J Mol Model (2005) 11: 516–524
DOI 10.1007/s00894-005-0273-6

ORIGINAL PAPER

Prosenjit Bhattacharya · J. Thomas Leonard
Kunal Roy

Exploring 3D-QSAR of thiazole and thiadiazole derivatives as potent and selective human adenosine A_3 receptor antagonists+

Received: 27 September 2004 / Accepted: 18 January 2005 / Published online: 1 June 2005

Abstract Binding affinity data [Bioorg Med Chem (2004) 12:613–623] of thiazole and thiadiazole derivatives (n = 30) for the human adenosine A_3 receptor subtype have been subjected to 3D-QSAR (Quantitative structure–activity relationships) analyses by molecular shape analysis (MSA) and molecular field analysis (MFA) techniques using Cerius2 Version 4.8. In the case of the MSA, the major steps were (1) generation of conformers and energy minimization; (2) hypothesizing an active conformer (global minimum of the most active compound); (3) selecting a candidate shape-reference compound (based on the active conformation); (4) performing pairwise molecular superimposition using the maximum common subgroup (MCSG) method; (5) measuring molecular shape commonality using MSA descriptors; (6) determining other molecular features by calculating spatial, electronic and conformational parameters; (7) selection of conformers; (8) generation of QSAR equations by genetic function algorithm (GFA) or stepwise regression. The best 3D-QSAR equation (MSA) obtained from GFA technique shows 70.0% predicted variance (leave-one-out) and 77.7% explained variance. This equation shows the importance of Jurs descriptors (atomic charge weighted positive surface area, relative negative charge and relative positive charge surface area), partial moment of inertia, energy of the most stable conformer and the ratio of common overlap steric volume to volume of individual molecules. In the case of stepwise regression, the best relation showed 46.1% predicted variance and 72.3% explained variance. In the case of MFA, the major steps were (1) generating conformers and energy minimization; (2) matching atoms using a maximum common substructure (MCS) search and aligning molecules using the default options; (3) setting MFA preferences (rectangular grid with 2 Å step size, charges by the Gasteiger algorithm, H^+ and CH_3 as probes); (4) creating the field; (5) analysis by the Genetic partial least squares (G/PLS) method. The equation obtained was of excellent statistical quality: 96.1% explained variance and 71.6% predicted variance. Statistically reliable 3D-QSAR models obtained from this study suggest that these techniques could be useful to design potent A_3 receptor antagonists.

Keywords QSAR · MSA · MFA · Thiazole · Thiadiazole · Adenosine A_3 receptor

Abbreviations QSAR: Quantitative structure–activity relationships · GFA: Genetic function approximation · G/PLS: Genetic partial least squares · MSA: Molecular shape analysis · MFA: Molecular field analysis

Introduction

Adenosine, a metabolite of adenine nucleotides, is a physiological regulator of several cellular activities and cellular metabolism. It acts as an autacoid and activates G protein-coupled membrane receptors (GPCRs), designated as A_1, A_{2A}, A_{2B}, and A_3. Adenosine receptors are present on virtually every cell. However, receptor subtype distribution and densities vary greatly. Adenosine plays an important role in many pathophysiological conditions in the CNS as well as in peripheral organs and tissues [1]. The multiple effects of extracellular adenosine observed in many tissues depend on its ability to bind and activate GPCRs. Adenosine receptors have been considered promising therapeutic targets for treating conditions of the cardiovascular, renal, respiratory,

Presented in 4th Indo-US Workshop on Mathematical Chemistry, Pune, India (8–12 January 2005)

P. Bhattacharya · J. T. Leonard · K. Roy (✉)
Drug Theoretics and Cheminformatics Laboratory,
Division of Medicinal and Pharmaceutical Chemistry,
Department of Pharmaceutical Technology,
Jadavpur University, Kolkata, 700 032, India
E-mail: kunalroy_in@yahoo.com
Tel.: +91-33-28670786

immune, gastrointestinal, and central nervous systems [2]. Adenosine has a wide range of anti-inflammatory properties, mediated mainly by signals transduced via its receptor [3]. Adenosine mediates diverse physiological effects including stimulation of gluconeogenesis [1], suppression of cardiac rate and contractility [4], and protection of heart from hypoxic damage [5].

The A_1 adenosine receptor activation inhibits inflammation, necrosis, and apoptosis after renal ischaemia reperfusion injury in mice [6]. Its activation in CNS leads to neuroprotective effects through the blockade of neurotransmitter release, whereas, in the heart, it is a potential target for cardioprotective and anti-infarct agents [7]. Some A_1 antagonists are undergoing clinical trials as renal protective agents [7].

Specific A_{2A} agonists promote wound healing in both normal animals and in animals with impaired wound healing [8]. The A_{2A} antagonists are being developed as novel therapeutic agents for Parkinson disease based on their capacity to enhance motor function [9]. Activation of A_{2A} also leads to the control of CNS excitability [10]. The A_{2B}receptor has been found to mediate vasodilation, inhibit vascular smooth muscle growth and collagenase expression, stimulate cytokine synthesis, and modulate intestinal functions and neurosecretions [11]. The presence of adenosine A_{2B} receptors in human lung mast cells mediates adenosine-induced bronchoconstriction in asthmatics [11].

