Bioorganic & Medicinal Chemistry 19 (2011) 6853–6859

Contents lists available at SciVerse ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

Affinity prediction on A_3 adenosine receptor antagonists: The chemometric approach

Feng Luan [a], André Melo [a], Fernanda Borges [b], M. Natália D.S. Cordeiro [a,*]

[a] REQUIMTE, Department of Chemistry and Biochemistry, Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal
[b] CIQ, Department of Chemistry and Biochemistry, Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal

ARTICLE INFO

Article history:
Received 6 June 2011
Revised 17 September 2011
Accepted 19 September 2011
Available online 22 September 2011

Keywords:
Human A_3 adenosine receptor
Quantitative Structure–Activity Relationships
Multiple linear regression
Antagonists

ABSTRACT

Potent and selective ligands with a nucleoside skeleton are generally thought as agonists of the human A_3 adenosine receptor (AR), however, some of them can also act as full antagonists. This work reports a Quantitative Structure–Activity Relationship (QSAR) study for predicting the binding affinity of such type of compounds towards the A_3 AR. Several different theoretical molecular descriptors, calculated only on the basis of knowledge of the molecular structure and an efficient variable selection procedure, such as forward stepwise regression, led to models with satisfactory accuracy and predictive ability. But the best-final QSAR model is based on the Molecule Representation of Structures based on Electron diffraction (3D-MoRSE) descriptors capturing a reasonable interpretation. This QSAR model is able to explain more than 85% of the variance in the experimental affinity and manifests good predictive ability as indicated by the higher Q^2s of cross- and external-validations. The model obtained in this study may provide guidance for future design of new potent and selective human A_3 AR full antagonists with a nucleoside skeleton.

1. Introduction

During the last decades, several pharmacologists, organic chemists as well as other researchers focused their work on the adenosine receptors (AR) and in particular, on the human A_3 adenosine receptor by using an interdisciplinary approach to speed the discovery and structural refinement of new potent and selective agonists and antagonists.[1–3] The A_3 adenosine receptor, the most recently identified AR subtype, is thought to be a good therapeutic target for the development of clinically efficacious drug candidates.[4] A_3 AR is expressed in relatively high densities in some organics including lung, liver, neutrophils, macrophages, and glial cells. In the heart and brain, there are distinct effects on cell survival, but the A_3 AR is expressed at a much lower level. Its presence on neurons in the central nervous system has been controversial for a while, but now is well established.[5,6]

Due to its duality, the A_3 adenosine receptor can even play its role either through agonists or by antagonists.[7] In pharmacology, the term agonist–antagonist is used to refer to a drug which exhibits some properties of an agonist (a substance that fully activates the neuronal receptor that it attaches to) and some properties of an antagonist (a substance that attaches to a receptor but does not activate it or if it displaces an agonist at that receptor it seemingly deactivates it thereby reversing the effect of the agonist). Up to now, researches are performed continually on both of agonist and antagonists in the field of A_3 AR study. A_3 AR agonists have been shown to be involved in inflammation,[8] cancer[9] and cardioprotection.[10,11] The pharmaceutical development of A_3 AR antagonists has led to preclinical candidates for the treatment of glaucoma[12] and cancer.[13] It is worth noting that both A_3 AR agonists and antagonists are being explored for cancer treatment. From past to present, worldwide efforts have been made to discover potent and selective human A_3 AR agonists and antagonists on the basis of the structure of adenosine.[14,15]

Adenosine receptor antagonists, including A_3 selective compounds, have been extensively reviewed in the literature.[16–18] As some authors concluded that all known A_3 receptor antagonists can be subdivided into two major groups from a chemical point of view: purines and structurally related compounds, and nonpurine compounds. Within the 'nonpurine derivatives' class, a variety of different heterocyclic nuclei have been identified as potential A_3 adenosine antagonists that can be classified into six major chemical classes including flavonoids, 1,4-dihydropyridines and pyridines, triazoloquinazolines, isoquinoline and quinazolines, pyrazolotriazolopyrimidines, and several other minor classes.

The same as the antagonists, a number of pharmacologically important agonists displaying A_3 AR selectivity have also been synthesized. For example, IB-MECA is currently in Phase II trials for the treatment of colorectal cancer and the structurally similar Cl-IB-MECA is planned to be used in conjunction with chemotherapy.[19] MRS3558 is in preclinical development for the treatment of arthritis, whereas CP-532, 903 recently demonstrated protection against myocardial ischemia/reperfusion injury via

* Corresponding author. Tel.: +351 220402503; fax: +351 22402659.
E-mail address: ncordeir@fc.up.pt (M. N.D.S. Cordeiro).

0968-0896/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2011.09.032

sarcolemmal KATP channels, as well as anti-inflammatory effects.[19–21] Looking at the structure of these compounds, one can see that they exhibit a number of structural similarities. That is to say, they are all nucleosides possessing a 5′-*N*-methylcarboxamido group, a *N*′-benzyl group and a *meta*-substituted halogen on the latter benzyl group. And all of these modifications typically confer A_3AR selectivity.

As mentioned above, agonist ligands for the A_3 AR are almost exclusively nucleoside derivatives. Indeed nowadays, Structure–Activity Relationships (SARs) of adenine nucleoside derivatives as agonists over A_3 AR binding are well-established. The positions on the structure of adenosine suitable for derivatisation have been extensively modified. Thus, nucleoside derivatives that are highly selective A_3 AR agonists have been accepted widely. However, some bias exists when taking as agonists the potent and selective ligands with a nucleoside skeleton. In fact, in the process of studying the SARs of adenosine derivatives at the A_3 AR, it has been discovered that this kind of compounds can act also as partial agonists, and antagonists.[22–24]

