Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry 16 (2008) 1658–1675
Bioorganic & Medicinal Chemistry

New QSAR combined strategy for the design of A_1 adenosine receptor agonists

Maykel Pérez González,[a,b,*] Pedro Besada,[b] Maria José González Moa,[b] Marta Teijeira[b] and Carmen Terán[b,*]

[a]*Molecular Simulation and Drug Design Group, Chemical Bioactive Center, Central University of Las Villas, Santa Clara, C.P. 54830 Villa Clara, Cuba*
[b]*Department of Organic Chemistry, Vigo University, C.P. 36310, Vigo, Spain*

Received 14 August 2007; revised 7 November 2007; accepted 8 November 2007
Available online 17 November 2007

Abstract—Combined discriminant and regression analysis was carried out on a series of 167 A_1 adenosine receptor agonists to identify the best linear and nonlinear models for the design of new compounds with a better biological profile. On the basis of the best linear discriminant analysis and both linear and nonlinear Multi Layer Perceptron neural networks regression, we have designed and synthesized 14 carbonucleoside analogues of adenosine. Their biological activities were predicted and experimentally measured to demonstrate the capability of our model to avoid the prediction of false positives. A good agreement was found between the calculated and observed biological activity.

1. Introduction

Adenosine is an endogenous nucleoside that regulates a wide range of physiological effects in the different mammalian organ systems, in particular in the cardiac, nervous, and immune systems.[1] Biochemical and pharmacological studies have demonstrated that most adenosine actions are mediated by extracellular receptors, of which there are four subtypes: A_1, A_{2A}, A_{2B}, and A_3.[2,3]

The A_1 adenosine receptors (A_1 ARs) are the best studied within the AR family. Many physiological responses to adenosine are mediated by activation of this receptor subtype: in the central nervous system (CNS), a neuroprotective effect produced during hypoxia and ischemic conditions; in the heart, negative chronotropic, dromotropic, and inotropic effects; or in the kidney, vasoconstriction, inhibition of renin secretion, diuresis, natriuresis, and other effects.[4]

Agonists of the A_1 AR have therapeutic potential as analgesics, antiepileptics or neuroprotective agents, and have received attention as anti-arrhythmic agents or, more recently, as anti-lipolytic agents. A_1 antagonists are being developed as kidney-protective, diuretics and for the treatment of dementias, depression, or asthma.[5–7]

During the last two decades a large number of A_1 ARs agonists and antagonist have been developed.[7] However, new investigations of structure–activity relationship (SAR) are necessary to find more potent and selective candidates.[8–10]

At present the pharmaceutical industry is under increasing pressure to discover new drug leads in faster and more efficient ways than in the past. In this sense, the rational drug design strategies, especially the computer-aided-drug design approaches, have emerged as a promising solution for the efficient search for new lead compounds.[11–16]

In this connection, the search for new adenosine analogues that selectively activate A_1 ARs in a specific target tissue has always been on the agenda of medicinal chemists. Nevertheless, only a limited number of computational approaches have been developed for this type of

Keywords: QSAR; A_1 adenosine receptor agonists; Discriminant analysis; Neural networks.
* Corresponding authors. Tel.: +53 42 281473; fax: +53 42 281130 (M.P.G.), tel.: +34 986 812276; fax: +34 986 812262 (C.T.); e-mail addresses: mpgonzalez76@yahoo.es; mcteran@uvigo.es

0968-0896/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2007.11.026

receptor with this aim.[17–20] In addition, these works only have a few compounds in their training set, which could be considered as an obstacle in the practical application of these approaches.[21,22]

Recently, novel applications of the Radial Distribution Function (RDF) to the rational design and selection of molecules have been introduced by our research group.[15,23–26] The RDF approach is very useful for the selection of novel subsystem of chemicals having a desired property. These subsystems can be further optimized by using some of the many molecular modeling techniques at the disposition of the medicinal chemists.

In this work we explore several kinds of descriptors to find a prediction system that discriminates active A_1 adenosine receptor agonists from inactive ones. Subsequently, the activity at this receptor subtype is predictable using different regression techniques. In addition, a simulated experiment of virtual screening for the search of novel active compounds is conducted using the models developed here. Finally, the design, synthesis, and biological evaluation of a new series of nucleosides are presented.

2. Results and discussion

Virtual screening of chemical databases is a well established method for finding new hit candidates in the drug discovery process.[27–29] In the present paper, we have developed an integrated approach incorporating the combination of discriminant and regression analysis. The result of the combination of both statistical techniques permits us to classify new adenosine analogues as active or inactive, and subsequently, once the inactive compounds are discarded, to predict the affinity of the active ones.

The majority of the excellent QSAR tools developed by other laboratories do not take into account the important information of the inactive compounds in the design of new analogues. Thus, only the active compounds are reported in the training sets and only these are the focus of sequential analyses.

A graphical representation of our integrated system of prediction is shown in Figure 1. The new classified compounds, once their affinity in the biological laboratories was determined, are incorporated in the prediction system for increasing its predictive capability.

Different variable selection methods (forward stepwise, Genetic algorithm, and backward stepwise) were used for selecting the most significant variables and also various linear and nonlinear (neural networks) statistical techniques were used for obtaining the discriminant and regression models.

2.1. Developing the classification function

The development of a discriminate function, which permits the classification of a compound as active or inactive, for the design and discovery of new A_1 ARs is the key of the present approach. Thus, it is necessary to select a training data set of analogues of adenosine with high structural variability. The quality of the classification model depends to a large degree on the quality of the training data set selected.[12]

In this study, a general data set of 167 adenosine analogues has been considered, 136 of which are considered active and the remainder being inactive at the A_1 AR.[30–35] This data set was divided into two subsets: the first set containing 134 compounds (109 active, 25 inactive) was used as a training set; the other set of 33 compounds (27 active, 6 inactive) was used for the external validation of the developed model. The selection was carried out using a *K*-Means Cluster Analysis (*k*-MCA). This technique has been used recently by us with excellent results.[12,13,36–38] The main idea consists of carrying out a partition of either an active or an inactive series of compounds in several statistically representative groups of substances, and then to select from these representative groups the members of training and predicting series. This procedure ensures that any chemical group (as determined by the clusters derived from *k*-MCA) will be represented in both compound series (training and predicting). Figure 2 graphically illustrates the above-described procedure where two independent cluster analyses (one for the active compounds and the other for the inactive ones) were carried out to select a representative sample for the prediction and training sets.

An initial *k*-MCA (*k*-MCA1) step split the active compounds into four clusters of 39, 46, 29, and 22 members and with standard deviations of 1.22, 1.12, 0.87, and 0.96, respectively. On the other hand, the series of inactive compounds were partitioned into three clusters (*k*-MCA2) of 8, 10, and 13 members and with standard deviations of 0.65, 0.42, and 0.27, respectively. The training and prediction sets were constituted by selecting, in a random manner, compounds that belong to each cluster (Fig. 2).

To ensure a statistically acceptable data partition into several clusters, we considered the number of members in each cluster and the standard deviation of the variables in the cluster (as low as possible). We also made an inspection between SS and within SS (standard deviation between and within clusters), the respective Fisher ratio, and their *p*-level of significance considered to be lower than 0.05. The set of the most relevant descriptors was used in both analyses.

There is a main conclusion that may be deduced from *k*-MCA analysis: the structural diversity of several adenosine analogues known to date may be described by at least four statistically homogeneous clusters of chemicals. In any case, further conclusions about the mechanistic and/or pharmacological signification of these clusters seem to be speculative. However, the use of the *k*-MCA analysis here points to a structurally representative distribution of structures into training and predicting series. Finally, it is important to

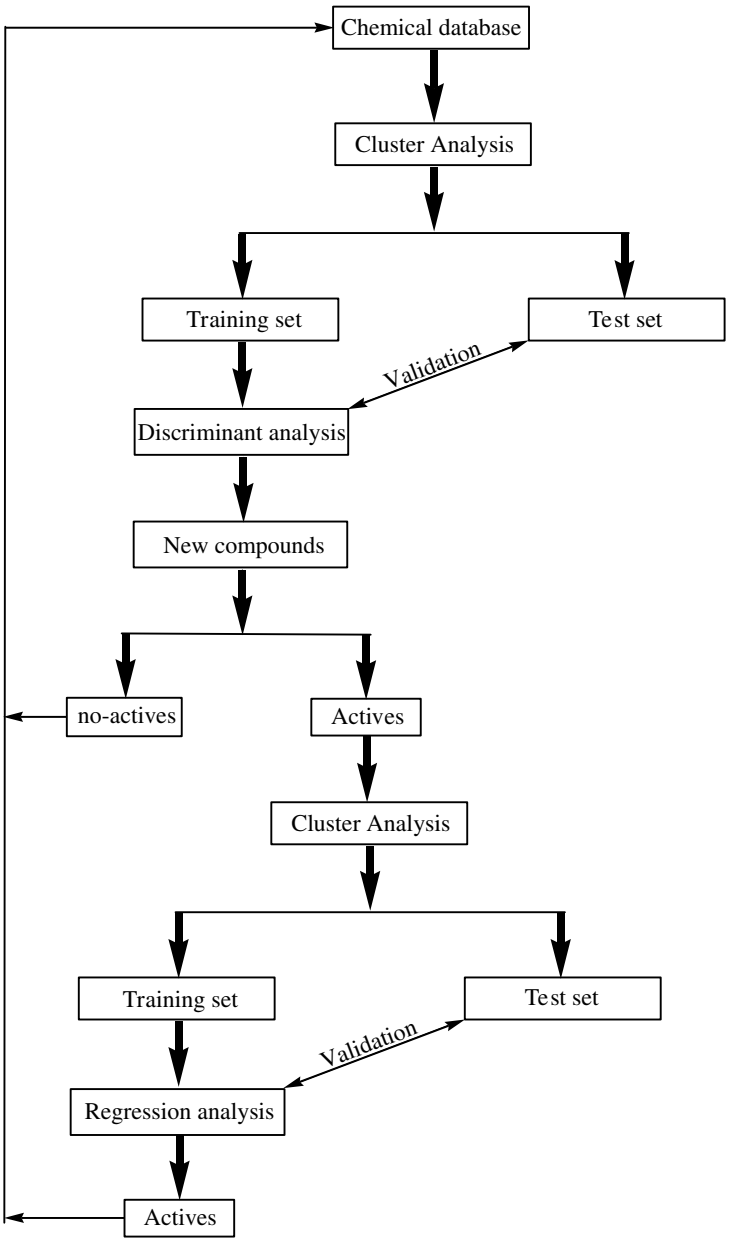


Figure 1. Decision tree illustrating the classification and prediction scheme for the QSAR strategy proposed.

highlight that the compounds in the predicting set were never used in the development of the classification model.

The following classification model was obtained with the use of the linear discriminant analysis (LDA) techniques implemented in the STATISTICA computer software[39] (Eq. 1)

$$\begin{aligned} A = {} & 20.795 \cdot \mathrm{R3m} - 8.279 \cdot \mathrm{J3D} - 284.771 \cdot \mathrm{R4v}^{+} \\ & + 88.116 \cdot \mathrm{HATS2e} + 10.903 \cdot \mathrm{H3u} - 15.648 \\ & \times \mathrm{H3v} + 7.589 \cdot \mathrm{Mor11v} + 9.369 \cdot 10^{-5} \\ & \times \mathrm{GMTIV} - 23.600, \end{aligned} \quad (1)$$

$N = 134$, $D^2 = 7.41$, $\lambda = 0.3512$, $F(8,125) = 28.861$, $p < 10^{-7}$,

where N is the number of cases in the model, F is the Fisher ratio, p is the significance of the model, λ is the Wilks λ, where the overall discrimination can take values in the range of 0 (perfect discrimination) to 1 (no discrimination). The Mahalanobis distance (D) indicates the separation of the respective groups; it shows if the model has an appropriate discriminatory power for differentiating between the two respective groups. In developing these classification functions the values of 1 and -1 were assigned to active and inactive compounds, respectively. We used the a posteriori probabilities in order to classify the compounds as active/inactive. This is the probability that the respective case belongs to a particular group (active or inactive). In addition, the compounds were considered unclassified (U) by the model when the differences in the percentage of classification between two groups did not differ by more than 5%.

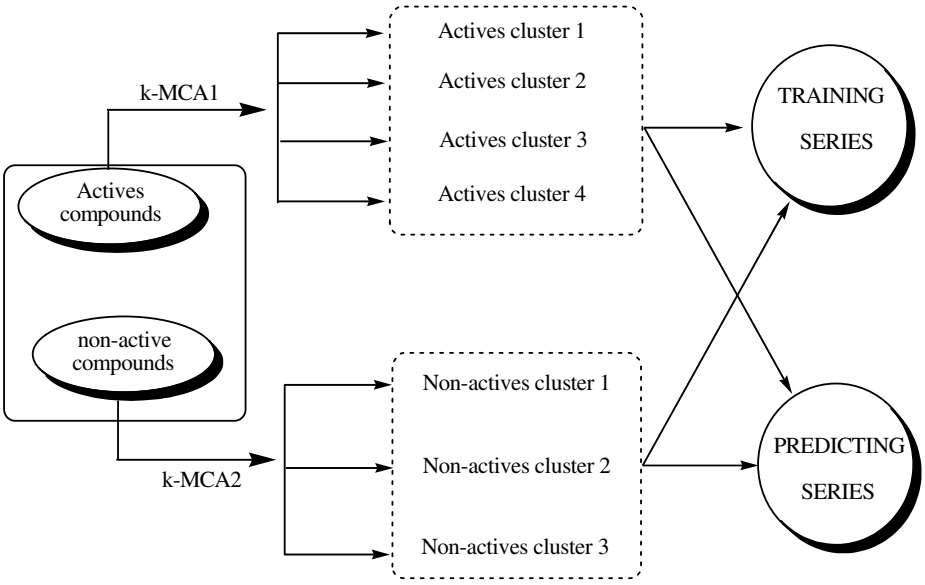


Figure 2. Training and predicting series design throughout *k*-MCA.