Activation of A_3 agonists causes stimulation of phospholipase D and the release of inflammatory mediators, such as histamine from mast cells, which are responsible for inflammation and hypotension [12]. Moreover, the A_3 adenosine receptor blocks ultraviolet (UV)-irradiation-induced apoptosis in mast-like cells [8]. Activation of A_3 also leads to enhancement of intellectual performance and various learning and memory paradigms [13].

Quantitative structure–activity relationship (QSAR) studies have been done on various derivatives acting on different adenosine receptors. Comparative molecular field analysis (CoMFA) has been used on xanthines [2, 14] styryl-xanthines [15] and oxyadenosines [16] to study the affinities for adenosine receptors. Multiple regression analysis was used on 1,3-dimethylxanthines [17], quinazolines [18], quinolines [19] and triazolopurine derivatives [20] for the QSAR study of binding affinities on various adenosine receptors. The present paper deals with 3D-QSAR analysis of the human A_3 receptor binding affinity data of thiazole and thiadiazole derivatives.

Materials and methods

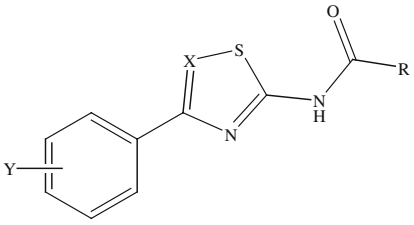
Adenosine A_3 binding affinity data reported by Jung et al. [21] has been used for the present QSAR study. The affinity data [K_i(nM)] of thiazole and thiadiazole derivatives (Table 1) for recombinant human A_3 receptors expressed in CHO (Chinese hamster ovary) cells have been converted to the logarithmic scale [pC(μM)] and then used for subsequent QSAR analyses as the response variable. Some of the compounds reported in the original papers were excluded in the present study because of their non-graded quantitative activity data, the presence of uncommon structural features or outlier behavior.

All computational experiments were conducted with Cerius^2 4.8 [22] version QSAR environment from Accelrys (San Diego, USA) on a Silicon Graphics O2 workstation running under the IRIX 6.5 operating system. Molecular shape analysis (MSA) and Molecular field analysis (MFA) were used as the 3D-QSAR techniques.

The MSA [23] is a formalism that deals with the quantitative characterization, representation and manipulation of molecular shape in the construction of a QSAR. The overall aim of MSA is to identify the biologically relevant conformation without knowledge of the receptor geometry and explain in a quantitative fashion the activity of a series of congeners. The major steps of MSA were (1) generation of conformers and energy minimization; (2) hypothesizing an active conformer (global minimum of the most active compound); (3) selecting a candidate shape reference compound (based on the active conformation); (4) performing pairwise molecular superimposition using the maximum common subgroup (MCSG) method; (5) measuring molecular shape commonality using MSA descriptors; (6) determining other molecular features by calculating spatial, electronic and conformational parameters; (7) selection of conformers; (8) generation of QSAR equations by genetic function algorithm (GFA) or stepwise regression. A complete list of descriptors used in MSA is given in Table 2. Multiple conformations of each molecule were generated using the Boltzmann jump as a conformational search method. The upper limit of the number of conformations per molecule was 150. Each conformer was subjected to an energy minimization procedure using the smart minimizer with the open force field (OFF) to generate the lowest energy conformation for each structure. A conformer of the most active antagonist **28**for the A_3 receptor was selected as a shape reference to which all the structures in the study compounds were aligned through pair-wise superpositioning. The method used for performing the alignment was maximum common subgroup (MCSG) [22]. This method looks at molecules as points and lines, and uses the techniques of graph theory to identify patterns. It finds the largest subset of atoms in the shape-reference compound that is shared by all the structures in the study table and uses this subset for alignment. A rigid fit of atom pairings was performed to superimpose each structure so that it overlays the shape-reference compound.

The major steps of MFA [23] were (1) generating conformers and energy minimization; (2) matching atoms using maximum common substructure (MCS) search and aligning molecules using the default options; (3) setting MFA preferences (rectangular grid with 2 Å

Table 1 Structural features, observed and calculated adenosine A_3 binding affinity data of thiazole and thiadiazole derivatives