Some researchers recently stated that an advantage of nucleoside-based A_3AR antagonists over other heterocyclic antagonists is their ability to achieve high affinity at murine species. For instance, nucleoside-based A_3 AR antagonists maintaining an intact ribose moiety were reported by Volpini et al.[25] with a series of 8-alkynyladenosine derivatives that exhibited A_3AR selectivity, but suffered from weak A_3 AR affinity. A spirolactam derivative, in which the 5′-alkyluronamide group was cyclized onto the 4′ carbon, was found to potently and selectively antagonize the A_3 AR.[12]

On the basis of the former work, with the aim of searching for agonists,[26–28] 'Click chemistry' was explored by Cosyn et al. to synthesize two series of 2-(1,2,3-triazolyl) adenosine derivatives. The biological results showed that N^6-substituted 2-(1,2,3-triazolyl)-adenosine analogues constitute a novel class of highly potent and selective nucleoside-based A_3AR antagonists, partial agonists, and agonists.[29] On the other hand, Jeong et al.[30] searched for novel nucleoside analogues and found that, among which on two compounds, appending an additional methyl group to the 5′-uronamide nitrogen converted them from A_3AR full agonists into potent and selective A_3 AR full antagonists. Also, the same authors synthesized novel D- and L-4′-thioadenosine derivatives lacking the 4′-hydroxymethyl moiety starting from D-mannose and D-gulonic γ-lactone, respectively, and found those to be potent and selective species-independent A_3 AR antagonists.[31] Therefore, one can judge that it is very interesting to probe the relationships between the molecular structure and binding affinity of this special kind of compounds as A_3 AR antagonists.

In the present study, we have examined the use of regression models, along with a feature selection algorithm, derived from a variety of molecular representations. For this training set, the 3D-MoRSE descriptors provided the best model and exhibited good quality and predictive power, as judged by extensive internal and external validation. Our final model might serve as a useful tool for the discovery and structural refinement of new potent and selective A_3 receptor full antagonists. To the best of our knowledge, no systematic theoretical study seems to have yet been reported for the molecular or electronic properties of this set of compounds.

2. Materials and methods

2.1. Data

The binding affinities for a set of 37 potent and selective compounds were taken from the literature.[29–31] This dataset comprises compounds that are all full antagonists with a nucleoside skeleton, even though this type of compounds is usually considered as agonists. The binding affinities for A_3 AR of such compounds, expressed as K_i (nM), have been determined by using the agonist radioligands [^{125}I] I-AB-MECA, and were log-transformed (pK_i) to be used in the following QSAR modelling. All these pK_i values are within the range from 5.82 to 8.89. Moreover, the data set was split into a training set (29 compounds) and a randomly chosen external test set (8 compounds). The training set has pK_i values in the range of 5.82–8.89, and was used to build the models, adjust the parameters and take into account the models' robustness. Regarding the external test set, it has pK_i values in the range of 6.06–8.43, and was used to evaluate the predictive ability of the models built. The values of both data sets averagely fall into the scope of the data set. All these compounds, including their structures and the experimental measured pK_i values as well as the calculated pK_i values with the final QSAR model, along with the meaning of each of the 3D-MoRSE descriptor included in such model, are shown in Table 1.

2.2. Chemometric approach

Firstly, the structures of all compounds were drawn with the aid of ChemBioOffice[32] and pre-optimized using the molecular mechanics force field method (MM+) available in the ChemBio 3D Ultra.[32] Then, their geometry was fully optimized without any symmetry restrictions, using the semi-empirical PM3 method implemented in the MOPAC package.[33] A gradient cutoff of 0.01 kcal/mol was used for all geometry optimizations.

Subsequently, the final optimized structures were brought into the E-Dragon software package[34] for computing different families of descriptors,[35] grouped as follows: constitutional descriptors (48), walk and path counts (47), information indices (47), edge adjacency indices (107), Gálvez charge topological indices (21), Randić molecular profiles (41), RDF descriptors (150), WHIM descriptors (99), functional group counts (154), topological descriptors (119), connectivity indices (33), 2D autocorrelations (96), Burden eigenvalues (64), eigenvalue based indices (44), geometrical descriptors (74), 3D-MoRSE descriptors (160), GETAWAY descriptors (197), atom-centered descriptors (120), and molecular properties (31). The total number of descriptors was 1652. Descriptors with constant values inside each family of descriptors were discarded.

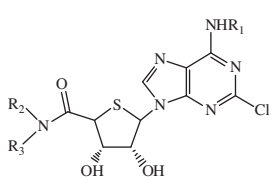
Forward stepwise multilinear regression (FS-MLR)[36,37] was applied to find the mathematical models that best describe the desired activity, that is to say, the pK_i values of the chosen training set compounds. All these calculations were carried out with the STATISTICA software package.[38]

Several diagnostic statistical tools were used for evaluating our model equations, in terms of the criteria *goodness-of-fit* and *goodness-of-prediction*. Measures of *goodness-of-fit* were estimated by standard statistics, namely the determination coefficient, R^2; the root mean square, *RMS*; the Fisher's statistics, *F*; as well as the ratio between number of compounds and the number of adjustable parameters, known as the ρ statistics. *Goodness-of-prediction* of the final models was assessed by means of internal cross-validation (CV), basically by leave-one-out (Q^2_{LOO}) and fivefold-out (Q^2_{FFO}), as well as by external validation (verified by Q^2_{Ext}). Good overall quality of the models is indicated by large *F* and ρ values as well as small *RMS* values (model significance), along with R^2 (goodness of fit) and Q^2 (predictivity) values close to one.