The parametrical assumptions (noncolinearity, normality, and homocedasticity) are very important factors in the application of linear multivariate statistic techniques to carry out any QSAR model. The validity and statistical significance of any QSAR model is strongly conditioned by the above mentioned factors.[40]

One of the mathematical forms of our discriminant model was chosen to be linear because, in the absence of prior information, it is the simplest mathematical form to assume for any model. This choice was supported by visual examination of the distribution of the residuals for the 134 cases (residuals against cases), that did not show any characteristic pattern. This observation was an indication that our model did not exhibit any nonlinear dependencies.[41]

Multi-colinearity is not a desirable phenomenon. Even if in any postulated model it is assumed the separated or independent influence of all the predictors (independent variables) over the activity under study (dependent variable), it could be the case that in any sample used to evaluate the model, some or all the predictive variables were highly correlated and it would not be feasible to separate their respective influences over the modeled activity. Multi-colinearity does not affect the predictability of the model, but the physical meaning of the resultant model, as well as the influence of the predictive variables may be misinterpreted and consequently, result in erroneous inferences about the activity under study.

The study of multi-colinearity of the variables in our model using the resultant matrix correlation (Table 1) showed the existence of significant intercorrelation between some of the variables. The most important ones are remarked in bold face.

In order to avoid colinearity, Randić's orthogonalization procedure was carried out.[42–46] The main philosophy of this approach is to allow the exclusion of descriptors on the basis of colinearity with other variables previously included in the model. It is known that the interrelatedness among the different descriptors can result in highly unstable regression coefficients, which makes it impossible to know the relative importance of an index and underestimates the utility of the regression coefficients in a given model.

The first step in orthogonalizing the molecular descriptors in model 1 was to select the appropriate order of orthogonalization, which in this case was the order in which the variables were selected in the forward stepwise search procedure of the discrimination analysis. In this sense, in Eq. 1 $J3D = {}^{1}\Omega(J3D)$ was used as the first

Table 1. Correlation coefficients among the eight most significant variables in the discriminant model (Marked correlations are significant at $p < .05000$)

	J3D	R4v+	R3m	HATS2e	H3u	H3v	Mor11v	GMTIV
J3D	1.00	0.49	−0.23	**0.63**	**−0.55**	**−0.59**	−0.35	**−0.54**
R4v+		1.00	0.23	**0.54**	**−0.62**	**−0.54**	−0.19	−0.39
R3m			1.00	−0.14	0.08	0.34	0.24	0.14
HATS2e				1.00	**−0.84**	**−0.83**	−0.47	**−0.62**
H3u					1.00	**0.93**	0.48	**0.69**
H3v						1.00	**0.54**	**0.74**
Mor11v							1.00	**0.64**
GMTIV								1.00

orthogonal variable. Afterwards, the successive residuals of the step-by-step regressions between each variable selected in the model and the others in order of statistical significance were calculated.[43] All these residuals were used as the remnant orthogonal variables in Eq. 1.[43] In this analysis the least squares method selected all orthogonal analogues of colinear variables. This ensured that, in spite of the colinearity of the variables, each variable had an amount of information not encoded in the others.[43,44]

Once the descriptors had been orthogonalized, Eq. 2 was obtained. The statistical parameters of Eq. 2 did not suffer high variation with respect to Eq. 1, only differences in the values and the sign of the coefficients were observed. These variations are due to the fact that the coefficients of the orthogonal model are more stable than the model without orthogonalization.

$$A = -2.717 \cdot {}^{1}\Omega\mathrm{J3D} - 3.252 \cdot {}^{2}\Omega\mathrm{R4v}^{+} + 2.546 \times {}^{3}\Omega\mathrm{R3m} + 1.436 \cdot {}^{4}\Omega\mathrm{HATS2e} - 2.104 \times {}^{5}\Omega\mathrm{H3v} - 2.745 \cdot {}^{6}\Omega\mathrm{GMTIV} + 1.769, \quad (2)$$

$N = 134$, $D^2 = 7.19$, $\lambda = 0.385$, $F(6,127) = 33.856$, $p < 10^{-7}$.

Nevertheless, as result of the orthogonalization process of model 1 were found two variables statistically not significant. These variables, H3u and Mor11v, were excluded in the final model (see Eqs. (1) and (2)). H3u was strongly correlated with H3v, HATS2e, and J3D containing 93.0%, 84.0%, and 55.0% of duplication, respectively. J3D was the variable taken as the first orthogonal descriptor, and the rest of the descriptors were orthogonalized with respect to it. Therefore, those variables that have elevated correlation with it will have a high probability to leave the model. Similarly to the previous case, the Mor11v descriptor was correlated with H3v, and GMTIV variables, with coefficients of correlation of 54.0% and 64.0%, respectively. This was probably the main cause for the exclusion of Mor11v in the description of the affinity for ARs.

The final selected descriptors were principally 3D descriptors which belong to the family of GETAWAY ($R4v^{+}$, R maximal autocorrelation of lag 4/weighted by atomic van der Waals volumes; R3m, R autocorrelation of lag 3/weighted by atomic masses; HATS2e, leverage-weighted autocorrelation of lag 2/weighted by atomic Sanderson electronegativities; H3v, H autocorrelation of lag 3/weighted by atomic van der Waals volumes) and Geometrical descriptors (J3D, 3D-Balaban index). Only one significant 2D descriptor, from statistical point of view, was included in this QSAR model (GMTIV, Gutman MTI by valence vertex degrees). It seems the conformation and the flexibility of nucleosides are very important in the affinity by A_1 adenosine receptors, even to a larger extent than the topology of these compounds.

After the orthogonalization of descriptors, verification of other important parametrical assumptions, as normality and homocedasticity, was carried out. An important aspect of the description of a variable is the shape of its distribution, which tells you the frequency of values from different ranges of the variable. Our interest here is to check how well the distribution can be approximated by the normal distribution. Precise information can be obtained by performing the tests of normality to determine the probability that the sample came from a normally distributed population of observations. The tests of normality (Kolmogorov–Smirnov)[47] carried out for our model did not find statistically significant differences ($p > 0.02$) from the respective value (d), assuming the hypothesis of normal distribution of the explored variables. Normality can also be evaluated by a visual examination of the data using a histogram showing the frequency distribution of the variables in the model. In our model, these representations showed the variables as normally distributed (Fig. 3). This result was in agreement with the normality test and demonstrated the linear distribution of these variables.

The last parametrical assumption is the homocedasticity. The heterocedasticity is a potentially serious problem for the researcher to consider. If we continue employing the usual statistical test methods in the presence of heterocedasticity, independently of the conclusions or inferences made by using the model, we may commit serious errors. This is due to the overestimation or underestimation of the coefficients of the variables included in the model, associated with the determination of these coefficients by the usual least squares method in the presence of heterocedasticity. This implies variations in the confidence intervals with which it is working. As a result, Student and Fisher tests will possibly be the source of inaccurate results due to a high variance of the coefficients. Consequently, at first sight, statistically nonsignificant coefficients could be statistically significant if the proper confidence intervals would be established based on general least squares method.[40]

A simple method to detect heterocedasticity in large data sets is a graphic method based on the examination of the residuals of the variables included in the model. In reality, if any prior or empiric information concerning to the heterocedasticity does not exist, it is possible to develop our model under the assumption of homocedasticity, and then, to carry out a *post mortem* analysis of the square residuals of the variables included in the model in order to detect any systematic or consistent pattern.[40]

As can be noted, in Figure 4, from the scatter plots for the six predictive variables against their respective square residuals, no systematic pattern is observed. So, we can assume that the homocedasticity assumption is valid with respect to our model and consequently the resultant inferences can be used to derive conclusion based on model predictions.

Once the parametrical assumptions were validated, an interpretation of the model obtained in Eq. 2 was carried out. This model classified correctly 91.8% of the compounds in the training set (11 misclassifications among 134 cases) and 84.9% of the compounds in the prediction set (5 misclassifications among 33 cases).

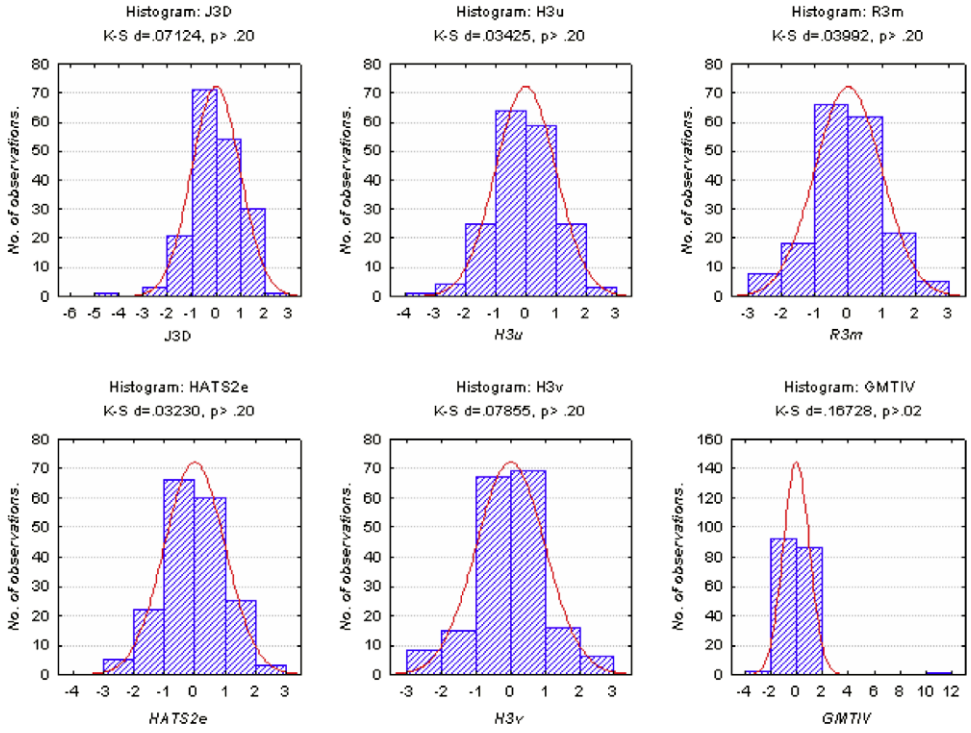


Figure 3. Histograms showing the frequency distribution of the variables within the model reported in Eq. (2).

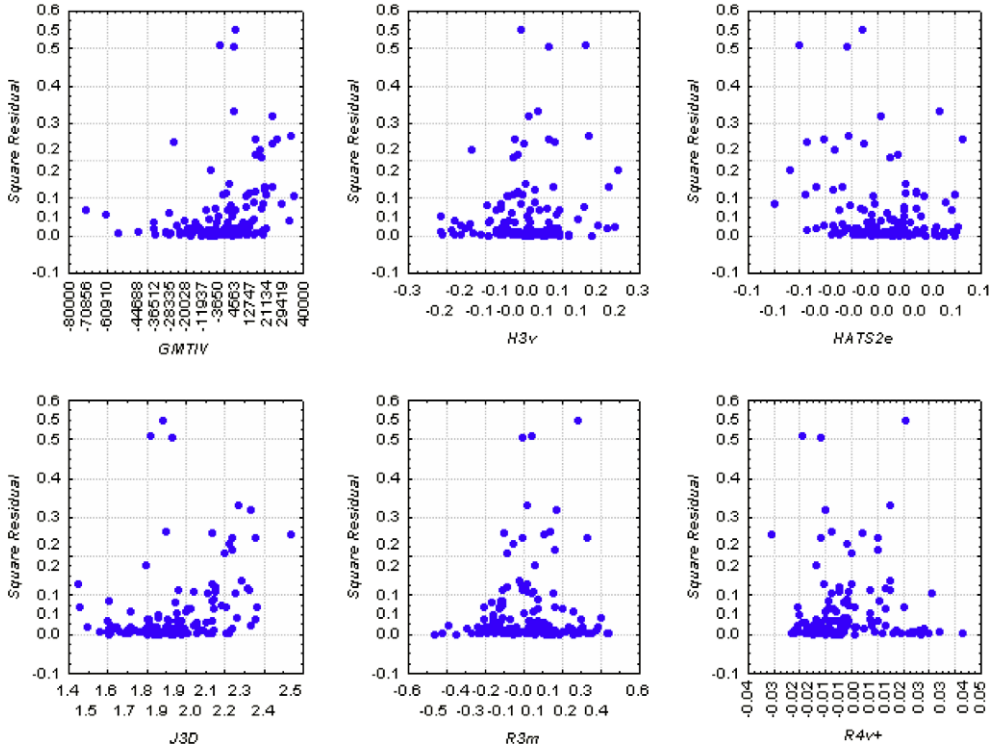


Figure 4. Scatter plots of the four predictive variables against their respective square residuals.

The percentage of false actives in the training set was only 2.24% (3 inactive compounds were classified as active among 134 cases), and 5.97% were false inactives (8 active compounds were classified as inactive). The percentages for the prediction set were 3.03% of the false actives and 6.06% of the false inactives. It is desirable that the number of false active compounds be as low as possible because this number represents inactive com-

pounds that will be subjected to the biological assays with the consequent loss of time and resources. The results of classification of the compounds used in this study, in both training and external set, are included in the Supporting Information. In the prediction set there are two compounds for which the percentage of classification as active and inactive is not more than 5%. These compounds are considered as unclassified and they are designated with 'U' in the Supporting Information. In the training set no unclassified compounds were found.

