CrossMark

FULL-LENGTH PAPER

Latest QSAR study of adenosine A₂B receptor affinity of xanthines and deazaxanthines

Alfonso Pérez-Garrido[1,2] · Virginia Rivero-Buceta[3,4] · Gaspar Cano[5] · Sanjay Kumar[6] · Horacio Pérez-Sánchez[1] · Marta Teijeira Bautista[3,4]

Abstract Adenosine, a widespread and endogenous nucleoside that acts as a powerful neuromodulator in the nervous system, is a promising therapeutic target in a wide range of conditions. The structural similarity between xanthine derivatives and neurotransmitter adenosine has led to the derivatives of the heterocyclic ring being among the most abundant chemical classes of ligand antagonists of adenosine receptor subtypes. Small changes in the xanthine scaffold have resulted in a wide array of adenosine receptor antagonists. In this work, we developed a QSAR model for the A₂B subtype, which is, as yet, not well characterized, with two purposes in mind: to predict adenosine A₂B antagonist activity and to offer a substructural interpretation of this group of xanthines. The QSAR model provided good classifications of both the test and external sets. In addition, most of the contributions to adenosine A₂B receptor affinity derived by subfragmentation of the molecules in the training set agree with the relationships observed in the literature. These two factors mean that this QSAR ensemble could be used as a model to predict future adenosine A₂B antagonist candidates.

Keywords Xanthine · Adenosine A₂B · Antagonist · QSAR · Ensemble

✉ Alfonso Pérez-Garrido
Aperez@ucam.edu

1 Bioinformatics and High Performance Computing Research Group, Computer Science Department, Universidad Católica de Murcia, 30107 Guadalupe, Murcia, Spain

2 Departamento de Tecnología de la Alimentación y Nutrición, Universidad Católica de Murcia, 30107 Guadalupe, Murcia, Spain

3 Departamento de Química Orgánica, Universidad de Vigo, 36320 Vigo, Pontevedra, Spain

4 Instituto de Investigación Biomédica (IVIB), Universidade de Vigo, Vigo, Spain

5 Computing Technology Department, University of Alicante, Ap. 99, E03080 Alicante, Spain

6 Bioinformatics Centre, Biotech Park, Sector G. Janpakipuram, Lucknow 226021, India

Introduction

Xanthine, a common structural framework within both natural and synthetic compounds, is of broad therapeutic interest. This scaffold is often found in compounds with applications in pathologies, such as Alzheimer's disease [1], peripheral vascular disease [2], diabetes [3], asthma [4], Parkinson's disease [5] and cancer [6], among others (Fig. 1).

Caffeine is a natural alkaloid and is frequently used as a central nervous system stimulant. Other methylxanthine derivatives are used therapeutically as antiasthmatic, e.g. theophylline, a natural adverse compound and doxofylline, a new generation xanthine bronchodilator with no major cardiovascular effects. Allopurinol is used to treat gout and certain types of kidney stones. Furthermore, pentoxifylline is used to treat symptoms of intermittent claudication resulting from peripheral artery diseases. Recently, linagliptin was approved by the FDA (Food and Drug Administration) and the EMA (European Medicines Agency) for type 2 diabetes treatment.

