−2.812 (0.852) q_5 + 2.493 (0.405) q_7−0.291 (0.115) $\log P$−1.459 (0.262) I_{OBu_t} + 0.879 (0.198) I_{Me_Et} + 4.047 (0.470)	0.636(0.449)	9.6(0.566)
(2)	5[a]	pK_i = 0.276 (0.030) $f2$ + 0.411 (0.021) $f3$−0.371 (0.028) $f4$−0.242 (0.022) $f6$−0.429 (0.027) $f7$ + 0.458 (0.039) $f8$ + 4.405 (0.030)	0.874(0.256)	3.3(0.332)
(3)	5[a]	pK_i = −1.969 (0.550) q_2−11.346 (0.833) q_8 −13.537 (1.040) q_9−0.039 (0.007) $[\log P]^2$−1.570 (0.109) I_{OBu_t} + 3.103 (0.279)	0.748(0.366)	6.7(0.473)
(4)	5[a]	pK_i = −1.919 (0.534) q_2−11.328 (0.822) q_8 −13.580 (1.171) q_9−0.315 (0.079) $\log P$−1.533 (0.115) I_{OBu_t} + 3.720 (0.320)	0.730(0.383)	7.1(0.486)

Q^2 denotes cross-validated R^2. Average Pres means average of absolute values of predicted residuals PRESS and SDEP denote predicted residual sum of squares and standard deviation error of predictions respectively.

[a] Compounds were deleted in 5 cycles in the following manner: (**1**, **6**, **11**,..., **26**), (**2**, **7**, **12**,..., **27**),..., (**5**, **10**, **15**,..., **30**).

hydrogen bonding with a serine residue (S181) in the receptor. From the regression coefficients of q_2, q_8 and q_9 and their t statistics, it appears that X_8 and S_9 are more important for the binding affinity in comparison to C_2 suggesting importance of the thiadiazole nucleus. Eqs. 3 and 4 show better predicted variance values (0.753 and 0.739) than that found in case of the best equation derived from FA. However, considerable intercorrelation was found between q_8 and q_9 (Table 4) in case of the GFA derived equations.

The calculated and leave-one-out predicted A_3 binding affinity values according to Eqs. 1–3 are given in Table 1. Leave-20%-out cross-validation was applied on Eqs. 1–4 and the results are given in Table 5. Table 5 shows that Eq. 3 is statistically better than Eq. 4 as evidenced from Q^2 and PRESS values.

4. Conclusion

The study suggests the importance of Wang-Ford charges of atoms C_2, C_5, C_7, X_8 and S_9 (Fig. 1), the last two playing the most important roles. Furthermore, the A_3 binding affinity increases with decrease of lipophilicity of the compounds and in the presence of methyl or ethyl substituent at R position. Again, the binding affinity decreases in the presence of *tert*-butyloxy group at R position. In summary, the phenyl ring substituents should be such that charges on C_2 and C_5 are more negative, while R substituent should not be more hydrophobic than methyl or ethyl group. Furthermore, thiadiazole nucleus is more preferred over thiazole nucleus for the binding. The best two equations derived from GFA show better predicted variance values than that found in case of the best equation derived from FA though q_8 and q_9 are considerably intercorrelated in the former case. The best equation statistics were obtained with the principal component regression analysis derived equation.

Acknowledgements

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

References and notes

1. http://www.ccc.nottingham.ac.uk/~mqzwww/adenosine.html.
2. Beukers, M. W.; Wanner, M. J.; Von Frijtag Drabbe Kunzel, J. K.; Klaasse, E. C.; Ijzerman, A. P.; Koomen, G. J. *J. Med. Chem.* **2003**, *46*, 1492.
3. Palmer, T. M.; Fruguson, G.; Watterson, K. R. *Drug Dev. Res.* **2003**, *58*, 302.
4. http://www.bioreagents.com/index.cfm/fuseaction/products.detail/Product/PA1.