Although the previous linear discriminant analysis showed excellent statistical parameters for both training and prediction sets, and with the aim of exploring other statistical techniques, several neural network classification models have been developed. A summary of the best model obtained for each methodology used is shown in Table 2, while Figure 5 shows the architecture of these neural networks.

All the network architectures used in the analysis have shown satisfactory performance on the training and predicting sets (Table 2). If the ratio between false actives and inactives in both training and prediction sets of these nonlinear models and the linear discriminant analysis is considered, the best model found was the one for the LDA reported in Eq. 2. This model presented the best equilibrium between the false actives and inactives in comparison with the models reported using neural network methodologies (Table 3). The neural networks presented a better classification in the training set; but a worse predictive ability than the LDA. Nowadays, it is known that the neural networks tend to overfit the training set, therefore, this kind of model cannot be considered as useless tool for QSAR studies. This specific case should not be considered as an overfitting problem but a high imbalance in the test set. However, considering the simplicity of the LDA contrasted to the neural networks, and the possibilities of physical interpretation offered by the LDA model, Eq. 2 was proposed as the more significant model presented in this work.

Table 2. Summary of the best neural network models obtained

Neural network model	Profile[a]	Train perf.[b]	Predict perf.[c]
Three Layer Perceptrons (MLP3)	7:7-7-1:1	0.947	0.878
Four Layer Perceptrons (MLP4)	8:8-8-9-1:1	0.947	0.878
Linear Network Training (LNT)	7:7-1:1	0.940	0.878
Probabilistic Neural Network (PNN)	4:4-7-1:1	0.962	0.696
Radial Basis Function (RBF)	2:2-6-1:1	0.917	0.878

[a] Network composition. For instance: 7:7-7-1:1 7 variables (inputs) in the MLP3 model: 7 neurons in the first layer—7 neurons in the second (hidden) layer—1 Output in the fourth layer: 1 Output.
[b] Performance of correctly classified compounds in training set.
[c] Performance of correctly classified compounds in prediction set.

2.2. Developing the regression function

The second part of our prediction system was the development of a regression model applying linear and nonlinear techniques with the aim of prediction quantitatively the affinity of new analogues, according to discriminant analysis described in the first part of our system (see Fig. 1), for A_1 ARs. The active set of analogues used in the discriminant analysis was split according to cluster analysis in two sets, for training and prediction. The training set was used for obtaining the regression model and the predicting set was utilized for validating the model.

The model selection was subjected to the principle of parsimony.[48] A function with high statistical significance but having as few parameters as possible was chosen. The linear regression model with eight variables was the best according to this rule, because it showed a major balance between variables and statistical parameters (Eq. 3).

$$\begin{aligned} -\log(K_i) = {} & 0.872 \cdot \mathrm{BELm3} - 0.236 \cdot \mathrm{X4Av} \\ & - 1.003 \cdot \mathrm{GGI8} + 0.673 \cdot \mathrm{ATS3m} \\ & + 0.510 \cdot \mathrm{RDF020u} - 0.169 \cdot \mathrm{Mor17m} \\ & + 0.215 \cdot \mathrm{Mor31v} + 0.405 \cdot \mathrm{C008} - 2.101, \quad (3) \end{aligned}$$

$N = 111$, $R^2 = 0.662$, $S = 0.622$, $F_{exp} = 24.954$, $p < 10^{-5}$, AIC = 0.455, FIT = 1.142,

where N is the number of compounds used, R^2 is the determination coefficient, S is the standard deviation of the regression, F_{exp} is the Fisher ratio at the 95% confidence level, p the significance of the variables in the model, AIC is the Akaike information criterion,[49,50] and FIT is the Kubinyi function.[51,52]

Although the above reported model (Eq. 3) presented acceptable statistical parameters, a step-by-step outlier extraction procedure could lead to different models with a better statistical profile according to the results obtained in Table 4.

An outlier to a QSAR is identified normally by having a large standard and deleted residual.[53] There are several reasons for their occurrence in QSAR studies, for example, some compounds might be acting by a different mechanism than the majority of the data set. It is also likely that outliers might be a result of random experimental error that might be significant when analyzing large data sets from different laboratories. Although it is acceptable to remove a small number of outliers from a QSAR, it is noted that it is not acceptable to remove outliers repeatedly from a QSAR analysis simply to improve a correlation. The limit value for outlier extraction accepted in the literature is 10% of the general data.[53] In our case six outliers were extracted, representing only 5.40% of the whole data set. The compounds 1, 9, 10, 64, 133, and 142 displayed the largest residuals and were deleted from the entire data set and should be considered as outliers (see Supporting Information).

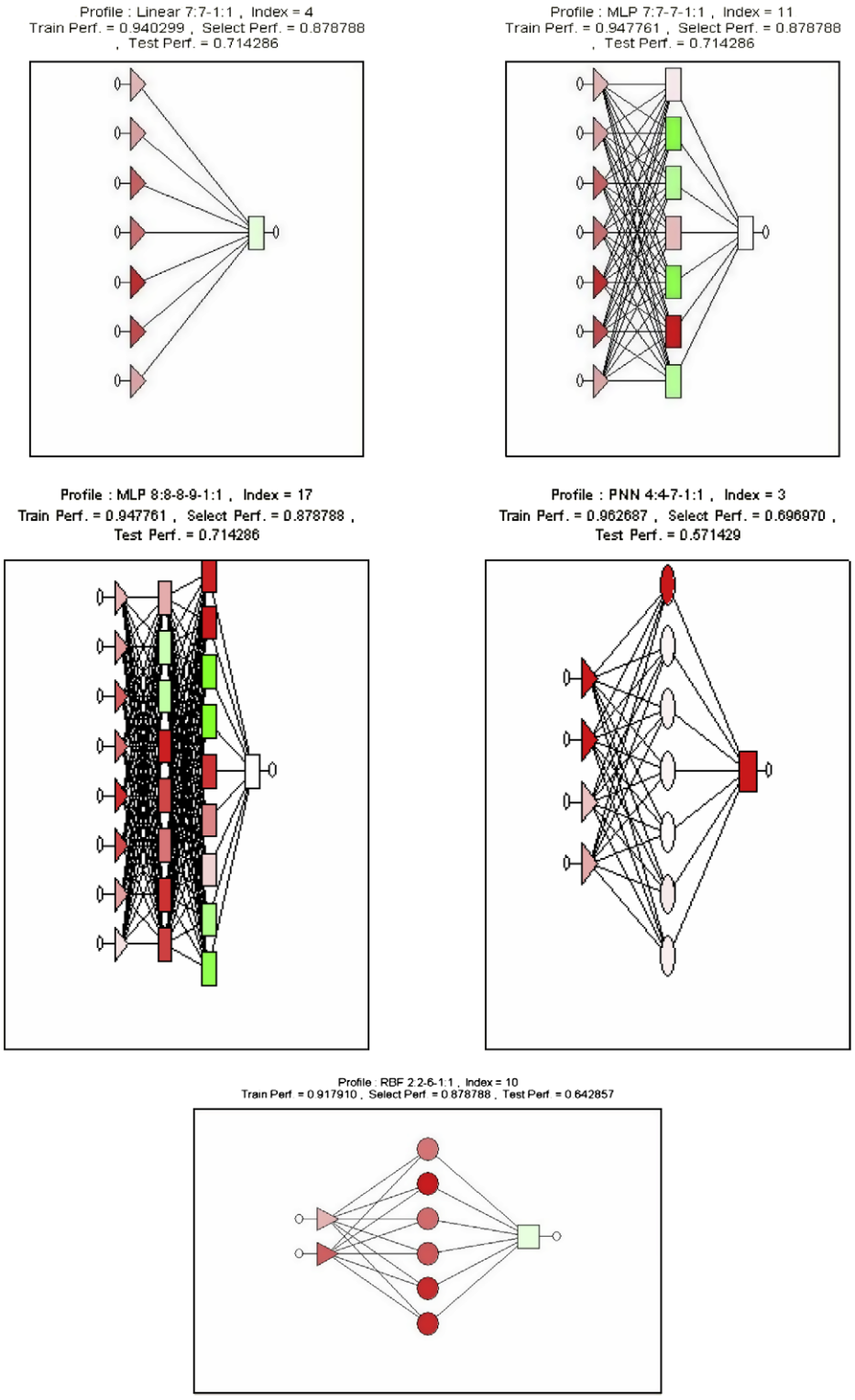


Figure 5. Different neural networks architectures used in the discriminant analysis.

After removal of these compounds from the training set, Eq. 4 was obtained. As can be seen, the statistical parameters of the equation improved significantly from a statistical point of view, in comparison with those reported for Eq. 3. A conclusive proof to find the best model by statistical criteria is given by the parameters AIC and FIT. The model that produces the minimum value of the AIC and the highest FIT statistics should potentially be considered the most useful. These criterions were better for the new model (Eq. 4; AIC = 0.331, FIT = 1.760), than the reported values in Eq. 3, in 27.25% and 35.11%, respectively. Therefore, this was new evidence that compounds 1, 9, 10, 64, 133, and 142 should be considered as potential outliers.

Table 3. Summary of the best models obtained according to the prediction of false active and inactive compounds in the training and prediction sets

Models	False actives[a]	False inactives[a]	False actives[b]	False inactives[b]
LDA	2.24[*]	5.97	3.03	6.06
Linear NN	5.22	0.74	12.12	0
RBF	5.22	2.99	9.09	3.03
MLP 3	5.22	0	12.12	0
MLP 4	5.22	0	12.12	0
PNN	0	3.73	9.09	21.21

[a] Training set.
[b] Prediction set.
[*] The values are expressed in percentage.

Table 4. Statistical parameters of the models obtained by step-by-step outlier extraction procedure in the regression training set

Models	Outliers	R^2	S	F_{exp}	AIC	FIT
4	1	0.695	0.589	28.760	0.409	1.323
5	1, 142	0.708	0.570	30.296	0.384	1.402
6	1, 142, 133	0.721	0.559	31.958	0.369	1.487
7	1, 142, 133, 10	0.732	0.547	33.844	0.355	1.565
8	1, 142, 133, 10, 64	0.746	0.537	35.535	0.342	1.676
9	1, 142, 133, 10, 64, 9	0.756	0.528	37.300	0.331	1.760

$$-\log(K_i) = 0.818 \cdot \mathrm{BELm3} - 0.301 \cdot \mathrm{X4Av} - 1.060 \cdot \mathrm{GGI8} + 0.729 \cdot \mathrm{ATS3m} + 0.556 \cdot \mathrm{RDF020u} - 0.200 \cdot \mathrm{Mor17m} + 0.228 \cdot \mathrm{Mor31v} + 0.455 \cdot \mathrm{C008} - 2.072, \quad (4)$$

$N = 105$, $R^2 = 0.756$, $S = 0.528$, $F_{exp} = 37.300$, $p < 10^{-5}$, AIC = 0.331, FIT = 1.760.

The final selected variables were 2D and 3D descriptors which belong to the family of Burden eigenvalues (BELm3: lowest eigenvalue n 3 of Burden matrix/weighted by atomic masses), connectivities indices (X4Av, average valence connectivity index chi-4), 2D autocorrelations (ATS3m, Broto-Moreau autocorrelation of a topological structure—lag 3/weighted by atomic masses), topological (GGI8, topological charge index of order 8), RDF (RDF020u, Radial Distribution Function—2.0/unweighted), and 3D-MoRSE descriptors (Mor17m, 3D-MoRSE—signal 17/weighted by atomic masses; Mor31v, 3D-MoRSE—signal 31/weighted by atomic van der Waals volumes). It seems the conformation and the flexibility of nucleosides are as important as the topology in the determination of the quantitative affinity by A_1 adenosine receptors.

On the other hand, several nonlinear regression models using neural networks were obtained. A summary of the best model obtained for each methodology used and the architectures of these neural networks can be observed in Table 5 and Figure 6, respectively.

Table 5. Summary of the best neural network models obtained

Neural network model	Profile[a]	Train perf.[b]
Multi Layer Perceptrons (MLP)	8:8-6-1:1	0.720
Linear Network Training (LNT)	8:8-1:1	0.662
Generalized Regression Neural Networks (GRNNs)	5:5-6-2-1:1	0.925
Radial Basis Function (RBF)	6:6-8-1:1	0.342

[a] Network composition. For instance: 8:8-6-1:1 8 variables (inputs) in the MLP model: 8 neurons in the first layer—6 neurons in the second (hidden) layer—1 neuron in the third (hidden) layer: 1 Output.
[b] Performance (R^2) of correctly classified compounds in training set.

According to the results obtained for the neural networks, all of them, but the Radial Basis Function (RBF) model, presented similar or better R^2 coefficients than the linear model reported in Eq. 3. These approximations have good fit capability, however as before for the linear model the elimination of some potentials outliers was carried out. The result of the models obtained after this process is shown in Table 6. As can be seen, all the models were improved significantly in determination coefficient by elimination of the potential outliers.

The Generalized Regression Neural Network (GRNN) presented the best determination coefficient (Table 6). However, the results of the validation process showed a very different behavior (Table 7). The GRNN model reported a poor performance in the prediction set and based on the nature of this model it can be concluded that the model is overfitting. All models require validation to avoid this type of effect and the only appropriate approach is to use an external prediction set. If a QSAR model cannot be used to make predictions, then it is not of practical use. Statistical fit should not be confused with the ability of a model to make predictions.

Therefore, based on this hypothesis, the linear model reported in Eq. 4 and the Multi Layer Perceptrons (MLP) neural network present the best profile for predicting new candidates in the second part of our system.

2.3. Prediction activity of new agonists for A_1 adenosine receptors

Virtual screening has emerged as a worthwhile alternative to high-throughput library screening.[13,54,55] It mainly consists of the handling and screening of large databases in order to find a reduced set of compounds for which a prediction of a specific biological activity has been made using clustering and similarity searching techniques. After these techniques are applied, other molecular modeling techniques, such as docking and molecular superimposition, are used to define pharmacophores and to refine the designs.