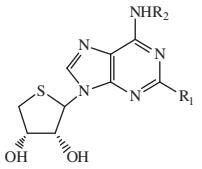
3. Results and discussion

The task of building the best QSAR model for a particular endpoint is a search for the best subset of descriptors that characterize the endpoint and the selection of the type and optimal parameters of the model. Of the nineteen families of different descriptors available from DRAGON, the best model attained from the FS-MLR

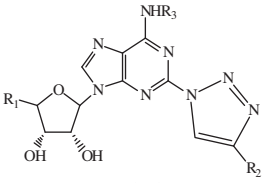
Table 1
Experimental and predicted values of pK_i for the training (T) and external (E) compounds of the dataset



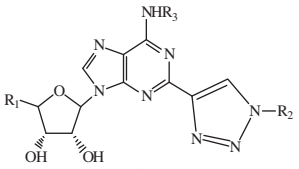
1-19



20-29



30-34



35-37

No.	R_1	R_2	R_3	Type	Descriptors						Experimental pK_i	Calculated pK_i	Residuals
					*Mor*01*p*	*Mor*13*v*	*Mor*26*v*	*Mor*04*m*	*Mor*31*m*	*Mor*09*u*			
1	3-F-Bn	Me	Me	*E*	+596.52	+0.21	+0.31	−3.28	+0.25	−0.37	6.92	6.995	+0.009
2	3-Cl-Bn	Me	Me	*T*	+627.38	+0.46	+0.20	+0.94	+0.01	−0.52	7.67	7.542	+0.128
3	3-Br-Bn	Me	Me	*T*	+644.49	+0.13	+0.19	+0.77	+0.42	−0.23	8.03	7.731	+0.299
4	3-I-Bn	Me	Me	*T*	+689.72	+0.25	+0.30	−3.80	+0.29	−0.45	7.81	6.829	+0.981
5	Me_2	Me	Me	*T*	+415.30	+0.34	+0.19	−2.30	+0.05	+0.19	6.87	7.349	−0.479
6	2-MeO-Et	Me	Me	*E*	+481.79	+0.38	+0.00	−4.38	+0.04	+0.78	6.37	7.047	−0.677
7	Cp	Me	Me	*T*	+444.54	−0.06	+0.21	−2.25	−0.07	−0.41	6.96	7.270	−0.310
8	CpMe	Me	Me	*T*	+498.70	+0.11	+0.04	−3.79	+0.07	+0.11	7.51	7.426	+0.084
9	Cyclobutyl	Me	Me	*T*	+498.70	+0.35	+0.18	−1.22	−0.11	+0.23	6.94	6.845	+0.095
10	Cyclopentyl	Me	Me	*T*	+555.96	+0.45	+0.07	−2.00	−0.28	+0.20	6.08	6.764	−0.684
11	3-I-Bn	Me	Pr	*T*	+827.28	+0.59	+0.41	−2.41	+0.01	−0.67	6.14	5.982	+0.158
12	3-I-Bn	Me	CH_2CH_2OH	*T*	+774.87	+0.35	+0.26	+0.05	−0.04	−0.70	6.90	6.711	+0.189
13	3-I-Bn	Et	Ph	*T*	+997.99	+0.52	+0.77	+1.78	+0.37	−2.09	6.40	6.427	−0.027
14	3-I-Bn	Piperidine		*T*	+911.11	+0.19	+0.21	+0.46	+0.11	−0.39	6.25	6.282	−0.032
15	3-I-Bn	4-Me-piperazine		*T*	+954.78	+0.25	+0.30	+0.88	+0.07	−0.92	6.18	6.328	−0.148
16	3-I-Bn	Azetidine		*T*	+766.49	+0.36	+0.33	−3.64	+0.43	−1.57	7.36	6.861	+0.069
17	3-I-Bn	Pyrrolidine		*E*	+837.25	+0.25	+0.49	−2.06	−0.06	−2.21	6.93	7.270	−0.310
18	3-I-Bn	4-Hydroxypiperidine		*T*	+963.62	+0.21	+0.38	−0.24	+0.21	−0.80	5.82	6.010	−0.190
19	3-I-Bn	Thiomorpholine		*E*	+937.95	+0.32	+0.40	−2.35	−0.06	−2.21	6.06	6.852	−0.792
20	Cl	Me		*E*	+221.71	+0.28	+0.08	−2.10	+0.19	−0.11	8.43	9.006	−0.576
21	Cl	3-F-Bn		*E*	+357.78	+0.25	+0.26	−2.50	+0.17	−1.00	8.13	8.642	−0.512
22	Cl	3-Cl-Bn		*T*	+382.51	+0.79	+0.22	−2.35	+0.25	−0.75	8.78	8.994	−0.214
23	Cl	3-Br-Bn		*T*	+395.80	+0.34	+0.33	+0.61	+0.34	−0.46	8.05	8.344	−0.294
24	Cl	3-I-Bn		*T*	+430.91	+0.42	+0.36	−2.02	+0.45	−0.93	8.38	8.611	−0.231
25	H	Me		*T*	+204.33	+0.28	+0.21	−1.23	−0.22	−0.20	8.32	7.994	+0.326
26	H	3-F-Bn		*T*	+335.52	+0.28	+0.29	−2.30	−0.02	−0.91	8.14	8.209	−0.069
27	H	3-Cl-Bn		*T*	+359.46	+0.42	+0.25	−1.37	−0.04	−0.86	8.82	8.363	+0.457
28	H	3-Br-Bn		*T*	+372.32	+0.29	+0.33	−1.30	−0.05	−0.82	8.17	7.872	+0.298
29	H	3-I-Bn		*T*	+406.31	+0.32	+0.34	−0.54	+0.07	−1.04	8.60	8.241	+0.359
30	CH_2OH	Butyl	Me	*T*	+530.79	+0.46	+0.25	−2.95	+0.29	−0.87	7.93	8.246	−0.316
31	CH_2OH	Bn	Me	*T*	+545.92	+0.30	+0.51	−2.74	+0.29	−1.76	8.02	8.026	−0.006
32	CH_2OH	CyclopentylMe	Me	*T*	+624.46	+0.82	+0.17	−2.58	+0.14	−1.37	8.89	8.717	+0.173
33	CH_2OH	CyclohexylMe	Me	*T*	+688.28	+0.54	+0.17	−2.95	+0.41	−0.27	7.67	7.638	+0.032
34	CH_2OH	CyclopentylMe	2-Cl-5-MeO-Bn	*E*	+1054.35	+0.66	+0.12	−2.97	+0.45	−1.80	7.74	8.184	−0.444
35	CH_2OH	Bn	Me	*E*	+545.92	+0.30	+0.51	−2.74	+0.29	−1.76	7.27	8.026	−0.756
36	CH_2OH	3-Cl-Bn	Me	*T*	+574.28	+0.23	+0.57	−2.51	+0.28	−1.22	7.09	7.104	−0.014
37	CH_2OH	3-MeO-Bn	Me	*T*	+621.71	+0.19	+0.47	−2.40	+0.20	−1.78	7.07	7.701	−0.631