Sl. No.	Structural features			Adenosine A_3 receptor binding affinity			
	R	X	Y	Obs.[a]	Calc.[b]	Calc.[c]	Calc.[d]
1	CH_3	CH	H	4.738	4.517	4.472	4.888
2	$(CH_3)_3CO$	CH	H	2.292	2.572	2.783	2.216
3	$NCCH_2$	CH	H	3.690	4.332	4.149	3.498
4	CH_3	CH	4-Cl	4.293	4.302	4.482	4.479
5	$C_6H_5CH_2$	CH	4-Cl	4.000	3.719	4.066	4.073
6	CH_3	CH	4-OCH_3	5.523	4.952	5.007	5.245
7	CH_3	CH	3-OCH_3	5.387	4.707	5.221	5.387
8	CH_3	CH	2-OCH_3	4.086	5.108	4.748	4.061
9	CF_3	CH	4-OCH_3	3.276	3.020	3.190	3.347
10	CH_3CH_2	CH	4-OCH_3	5.620	5.065	5.291	5.262
11	$CH_3CH_2CH_2$	CH	4-OCH_3	5.108	5.025	5.149	5.277
12	$(CH_3)_2CH$	CH	4-OCH_3	4.788	4.932	5.055	4.800
13	$NCCH_2$	CH	4-OCH_3	4.614	4.602	4.630	4.568
14	$(CH_3)_3C$	CH	4-OCH_3	4.496	4.699	4.656	4.666
15	$(CH_3)_3CO$	CH	4-OCH_3	2.487	2.705	2.440	2.610
16	C_6H_5	CH	4-OCH_3	4.542	4.717	4.678	4.620
17	$C_6H_5CH_2$	CH	4-OCH_3	4.848	4.265	4.558	4.846
18	C_6H_5 CH_2CH_2	CH	4-OCH_3	4.536	4.182	4.178	4.512
19	p-$CH_3OC_6H_4CH_2$	CH	4-OCH_3	2.936	3.953	3.796	2.967
20	p-$CH_3OC_6H_4CH_2CH_2$	CH	4-OCH_3	4.544	3.887	3.939	4.489
21	$(C_6H_5)_2CH$	CH	4-OCH_3	3.279	3.362	3.596	3.286
22	$(C_6H_5)_2CHCH_2$	CH	4-OCH_3	3.398	3.301	3.128	3.395
23	2-Furan	CH	4-OCH_3	4.502	4.264	4.507	4.403
24	Thiophene-2-CH_2	CH	4-OCH_3	4.491	4.290	4.027	4.469
25	2-Thiophene	CH	4-OCH_3	4.159	4.611	4.206	4.375
26	CH_3	N	H	5.638	5.226	5.122	5.622
27	$C_6H_5CH_2$	N	H	4.102	4.341	3.638	3.917
28	CH_3	N	4-OCH_3	6.102	6.147	6.094	6.076
29	$C_6H_5CH_2$	N	4-OCH_3	4.623	4.381	4.254	4.590
30	CH_3CH_2	N	4-OCH_3	5.945	5.860	5.972	6.100

[a]Reference [21]; Obs. = Observed; Calc.= Calculated

[b]From Equation (1)

[c]From Equation (2)

[d]From Equation (4)

step size, charges by Gasteiger algorithm, H^+ and CH_3 as probes); (4) creating the field; (5) analysis by the Genetic partial least squares (G/PLS) method. The MFA models are predictive and sufficiently reliable to guide the chemist in the design of novel compounds. This approach is effective for the analysis of data sets where activity information is available but the structure of the receptor site is unknown. The MFA attempts to postulate and represent the essential features of a receptor site from the aligned common features of the molecules that bind to it. This method generates multiple models that can be checked easily for validity. The MFA formalism calculates probe interaction energies on a rectangular grid around a bundle of active molecules. The surface is generated from a ''Shape Field''. The atomic coordinates of the contributing models are used to compute field values on each point of a 3D grid. Grid size was adjusted to default 2.00 Å . The MFA evaluates the energy between a probe (H^+ and CH_3) and a molecular model at a series of points defined by a rectangular grid. Fields of molecules are represented using grids in MFA and each energy associated with an MFA grid point can serve as input for the calculation of a QSAR. These energies were added to the study table to form new columns headed according to the probe type.

Statistical analysis of data was done using techniques like genetic function approximation (GFA) and stepwise regression for MSA and G/PLS for MFA using QSAR + environment of Cerius^2 software [22].

The GFA technique [24, 25] was used to generate a population of equations rather than one single equation for correlation between biological activity and

Table 2 A complete list of descriptors used in MSA

Sl No.	Spatial parameters	Electronic parameters	Molecular shape analysis parameters	Conformational parameters
1	Vm	LUMO	DIFFV	Energy
2	RadOfGyration	HOMO	COSV	
3	Density	Dipole	Fo	
4	PMI		NCOSV	
5	Area		ShapeRMS	
6	Sxy		SRVol	
7	Syz			
8	Sxz			
9	(*Sxy*, *f*)			
10	(*Syz*, *f*)			
11	(*Sxz*, *f*)			
12	Lx			
13	Ly			
14	Lz			
15	η			
16	JursPPSA_1			
17	JursPPSA_2			
18	JursPPSA_3			
19	JursPNSA_1			
20	JursPNSA_2			
21	JursPNSA_3			
22	JursDPSA_1			
23	JursDPSA_2			
24	JursDPSA_3			
25	JursFPSA_1			
26	JursFPSA_2			
27	JursFPSA_3			
28	JursFNSA_1			
29	JursFNSA_2			
30	JursFNSA_3			
31	JursWPSA_1			
32	JursWPSA_2			
33	JursWPSA_3			
34	JursWNSA_1			
35	JursWNSA_2			
36	JursWNSA_3			
37	JursRPCG			
38	JursRNCG			
39	JursRPCS			
40	JursRNCS			
41	JursTPSA			
42	JursTASA			
43	JursRPSA			
44	JursRASA			
45	JursSASA			

physicochemical properties. The GFA involves the combination of the multivariate adaptive regression splines (MARS) algorithm with a genetic algorithm to evolve a 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 encourages the use of splines as a powerful tool for non-linear modeling. The GFA is done as follows: (1) an initial population of equations is generated by random choice of descriptors; (2) pairs from the population of equations are chosen at random and ''crossovers'' are performed and progeny equations are generated; (3) it is better at discovering combinations of features that take advantage of correlations between multiple features; (4) the fitness of each progeny equation is assessed by the LOF measure; (5) it can use a larger variety of equation-term types in construction of its models; (6) if the fitness of a new progeny equation is better, then it is preserved. The model with a proper balance of all statistical terms will be used to explain the variance of 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 G/PLS algorithm may be used as an alternative to a GFA calculation. The G/PLS 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. The PLS is a generalization of regression, which can handle data with strongly correlated and/or noisy or numerous X variables [26]. It gives a reduced solution

that is statistically more robust than multiple linear regression (MLR). 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 [26]. The use of G/PLS thus allows the construction of larger QSAR equations while still avoiding overfitting and eliminating most variables.