Applicability of the models is the major bottleneck in QSAR. As outlined in Section 1, almost all adenosine agonist QSAR models have been derived from a reduced set of compounds; for this reason these models cannot predict the potential of any organic chemical to act as an AR agonist irrespective of its molecular structure.

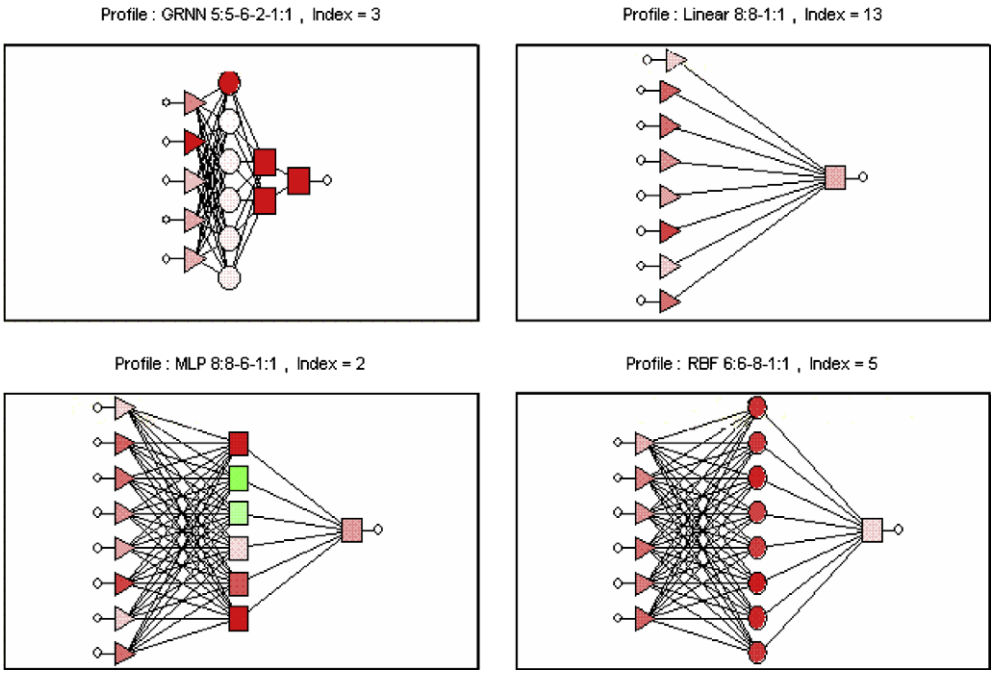


Figure 6. Different neural networks architectures used in the regression analysis.

Table 6. Summary of the best neural network models obtained without six potential outliers

Neural network model	Profile[a]	Train perf.[b]
Multi Layer Perceptrons (MLP)	8:8-6-1:1	0.868
Linear Network Training (LNT)	8:8-1:1	0.737
Generalized Regression Neural Networks (GRNNs)	5:5-6-2-1:1	0.968
Radial Basis Function (RBF)	6:6-8-1:1	0.471

[a] Network composition. For instance: 8:8-6-1:1 8 variables (inputs) in the MLP model: 8 neurons in the first layer—6 neurons in the second (hidden) layer—1 neuron in the third (hidden) layer: 1 Output.
[b] Performance (R^2) of correctly classified compounds in training set.

Considering that our proposed model was derived from a diverse series of adenosine analogues, this model could be used as generally applicable for the computational discovery of A_1 AR agonists. To demonstrate the power of the present model to recognize as active or inactive different adenosine analogues, a virtual screening experiment was carried out.

Table 7. Summary of the validation process for the regression models obtained

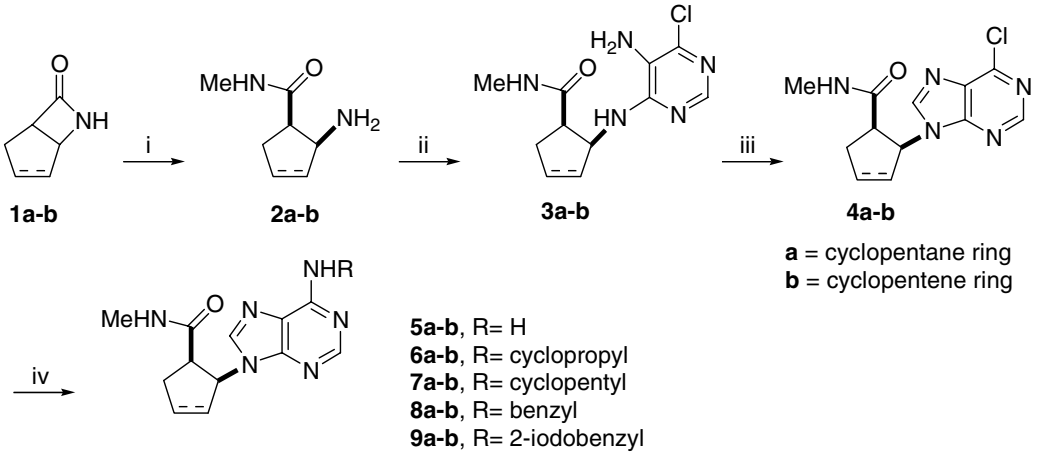
Regression models	Profile[a]	Predict perf.[b]
Multiple Linear Regression	Linear	0.745
Multi Layer Perceptrons (MLP)	8:8-6-1:1	0.729
Linear Network Training (LNT)	8:8-1:1	0.701
Generalized Regression Neural Networks (GRNNs)	5:5-6-2-1:1	0.518
Radial Basis Function (RBF)	6:6-8-1:1	0.080

[a] Network composition. For instance: 8:8-6-1:1 8 variables (inputs) in the MLP model: 8 neurons in the first layer—6 neurons in the second (hidden) layer—1 neuron in the third (hidden) layer: 1 Output.
[b] Performance (R^2) of correctly classified compounds in prediction set.

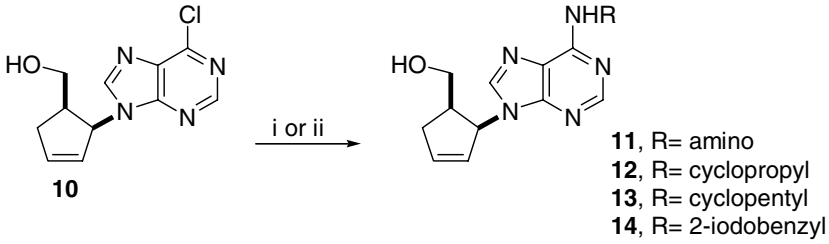
A virtual search of new adenosine analogues with affinity for A_1 ARs was performed by using the obtained discriminant and regression functions. A total of 14 carbocyclic analogues of adenosine derivatives, 1,2-disubstituted with a cyclopentane or a cyclopentene ring (compounds **5**–**9** and **11**–**14**, Schemes 1 and 2), were predicted using the model. These compounds were never used in either the training or predicting series.

Here, it is important to highlight that our main aim is to reduce as much as possible the false positives in order to reduce the number of compounds that would be sent unnecessarily to biological testing. Our effort was focused on the design of inactive compounds and active ones that would be at the boundary of active and inactive, according to the model reported in Eq. 2, to corroborate in a definitive manner the predictive potential of our system and because the majority of the existing models fail in the prediction of this latter class of compounds.

The synthesis of new adenosine analogues **5**–**9** and **11**–**14** was carried out according to the strategies given in Schemes 1 and 2. Compounds **5**–**9** were synthesized starting from the appropriate β-lactam **1**, which was transformed in the corresponding aminoamide **2** by direct amidation with methylamine in methanol. Reaction of **2a–b** with 5-amino-4,6-dichloropyrimidine gave intermediates **3** with a global yield of 40% (**3a**) and 63% (**3b**). The treatment of compound **3** with ethyl orthoformate under reflux, followed by acidic treatment, gave **4a–b** in good yields. Finally, the 6-chloro group of **4a–b** was displaced by aqueous ammonia to give **5a–b** or by an alkylamine in ethanol to give compounds **6**–**9** (Scheme 1). Similarly, the target compounds **11**–**14** were also obtained by displacement of the 6-chloro atom of **10**[56] with the appropriate nucleophile (Scheme 2).



Scheme 1. Reagents and conditions: (i) $MeNH_2$ 2 M in MeOH, MeOH, 90 °C, 29 h; (ii) 5-amino-4,6-dichloropyrimide, *n*-BuOH, Et_3N, reflux, 15 h, 40% (**3a**), 63% (**3b**); (iii) $CH(OEt)_3$, HCl 12 M, rt, 14 h, 99% (**4a**), 89% (**4b**); (iv) NH_4OH 25%, reflux, 6 h, 76% (**5a**), 46% (**5b**) or RNH_2, EtOH, reflux, 98% (**6a**), 99% (**6b**), 100% (**7a**), 98% (**7b**), 100% (**8a**), 95% (**8b**), 81% (**9a**), 93% (**9b**).



Scheme 2. Reagents and conditions: (i) 25% NH_4OH reflux, 84% (**11**); (ii) RNH_2, EtOH, reflux, 99% (**12**), 99% (**13**), 92% (**14**).

All the synthesized compounds (**5–9** and **11–14**) were tested in radioligand binding assays for affinity to rat brain A_1 receptors. The results of experimental pharmacological evaluation together with the predictions carried out with the discriminant models obtained are given in Table 8. According to the experimental results, the majority of compounds tested are inactive, and only two compounds (**9b**, **13**) present some activity. The linear model described by Eq. 2 gave the best predictions. This behavior was already indicated when the model was validated with the prediction set. Table 9 illustrates the general predictions of the models.

Even if apparently the models presented a good overall classification, only the linear model showed a good balance between the false active and inactive compounds

Table 8. Affinities of Adenosine Analogues in Radioligand Assays at Rat Brain A_1 receptors and their respective predictions according to the discriminant models obtained

Compound	K_i[a] (μM) or % inhibition at 10^{-4} M[b]	Linear[c]	LNT[d]	RBF[d]	MLP 3[d]	MLP 4[d]	PNN[d]
5a	25%	−1	−1	−1	−1	−1	−1
6a	<10%	−1	1	1	1	1	1
7a	31%	−1	−1	1	−1	−1	−1
8a	26%	−1	−1	−1	−1	−1	−1
9a	<10%	−1	−1	−1	−1	−1	−1
5b	22%	−1	−1	−1	−1	−1	−1
6b	50%	−1	−1	1	−1	−1	1
7b	<10%	1	1	−1	1	1	1
8b	<10%	−1	−1	−1	−1	−1	1
9b	8.12 ± 2.18	1	−1	−1	−1	−1	−1
11	29%	−1	−1	−1	−1	−1	−1
12	32.3 ± 5.9	−1	−1	1	−1	−1	−1
13	3.0 ± 0.70	1	−1	1	−1	−1	−1
14	25.5 ± 10.5	−1	−1	−1	−1	−1	−1

[a] Inhibition of specific [^{3}H]R-PIA binding in rat brain membranes, expressed as K_i ± SEM in μM (n = 3–5).
[b] When a percentage is given, it refers to the percent displacement of inhibition binding at 10^{-4} M.
[c] Classification according to the linear discriminant analysis of Eq. 2.
[d] Classification according to the respective neural networks discriminant analysis.

Table 9. Summary of the general classification process for the new compounds obtained

Models	Profile	Overall performance*
Linear discriminant analysis	Linear	0.928
Three Layer Perceptrons (MLP3)	7:7-7-1:1	0.714
Four Layer Perceptrons (MLP4)	8:8-8-9-1:1	0.714
Linear Network Training	7:7-1:1	0.714
Probabilistic Neural Network (PNN)	4:4-7-1:1	0.571
Radial Basis Function (RBF)	2:2-6-1:1	0.642

* The values are expressed in percentage.

Table 10. Summary of the prediction of false actives and inactives of the new compounds tested

Models	False actives[a]	False inactives[a]
Linear discriminant analysis	8.33*	0
Linear Network Training	14.28	100
Radial Basis Function	28.57	50
Three Layer Perceptrons	14.28	100
Four Layer Perceptrons	16.67	100
Probabilistic Neural Network	28.57	100

[a] Performance of correctly classified compounds in the new compounds obtained.
* The values are expressed in percentage.

(Table 10). In this connection, only the linear model recognized the two active compounds and classified an inactive one as active, for an overall rate of correct classification of 92.86%, which corroborates our previous results.

Once the compounds were classified as active or inactive with the best discriminant model, and according to our scheme of prediction proposed in Figure 1, the K_i values of the active compounds were predicted using the linear model reported by Eq. 4 and MLP neural networks, which were the best regression models obtained above. The models reported an excellent prediction of the activities of these compounds (**9b**, **13**). Very small residual was obtained and for that reason a precise prediction was achieved using both regression techniques (Table 11).

3. Conclusions

Although biological phenomena are complex by nature, in this work a high or low affinity for the A_1 AR of a set of 167 compounds was successfully modeled through LDA. This statistical technique has demonstrated its superiority to other nonlinear techniques in the first phase of our prediction system. In the second phase, the linear and nonlinear MLP neural networks regression showed the best results to predict the quantitative affinity of new compounds for the A_1 AR. Consequently, we consider that this strategy is very promising for screening of large chemical libraries, and to speed the identification of new hit compounds.

QSAR methods are very useful tools in Medicinal Chemistry and specifically in the nucleoside field have not yet found general use. They can aid undoubtedly in the design of new analogues having better biological and pharmacokinetic profiles. Other investigations in this field to obtain better models than those published previously (using different descriptors and statistical techniques) and for obtaining new models with biological data from the other receptor subtypes are in progress in our laboratory.