Accordingly, the biological effects of xanthine derivatives cover a wide range of pharmacological targets. The most important of these targets are xanthine oxidase [7], cholinesterases [1], phosphodiesterases [8] and dipeptidyl
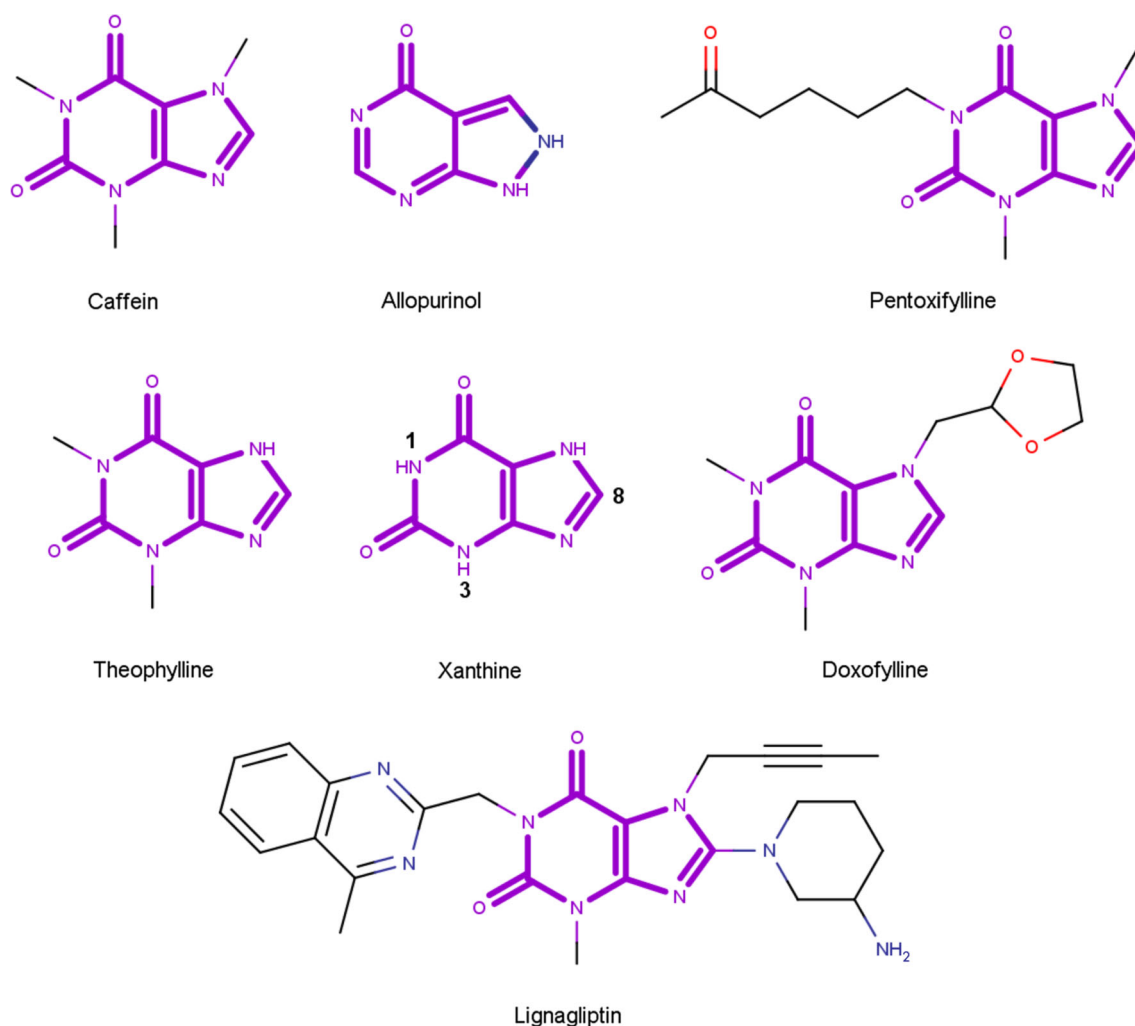
Fig. 1 Xanthine derivatives already used in a variety of medical therapies
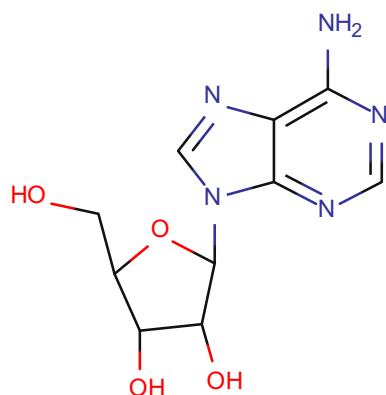