The statistical qualities of the MLR equations [27] 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 (d.f.). All accepted MLR equations have regression coefficients and F ratios significant at 95 and 99% levels, respectively, if not stated otherwise. For PLS equations R_a^2, R^2 and least square error (LSE) were taken as statistical measures while LOF was noted for the GFA-derived equations. The 3D-QSAR equations generated were validated by PRESS (leave-one-out) [28, 29] and bootstrap statistics which were calculated using the QSAR + module of the Cerius^2 software [22] 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). Both the model development process and finally developed models were subjected to randomization tests for validation purposes. Additionally, the final models were subjected to leave-20%-out crossvalidation with 15 trials in each case.

Results and discussion

Molecular shape analysis

A view of aligned molecules studied is shown in Fig. 1. The values of important descriptors used in MSA-derived equations are given in Table 3. The best equation obtained from stepwise regression (F value for inclusion of variables was set to 4) is the following:

$$\begin{aligned} pC = & -0.010(\pm 0.006)\text{JursWPSA\_1} \\ & -17.170(\pm 8.925)\text{JursRPCG} \\ & -0.007(\pm 0.004)\text{NCOSV} + 0.005(\pm 0.004)\text{Energy} \\ & +1.712(\pm 1.457)\text{LUMO} + 7.900 \\ & n = 30, R_a^2 = 0.723, R^2 = 0.771, R = 0.878, \\ & F = 16.2(d.f.5,24), s = 0.252, \\ & Q^2 = 0.461, \text{SDEP} = 0.688, S_{\text{PRESS}} = 0.769, \\ & \text{PRESS} = 14.2, bsr^2(\pm\text{SD}) = 0.772(\pm 0.009) \end{aligned} \quad (1)$$

The 95% confidence intervals of regression coefficients are given within parentheses. Equation 1 could explain 72.3% of the variance and predict 46.1% of the variance. The theoretical F value at probability level of 0.01 (d.f. 5, 20) being 4.1, the variance ratio of Eq. 1 is significant at the 99% level, indicating stability of the regression coefficients. The model-development process was subjected to a randomization test with 99 random trials (Table 4). The mean value of the random R's is 0.445 while the value of R from the non-random model is 0.878. In the case of 98 of 99 random trials, the values of random R's were less than the R-value of the non-random model. The final model (Eq. 1) was also subjected to a randomization test with 99 random trials (Table 5). In all cases the values of random R's (mean 0.401) were less than that of the non-random model. The calculated values of binding affinity according to Eq. 1 are given in Table 1. Figure 2a shows a scatter plot of observed versus leave-one-out predicted binding-affinity values. The model was also subjected to a leave-20%-out cross-validation test with 15 trials and the R^2 value between the observed and predicted values was found to be 0.570 (Table 6).

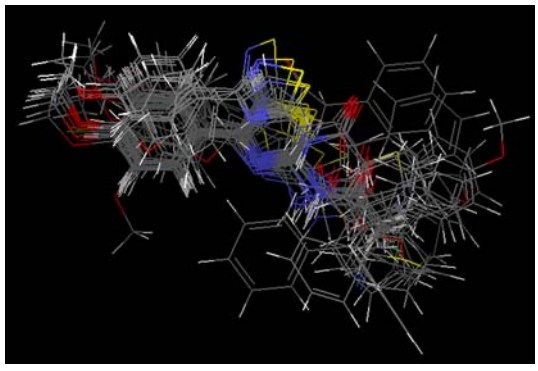


Fig. 1 View of aligned study compounds in MSA

The negative coefficient of JursWPSA_1 (obtained by multiplying the sum of the solvent-accessible surface areas of all positively-charged atoms with the total molecular solvent-accessible surface area and dividing by 1000) in Eq. 1 indicates that the surface-weighted charged partial surface area is detrimental to the binding affinity. Higher values of JursWPSA_1 are observed for compounds **21** and **22**, which have diphenylmethyl and 2-(diphenyl)ethyl substituents, respectively, at the R position and these compounds show less binding affinity. Again, the negative coefficient of JursRPCG indicates the significant negative contribution of relative positive charge (charge of the most positive atom divided by the total positive charge). Compound **9**, with a trifluoromethyl group at the R position, has a higher value of JursRPCG than compound **6**, with a methyl substituent at R position, and thus shows lower A_3 binding affinity than the latter. The negative coefficient of non-common overlap steric volume (NCOSV) indicates that the non-common overlap steric volume is also detrimental to the binding affinity. This means the presence of substituents larger than those present in the shape-reference compound lowers the binding affinity. Compound **26** (R = methyl), which has a

Table 3 List of values of selected descriptors used in MSA for compounds **1–30**