4. Experimental

4.1. Chemistry

Compounds **10** and **11** were synthesized as reported.[56] Melting points were determined in capillary tubes using a Gallenkamp apparatus. ^{1}H and ^{13}C NMR spectra were recorded on a Bruker ARX-400 or Bruker AMX-500 instruments. The chemical shifts δ are expressed as ppm downfield from TMS. Coupling constants J are given in Hz. Complete assignment of the signals was performed by NOE, DEPT, HMQC, or HMBC experiments. Mass spectra were recorded using a Hewlet–Packard 5988A spectrometer. Microanalyses were performed using a Perkin-Elmer 240B elemental analyzer (C, H, N) and were within ±0.4% of theoretical values. Flash chromatography (FC) was carried out using silica gel Merck 60 (230–400 mesh). Analytical TLC was performed on plates precoated with silica gel Merck 60 F254 (0.25 mm).

Table 11. Biological activity, predicted, and residual for the active compounds according to the linear and Multi Layer Perceptrons neural networks models

Compound	K_i (μM)[a]	$-\log(K_i)$[b]	Predicted	Residual
Linear regression model				
9b	8.12 ± 2.18	−3.910	−3.932	0.022
13	3.0 ± 0.70	−3.477	−3.235	−0.242
Multi Layer Perceptrons neural networks				
9b	8.12 ± 2.18	−3.910	−3.867	−0.043
13	3.0 ± 0.70	−3.477	−3.417	−0.060

[a] Displacement of specific [^{3}H]R-PIA binding in rat brain membranes, expressed as K_i ± SEM in μM (n = 3–5).
[b] The K_i is expressed in nM concentration.

4.1.1. (±)-*cis*-2-[(5-Amino-6-chloro-4-pyrimidinyl)amino]-*N*-methylcyclopentanecarboxamide 3a. To a solution of **1a**[57] (245 mg, 2.2 mmol) in MeOH (1 mL) was added dropwise $MeNH_2$ (2 M in MeOH, 27 mL). The reaction mixture was stirred at 90 °C in a sealed tube under an argon atmosphere for 24 h. The solvent was evaporated in vacuo, and the crude residue that contains the 2-amino-*N*-methylcarboxamide **2a** was re-dissolved in *n*-BuOH (8 mL), treated with 5-amino-4,6-dichloropyrimidine (384 mg, 2.34 mmol) and Et_3N (0.68 mL), and refluxed under an argon atmosphere for 13 h. The solvent was removed in vacuo, and the residue was purified by FC (CH_2Cl_2–MeOH, 98:2) producing **3a** (239 mg, 40%) as a white solid: mp 233–235 °C; ^{1}H NMR (400 MHz, DMSO-d_6) δ: 1.51 (1H, m, HC*H*), 1.99–1.72 (5H, m, $2CH_2$, *H*CH), 2.39 (3H, d, J = 4.6, CH_3), 2.90 (1H, m, H-1), 4.56 (1H, m, H-2), 5.06 (2H, s, NH_2), 6.56 (1H, d, J = 7.6, NH), 7.42 (1H, m, N*H*CH_3), 7.71 (1H, s, pyrimidine); ^{13}C NMR (100 MHz, DMSO-d_6) δ: 22.1 (C-4), 25.3 (CH_3), 27.4 (C-5), 31.2 (C-3), 46.6 (C-1), 54.5 (C-2), 123.4, 136.7, 145.5, 151.7, 173.1 (CO); MS (EI) *m*/*z* (%): 271 ([M+2]$^+$, 27), 269 (M$^+$, 84), 238 (M$^+$–CH_5N, 10), 213 (M$^+$-C_2H_2NO, 32), 211 (M$^+$-C_2H_4NO, 100), 184 (M$^+$-C_4H_7NO, 24), 171 (M$^+$-C_5H_8NO, 16), 157 (M$^+$-$C_6H_{10}NO$, 12), 144 (M$^+$-$C_7H_{11}NO$, 14), 126 (M$^+$-$C_4H_4N_4Cl$, 34), 95 (16), 86 (31), 69 (19), 67 (16); Anal. calcd for $C_{11}H_{16}ClN_5O$: C, 48.98; H, 5.98; N, 25.96. Found: C, 48.88; H, 5.90; N, 25.83.

4.1.2. (±)-*cis*-2-[(5-Amino-6-chloro-4-pyrimidinyl)amino]-N-methyl-3-cyclopentenecarboxamide 3b. Prepared from **1b**[56] (175 mg, 1.61 mmol) following the same procedure as for **3a** from **1a**. The residue was purified by FC (hexane–AcOEt, 6:1 → 1:3) producing **3b** (269 mg, 63%) as a white solid: mp 183–185 °C; ^{1}H NMR (400 MHz, $CDCl_3$) δ: 2.54 (1H, m, 1H-5), 2.55 (3H, d, J = 4.9, CH_3), 2.91 (1H, m, 1H-5), 3.32 (1H, m, H-1), 3.96 (2H, s, NH_2), 5.56 (1H, m, H-2), 5.71 (1H, m, H-3), 5.84 (1H, d, J = 8.6, NH), 6.00 (1H, m, H-4), 6.06 (1H, m, N*H*CH_3), 7.99 (1H, s, pyrimidine); ^{13}C NMR (100 MHz, $CDCl_3$) δ: 26.2 (CH_3), 35.0 (C-5), 47.2 (C-1), 58.2 (C-2), 122.6, 129.5 (C-3), 134.1 (C-4), 141.8, 148.4, 153.4, 173.4 (CO); MS (CI) *m*/*z* %: 269 ([M+2]$^+$, 9), 267 (M$^+$, 50), 236 (M$^+$-CH_5N, 27), 211 (M$^+$-C_2H_2NO, 24), 209 (M$^+$-C_2H_4NO, 100), 208 (37), 162 (13), 144 (M$^+$-C_7H_9NO, 40), 67 (62); Anal. calcd for $C_{11}H_{14}ClN_5O$: C, 49.35; H, 5.27; N, 26.16. Found: C, 49.24; H, 5.16; N, 26.05.

4.1.3. (±)-*cis*-6-chloro-9-[2-(*N*-methylcarbamoyl)cyclopentyl]purine 4a. A mixture of compound **3a** (55 mg, 0.2 mmol), triethyl orthoformate (1.1 mL, 6.61 mmol), and 12 M HCl (0.01 mL) was stirred at room temperature for 16 h. The solvent was removed in vacuo, and the residue was dissolved in THF (3 mL) and treated with 0.5 M HCl (4 mL) for 2 h at rt. The mixture was neutralized with 0.25 M NaOH and the solvent was evaporated in vacuo. The residue was purified by FC (CH_2Cl_2–MeOH, 98:2) producing **4a** (57 mg, 100%) as a white solid: mp 138–139 °C; ^{1}H NMR (400 MHz, DMSO-d_6) δ: 1.73 (1H, m, 1H-4′), 1.91 (1H, m, 1H-3′), 2.07 (2H, m, 1H-3′, 1H-4′), 2.15 (3H, d, J = 4.6, CH_3), 2.28 (1H, m, 1H-5′), 2.41 (1H, m, 1H-5′), 3.07 (1H, q, J = 7.6, H-2′), 5.23 (1H, q, J = 7.6, H-1′), 7.67 (1H, m, NH), 8.53 (1H, s, H-8), 8.76 (1H, s, H-2); ^{13}C NMR (100 MHz, DMSO-d_6) δ: 22.5 (C-4′), 25.2 (CH_3), 27.3 (C-3′), 30.3 (C-5′), 48.0 (C-2′), 57.5 (C-1′), 130.5 (C-5), 146.4 (C-8), 148.7, 151.1 (C-2), 152.2, 171.2 (CO); MS (EI) *m*/*z* %: 281 ([M+2]$^+$, 13), 279 (M$^+$, 41), 238 (19), 223 (20), 221 (M$^+$-C_2H_4NO, 63), 207 (M$^+$-C_3H_6NO, 23), 194 (M$^+$-C_4H_7NO, 19), 181 (M$^+$-C_5H_8NO, 28), 157 (M$^+$-C_7H_8NO, 32), 155 (M$^+$-$C_7H_{10}NO$, 100), 126 (M$^+$-$C_5H_2N_4Cl$, 41), 119 (14), 99 (20), 95 (23), 86 (21), 69 (16), 67 (47); Anal. calcd for $C_{12}H_{14}ClN_5O$: C, 51.52; H, 5.04; N, 25.04. Found: C, 51.40; H, 4.90; N, 25.16.

4.1.4. (±)-*cis*-6-Chloro-9-[2-(*N*-methylcarbamoyl)-4-cyclopentenyl]purine 4b. Prepared from **3b** (360 mg, 1.35 mmol) following the same procedure as for **4a** from **3a**. The residue was purified by FC (Cl_2CH_2-MeOH, 95:5) producing **4b** (333 mg, 89%) as a white solid: mp 200–201 °C; ^{1}H NMR (400 MHz, DMSO-d_6) δ: 2.01 (3H, d, J = 4.5, CH_3), 2.55 (1H, m, 1H-3′), 2.95 (1H, m, 1H-3′), 3.46 (1H, q, J = 8.2, H-2′), 5.88 (1H, m, H-5′), 5.93 (1H, m, H-1′), 6.40 (1H, m, H-4′), 7.93 (1H, m, NH), 8.24 (1H, s, H-8), 8.77 (1H, s, H-2); ^{13}C NMR (100 MHz, DMSO-d_6) δ: 25.0 (CH_3), 33.9 (C-3′), 46.5 (C-2′), 61.0 (C-1′), 126.1 (C-5′), 130.7 (C-5), 138.7 (C-4′), 146.6 (C-8), 148.6, 151.2 (C-2), 151.7, 169.7 (CO); MS (CI) *m*/*z* %: 279 ([M+2]$^+$, 19), 277 (M$^+$, 54), 246 (M$^+$-CH_5N, 10), 221 (M$^+$-C_2H_2NO, 18), 192 (34), 157 (37), 124 (M$^+$-$C_5H_2N_4Cl$, 36), 123 (M$^+$-$C_5H_3N_4Cl$, 86), 119 (20), 93 (10), 67 (82), 66 (100), 65 (31); Anal. calcd for $C_{12}H_{12}ClN_5O$: C, 51.90; H, 4.36; N, 25.22. Found: C, 51.75; H, 4.50; N, 25.11.

4.1.5. (±)-*cis*-9-[2-(*N*-Methylcarbamoyl)cyclopentyl]adenine 5a. A mixture of 6-chloropurine **4a** (55 mg, 0.2 mmol) and 25% aqueous solution of NH_4OH (10 mL) was refluxed for 7 h. The solvent was removed in vacuo, and the residue was purified by FC (CH_2Cl_2–MeOH, 94:6) producing **5a** (39 mg, 76%) as a white solid: mp 270–272 °C; ^{1}H NMR (400 MHz, DMSO-d_6) δ: 1.69 (1H, m, 1H-4′), 1.89 (1H, m, 1H-3′), 2.04 (2H, m, 1H-3′, 1H-4′), 2.17 (1H, m, 1H-5′), 2.21 (3H, d, J = 4.6, CH_3), 2.34 (1H, m, 1H-5′), 3.02 (1H, m, H-2′), 5.07 (1H, q, J = 7.9, H-1′), 7.13 (2H, s, NH_2), 7.58 (1H, m, NH), 7.93 (1H, s, H-8), 8.13 (1H, s, H-2); ^{13}C NMR (100 MHz, DMSO-d_6) δ: 22.3 (C-4′), 25.2 (CH_3), 27.1 (C-3′), 30.6 (C-5′), 47.7 (C-2′), 56.4 (C-1′), 118.3 (C-5), 139.5 (C-8), 149.8 (C-4), 152.1 (C-2), 155.8 (C-6), 171.8 (CO); MS (CI) *m*/*z* %: 261 ([M+1]$^+$, 9), 260 (M$^+$, 56), 219 (15), 203 (13), 202 (M$^+$-C_2H_4NO, 90), 188 (M$^+$-C_3H_6NO, 16), 175 (M$^+$-C_4H_7NO, 10), 162 (M$^+$-C_5H_8NO, 36), 148 (M$^+$-$C_6H_{10}NO$, 5), 136 (M$^+$-$C_7H_{10}NO$, 77), 135 (M$^+$-$C_7H_{11}NO$, 100), 126 (M$^+$-$C_5H_4N_5$, 7), 108 (21), 95 (10), 69 (8), 67 (16); Anal. calcd for $C_{12}H_{16}N_6O$: C, 55.37; H, 6.20; N, 32.29. Found: C, 55.26; H, 6.32; N, 32.17.

4.1.6. (±)-*cis*-9-[2-(*N*-Methylcarbamoyl)-4-cyclopentenyl]adenine 5b. Prepared from **4b** (50 mg, 0.18 mmol) following the same procedure as for **5a** from **4a**. The residue was purified by FC (CH_2Cl_2–MeOH, 94:6)

producing **5b** (21 mg, 46%) as a white solid: mp 273–274 °C; ^{1}H NMR (400 MHz, DMSO-d_6) δ: 2.06 (3H, d, J = 4.6, CH_3), 2.50 (1H, m, 1H-3′), 2.95 (1H, m, 1H-3′), 3.41 (1H, m, H-2′), 5.82 (2H, m, H-5′, H-1′), 6.31 (1H, m, H-4′), 7.11 (2H, s, NH_2), 7.64 (1H, s, H-8), 7.77 (1H, m, NH), 8.14 (1H, s, H-2); ^{13}C NMR (100 MHz, DMSO-d_6) δ: 25.1 (CH_3), 33.6 (C-3′), 46.6 (C-2′), 59.8 (C-1′), 118.5 (C-5), 127.1 (C-5′), 137.4 (C-4′), 139.7 (C-8), 149.3 (C-4), 152.0 (C-2), 155.8 (C-6), 169.7 (CO); MS (EI) m/z %: 259 ([M+1]$^+$, 11), 258 (M$^+$, 69), 200 (M$^+$-C_2H_4NO, 36), 173 (15), 136 (M$^+$-C_7H_8NO, 36), 135 (M$^+$-C_7H_9NO, 100), 124 (M$^+$-$C_5H_4N_5$, 5), 108 (19), 67 (19), 66 (14); Anal. calcd for $C_{12}H_{14}N_6O$: C, 55.80; H, 5.46; N, 32.54. Found: C, 55.68; H, 5.57; N, 32.43.