Cp, cyclopropyl; Me, methyl; Et, ethyl; Ph, phenyl; Bn, benzyl.
*Mor*01*p*, 3D-MoRSE—signal 1/weighted by atomic polarizabilities; *Mor*13*v*, 3D-MoRSE—signal 13/weighted by atomic van der Waals volumes; *Mor*26*v*, 3D-MoRSE—signal 26/weighted by atomic van der Waals volumes; *Mor*04*m*, 3D-MoRSE—signal 4/weighted by atomic masses; *Mor*31 m, 3D-MoRSE—signal 31/weighted by atomic masses; *Mor*09*u*, 3D-MoRSE—signal 09/unweighted.

analysis for the endpoint selected was the one based on the 3D-Morse descriptors (see the results summarized in Supplementary data).

As can be judged, the 3D-Morse descriptors-based model is better in terms of both accuracy and predictivity, as manifested by the higher R^2 (model fit) and Q^2_{Ext} (external validation) values, respectively. While this model is able to predict >85% of the experimental variances of external set, the other models explain <65%.

The optimum QSAR equation obtained based on the 3D-Morse descriptors, together with the statistical parameters of the regression, are as follows:

$$pK_i = 9.636 - 0.004 \times Mor01p + 0.715 \times Mor13v + 1.848 \times Mor31m + 0.087 \times Mor04m - 1.010 \times Mor09u - 3.721 \times Mor26v \quad (1)$$

$N = 29$; $R^2 = 0.860$; $RMS = 0.3350$; $F = 22.49$; $p < 10^{-5}$; $\rho = 4.83$, $Q^2_{LOO} = 0.767$; $RMS_{LOO} = 0.4357$; $F_{LOO} = 88.65$.

The large F and small RMS, p values are indicative of the statistical significance of the model. Besides the calculated ρ statistics (=4.83) of the model is higher than the reference value 4,[39] indicating that the model is proper. In addition, the values of the coefficients for the regression (R^2), and for LOO-CV (Q^2_{LOO})—measures of the model predictivity, all show that the models displayed an adequate goodness-of-fit and prediction. The latter property was further checked by fivefold-out cross validation (FFO). In present study, the training set was split into five subsets: A (1,6,11,...), B (2,7,10,...), C (3,8,13,...), D (4,9,14,...), and E(5,10,15,...), and the procedure was run five times. In each run, one of these subsets was kept apart, while the other four were used to construct the MLR models. The statistical results of FFOCV obtained on each of the five training-test set splits are shown in Table 2. As can be seen, the results disclose an average training quality of $R^2 = 0.8707$; $RMS = 0.3828$, and an average predicting quality of $Q^2_{FFO} = 0.8049$; $RMS_{FFO} = 0.4112$ for the model, showing that it has a satisfactory statistical stability and validity.

For a final validation, prediction results were calculated for the model using the external set referred to above. The statistical parameters were $Q^2_{Ext} = 0.8700$; $RMS_{Ext} = 0.5551$; $F_{Ext} = 40.24$, confirming once more the adequate predictive ability of the model. The quality of this model in terms of its fitting and prediction ability can also be judged by inspection of Figure 1, where a plot of the predicted pK_i values versus the experimental values is shown for all compounds, as well as by analysis of the same data in Table 1.

However, further analysis of this regression model should only be pursued after checking the reliability of the pre-adopted assumptions. MLR analysis establishes a linear, additive relation between the molecular descriptors and the underlying biological activity (i.e., the binding affinity for A3 AR), and, in fact, this is the simplest mathematical form that might be envisaged for the model without any a priori information. Nevertheless, by looking at the distribution of the standardized residuals for all cases (Fig. 2a), no specific pattern is seen, thereby reinforcing the idea that the model does exhibit a linear dependence.

The hypothesis of normally distributed residuals can be verified from the results from the Kolmogorov–Smirnov (D) and Shapiro–Wilk (W) tests. Both statistics, $D = 0.130$, $p = 0.200$ and $W = 0.947$, $p = 0.153$, are not significant ($p > 0.05$), thus the hypothesis that the distribution of residuals of the present model is normal should be accepted.

Table 2
Fivefold out cross-validation results for the best QSAR model found

Training subsets	R^2	RMS	Test	Q^2_{FFO}	RMS_{FFO}
A+B+C+D	0.8411	0.4225	E	0.9428	0.2427
A+B+C+E	0.8619	0.4118	D	0.8204	0.3286
A+B+D+E	0.8933	0.3372	C	0.7315	0.5633
A+C+D+E	0.8529	0.4347	B	0.9087	0.2120
B+C+D+E	0.9044	0.3078	A	0.6213	0.7096
Average	0.8707	0.3828		0.8049	0.4112

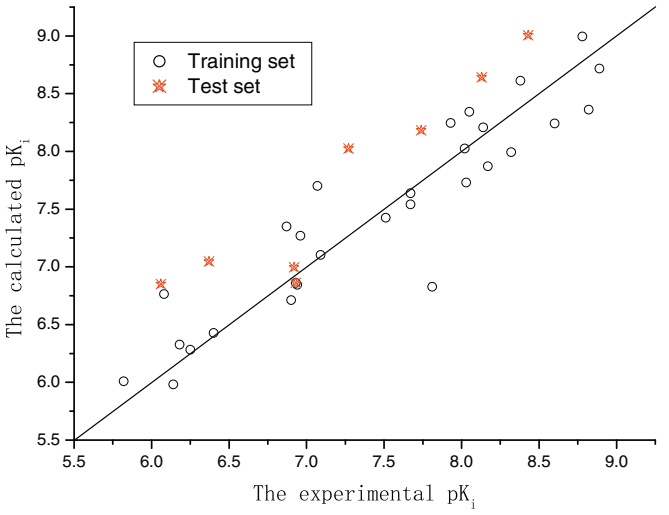


Figure 1. Predicted pK_i values by the MLR model versus experimental values for the whole set of compounds.