Sl No.	Fo	NCOSV	Energy	LUMO	JursSASA	JursPPSA_3	JursWPSA_1	JursRPCG	JursRNCG	JursRPCS	PMI_mag
1	0.033	182.71	−53.549	1.505	423.954	35.896	105.981	0.197	0.198	1.383	534.102
2	0.042	237.779	−177.999	1.413	524.475	51.08	184.654	0.199	0.155	1.154	1021.984
3	0.068	192.248	−57.446	1.38	457.421	34.771	108.015	0.19	0.191	0.962	692.608
4	0.039	195.316	−61.434	1.202	447.564	36.906	103.722	0.174	0.194	1.091	846.84
5	0.042	263.605	−39.393	1.165	566.58	42.585	166.284	0.147	0.173	0.498	1621.989
6	0.038	206.822	−36.457	1.563	472.121	43.373	135.417	0.156	0.187	1.096	825.027
7	0.065	201.305	−45.195	1.446	471.993	43.246	135.84	0.159	0.19	1.034	690.074
8	0.017	211.599	−24.347	1.613	457.761	42.515	133.31	0.155	0.182	1.011	579.405
9	0.023	225.487	134.735	1.108	499.317	37.127	111.519	0.279	0.139	0	1284.913
10	0.053	220.458	14.814	1.552	501.713	48.916	161.302	0.143	0.166	0.485	955.514
11	0.055	235.619	10.963	1.576	529.436	54.547	187.821	0.126	0.153	0.426	1179.29
12	0.025	243.097	34.266	1.549	529.125	51.317	182.951	0.135	0.152	0.13	1055.939
13	0.044	222.122	−40.744	1.381	499.834	41.362	134.362	0.152	0.181	0.66	1061.992
14	0.054	251.907	24.452	1.515	540.935	52.234	197.724	0.13	0.145	0.095	1197.389
15	0.018	268.847	−159.124	1.459	563.727	57.85	217.523	0.17	0.132	0.905	1432.314
16	0.064	252.997	30.155	1.538	556.155	46.602	182.551	0.141	0.172	0.956	1511.742
17	0.061	269.731	−12.142	1.558	584.469	48.781	210.486	0.135	0.165	0.489	1685.893
18	0.084	278.744	5.882	1.489	600.245	52.148	231.121	0.123	0.15	0.505	1668.26
19	0.033	302.116	0.475	1.513	631.812	54.851	250.752	0.117	0.141	0.366	2262.151
20	0.089	300.777	15.865	1.444	645.571	60.187	272.659	0.106	0.129	0.385	1876.351
21	0.051	340.963	13.489	1.559	693.942	55.841	298.664	0.116	0.146	0.787	2275.538
22	0.03	364.042	14.323	1.506	677.229	53.017	286.076	0.113	0.138	0.435	2109.69
23	0.087	230.175	85.142	1.502	527.238	47.808	163.38	0.142	0.161	1.131	1384.407
24	0.038	267.265	−30.613	1.479	567.508	45.375	188.076	0.134	0.165	0.486	1769.541
25	0.578	110.356	46.436	0.864	545.182	44.007	164.811	0.153	0.174	1.185	1537.805
26	0.622	69.85	−49.444	1.313	420.523	37.791	99.924	0.188	0.192	1.632	532.675
27	0.324	173.695	−23.696	1.283	542.077	41.434	168.593	0.163	0.17	1.258	1073.924
28	0.926	15.564	−29.936	1.365	468.231	45.838	129.893	0.15	0.179	1.343	825.21
29	0.277	204.26	−7.855	1.306	590.836	50.058	206.387	0.133	0.158	0.903	1596.654
30	0.618	87.026	21.683	1.364	497.099	51.961	157.295	0.137	0.162	0.927	994.467

Table 4 Results of randomization test applied on model development process

Equation No.	1	2	3	4
3D QSAR method	MSA	MSA	MSA	MFA
Modeling technique	Stepwise regression	GFA	GFA	G/PLS
R from non-random model	0.878	0.901	0.907	0.990
No. of random trials	99	9[a]	9[a]	9[a]
No. of random *R*'s less than non-random *R*	98	9	9	8
No. of random *R*'s more than non-random *R*	1	0	0	1
Confidence level	98%	90%	90%	80%
Mean value of *R* from random trials ± SD	0.445 ± 0.127	0.568 ± 0.119	0.568 ± 0.119	0.961 ± 0.024

[a] In case of each trial, 50,000 crossovers were performed

Table 5 Results of randomization test applied on the developed models

Equation No.	1	2	3	4
3D QSAR method	MSA	MSA	MSA	MFA
Modeling technique	Stepwise regression	GFA	GFA	G/PLS
R from non-random model	0.878	0.901	0.907	0.990
No. of random trials	99	99	99	99
No. of random *R*'s less than non-random *R*	99	99	99	99
No. of random *R*'s more than non-random *R*	0	0	0	0
Confidence level	99%	99%	99%	99%
Mean value of *R* from random trials ± SD	0.401 ± 0.123	0.442 ± 0.099	0.448 ± 0.117	0.101 ± 0.223

comparatively smaller value of NCOSV, shows a binding affinity close to that of the shape-reference compound **28**, while compound **22** [R = 2-(diphenyl)ethyl], which has a higher value of NCOSV, has less A_3 binding affinity. In a recent paper [30] it was shown that larger and more hydrophobic substituents than methyl or ethyl at the R