4.1.7. General procedure for the preparation of N^6-cyclopropyladenines 6, 12. A solution of the appropriate 6-chloropurine (0.16 mmol) in EtOH (3.8 mL) was treated with cyclopropylamine (3.46 mmol). The reaction mixture was refluxed under an argon atmosphere for 1 h. The solvent was removed in vacuo, and the residue was purified by FC (CH_2Cl_2–MeOH, 97:3) affording the desired compounds **6a–b** and **12**.

4.1.7.1. (±)-*cis*-N^6-Cyclopropyl-9-[2-(*N*-methylcarbamoyl)cyclopentyl]adenine 6a. White solid (47 mg, 98%): mp 173–175 °C; ^{1}H NMR (400 MHz, DMSO-d_6) δ: 0.60 (2H, m, H-2″, H-3″), 0.71 (2H, m, H-2″, H-3″), 1.70 (1H, m, 1H-4′), 1.90 (1H, m, 1H-3′), 2.03 (2H, m, 1H-3′, 1H-4′), 2.17 (1H, m, 1H-5′), 2.21 (3H, d, J = 4.6, CH_3), 2.33 (1H, m, 1H-5′), 3.02 (2H, m, H-2′, H-1″), 5.08 (1H, q, J = 7.9, H-1′), 7.58 (1H, m, N*H*CH$_3$), 7.78 (1H, s, NH), 7.93 (1H, s, H-8), 8.23 (1H, s, H-2); ^{13}C NMR (100 MHz, DMSO-d_6) δ: 6.3 (C-2″, C-3″), 22.3 (C-4′), 23.9 (C-1″), 25.3 (CH_3), 27.1 (C-3′), 30.6 (C-5′), 47.8 (C-2′), 56.4 (C-1′), 118.7 (C-5), 139.3 (C-8), 149.5 (C-4), 152.0 (C-2), 155.4 (C-6), 171.8 (CO); MS (EI) m/z %: 301 ([M+1]$^+$, 12), 300 (M$^+$, 63), 299 ([M−1]$^+$, 15), 286 (M$^+$-CH_2, 17), 285 (M$^+$-CH_3, 100), 273 (M$^+$-C_2H_3, 10), 242 (M$^+$-C_2H_4NO, 5), 176 (M$^+$-$C_7H_{10}NO$, 11), 174 (M$^+$-$C_7H_{12}NO$, 32), 160 (M$^+$-$C_8H_{14}NO$, 40), 147 (M$^+$-$C_9H_{15}NO$, 12), 120 (M$^+$-$C_{10}H_{16}N_2O$, 7), 93 (5), 78 (6), 69 (9), 67 (10); Anal. calcd for $C_{15}H_{20}N_6O$: C, 59.98; H, 6.71; N, 27.98. Found: C, 59.82; H, 6.81; N, 27.85.

4.1.7.2. (±)-*cis*-N^6-Cyclopropyl-9-[2-(*N*-methylcarbamoyl)-4-cyclopentenyl]adenine 6b. White solid (47 mg, 99%): mp 173–174 °C; ^{1}H NMR (400 MHz, DMSO-d_6) δ: 0.60 (2H, m, 1H-2″, 1H-3″), 0.70 (2H, m, 1H-2″, 1H-3″), 2.07 (3H, d, J = 4.5, CH_3), 2.56–2.42 (1H, m, 1H-3′), 2.95 (1H, m, 1H-3′), 3.04 (1H, m, H-1″), 3.41 (1H, m, H-2′), 5.82 (2H, m, H-5′, H-1′), 6.31 (1H, m, H-4′), 7.64 (1H, s, H-8), 7.78 (2H, m, 2NH), 8.23 (1H, s, H-2); ^{13}C NMR (100 MHz, DMSO-d_6) δ: 6.3 (C-2″, C-3″), 23.8 (C-1″), 25.1 (CH_3), 33.6 (C-3′), 46.6 (C-2′), 59.8 (C-1′), 118.8 (C-5), 127.1 (C-5′), 137.4 (C-4′), 139.5 (C-8), 148.8 (C-4), 152.0 (C-2), 155.3 (C-6), 169.7 (CO); MS (CI) m/z %: 299 ([M+1]$^+$, 14), 298 (M$^+$, 69), 283 (M$^+$-CH_3, 52), 271 (M$^+$-C_2H_3, 11), 240 (M$^+$-C_2H_4NO, 6), 175 (M$^+$-C_7H_9NO, 63), 174 (M$^+$-$C_7H_{10}NO$, 100), 160 (M$^+$-$C_8H_{12}NO$, 99), 147 (M$^+$-$C_9H_{13}NO$, 22), 120 (M$^+$-$C_{10}H_{14}N_2O$, 15), 93 (13), 81 (14), 69 (29), 67 (28); Anal. calcd for $C_{15}H_{18}N_6O$: C, 60.39; H, 6.08; N, 28.17. Found: C, 60.49; H, 6.19; N, 28.32.

4.1.7.3. (±)-*cis*-N^6-Cyclopropyl-9-[(2-hydroxymethyl)-4-cyclopentenyl]adenine 12. White solid (43 mg, 99%): mp 145–146 °C; ^{1}H NMR (400 MHz, DMSO-d_6) δ: 0.61 (2H, m, 1H-2″, 1H-3″), 0.72 (2H, m, 1H-2″, 1H-3″), 2.43 (1H, m, 1H-3′), 2.54 (1H, m, 1H-3′), 2.71 (1H, sext, J = 7.5, H-2′), 2.92 (1H, m, 1H-6′), 2.98 (2H, m, 1H-6′, H-1″), 4.37 (1H, t, J = 4.8, OH), 5.63 (1H, m, H-1′), 5.90 (1H, m, H-5′), 6.29 (1H, m, H-4′), 7.83 (1H, s, H-8), 7.88 (1H, br s, NH), 8.24 (1H, s, H-2); ^{13}C NMR (100 MHz, DMSO-d_6) δ: 6.4 (C-2″, C-3″), 23.8 (C-1″), 34.9 (C-3′), 43.2 (C-2′), 59.4 (C-1′), 60.2 (C-6′), 119.2 (C-5), 128.3 (C-5′), 137.2 (C-4′), 139.7 (C-8), 149.1 (C-4), 152.2 (C-2), 155.5 (C-6); MS (CI) m/z %: 272 ([M+1]$^+$, 18), 271 (M$^+$, 93), 256 (30), 176 (M$^+$-C_6H_7O, 20), 175 (M$^+$-C_6H_8O, 100), 174 (M$^+$-C_6H_9O, 77), 160 (M$^+$-$C_7H_{11}O$, 85), 147 (M$^+$-$C_8H_{12}O$, 15), 134 (M$^+$-$C_9H_{13}O$, 7), 120 (M$^+$-$C_9H_{13}NO$, 15), 93 (8), 79 (6), 67 (10); Anal. calcd for $C_{14}H_{17}N_5O$: C, 61.98; H, 6.32; N, 25.81%. Found: C, 62.09; H, 6.20; N, 25.95.

4.1.8. General procedure for the preparation of N^6-cyclopentyladenines 7, 13. A solution of the appropriate 6-chloropurine (0.16 mmol) in EtOH (3.8 mL) was treated with cyclopentylamine (1.01 mmol). The reaction mixture was refluxed under an argon atmosphere for 2 h. The solvent was removed in vacuo, and the residue was purified by FC (CH_2Cl_2–MeOH, 97:3) affording the desired compounds **7a–b** and **13**.

4.1.8.1. (±)-*cis*-N^6-Cyclopentyl-9-[2-(N-methylcarbamoyl)cyclopentyl]adenine 7a. White solid (52 mg, 99%): mp 153–154 °C; ^{1}H NMR (400 MHz, DMSO-d_6) δ: 1.55 (4H, m, 1H-2″, 1H-3″, 1H-4″, 1H-5″), 1.71 (3H, m, 1H-4′, 1H-3″, 1H-4″), 1.92 (3H, m, 1H-3′, 1H-2″, 1H-5″), 2.04 (2H, m, 1H-3′, 1H-4′), 2.17 (1H, m, 1H-5′), 2.22 (3H, d, J = 4.6, CH_3), 2.33 (1H, m, 1H-5′), 3.02 (1H, m, H-2′), 4.54 (1H, br s, H-1″), 5.07 (1H, q, J = 7.8, H-1′), 7.52 (1H, br s, NH), 7.58 (1H, m, N*H*CH$_3$), 7.93 (1H, s, H-8), 8.18 (1H, s, H-2); ^{13}C NMR (100 MHz, DMSO-d_6) δ: 22.3 (C-4′), 23.5 (C-3″, C-4″), 25.3 (CH_3), 27.1 (C-3′), 30.6 (C-5′), 32.2 (C-2″, C-5″), 47.8 (C-2′), 51.5 (C-1″), 56.4 (C-1′), 118.4 (C-5), 139.1 (C-8), 149.1 (C-4), 152.0 (C-2), 154.1 (C-6), 171.8 (CO); MS (CI) m/z %: 329 ([M+1]$^+$, 12), 328 (M$^+$, 57), 299 (M$^+$-C_2H_5, 12), 287 (M$^+$-C_3H_5, 16), 270 (M$^+$-C_2H_4NO, 35), 260 (M$^+$-C_5H_8, 100), 230 (M$^+$-$C_6H_{12}N$, 15), 204 (M$^+$-$C_7H_{10}NO$, 16), 202 (M$^+$-$C_7H_{12}NO$, 26), 174 (M$^+$-$C_9H_{16}NO$, 21), 162 (M$^+$-$C_{10}H_{16}NO$, 18), 136 (M$^+$-$C_{12}H_{18}NO$, 24), 135 (M$^+$-$C_{12}H_{19}NO$, 48), 95 (7), 84 (5), 69 (12), 67 (13); Anal. calcd for $C_{17}H_{24}N_6O$: C, 62.17; H, 7.37; N, 25.59%. Found: C, 62.29; H, 7.18; N, 25.70.

4.1.8.2. (±)-*cis*-N^6-Cyclopentyl-9-[2-(*N*-methylcarbamoyl)-4-cyclopentenyl]adenine 7b. White solid (51 mg, 98%): mp 175–177 °C; ^{1}H NMR (400 MHz, DMSO-d_6) δ: 1.55 (4H, m, 1H-2″, 1H-3″, 1H-4″, 1H-5″), 1.71

(2H, m, 1H-3″, 1H-4″), 1.92 (2H, m, 1H-2″, 1H-5″), 2.07 (3H, d, $J = 4.6$, CH_3), 2.49 (1H, m, 1H-3′), 2.95 (1H, m, 1H-3′), 3.40 (1H, m, H-2′), 4.50 (1H, br s, H-1″), 5.81 (2H, m, H-5′, H-1′), 6.31 (1H, m, H-4′), 7.50 (1H, d, $J = 6.3$, NH), 7.63 (1H, s, H-8), 7.77 (1H, m, N*H*CH_3), 8.18 (1H, s, H-2); ^{13}C NMR (100 MHz, DMSO-d_6) δ: 23.4 (C-3″, C-4″), 25.1 (CH_3), 32.2 (C-2″, C-5″), 33.6 (C-3′), 46.6 (C-2′), 51.5 (C-1″), 59.7 (C-1′), 118.6 (C-5), 127.2 (C-5′), 137.4 (C-4′), 139.3 (C-8), 148.7 (C-4), 152.0 (C-2), 154.1 (C-6), 169.7 (CO); MS (CI) *m*/*z* %: 327 ([M+1]$^+$, 20), 326 (M$^+$, 75), 268 (M$^+$-C_2H_4NO, 20), 258 (M$^+$-C_5H_8, 53), 203 (M$^+$-C_7H_9NO, 60), 202 (M$^+$-$C_7H_{10}NO$, 100), 174 (M$^+$-$C_9H_{14}NO$, 23), 160 (M$^+$-$C_{10}H_{16}NO$, 18), 135 (M$^+$-$C_{12}H_{17}NO$, 95), 124 (M$^+$-$C_{10}H_{12}N_5$, 35), 93 (6), 84 (7), 67 (25), 66 (9); Anal. calcd for $C_{17}H_{22}N_6O$: C, 62.56; H, 6.79; N 25.75. Found: C 62.43; H, 6.83; N, 25.62.