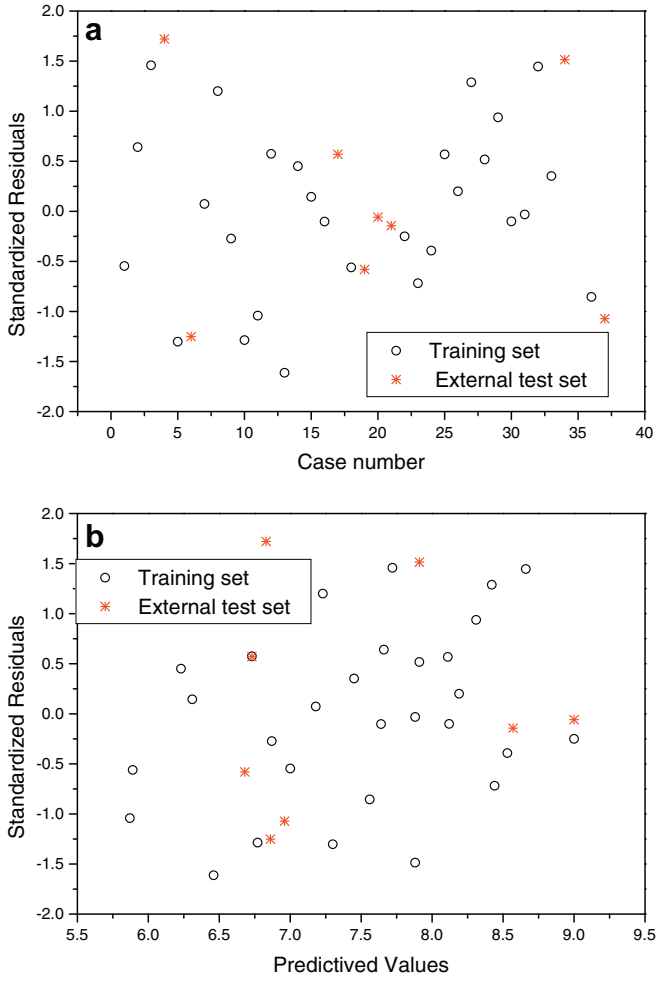


Figure 2. Checking the validity of the MLR assumptions. (a) Linearity of the model: plot of standardized residuals versus cases. (b) Homocedasticity: plot of standardized residuals versus predicted values.

With regard to the hypothesis of homoscedasticity of residuals, Figure 2b shows the plot of standardized residuals versus the predicted pK_i values. As can be seen, the points are scattered throughout and do not seem to cluster in any significant way, thus confirming the homogeneity of variance. This plot also provides a check for the no autocorrelation of the residuals.

Another aspect deserving special attention is the degree of collinearity of the variables of the model, which can readily be diagnosed by analyzing the cross-correlation matrix. Table 3 shows this matrix for the six variables included in the MLR model based on the data set. As can be seen, pair correlations among the variables are always less than 0.80, meaning that the descriptors can be considered independent, which avoid serious overestimation of chance correlation effects.[40]

Finally, one should consider the applicability domain of the model according to the OECD QSAR validation principles.[41] There are some methods to define the applicability domain. The leverage test, a distance method, has been recommended for assessing QSAR applicability.[42–44] To identify chemicals that are outside the applicability domain (AD) of our MLR model, we have used a Williams plot, that is to say, a plot of the standardized residuals (y-axis) versus leverages (x-axis) for each compound of the training set.[45] From this plot, the AD was established inside a squared area within ±3 standard deviations and a leverage threshold, h^* ($h^* = 3m/N$, being m the number of model descriptors and N the number of training compounds used to build the model; $h^* = 0.6207$). In this plot, the horizontal and vertical straight lines represent the limits of normal values of response outliers (Y outliers) and structurally influential chemicals (X outliers), respectively. The resulting Williams plot for the data set including the training and test set are shown in Figure 3.

From this plot, one can see that all the compounds in the training set are in the applicability domain, but one of the compounds of the external test set can be considered a structurally influential chemical (compound **34**). Here, one should emphasize that the predicted data should be considered reliable only for those chemicals that fall within the AD of the model, and should not be extrapolated to other classes of chemical classes, to avoid making uncertain predictions in very different conditions from those relating to the data set used for fitting the present QSAR model. So, predictions related to such compound should be taken with much care.

Interpreting a QSAR model in terms of the specific contribution of substituents and other molecular features to the modelled activity is always a difficult task.[46] The present MLR model is based on the 3D-Morse descriptors (Molecule Representation of Structures based on Electron diffraction) proposed by Gasteiger and co-workers'.[35] These descriptors have the merit to take into account the 3D arrangement of the atoms without ambiguities (in contrast with those coming from chemical graphs), and also they do not depend on the molecular size, thus being applicable to a large number of molecules with great structural variance. This type of indices are based on the idea of obtaining information from the 3D atomic coordinates by the transform used in electron diffraction studies for preparing theoretical scattering curves. A generalized scattering function, called the molecular transform, can be used as the functional dependence for

Table 3
Intercorrelation among the six descriptors used in the model

	*Mor*01*p*	*Mor*13*v*	*Mor*31*m*	*Mor*04*m*	*Mor*09*u*	*Mor*26*v*
*Mor*01*p*	1.000					
*Mor*13*v*	0.003	1.000				
*Mor*31*m*	−0.244	−0.060	1.000			
*Mor*03*m*	−0.340	0.007	0.207	1.000		
*Mor*09*u*	0.069	0.279	0.140	−0.190	1.000	
*Mor*26*v*	−0.047	0.229	−0.146	−0.270	0.759	1.000