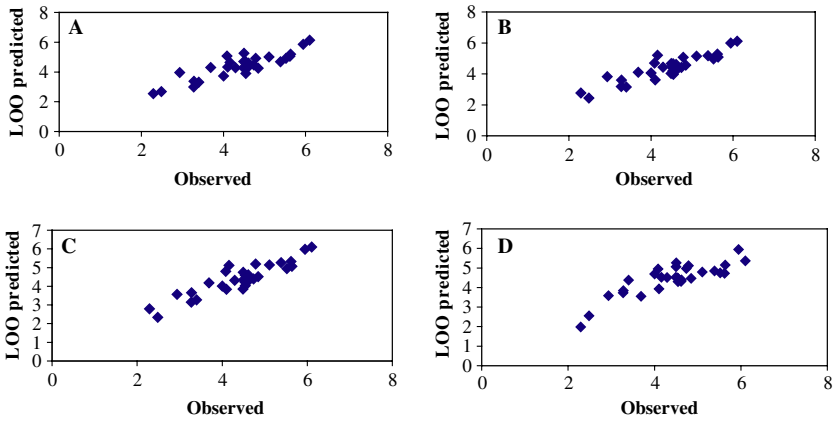


Fig. 2 Scatter plots of observed versus leave-one-out predicted binding affinity values according to **a** Eq. 1, **b** Eq. 2, **c** Eq. 3, and **d** Eq. 4

Table 6 Results of leave-20%-out cross-validation on the developed models

Equation No.	1	2	3	4
3D QSAR method	MSA	MSA	MSA	MFA
Modeling technique	Stepwise regression	GFA	GFA	G/PLS
No. of trials	15	15	15	15
R^2 between observed and predicted values	0.570	0.806	0.796	0.940

position decrease the binding affinity, which was shown further from the negative coefficient of log *P*. This is further corroborated by the docking study of Jung et al. [21], which shows that there is a serine residue (S170) in close proximity to the *R* group, suggesting preference for a relatively hydrophilic group. The positive coefficient of LUMO indicates that the electrophilicity of the molecules favors the binding affinity. Compound **25** (*R* = 2-thiophene), which has a lower value of LUMO, shows less A_3 binding affinity than compound **26** (*R* = thiophene-2-CH_2−), which has a higher value. The energies of the selected conformations also favor the binding affinity. Compounds **2** and **15** (both having *tert*-butyloxy substituent at *R* position) having highly negative values of Energy show low A_3 binding affinity.

The following two equations were among those obtained from the GFA (50,000 crossovers and other default settings):

$$\begin{aligned} pC = {} & 0.140(\pm 0.052)\text{JursPPSA\_3} \\ & + 61.920(\pm 20.616)\text{JursRNCG} \\ & - 1.024(\pm 0.766)\text{JursRPCS} \\ & - 0.006(\pm 0.004)\text{JursSASA} \\ & + 0.006(\pm 0.004)\text{Energy} + 1.645(\pm 1.022)Fo - 8.735 \\ & n = 30, R_a^2 = 0.764, R^2 = 0.812, R = 0.901, \\ & F = 16.6(d.f.\,6,23), \text{LOF} = 0.458, s = 0.215, \\ & Q^2 = 0.705, \text{SDEP} = 0.509, S_{\text{PRESS}} = 0.582, \\ & \text{PRESS} = 7.8, bsr^2(\pm\text{SD}) = 0.813(\pm 0.007) \end{aligned} \quad (2)$$

$$\begin{aligned} pC = {} & 0.130(\pm 0.048)\text{JursPPSA\_3} \\ & + 61.291(\pm 20.009)\text{JursRNCG} \\ & - 1.056(\pm 0.749)\text{JursRPCS} + 0.007(\pm 0.004)\text{Energy} \\ & + 1.701(\pm 0.983)Fo - 0.001(\pm 0.000)\text{PMI\_mag} - 10.180 \\ & n = 30, R_a^2 = 0.777, R^2 = 0.823, R = 0.907, \\ & F = 17.8(d.f.\,6,23), \text{LOF} = 0.432, s = 0.203, \\ & Q^2 = 0.700, \text{SDEP} = 0.513, S_{\text{PRESS}} = 0.586, \\ & \text{PRESS} = 7.9, bsr^2(\pm\text{SD}) = 0.824(\pm 0.006) \end{aligned} \quad (3)$$

Equations 2 and 3 are close in statistical quality and superior to Eq. 1. The theoretical *F* value at a probability level of 0.01 (d.f. 6, 20) being 3.9, variance ratios of Eqs. 2 and 3 are significant at the 99% level, indicating stability of the regression coefficients. The model-development process was subjected to a randomization test (Table 4) with nine trials in each of which 50,000 crossovers were performed. The mean value of random *R*'s was found to be 0.568 and in all cases the random *R*'s were less than those from the non-random models. Again, the models developed (Eqs. 2 and 3) were subjected to a randomization test with 99 trials in each case (Table 5) and mean values of random *R*'s were found to be 0.442 and 0.448, respectively, for Eqs. 2 and 3. In all cases the values of the random *R*'s were lower than those of the non-random models. The calculated values of binding affinity according to Eq. 2 are given in Table 1. Figure 2b, c shows scatter plots of observed versus leave-one-out predicted binding-affinity values for Eqs. 2 and 3, respectively. The models were also

subjected to a leave-20%-out cross-validation test with 15 trials (Table 6) and the R^2 values between the observed and predicted values were found to be 0.806 and 0.796 for Eqs. 2 and 3, respectively.