Fig. 2 Adenosine

peptidase 4 [9]; however, it should be emphasized that adenosine receptors (ARs) are those where more active ligands can be found [10].
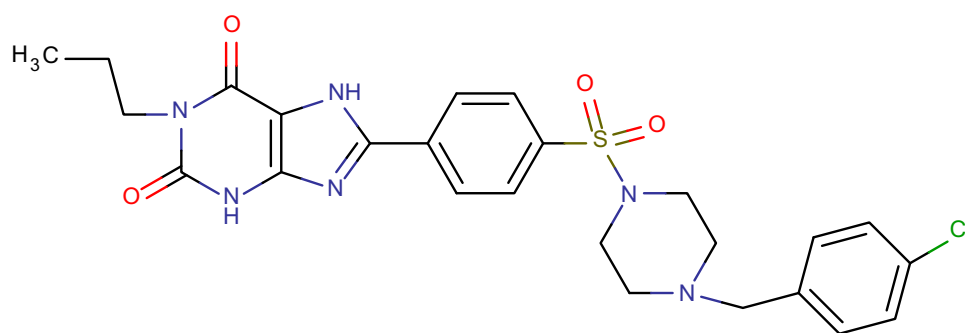
Adenosine (Fig. 2) is a widespread and endogenous nucleoside that acts as a powerful neuromodulator in the nervous system. It is also a known intermediate of several metabolic pathways, some of which are critical in maintaining ATP levels in all cell types, including neurons. Most of the physiological effects of adenosine are believed to be mediated by interaction with specific extracellular ARs, which have been classified into four subtypes: A_1, A_{2A}, A_{2B} and A_3 [11].

The structural similarity between xanthine derivatives and neurotransmitter adenosine has led to derivatives of the heterocyclic ring to be considered among the most abundant chemical classes of ligand antagonists of adenosine receptor subtypes. Small changes in the xanthine scaffold have resulted in a wide array of adenosine receptor antagonists, including some compounds with selectivity for a particular receptor subtype.

Among all adenosine receptor subtypes, there are few specific ligands of A_{2B} adenosine receptor subtype, perhaps because this subtype is the least well characterized. Furthermore, this subtype is defined as "low affinity" because it requires high micromolar concentrations of adenosine to be activated [12]. Nevertheless, A_{2B} AR regulates a number of pathological and physiological processes in vital organs,

Fig. 3 Xanthine derivative recently patented for the treatment of inflammatory diseases [20]

including lungs, kidneys, brain, eyes, bladder, liver, adipose and mast cells [12–17].

Therefore, the search for ligands with affinity for A_{2B} has been intense, the family of xanthine derivatives being considered of particular interest in this respect.

Among natural compounds, the adenosine A_{2B} antagonists, caffeine and theophylline, only show moderate binding affinities to this receptor. Nonetheless, modifications on the 1, 3 and 8 positions of xanthine with certain alkyl or aryl groups have given rise to many compounds with affinity for this receptor subtype [16,18,19] (Figs. 1, 3), and several of them have been patented for the treatment of inflammatory diseases [20].

Although a large number of xanthine derivative ligands have been synthesized and studied, there are still no guidelines available for the rational design of new potent and selective A_{2B} AR antagonists. To be able to prioritize the synthesis and testing of a compound, it is of great value to be able to predict the binding affinity of a given ligand-receptor. With this aim, quantitative structure-activity relationships (QSAR) methods are considered useful, and several studies on A_{2B} antagonists of ARs using this technique can be found in the literature [21–27]. Most QSAR models for the xanthine scaffold are based on regression analysis of active molecules or use "black box" techniques. In the first case, the substances used are usually the most active, while those substances which have very low activity are discarded. As a result, valuable information is lost and the applicability domain of the model is reduced. In the second case, it is not possible to obtain a relationship between activity and structure. The aim of this study was to obtain a QSAR model capable of predicting the A_{2B} antagonist activity of xanthines and to obtain an interpretable relationship between their activity and structure.

Materials and methods

Dataset

The data used were taken from the literature [28–37]. Among all of the binding data for specific targets, only measurements using human A_{2B} adenosine receptor subtype cloned in HEK-293 cells and [3H]DPCPX, such as radiolabeled ligands, were considered. A rigorous curation of structural data and elimination of questionable data points was performed. This included the removal of duplicates, the detection of valence violations, ring aromatization and the standardization of tautomeric forms. Finally, in order to obtain a reliable dataset, incomplete or unclear structures were not considered. Of 419 xanthine derivatives considered, only six structures were eliminated after the cleaning process. We used the remaining 413 chemicals. The output set of structures contains xanthines and deazaxanthines (Table S1 in Online Resource 1).