4.1.8.3. (±)-*cis*-N^6-cyclopentyl-9-[(2-hydroxymethyl)-4-cyclopentenyl]adenine (13). White solid (47 mg, 99%): mp 110–112 °C; ^{1}H NMR (400 MHz, DMSO-d_6) δ: 1.58 (4H, m, 1H-2″, 1H-3″,1H-4″, 1H-5″), 1.71 (2H, m, 1H-3″, 1H-4″), 1.94 (2H, m, 1H-2″, 1H-5″), 2.43 (1H, m, 1H-3′), 2.54 (1H, m, 1H-3′), 2.71 (1H, sext, $J = 7.5$, H-2′), 2.93 (1H, m, 1H-6′), 3.01 (1H, m, 1H-6′), 4.35 (1H, t, $J = 4.8$, OH), 4.51 (1H, br s, C*H*NH), 5.62 (1H, m, H-1′), 5.89 (1H, m, H-5′), 6.29 (1H, m, H-4′), 7.58 (1H, d, $J = 7.7$, NH), 7.81 (1H, s, H-8), 8.19 (1H, s, H-2); ^{13}C NMR (100 MHz, DMSO-d_6) δ: 23.5 (C-3″, C-4″), 32.2 (C-2″, C-5″), 34.9 (C-3′), 43.2 (C-2′), 51.5 (C-1″), 59.4 (C-1′), 60.3 (C-6′), 119.0 (C-5), 128.3 (C-5′), 137.2 (C-4′), 139.4 (C-8), 149.0 (C-4), 152.3 (C-2), 154.2 (C-6); MS (CI) *m*/*z* %: 300 ([M+1]$^+$, 5), 299 (M$^+$, 17), 231 (12), 204 (M$^+$-C_6H_7O, 28), 203 (M$^+$-C_6H_8O, 30), 202 (M$^+$-C_6H_9O, 36), 175 (M$^+$-$C_8H_{12}O$, 12), 174 (M$^+$-$C_8H_{13}O$, 22), 162 (M$^+$-$C_9H_{13}O$, 16), 160 (M$^+$-$C_9H_{15}O$, 23), 136 (M$^+$-$C_{11}H_{15}O$, 23), 135 (M$^+$-$C_{11}H_{16}O$, 100), 120 (M$^+$-$C_{11}H_{17}NO$, 10), 108 (9), 84 (8), 79 (10), 67 (19); Anal. calcd for $C_{16}H_{21}N_5O$: C, 64.19; H, 7.07; N, 23.39. Found: C, 64.25; H, 6.90; N, 23.28.

4.1.9. General procedure for the preparation of N^6-benzyladenines 8. A solution of the appropriate 6-chloropurine (0.14 mmol) in EtOH (4 mL) was treated with benzylamine (0.18 mmol). The reaction mixture was refluxed under an argon atmosphere for 6 h. The solvent was removed in vacuo, and the residue was purified by FC (CH_2Cl_2-MeOH, 96:4) affording the desired compounds **8a–b**.

4.1.9.1. (±)-*cis*-N^6-Benzyl-9-[2-(*N*-methylcarbamoyl) cyclopentyl]adenine 8a. White solid (49 mg, 99%): mp 152–154 °C; ^{1}H NMR (500 MHz, 323K, DMSO-d_6) δ: 1.71 (1H, m, 1H-4′), 1.91 (1H, m, 1H-3′), 2.05 (2H, m, 1H-3′, 1H-4′), 2.19 (1H, m, 1H-5′), 2.22 (3H, d, $J = 4.6$, CH_3), 2.33 (1H, m, 1H-5′), 3.05 (1H, m, H-2′), 4.78 (2H, br s, C*H*$_2$NH), 5.10 (1H, m, H-1′), 7.20 (1H, m, Ar-H), 7.28 (2H, m, Ar-H), 7.34 (2H, m, Ar-H), 7.44 (1H, m, N*H*CH_3), 8.02 (2H, br s, H-8, NH), 8.20 (1H, s, H-2); ^{13}C NMR (125 MHz, 323 K, DMSO-d_6) δ: 22.3 (C-4′), 25.1 (CH_3), 27.1 (C-3′), 30.5 (C-5′), 42.9 (CH_2NH), 47.8 (C-2′), 56.6 (C-1′), 119.1 (C-5), 126.5, 127.1, 128.1, 139.8 (C-8), 140.0 (C-1″), 150.3 (C-4), 152.0 (C-2), 154.4 (C-6), 171.8 (CO); MS (CI) *m*/*z* %: 351 ([M+1]$^+$, 24), 350 (M$^+$, 100), 309 (5), 292 (M$^+$-C_2H_4NO, 19), 278 (M$^+$-C_3H_6NO, 5), 252 (M$^+$-C_5H_8NO, 13), 226 (M$^+$-$C_7H_{10}NO$, 24), 225 (M$^+$-$C_7H_{11}NO$, 31), 224 (M$^+$-$C_7H_{12}NO$, 33), 120 (M$^+$-$C_{14}H_{18}N_2O$, 9), 106 (26), 91 (32), 67 (7); Anal. calcd for $C_{19}H_{22}N_6O$: C, 65.12; H, 6.33; N, 23.98. Found: C, 65.23; H, 6.21; N, 24.05.

4.1.9.2. (±)-*cis*-N^6-Benzyl-9-[2-(*N*-methylcarbamoyl)-4-cyclopentenyl]adenine 8b. White solid (47 mg, 96%): mp 179–180 °C; ^{1}H NMR (400 MHz, DMSO-d_6) δ 2.06 (3H, d, $J = 4.6$, CH_3), 2.50 (1H, m, 1H-3′), 2.96 (1H, m, 1H-3′), 3.42 (1H, m, H-2′), 4.72 (2H, br s, C*H*$_2$NH), 5.84 (2H, m, H-5′, H-1′), 6.31 (1H, m, H-4′), 7.19 (1H, m, Ar-H), 7.29 (4H, m, Ar-H), 7.67 (1H, s, H-8), 7.76 (1H, m, N*H*CH_3), 8.19 (2H, s, H-2, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ: 25.1 (CH_3), 33.6 (C-3′), 42.9 (CH_2NH), 46.6 (C-2′), 59.9 (C-1′), 118.7 (C-5), 126.5, 127.1 (C-5′, Ar-C), 128.1, 137.4 (C-4′), 139.7 (C-8), 140.2 (C-1″), 148.8 (C-4), 152.0 (C-2), 154.2 (C-6), 169.7 (CO); MS (CI) *m*/*z* %: 349 ([M+1]$^+$, 18), 348 (M$^+$, 76), 290 (M$^+$-C_2H_4NO, 9), 225 (M$^+$-C_7H_9NO, 74), 224 (M$^+$-$C_7H_{10}NO$, 100), 120 (M$^+$-$C_{14}H_{16}N_2O$, 14), 106 (40), 91 (38), 69 (18), 67 (20); Anal. calcd for $C_{19}H_{20}N_6O$: C, 65.50; H, 5.79; N, 24.12. Found: C, 65.66; H, 5,62; N, 24.06.

4.1.10. General procedure for the preparation of N^6-(3-iodobenzyl)adenines 9, 14. A solution of the appropriate 6-chloropurine (0.14 mmol) in EtOH (4 mL) was treated with 3-iodobenzylamine hydrochloride (0.15 mmol) and Et_3N (0.04 mL). The reaction mixture was refluxed under an argon atmosphere for 34 h. The solvent was removed in vacuo, and the residue was purified by FC (hexane–ethyl acetate 1:20 → AcOEt) affording the desired compounds **9a–b** and **14**.

4.1.10.1. (±)-*cis*-N^6-(3-Iodobenzyl)-9-[2-(*N*-methylcarbamoyl)cyclopentyl]adenine 9a. Yellow solid (54 mg, 81%): mp 147–149 °C; ^{1}H NMR (500 MHz, 353 K, DMSO-d_6) δ: 1.72 (1H, m, 1H-4′), 1.93 (1H, m, 1H-3′), 2.07 (2H, m, 1H-3′, 1H-4′), 2.23 (1H, m, 1H-5′), 2.24 (3H, d, $J = 4.5$, CH_3), 2.35 (1H, m, 1H-5′), 3.16 (1H, m, H-2′), 4.76 (2H, s, C*H*$_2$NH), 5.13 (1H, m, H-1′), 7.09 (1H, t, $J = 7.7$, H-5″), 7.25 (1H, m, N*H*CH_3), 7.38 (1H, d, $J = 7.6$, H-6″), 7.57 (1H, d, $J = 7.7$, H-4″), 7.74 (1H, s, H-2″), 7.89, (1H, br s, H-8), 8.14 (1H, br s, NH), 8.22 (1H, s, H-2); ^{13}C NMR (125 MHz, 353 K, DMSO-d_6) δ: 22.0 (C-4′), 24.8 (CH_3), 27.0 (C-3′), 30.3 (C-5′), 42.6 (*C*H$_2$NH), 47.6 (C-2′), 56.4 (C-1′), 93.8 (C-3″), 119.6 (C-5), 126.4 (C-6″), 129.9 (C-5″), 135.0 (C-4″), 135.6 (C-2″), 140.0 (C-8), 142.6 (C-1″), 151.6 (C-2), 154.1 (C-6), 171.4 (CO); MS (EI) *m*/*z* %:477 ([M+1]$^+$, 23), 476 (M$^+$, 100), 418 (M$^+$-C_2H_4NO, 20), 378 (M$^+$-C_5H_8NO, 13), 352 (M$^+$-$C_7H_{10}NO$, 21), 351 (M$^+$-$C_7H_{11}NO$, 24), 350 (M$^+$-$C_7H_{12}NO$, 15), 232 (M$^+$-$C_{12}H_{14}N_5O$, 11), 224 (10), 217 (M$^+$-$C_{12}H_{15}N_6O$, 11), 120 (8), 90 (8), 67 (6); Anal. calcd for $C_{19}H_{21}IN_6O$: C, 47.91; H, 4.44; N, 16.64. Found: C, 47.82; H; 4.56; N, 16.52.

4.1.10.2. (±)-*cis*-N^6-(3-Iodobenzyl)-9-[2-(*N*-methylcarbamoyl)-4-cyclopentenyl]adenine 9b. Yellow solid (58 mg, 87%): mp 219–220 °C; ^{1}H NMR (400 MHz, DMSO-d_6) δ: 2.04 (3H, d, J = 4.3, CH_3), 2.50 (1H, m, 1H-3′), 2.96 (1H, m, 1H-3′), 3.39 (1H, m, H-2′), 4.64 (2H, br s, CH_2NH), 5.83 (2H, m, H-5′, H-1′), 6.31 (1H, m, H-4′), 7.10 (1H, t, J = 7.7, H-5″), 7.33 (1H, d, J = 7.6, H-6″), 7.57 (1H, d, J = 7.6, H-4″), 7.67 (1H, s, H-8), 7.68 (1H, s, H-2″), 7.78 (1H, m, NHCH$_3$), 8.20 (1H, s, H-2), 8.30 (1H, br s, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ: 25.1 (CH_3), 33.6 (C-3′), 42.2 (CH_2NH), 46.5 (C-2′), 59.9 (C-1′), 94.7 (C-3″), 118.8 (C-5), 126.6 (C-6″), 127.0 (C-5′), 130.4 (C-5″), 135.2 (C-4″), 135.6 (C-2″), 137.5 (C-4′), 139.9 (C-8), 143.0 (C-1″), 148.8 (C-4), 152.0 (C-2), 154.1 (C-6), 169.7 (CO); MS (EI) m/z %: 475 ([M+1]$^+$, 18), 474 (M$^+$, 81), 416 (M$^+$-C_2H_4NO, 7), 351 (M$^+$-C_7H_9NO, 59), 350 (M$^+$-$C_7H_{10}NO$, 52), 232 (M$^+$-$C_{12}H_{12}N_5O$, 15), 224 (16), 217 (M$^+$-$C_{12}H_{13}N_6O$, 10), 131 (100), 129 (71), 103 (48), 102 (48), 101 (43), 78 (87), 63 (85); Anal. calcd for $C_{19}H_{19}IN_6O$: C, 48.11; H, 4.04; N, 26.76. Found: C, 48.23; H, 3.95; N, 26.65.

4.1.10.3. (±)-*cis*-N^6-(3-Iodobencyl)-9-[(2-hydroxymethyl)-4-cyclopentenyl]adenine 14. Eluent hexane–ethyl acetate 1:6; yellow solid (58 mg, 92%): mp 154–156 °C; ^{1}H NMR (400 MHz, DMSO-d_6) δ: 2.45 (1H, m, 1H-3′), 2.54 (1H, m, 1H-3′), 2.72 (1H, sext, J = 7.5, H-2′), 2.98 (2H, m, H-6′), 4.33 (1H, br s, OH), 4.66 (2H, br s, CH_2NH), 5.64 (1H, m, H-1′), 5.90 (1H, m, H-5′), 6.29 (1H, m, H-4′), 7.11 (1H, t, J = 7.8, H-5″), 7.37 (1H, d, J = 7.8, H-6″), 7.58 (1H, d, J = 7.9, H-4″), 7.73 (1H, s, H-2″), 7.87 (1H, s, H-8), 8.21 (1H, s, H-2), 8.33 (1H, br s, NH); ^{13}C NMR (100 MHz, DMSO-d_6) δ: 34.8 (C-3′), 42.2 (CH_2NH), 43.1 (C-2′), 59.5 (C-1′), 60.2 (C-6′), 94.7 (C-3″), 119.1 (C-5), 126.7 (C-6″), 128.3 (C-5′), 130.4 (C-5″), 135.3 (C-4″), 135.8 (C-2″), 137.1 (C-4′), 140.0 (C-8), 143.0 (C-1″), 149.1 (C-4), 152.2 (C-2), 154.1 (C-6); MS (CI) m/z %: 448 ([M+1]$^+$, 12), 447 (M$^+$, 56), 352 (M$^+$-C_6H_7O, 47), 351 (M$^+$-C_6H_8O, 96), 350 (M$^+$-C_6H_9O, 48), 273 (27), 232 (M$^+$-$C_{11}H_{11}N_4O$, 35), 224 (35), 217 (M$^+$-$C_{12}H_{14}N_4O$, 26), 164 (87), 119 (M$^+$-$C_{13}H_{15}INO$, 56), 105 (M$^+$-$C_{11}H_{11}IN_4O$, 100), 91 (30), 77 (22), 63 (28); Anal. calcd for $C_{18}H_{18}IN_5O$: C, 48.34; H, 4.06; N, 15.66. Found: C, 48.45; H, 3.95; N, 15.53.