The positive coefficient of JursPPSA_3 indicates that the atomic-charge weighted positive surface area (sum of the product of the solvent-accessible surface area times the partial charge for all positively charged atoms) contributes significantly to the binding affinity. Again, the relative negative charge is conducive to the binding affinity, as shown by the positive coefficient of JursRNCG (charge of the most negative atom divided by the total negative charge). Compounds **19–22** having lower values of JursRNCG show less A_3 binding affinity. Total molecular solvent-accessible surface area (JursSASA) contributes negatively to the binding affinity. Compounds **19–22** [with substituents like *p*-methoxybenzyl, *p*-methoxyphenylethyl, diphenylmethyl, and 2-(diphenyl)ethyl] have higher values of JursSASA and thus have less binding affinity. Again, it is to be noted that JursPPSA_3 has a positive coefficient in the presence of variables JursRNCG and JursSASA. Among compounds **19–22** (which have lower values of JursRNCG and higher values of JursSASA), compound **20** has the highest value of JursPPSA_3 and has a higher binding affinity than the remaining three. The relative positive charge surface area (JursRPCS) is detrimental to the activity. This parameter is obtained by dividing the solvent-accessible surface area of the most positive atom by JursRPCG. Compound **24** (R = thiophene-2-CH_2) has a smaller value of JursRPCS than **25** (R = 2-thiophene) and thus the former shows higher binding affinity. The positive coefficient of Energy indicates that the conformational energy of the molecules is conducive to the binding affinity. Compounds **2** and **15** (both with a *tert*-butyloxy substituent at the R position) with highly negative values of Energy show low A_3 binding affinity. Common overlap volume ratio (obtained by dividing the common overlap steric volume by the volume of the individual molecule) also contributes significantly to the binding affinity. As *Fo* shows positive coefficients in Eqs. 2 and 3, compounds having higher values of Fo (e.g., compounds **26** and **30**) show higher binding affinity. Higher values of the principal moment of inertia are detrimental to the binding affinity, as shown by the negative coefficient of PMI_mag in Eq. 3. Compounds **19** and **21** have the highest values of PMI_mag and show low binding affinity.

Molecular field analysis

The generated field was of the rectangular type. The probes used were H^+ and CH_3. The charge method used was Gasteiger and the energy cutoff was kept at −30 to +30 kcal. The QSAR equation was generated using the G/PLS method. The number of iterations was set to 50,000 to obtain the final equation. The mutation probabilities were set to the system defaults. The final result was obtained with the number of components at four. A view of aligned molecules studied in the field is shown in Fig. 3. The following equation was obtained from the MFA:

$$\begin{aligned} pC = & -0.050\mathrm{H}^+/283 - 0.028\mathrm{H}^+/366 - 0.031\mathrm{H}^+/427 \\ & - 0.019\mathrm{H}^+/448 - 0.054\mathrm{H}^+/466 - 0.052\mathrm{H}^+/534 \\ & - 0.030\mathrm{H}^+/608 - 0.021\mathrm{H}^+/618 - 0.020\mathrm{H}^+/687 \\ & + 0.033\mathrm{H}^+/757 - 0.010C\mathrm{H}_3/417 - 0.052C\mathrm{H}_3/474 \\ & + 0.025C\mathrm{H}_3/529 - 0.022C\mathrm{H}_3/756 + 5.590 \\ & n = 30, R_a^2 = 0.961, R^2 = 0.980, R = 0.990, \\ & \mathrm{LSE} = 0.017, Q^2 = 0.716, \\ & SDEP = 0.500, S_{\mathrm{PRESS}} = 0.707, \\ & \mathrm{PRESS} = 7.5, bsr^2(\pm\mathrm{SD}) = 0.905(\pm 0.233) \end{aligned} \quad (4)$$

In Eq. 4, H^+ /283, H^+ /366..., and so on are the probes and their numbering (corresponding to spatial positions as shown in Fig. 3); i.e., these represent interactions at points 283 by H^+, 366 by H^+, etc. Equation 4 is of excellent statistical quality. It shows 96.1% explained variance while leave-one-out cross-validation R^2 is found to be 71.6%. The model-development process was subjected to a randomization test (Table 4) with nine trials in each of which 50,000 crossovers were performed. The mean value of random R's was found to be 0.961 and in eight out of nine cases random R's were less than those from the non-random model. However, the difference between the R-value of the deterministic model and mean of those of the random models is small. This shows the impressive ability of flexible modeling techniques such as G/PLS to model even noise (i.e., permuted responses) and warns one to use commercial modeling packages with sufficient validation strategies. Again, the model developed (Eq. 4) was subjected to a randomization test with 99 trials (Table 5) and mean values of random R's were found to be 0.101. In all cases, the values of random R's were less than those of the non-random model. The calculated values of binding affinity according to Equation (4) are given in Table 1.

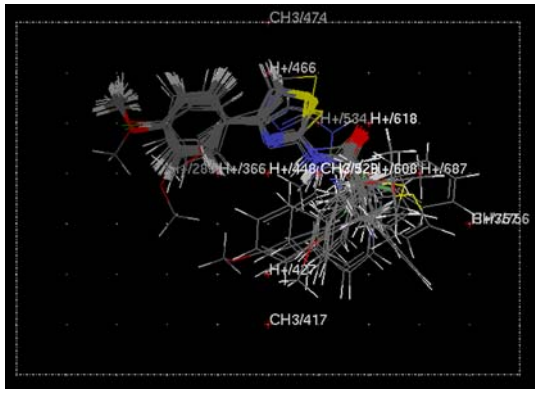


Fig. 3 View of aligned study compounds in the field (MFA)

Figure 2d shows scatter plots of observed versus leave-one-out predicted binding-affinity values for Eq. 4. The model was also subjected to a leave-20%-out cross-validation test with 15 trials (Table 6) and the R^2 values between the observed and predicted values were found to be 0.940.