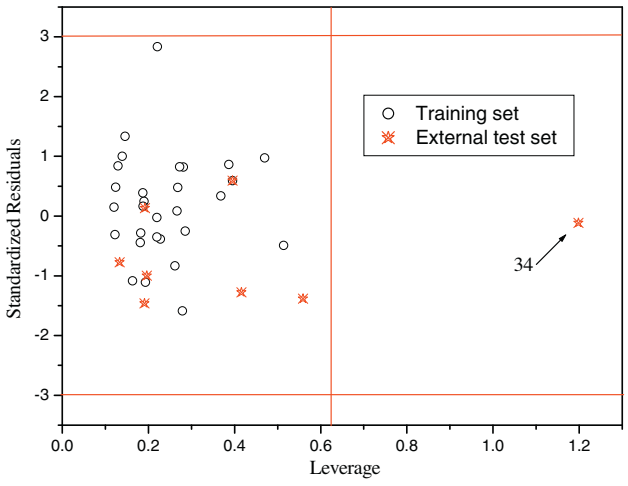


Figure 3. Williams plot, that is, plot of standardized residuals versus leverage values, with a warning leverage of 0.621 and taking into account the whole set of compounds.

deriving, from a known molecular structure, the specific analytic relationship of both X-ray and electron-diffraction. In order to take into account at a prescribed scattering angle, the specific contributions of the atoms to the property being studied, different atomic properties such as the atomic mass, polarizability, electronegativity, etc. can be employed as weighting schemes. For example, in the case of *Mor*05*p*, the scattering angle is 5 $Å^{-1}$ and atomic polarizabilities are employed as the weighting scheme. Therefore, 3D-MoRSE descriptors can reveal the skeleton and substituent information for a molecule.

As seen from the selected descriptors (*Mor*01*p*, *Mor*13*v*, *Mor*31*m*, *Mor*04*m*, *Mor*09*u*, and *Mor*26*v*), atomic polarizabilities (*p*), van der Waals volumes (*v*), and masses (*m*) do have influence on the binding affinity towards the human A_3 adenosine receptor of this set of compounds. Moreover, the larger is the absolute value of a standardized regression coefficient, the greater the weight of the variable in the model, which therefore leads to the following ranking of contributions to pK_i (see Fig. 4): *Mor*01*p* > *Mor*09*u* ~ *Mor*26*v* > *Mor*31*m* > *Mor*13*v* ~ *Mor*04*m*. All these things thought-out, one might focus the analysis on the weighting variables of our model, since those are the most meaningful descriptors. To estimate the significance of such descriptors, comparisons between particular groups of the studied compounds were made (see Tables 1 and 4).

As can be seen in Table 4, the higher polarizability of the R_1 substituent in the first group of 4′-thioadenosine compounds (**1–19**)

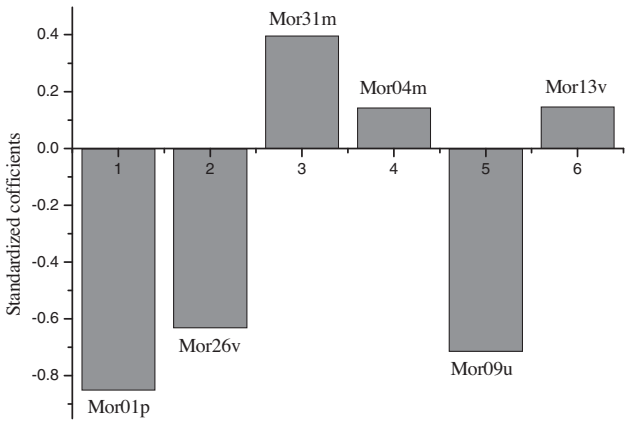


Figure 4. Standardized coefficients versus descriptors in the MLR model.

Table 4
Descriptor values[a] along with the experimental and predicted values of pK_i for a sample of compounds from the dataset

No.	*Mor*01*p*	*Mor*26*v* + *Mor*13*v*	*Mor*31*m* + *Mor*04*m*	Exp pK_i	Pred pK_i	Residuals
1	−2.39	−1.00	+0.18	6.92	7.00	+0.01
2	−2.51	−0.42	+0.10	7.67	7.54	+0.13
3	−2.58	−0.61	**+0.84**	8.03	7.73	+0.30
7	−1.78	−0.82	**−0.33**	6.96	7.27	−0.31
9	−1.99	−0.42	**−0.31**	6.94	6.85	+0.10
13	−3.99	**−2.49**	+0.84	6.40	6.43	−0.03
22	−1.53	**−0.25**	+0.26	8.78	8.99	−0.21
24	−1.58	−0.98	+0.68	8.05	8.34	−0.29
27	−1.44	**−0.63**	−0.19	8.82	8.36	+0.46
28	−1.49	−1.02	−0.21	8.17	7.87	+0.30
31	−2.18	−1.68	+0.30	8.02	8.03	−0.01
32	−2.50	**−0.05**	+0.26	8.89	8.72	+0.17

[a] Important contributions are highlighted in bold.

decreases the binding affinity towards A_3 AR. Nevertheless, that is compensated by the positive contribution of the mass factors (*Mor*31*m* + *Mor*04*m*) giving rise to better binding affinity for the N^6-(3-halobenzyl) series than larger N^6-alkyl series (compounds **1–3** vs **7–9**). Within the N^6-(3-halobenzyl) series, the polarizability factors are overall rather similar, being the rank binding order mainly due to a more favourable mass factor of 3-bromo (compound **3**) in relation to the others halogens (compounds **1** and **2**). Moreover, within the same N^6-(3-halobenzyl) series, substituting instead the methyl R_2 or R_3 substituents by larger alkyls lowers the A_3 AR binding affinity and that can be attributed to more unfavourable polarizability, but especially to van der Waals factors (e.g., compounds **1–3** vs **13**). Regarding the group of 4′-thioadenosine compounds without the 5′-uronamide group (compounds **20–29**), the better binding affinity of the N^6-(3-chlorobenzyl) derivative with respect to the other halo-substituents is due to lesser negative contributions of the van der Waals factors (e.g., compound **22** vs **23** and **27** vs **28**). The same seems to apply for the range of binding affinities of the 5′-OH (1,2,3-triazol)adenosine derivatives (compounds **30–37**), judging from the more favourable contributions of the van der Waals weighted descriptors of the most potent compound with the 4-cyclopentylmethyl substitution at the triazolyl ring (compound **32**) in comparison to other substituents at position 2 (e.g., compound **31**). Furthermore, this is in agreement with the results from a molecular docking study of compound **32** in an A_3 AR model previously performed.[29]