The compounds were first classified into two groups according to their Ki values. The first group, designated as potent or moderately potent antagonists, includes compounds with Ki < 100 nM (pKi > 7). The second group, designated weak antagonists, includes those compounds with Ki ≥ 100 nM (pKi ≤ 7). As a result of this process, 306 compounds were classified as potent or moderately potent antagonists and the remainder 107 compounds were classified as weak antagonists.

To evaluate the predictive accuracy of the model, the data were partitioned into three sets (training, test and external) by k-means cluster analysis (k-MCA). This technique is explained in the following sections.

Molecular descriptors

2D molecular descriptors available in DRAGON (version 6) [38] and MOE (version 2008.10) [39] were used in the present work. The descriptors were calculated for each molecule from the SMILES (Simplified Molecular Input Line Entry Specification) input of the chemical structures. Of all the calculated descriptors, those with constant or near constant values within each class were removed. After their removal, 406 DRAGON descriptors and 146 MOE descriptors remained.

K-means cluster analysis

The development of methods to select compounds for training and test sets is an active area of research, and many

methods are commonly applied, ranging from straightforward random selection [40] to various clustering techniques [41].

Using k-means (k-MCA), data partitions can be made that are significantly representative of the classes of compounds present in a database.

The number of variables used in k-MCA was reduced using principal component analysis (PCA) to maintain the maximum information with the minimum number of variables. A total of ten principal components for k-MCA was chosen. In PCA reduction, we used all previously standardized descriptors.

In order to avoid an unbalanced training set, the partitions were made according to the number of non-active ligands, following the ratio 60:20:20 for training, test and external sets, respectively.

Finally, the number of compounds in each group was training set 130 (65/65), test set 142 (121/21) and external set 141 (120/21). The numbers in parentheses represent the ratio of active to non-active ligands.

Variable selection

Linear discriminant analysis of two groups was used to develop the models. A wide range of molecular descriptors are currently available to model the activity of interest, so that the search for the most appropriate subset of variables is quite complicated and time consuming because of the many possible combinations. The variable selection method used to develop the QSAR models was the replacement method (RM) [42]. The Statistica package (v8.0) [43], in which we incorporated RM, was used to develop our QSAR model.

This technique begins by selecting a number of d descriptors before replacing each, one at a time, (the greatest relative error is in their coefficient) by another descriptor of the pool that minimizes the ROCFIT parameter. The process was repeated until the set of descriptors remains unchanged. It has been shown that this technique produces models with better statistical parameters than the Forward Stepwise procedure and similar to genetic algorithms [42]. The mathematical definition of the ROCFIT parameter has previously been described [44], but briefly, the parameter is based on the classification obtained and represented in a receiver operating characteristic (ROC) chart. On a ROC graph, the most precise and accurate classifier is one whose values are located in the upper left triangle, as close as possible to the corner. The ROCFIT parameter links the distances of the model presented in a ROC chart (Eq. 1) with the value of a perfect classifier (point 0,1). Thus, we see that better classifications for both the training and prediction sets are obtained at low values of this parameter:

$$\text{ROCFIT} = \frac{\text{ROCED}}{\text{FIT}(\lambda)} = \frac{(|d_1 - d_2| + 1)(d_1 + d_2)(d_2 + 1)}{\text{FIT}(\lambda)} \quad (1)$$

where the d_1 and d_2 represent the measures of Euclidean distances between the point of the classifier point and the perfect classifier point (0, 1) of the training and test set, respectively.

This parameter has been successfully used by several authors to evaluate the classification performance [45–48].

Model validation

The model performance was evaluated using two types of statistical tools: the so-called goodness of fit and goodness of prediction. The properties of fit of the model were those taken into account using the training set, while the predictive ability of the model for activities of compounds that were not used for training was chosen using the test and external sets.

A standards statistic such as the Mahalanobis distance (D^2), Wilks' lambda (λ), Fisher's statistic (F) and the corresponding p-level (p) as well as the percentage of correct classifications (accuracy, sensitivity and specificity) were taken as measurements of goodness of fit. The higher the value of the Mahalanobis distance, the greater the discriminatory power, whereas Wilks' lambda takes values in the range of zero (perfect discrimination) to one (no discrimination at all).