Conclusions

The present 3D-QSAR analysis explores the spatial, shape and charge requirements for the binding affinity of thiazole and thiadiazole derivatives for the adenosine receptor A_3 receptor. The MSA-derived equations show the importance of Jurs descriptors (atomic charge weighted positive surface area, relative negative charge and relative positive charge surface area), partial moment of inertia, energy of the most stable conformer, and the ratio of common overlap steric volume to the volume of individual molecules. The MFA-derived equation shows interaction energies at different grid points. In summary, this analysis shows the importance of charges and surface area for binding with the adenosine A_3 receptor. Statistically reliable 3D-QSAR models obtained from this study suggest that these techniques could be useful to design potent A_3 receptor antagonists.

Acknowledgements One of the authors (JTL) thanks the AICTE, New Delhi for a QIP fellowship. KR thanks the AICTE, New Delhi for a financial grant under the Career Award for Young Teachers scheme. The authors are thankful for the valuable suggestions of the Referee, which helped the authors to improve the manuscript substantially.

References

1. Palmer TM, Ferguson G, Watterson KR (2003) Drug Dev Res 58:302–314
2. Song Y, Coupar IM, Iskander MN (2001) Quant Struct Act Relat 20:23–30
3. Nakazawa T, Koshiba M, Kosaka H, Tsuji G, Nakamachi Y, Saura R, Kurosaka M, Tanaka Y, Kumagai S (2003) Drug Dev Res 58:368–376
4. Inbe H, Watanabe S, Miyawaki M, Tanabe E, Encinas JA (2004) J Biol Chem 279:19790–19799
5. Shneyvays V, Safran N, Halili-Rutman I, Shainberg A (2000) Drug Dev Res 50:324–337
6. Lee HT, Gallos G, Nasr SH, Emala CW (2004) J Am Soc Nephrol 15:102–111
7. Gutierrez-de-Teran H, Centeno NB, Pastor M, Sanz F (2004) Proteins 54:705–715
8. Merighi S, Mirandola P, Milani D, Varani K, Gessi S, Klotz KN, Leung E, Baraldi PG, Borea PA (2003) Drug Dev Res 58:377–385
9. Chen JF, Schwarzschild MA (2003) Drug Dev Res 58:354–367
10. Phillis JW (2001) Drug Dev Res 52:331–336
11. Feoktistov I, Goldstein A, Sheller JR, Schwartz LB, Biaggioni I (2003) Drug Dev Res 58:461–471
12. Baraldi PG, Tabrizi MA, Fruttarolo F, Bovero A, Avitabile B, Preti D, Romagnoli R, Merighi S, Gessi S, Varani K, Borea PA (2003) Drug Dev Res 58:315–329
13. de Mendonca A, Ribeiro JA (2001) Drug Dev Res 52:283–290
14. Doytchinova I (2001) J Comput Aided Mol Des 15:29–39
15. Baraldi PG, Borea PA, Bergonzoni M, Cacclari B, Ongini E, Recanatini M, Spalluto G (1999) Drug Dev Res 46:126–133
16. Doytchinova I, Valkova I, Natcheva R (2001) Quant Struct Act Relat 20:124–129
17. El-Taher S, El-Sawy KM, Hilal R (2002) J Chem Inf Comput Sci 42:386–392
18. Roy K (2003) Indian J Chem 42B:1485–1496
19. Roy K (2003) QSAR Comb Sci 22:614–621
20. Roy K, Leonard JT, Sengupta C (2004) Bioorg Med Chem Lett 14:3705–3709
21. Jung K-Y, Kim S-K, Gao Z-G, Gross AS, Melman N, Jacobson KA, Kim Y-C (2004) Bioorg Med Chem 12:613–623
22. Cerius^2 Version 4.8 (2002) Accelrys Inc, San Diego, USA; http://www.accelrys.com/cerius2
23. Hopfinger AJ, Tokarsi JS (1997) In: Charifson PS (ed) Practical applications of computer-aided drug design. Marcel Dekker Inc., New York, pp 105–163
24. Rogers D, Hopfinger AJ (1994) J Chem Inf Comput Sci 34:854–866
25. Fan Y, Shi LM, Kohn KW, Pommier Y, Weinstein JN (2001) J Med Chem 44:3254–3263
26. Wold S (1995) In: van de Waterbeemd H (ed) Chemometric methods in molecular design. VCH, Weinheim, pp 195–218
27. Snedecor GW, Cochran WG (1967) Statistical methods. Oxford& IBH Publishing Co. Pvt. Ltd, New Delhi, pp 381–418
28. Wold S, Eriksson L (1995) In: van de Waterbeemd H (ed) Chemometric methods in molecular design. VCH, Weinheim, pp 312–317
29. Debnath AK (2001) Combinatorial library design and evaluation. Marcel Dekker Inc., New York, pp 73–129
30. Bhattacharya P, Leonard JT, Roy K (2005) Bioorg Med Chem 13:1159–1165. Doi http://dx.doi.org/10.1016/j.bmc.2004.11.022