Finally, one should emphasise here that it would be very interesting to apply the presently derived QSAR model forward quantitatively predicting the selectivity of the compounds against other adenosine receptor subtypes. However, due to the lack of such kind of information for most of the studied compounds, along with the fact that even when available, it is sometimes expressed in a different form than the present one (e.g., as % of inhibition and not as K_i values), precluded us to carry out additional predictions; but we hope to be able to report on that shortly.

4. Conclusions

In this study, we modelled the antagonist activity of compounds with a nucleoside skeleton, specifically the binding affinity of 36 compounds for the human A_3 adenosine receptor, determined using the agonist radioligands [^{125}I] I-AB-MECA. For such purpose, we employed multiple linear regression analysis and 3D-MoRSE descriptors.

The results produced by the methodology we proposed were superior to other DRAGON descriptors such as topological, RDF, WHIM, GETAWAY, etc. taking into account the statistical parameters of the model and the external-validation results. In addition, the model provided some useful insights into which descriptors are most related to the antagonist activity of such compounds, indicating that to be mainly explained by atomic polarizability, van der Waals volume, and mass factors.

Hopefully, the QSAR model obtained in this study may provide guidance for future design of new potent and selective human A_3 AR full antagonists with a nucleoside skeleton.

Acknowledgments

The authors acknowledge the Portuguese Fundação para a Ciência e a Tecnologia (FCT) for financial support (project PTDC/QUI/70359/2006 and grant SFRH/BPD/63666/2009).Moreover, this work has been further supported by FCT through grant no. PEst C/EQB/LA0006/2011.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2011.09.032.

References and notes

1. Moro, S.; Deflorian, F.; Spalluto, G.; Pastorin, G.; Cacciari, B.; Kim, S. K.; Jacobson, K. A. *Chem. Commun.* **2003**, *24*, 2949.
2. Moro, S.; Spalluto, G.; Jacobson, K. A. *Trends Pharmacol. Sci.* **2005**, *26*, 44.
3. Michielan, L.; Stephanie, F.; Terfloth, L.; Hristozov, D.; Cacciari, B.; Klotz, K. N.; Spalluto, G.; Gasteiger, J.; Moro, S. *J. Chem. Inf. Model.* **2009**, *49*, 2820.
4. Gessi, S.; Merighi, S.; Varani, K.; Leung, E.; Lennan, S. M.; Borea, P. A. *Pharmacol. Ther.* **2008**, *117*, 123.
5. Lopes, L. V.; Rebola, N.; Pinheiro, P. C.; Richardson, P. J.; Oliveira, C. R.; Cunha, R. A. *Neuroreport* **2003**, *14*, 1645.
6. Yaar, R.; Lamperti, E. D.; Toselli, P. A.; Ravid, K. *FEBS Lett.* **2002**, *532*, 267.
7. Jacobson, K. A.; Gao, Z.-G.; Tosh, D. K.; Sanjayan, G. J.; de Castro, S. In *A3 Adenosine Receptor Agonists: History and Future Perspectives A3 Adenosine Receptors from Cell Biology to Pharmacology and Therapeutics*; Borea, P. A., Ed.; Springer Dordrecht Heidelberg: London, 2009; pp 93–120. Chapter 6.
8. Bar-Yehuda, S.; Silverman, M. H.; Kerns, W. D.; Ochaion, A.; Cohen, S.; Fishman, P. *Expert Opin. Invest. Drug* **2007**, *16*, 1601.
9. Fishman, P.; Jacobson, K. A.; Ochaion, A.; Cohen, S.; Bar-Yehuda, S.; Immun, S. *Immunol. Endocr. Metab. Agents Med. Chem.* **2007**, *7*, 298.
10. Headrick, J. P.; Peart, J. *Vascul. Pharmacol.* **2005**, *42*, 271.
11. Murry, C. E.; Jennings, R. B.; Reimer, K. A. *Circulation* **1986**, *74*, 1124.
12. Yang, H.; Avila, M. Y.; Peterson-Yantorno, K.; Coca-Prados, M.; Stone, R. A.; Jacobson, K. A.; Civan, M. M. *Curr. Eye Res.* **2005**, *30*, 747.
13. Gessi, S.; Merighi, S.; Varani, K.; Leung, E.; Mac Lennan, S.; Borea, P. A. *Pharmacol. Ther.* **2008**, *117*, 123.
14. Jacobson, K. A.; Gao, Z. G. *Nat. Rev. Drug Disc.* **2006**, *5*, 247.
15. Devine, S. M.; Gregg, A.; Figler, H.; McIntosh, K.; Urmaliya, V.; Linden, J.; Pouton, C. W.; White, P. J.; Bottle, S. E.; Scammells, P. J. *Bioorg. Med. Chem.* **2010**, *18*, 3078.
16. Jacobson, K. A.; Klutz, A. M.; Tosh, D. K.; Ivanov, A. A.; Preti, D.; Baraldi, P. G. Medicinal Chemistry of the A3 Adenosine Receptor: Agonists, Antagonists, and Receptor Engineering, Adenosine Receptors in Health and Disease In *Handbook of Experimental Pharmacology*; Fredholm, B. B., Ed.; Springer, 2009; Vol. 193, pp 123–159.
17. Gessi, S.; Merighi, S.; Varani, K.; Leung, E.; Lennan, S. M.; Bore, P. A. *Pharmacol. Ther.* **2008**, *117*, 123.
18. Baraldi, P. G.; Tabrizi, M. A.; Fruttarolo, F.; Bovero, A.; Avitabile, B.; Preti, D.; Romagnoli, R.; Merighi, S.; Gessi, S.; Varani, K.; Borea, P. A. *Drug Dev. Res.* **2003**, *58*, 315.
19. Gao, Z. G.; Jacobson, K. A. *Expert Opin. Emerg. Drug* **2007**, *12*, 479.
20. van der Hoeven, D.; Wan, T. C.; Auchampach, J. A. *Mol. Pharmacol.* **2008**, *74*, 685.
21. Wan, T. C.; Ge, Z. D.; Tampo, A.; Mio, Y.; Bienengraeber, M. W.; Tracey, W. R.; Gross, G. J.; Kwok, W. M.; Auchampach, J. A. J. *Pharmacol. Exp. Ther.* **2008**, *324*, 234.
22. Gao, Z. G.; Kim, S. K.; Biadatti, T.; Chen, W.; Lee, K.; Barak, D.; Kim, S. G.; Johnson, C. R.; Jacobson, K. A. *J. Med. Chem.* **2002**, *45*, 4471.
23. Gao, Z. G.; Blaustein, J.; Gross, A. S.; Melman, N.; Jacobson, K. A. *Biochem. Pharmacol.* **2003**, *65*, 1675.
24. Gao, Z. G.; Mamedova, L.; Chen, P.; Jacobson, K. A. *Biochem. Pharmacol.* **2004**, *68*, 1985.
25. Volpini, R.; Constanzi, S.; Lambertucci, C.; Vittori, S.; Klotz, K. N.; Lorenzen, A.; Cristalli, G. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1931.
26. Elzein, E.; Palle, V.; Wu, ; Y.; Maa, T.; Zeng, D.; Zablocki, J. *J. Med. Chem.* **2004**, *47*, 4766.
27. Volpini, R.; Constanzi, S.; Lambertucci, C.; Taffi, S.; Vittori, S.; Klotz, K. N.; Cristalli, G. *J. Med. Chem.* **2002**, *45*, 3271.
28. Lewis, W. G.; Green, L. G.; Grynszpan, F.; Radić, Z.; Carlier, P. R.; Taylor, P.; Finn, M. G.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2002**, *41*, 1053.