4.2. Discriminant and regression analysis

4.2.1. Data set. A data set of 167 adenosine analogues for which their affinity were reported in the literature was used in the present study.[30–35] This data set was split into active and inactive compounds for the discriminant analysis taking into account their affinity for A_1 receptor subtype. Analogues with K_i less than 10 μM were considered as actives, the others were considered as inactive compounds. Second, the regression analysis was carried out with the active compounds under the same rule. Both set of compounds were divided into training and predicting sets using cluster analysis. For the regression analysis, the logarithm of the activity reported was used in order to guarantee the linear distribution of the biological data. The structures of compounds, their observed and predicted biological activities and residual are provided as Supporting Information.

4.2.2. Molecular descriptors. Geometry optimization calculations were carried out for each compound of this study using the quantum chemical semi-empirical method AM1[58] included in MOPAC 6.0.[59] The energy of all compounds was minimized several times using the previous procedure to found a conformation with a reasonable minimal energy. The DRAGON[60] computer software was employed to calculate the molecular descriptors. Descriptors with constant values were discarded. For the remaining descriptors pairwise correlation analysis was performed. The following descriptors exclusion methods were used to reduce, in a first step, the colinearity and correlation between descriptors. However, certain colinearity among the descriptors used in this study remains. For this reason an orthogonalization process was carried out with the aim of reducing this colinearity.

4.2.3. Statistical methods. The statistical processing to obtain the QSAR models was carried out by using Linear discriminant, Genetic Algorithm (GA), and neural networks analysis using the Statistica 6.0 software.[39] GA is a class of methods based on biological evolution rules. The first step was to create a population of linear regression models. These regression models mate with each other, mutate, cross-over, reproduce, and then evolve through successive generations towards an optimum solution. The GA simulation conditions were 10,000 generations, number of cross-overs = 5000, smoothness factor = 1, mutation probability for adding new term = 50%, and 300 models populations. The models were linear combinations of eight descriptors. The GA procedure was repeated n-times to confirm that the selected descriptors were the most optimal descriptor set for describing the modeled property. Examining the regression coefficients, the standard deviations, the significances, and the number of variables in the equation determined the quality of the models. Analysis of residuals and deleted residuals from the regression equations was used to identify outliers. Artificial neural networks (ANN)[61,62] classification models were also developed using the variables included in the best LDA model founded. The ANN models were performed using five different network architectures (Three Layer Perceptrons (MLP3), Four Layer Perceptrons (MLP4), Linear Network Training, Probabilistic Neural Network (PNN), and Radial Basis Function (RBF)) included in the *intelligent problem solver* analysis implemented in the neural network module of the STATISTICA 6.0 software package.

4.3. Radioligand binding studies

Binding of [^{3}H]*R*-N^6-phenylisopropyladenosine ([^{3}H]*R*-PIA Amersham, Chicago, IL) to A_1 receptors from rat cerebral cortical membranes was performed as described previously.[63] Adenosine deaminase (3 U/mL) was present during the preparation of the brain membranes, in a preincubation of 30 min at 30 °C, and during the incubation with the radioligand.

Compounds were initially dissolved in DMSO and diluted with buffer to the final concentration where the amount of DMSO never exceeded 2%.

Incubations were terminated by rapid filtration over Whatman GF/B filters, using a Brandell cell harvester (Brandell, Gaithersburg, MD). The tubes were rinsed three times with 3 mL buffer each.

At least five different concentrations of competitor, spanning three orders of magnitude adjusted appropriately for the IC_{50} of each compound, were used. IC_{50} values, calculated with the nonlinear regression method implemented in the InPlot program (Graph-PAD, San Diego, CA), were converted to apparent K_i values using the Cheng–Prusoff equation[64] and K_d values of 1.0 nM.

Acknowledgments

We thank Dr. Kenneth A. Jacobson (NIDDK, NIH, Bethesda, MD) for helpful discussions and Dr. Neli Melman for assistance with the binding assays. We acknowledge the Universidad de Vigo and Xunta de Galicia (PGIDT01PX130114PR and PGIDIT04BTF301031PR) for financial support.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2007.11.026.

References and notes

1. Poulsen, S. A.; Quinn, R. J. *Bioorg. Med. Chem.* **1998**, *6*, 619.
2. Fredholm, B. B.; Arslan, G.; Halldner, L.; Kull, B.; Schulte, G.; Wasserman, W. *Naunyn Schmiedebergs Arch. Pharmacol.* **2000**, *362*, 364.
3. Fredholm, B. B.; IJzerman, A. P.; Jacobson, K. A.; Klotz, K. N.; Linden, J. *Pharmacol. Rev.* **2001**, *53*, 527.
4. Dhalla, A. K.; Shryock, J. C.; Shreeniwas, R.; Belardinelli, L. *Curr. Top. Med. Chem.* **2003**, *3*, 369.
5. Muller, C. E. *Curr. Med. Chem.* **2000**, *7*, 1269.
6. Kiec-Kononowicz, K.; Drabczynska, A.; Pekala, E.; Michalak, B.; Muller, C. E.; Schumacher, B.; Karolak-Wojciechowska, J.; Duddeck, H.; Rockitt, S.; Wartchow, R. *Pure Appl. Chem.* **2001**, *73*, 1411.
7. Soudijn, W.; van Wijngaarden, I.; Ijzerman, A. P. *Curr. Top. Med. Chem.* **2003**, *3*, 355.
8. González, M. P.; Teran, C. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3077.
9. González, M. P.; Teran, C.; Teijeira, M.; Besada, P.; Gonzalez-Moa, M. J. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3491.
10. González, M. P.; Teran, C.; Teijeira, M.; Gonzalez-Moa, M. J. *Eur. J. Med. Chem.* **2005**, *40*, 1080.
11. González-Diaz, H.; Marrero, Y.; Hernandez, I.; Bastida, I.; Tenorio, E.; Nasco, O.; Uriarte, E.; Castanedo, N.; Cabrera, M. A.; Aguila, E.; Marrero, O.; Morales, A.; González, M. P. *Chem. Res. Toxicol.* **2003**, *16*, 1318.
12. González, M. P.; Dias, L. C.; Morales, A. H.; Rodriguez, Y. M.; de Oliveira, L. G.; Gomez, L. T.; Diaz, H. G. *Bioorg. Med. Chem.* **2004**, *12*, 4467.
13. González, M. P.; Gonzalez Diaz, H.; Molina Ruiz, R.; Cabrera, M. A.; Ramos de Armas, R. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1192.
14. Morales, A. H.; Cabrera Perez, M. A.; González, M. P.; Ruiz, R. M.; Gonzalez-Diaz, H. *Bioorg. Med. Chem.* **2005**, *13*, 2477.
15. González, M. P.; Teran, C.; Fall, Y.; Teijeira, M.; Besada, P. *Bioorg. Med. Chem.* **2005**, *13*, 601.
16. Cabrera, M. A.; Bermejo, M.; González, M. P.; Ramos, R. *J. Pharm. Sci.* **2004**, *93*, 1701.
17. González, M. P.; Teran, C. *Bioorg. Med. Chem.* **2004**, *12*, 2985.
18. González, M. P.; Teran, C.; Teijeira, M.; Besada, P. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2641.
19. González, M. P.; Teran, C. *Bull. Math. Biol.* **2004**, *66*, 907.
20. González, M. P.; Teran, C.; Teijeira, M.; Morales, A. H. *Bull. Math. Biol.* **2007**, *69*, 347.
21. Morales, A. H.; Cabrera, M. A.; Combes, R. D.; González, M. P. *Curr. Comput. Aided Drug Des.* **2005**, *1*, 237.
22. González, M. P.; Terán, C.; Teijeira, M.; Morales, A. H. *Curr. Med. Chem.* **2006**, *13*, 2253.
23. González, M. P.; Caballero, J.; Tundidor-Camba, A.; Morales, A. H.; Fernandez, M. *Bioorg. Med. Chem.* **2006**, *14*, 200.
24. González, M. P.; Teran, C.; Teijeira, M.; Morales, A. H. *Eur. J. Med. Chem.* **2006**, *41*, 56.
25. González, M. P.; Puente, M.; Fall, Y.; Gomez, G. *Steroids* **2006**, *71*, 510.
26. Morales, A. H.; Cabrera Perez, M. A.; González, M. P. *J. Mol. Model. (Online)* **2006**, *12*, 769.
27. Saiz-Urra, L.; Gonzalez, M. P.; Teijeira, M. *Bioorg. Med. Chem.* **2006**, *14*, 7347.
28. Saiz-Urra, L.; Gonzalez, M. P.; Fall, Y.; Gomez, G. *Eur. J. Med. Chem.* **2007**, *42*, 64.
29. Saiz-Urra, L.; González, M. P.; Teijeira, M. *Bioorg. Med. Chem.* **2007**, *15*, 3565.
30. Baraldi, P. G.; Cacciari, B.; Pineda de Las Infantas, M. J.; Romagnoli, R.; Spalluto, G.; Volpini, R.; Costanzi, S.; Vittori, S.; Cristalli, G.; Melman, N.; Park, K. S.; Ji, X. D.; Jacobson, K. A. *J. Med. Chem.* **1998**, *41*, 3174.
31. Gallo-Rodriguez, C.; Ji, X. D.; Melman, N.; Siegman, B. D.; Sanders, L. H.; Orlina, J.; Fischer, B.; Pu, Q.; Olah, M. E.; van Galen, P. J.; Stiles, G. L.; Jacobson, K. A. *J. Med. Chem.* **1994**, *37*, 636.
32. Jacobson, K. A.; van Galen, P. J.; Williams, M. *J. Med. Chem.* **1992**, *35*, 407.
33. Kim, H. O.; Ji, X. D.; Siddiqi, S. M.; Olah, M. E.; Stiles, G. L.; Jacobson, K. A. *J. Med. Chem.* **1994**, *37*, 3614.
34. Lee, K.; Ravi, G.; Ji, X. D.; Marquez, V. E.; Jacobson, K. A. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1333.
35. Siddiqi, S. M.; Jacobson, K. A.; Esker, J. L.; Olah, M. E.; Ji, X. D.; Melman, N.; Tiwari, K. N.; Secrist, J. A., 3rd; Schneller, S. W.; Cristalli, G., et al. *J. Med. Chem.* **1995**, *38*, 1174.
36. González, M. P.; Morales, A. H.; Cabrera, M. A. *Bioorg. Med. Chem.* **2005**, *13*, 1775.
37. Gonzalez-Diaz, H.; Gia, O.; Uriarte, E.; Hernadez, I.; Ramos, R.; Chaviano, M.; Seijo, S.; Castillo, J. A.; Morales, L.; Santana, L.; Akpaloo, D.; Molina, E.; Cruz, M.; Torres, L. A.; Cabrera, M. A. *J. Mol. Model. (Online)* **2003**, *9*, 395.
38. Molina, E.; Diaz, H. G.; González, M. P.; Rodriguez, E.; Uriarte, E. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 515.
39. Statsoft, I., 2002.

40. Stewart, J.; Gill, L. *Econometrics*; Prentice Hall: London, 1998.
41. Dillon, W. R.; Goldstein, M. *Multivariate Analysis: Methods and Applications*; John Wiley and Sons Inc: NY, 1984.
42. Klein, D. J.; Randić, M.; Babić, D.; Lučić, B.; Nikolić, S.; Trinajstić, N. *Int. J. Quant. Chem.* **1991**, *63*, 215.
43. Lučić, B.; Nikolić, S.; Trinajstić, N.; Jurić, D. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 532.
44. Randić, M. *J. Mol. Struct. (Theochem)* **1991**, *233*, 45.
45. Randić, M. *New J. Chem.* **1991**, *15*, 517.
46. Randić, M. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 311.
47. Van Waterbeemd, H. *Chemometric Methods in Molecular Design*; Wiley-VCH: New York, 1995.
48. Hawkins, D. M. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1.
49. Akaike, H. IEEE Transactions on Automatic Control **1974,** AC-19, 716.
50. Akaike, H. *Information Theory and an Extension of the Maximum Likelihood Principle*; Akademiai Kiado: Budapest, 1973.
51. Kubinyi, H. *Quant. Struct.-Act. Relat.* **1994**, *13*, 285.
52. Kubinyi, H. *Quant. Struct.-Act. Relat.* **1994**, *13*, 393.
53. Lipnick, R. L. *Sci. Total Environ.* **1991**, *109–110*, 131.
54. Hou, T.; Xu, X. *Curr. Pharm. Des.* **2004**, *10*, 1011.
55. Toledo-Sherman, L. M.; Chen, D. *Curr. Opin. Drug Discov. Dev.* **2002**, *5*, 414.
56. Besada, P.; González-Moa, M. J.; Terán, C.; Santana, L.; Uriarte, E. *Synthesis* **2002**, *2002*, 2445.
57. Teijeira, M. Ph.D.Thesis, Universidad de Santiago de Compostela, 1996.
58. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
59. Frank, J. Seiler Research Laboratory, US Air Force Academy, Colorado Springs CO, 1993.
60. Todeschini, R.; Consonni, V.; Pavan, M. *DRAGON for windows (Software for Molecular Descriptor Calculations)*. Version 5.4; Talete srl: 2002.
61. Havel, J.; Madden, E.; Haddad, P. R. *Chromatographia* **1999**, *49*, 481.
62. Zupan, J.; Gasteiger, J. *Neural Networks in Chemistry and Drug Design: An Introduction*; Wiley-VCH: Weinheim, 1999.
63. Schwabe, U.; Trost, T. *Naunyn Schmiedebergs Arch. Pharmacol.* **1980**, *313*, 179.
64. Cheng, Y. C.; Prusoff, W. H. *Biochem. Pharmacol.* **1973**, *22*, 3099.