5. Lankford, A. R.; Cemiway, R. J.; Regan, S. E.; Crawford, M. M.; Byford, A. M.; Matheme, G. M. *Drug Dev. Res.* **2003**, *58*, 439.
6. Teran, G. H.; Conteno, B. N.; Pastor, M.; Sanz, F. *Proteins: Structure, Function and Bioinformatics* **2003**, *54*, 705.
7. Doytchinova, I.; Valkova, I.; Natcheva, R. *Quant. Struct.—Act. Relat.* **2001**, *20*, 124.
8. Philliis, J. W. *Drug Dev. Res.* **2001**, *52*, 331.
9. Chen, J. F.; Schwarzschild, M. A. *Drug Dev. Res.* **2003**, *58*, 354.
10. Rubino, A.; Ralevic, V.; Burnstock, G. *Br. J. Pharmacol.* **1995**, *115*, 648.
11. Strohmeir, G. R.; Reppert, S. M.; Lencer, W. I.; Madara, J. L. *J. Biol. Chem.* **1995**, *270*, 2387.
12. Feoktistov, I.; Goldstein, A.; Sheller, J. R.; Schwartz, L. B.; Biaggioni, I. *Drug Dev. Res.* **2003**, *58*, 461.
13. Muller, C. E. *Drugs Future* **2000**, *25*, 1043.
14. Carrotti, A.; Stefanachi, A.; Ravina, E.; Sotelo, E.; Loza, M. I.; Cadavid, M. I.; Conteno, N. B.; Nicoltti, O. *Eur. J. Med. Chem.* **2004**, *39*, 879.
15. Feoktistov, I.; Biaggioni, I. *Pharmacol. Rev.* **1997**, *49*, 381.
16. Jacobson, K. A.; Ijzerman, A. P.; Linden, J. *Drug Dev. Res.* **1999**, *47*, 45.
17. Fozard, J. R. *Curr. Opin. Pharmacol.* **2003**, *3*, 264.
18. Baraldi, P. G.; Tabrizi, M. A.; Fruttarolo, F.; Bovero, A.; Avitabile, B.; Preti, D.; Romagnoli, R.; Merighi, S.; Katia, S. G.; Varani; Borea, P. A. *Drug Dev. Res.* **2003**, *58*, 315.
19. Lewerenz, A.; Hentschel, S.; Vissiennon, Z.; Michael, S.; Nieber, K. *Drug Dev. Res.* **2003**, *58*, 420.
20. Roy, K. *Indian J. Chem.* **2003**, *42B*, 1485.
21. Roy, K. *QSAR Comb. Sci.* **2003**, *22*, 614.
22. Roy, K.; Leonard, J. T.; Sengupta, C. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3705.
23. Jung, K.-Y.; Kim, S.-K.; Gao, Z.-G.; Gross, A. S.; Melman, N.; Jacobson, K. A.; Kim, Y.-C. *Bioorg. Med. Chem.* **2004**, *12*, 613.
24. Dewar, M. J. S.; Zoebisch, E. G.; Healey, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
25. Dewar, M. J. S.; Hwang, C. H.; Kuhn, D. R. *J. Am. Chem. Soc.* **1991**, *113*, 735.
26. Civcir, P. U. *J. Mol. Struct.: THEOCHEM* **2001**, *535*, 121.
27. CHEM 3D PRO VERSION 5.0 and CHEM DRAW ULTRA VERSION 5.0 are programs of CambridgeSoft Corporation, U.S.A.
28. Ghose, A. K.; Crippen, G. M. *J. Chem. Inf. Comp. Sci* **1987**, *27*, 21.
29. Franke, R. *Theoretical Drug Design Methods*; Elsevier: Amsterdam, 1984; pp 184–195.
30. Franke, R.; Gruska, A. In *Chemometric Methods in Molecular Design*; Waterbeemd, H. van de, Ed.; VCH: Weinheim, 1995; pp 113–163.
31. SPSS is statistical software of SPSS Inc, II, U.S.A.
32. Rogers, D.; Hopfinger, A. J. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854.
33. Fan, Y.; Shi, L. M.; Kohn, K. W.; Pommier, Y.; Weinstein, J. N. *J. Med. Chem.* **2001**, *44*, 3254.
34. CERIUS2 version 4.8 is a product of Accelyrs, Inc, San Diego, USA.
35. Snedecor, G. W.; Cochran, W. H. *Statistical Methods*; Oxford And IBH Publishing Co. Pvt. Ltd: New Delhi, 1967, pp 381–418.
36. Wold, S.; Eriksson, L. In *Chemometric Methods in Molecular Design*; Waterbeemd, H. van de, Ed.; VCH: Weinheim, 1995; pp 312–317.
37. Debnath, A. K. In *Combinatorial Library Design and Evaluation*; Ghose, A. K., Viswandhan, V. N., Eds.; Marcel Dekker: New York, 2001; pp 73–129.