29. Cosyn, L.; Palaniappan, K. K.; Soo-Kyung, K.; Duong, H. T.; Gao, Z. G.; Jacobson, K. A.; Calenbergh, S. V. *J. Med. Chem.* **2006**, *49*, 7373.
30. Jeong, L. S.; Lee, H. W.; Kim, H. O.; Tosh, D. K.; Pal, S.; Choi, W. J.; Gao, Z. G.; Patel, A. R.; Williams, W.; Jacobson, K. A.; Kim, H. D. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1612.
31. Jeong, L. S.; Pal, S.; Choe, S. A.; Choi, W. J.; Jacobson, K. A.; Gao, Z. G.; Klutz, A. M.; Hou, X.; Kim, H. O.; Lee, H. W.; Lee, S. K.; Tosh, D. K.; Moon, H. R. *J. Med. Chem.* **2008**, *51*, 6609.
32. ChemBioOffice 11.0, Cambridge Software, Cambridge, MA, 2008.
33. Stewart, J. P. P. MOPAC 6.0, Quantum Chemistry Program Exchange, QCPE, No. 455, Indiana University, Bloomington, IN, 1989.
34. http://www.vcclab.org/lab/edragon/.
35. Schuur, J. H.; Selzer, P.; Gasteiger, J. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 334.
36. Deconinck, E.; Coomans, D.; Heyden, Y. V. *J. Pharm. Biomed. Anal.* **2007**, *43*, 119.
37. Massart, D. L.; Vandeginste, B. G. M.; Buydens, L. M. C.; de Jong, S.; Lewis, P. J.; Smeyers-Verbeke, J. *Handbook of Chemometrics and Qualimetrics—Part A*; Elsevier Science: Amsterdam, 1997.
38. StatSoft, Inc., STATISTICA (data analysis software system), version 8.0, 2007. www.statsoft.com.
39. Garcia-Domenech, R.; Julian-Ortiz, D. J. V. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 445.
40. Topliss, J. G.; Edwards, R. P. *J. Med. Chem.* **1979**, *22*, 1238.
41. Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure–Activity Relationship Models. OECD, Paris, France, 2004. http://www.oecd.org/document/23/0,2340,en_2649_201185_33957015_1_1_1_1,00.html.
42. Netzeva, T. I.; Worth, A. P.; Aldenberg, T.; Benigni, R.; Cronin, M. T. D.; Gramatica, P.; Jaworska, J. S.; Kahn, S.; Klopman, G.; Marchant, C. A.; Myatt, G.; Nikolova-Jeliazkova, N.; Patlewicz, G. Y.; Perkins, R.; Roberts, D. W.; Schultz, T.; Stanton, W.; van de Sandt, D. T.; Tong, J. J. M.; Veith, W.; Yang, G. C. *ATLA* **2005**, *33*, 155.
43. González-Diaz, H.; Vilar, S.; Santana, L.; Podda, G.; Uriarte, E. *Bioorg. Med. Chem.* **2007**, *15*, 2544.
44. Helguera, A. M.; González, M. P.; Cordeiro, M. N. D. S.; Cabrera, M. A. *Toxicol. Appl. Pharmacol.* **2007**, *221*, 189.
45. Papa, E.; Kovarich, S.; Gramatica, P. *QSAR Comb. Sci.* **2008**, *28*, 790.
46. Abreu, R. M. V.; Ferreira, I. C. F. R.; Queiroz, M. J. R. P. *Eur. J. Med. Chem.* **2009**, *44*, 1952.