The ROCED and ROCFIT parameters were used mainly to evaluate the classification performance. Other common indices, such as the area under the ROC curve (AUC), sensitivity (Se), specificity (Sp) and accuracy (Ac), were also calculated for this purpose. All of these parameters were calculated for the training, test and external sets. The test and external sets were used to check the predictive ability of the model selected.

In addition, the leave-group-out (LGO) approach was conducted on the training set to assess the internal predictivity. In this approach, 20 % of the compounds of the training set are randomly omitted, and the statistical parameters are recalculated with the remaining substances. This process was conducted 300 times and the average values of the accuracy, sensitivity, specificity and ROCFIT for both training and test sets, as well as the mean Wilk's λ values (λ_{Cross}) and squared Mahalanobis distances (D^2_{Cross}) are reported. A low average λ_{Cross}, ROCFIT and high average squared Mahalanobis distances (D^2_{Cross}) in the bootstrap validation is indicative of a model's robustness.

In addition, a robustness test, Y-randomization, was conducted [49]. In this test, the response variable values of the training set are randomly exchanged among themselves 300 times and new models are generated for the new response variables. For a robustness assessment, we used the approach described by de Cerqueira et al. [50] with a slight modifica-

tion involving the R parameter that is used to evaluate the validity of the models in the Y-randomization process, which was calculated as follows:

$$R = 1 - F_{\text{rand}}$$

where F_{rand} is the ratio between the number of models built with randomization classes with high predictive values (values of CCR_{train} and $CCR_{\text{test}} \geq 0.7$ [50]) and the total number of models built with randomization classes (number of iterations performed). If $R \geq 0.9$ then the predictive models are considered reliable. CCR_{train} and CCR_{test} are the average between selectivity and specificity of the training and test sets, respectively. If $R < 0.9$, the QSAR model is not acceptable because the independent variables are randomly associated with the response variable.

The average of the parameters, sensitivity, specificity, accuracy, Wilk's lambda, etc., is reported for each model. All calculations were performed with the STATISTICA software [43].

In summary, the good overall quality of a model is indicated by small values of ROCFIT, $ROCFIT_{\text{Cross}}$, λ and λ_{Cross}, along with high values of AUC, $FIT(\lambda)$, D^2 and F.

After optimizing the models using RM, the best and different models were submitted to an ensemble modelling process. To this end, we first selected the models with the best performance and ordered them from the lowest to highest ROCFIT values. Then, the degree of similarity/diversity among the final models was checked through the Canonical Measure of Distance (CMD) including the chemicals of the test set [51]. The distance matrix containing the CMD index values for all the pairs of models was represented in a multidimensional scaling (MDS) [52] to visualize the similarity/diversity of the same

Ensemble model

In order to improve the predictions and include all chemical spaces, we constructed an ensemble model based on a series of combinations of the best and most diverse individual QSAR models. Combinations were performed in three ways [53]:

1. Majority vote (unanimity, simple majority and plurality): using two classes (active and non-active ligands), an accurate class label is given by the ensemble if (i) all classifier are in agreement (unanimity) (ii) at least 50 percent of the classifiers have +1 values (simple majority) or (iii) most of the classifiers (plurality) give correct answers [53].
2. Weighted majority vote: The classification of each model is weighted with its individual accuracy. The ensemble model then collects weighted votes from all the individual

models and gives the prediction that has a higher vote [53].

3. Naive bayes (NB): a naïve Bayesian classifier is a structural model based on conditional probabilities assuming that all molecular descriptors (attributes) are fully independent, given the class. The following equation is used to estimate the probability of a test compound x belonging to a class c:

$$P(c \,|x\,)_{\text{NB}} = P(c) \prod_{j=1}^{m} P(a_j \,|c) \qquad (2)$$

where m is the number of attributes, a_j is the jth attribute value of x, the prior probability $P(c)$ and the conditional probability $P(a_j|c)$. To construct naive Bayes classifiers, it is simply necessary to estimate the probability values of $P(c)$ and $P(a_j|c)$, $j = 1, 2, \ldots, $ m, from the training data [53].

All the possible combinations between models were made and the best ones were chosen based on their goodness of fit and goodness of prediction.

Applicability domain of the models

The interest of a QSAR model lies in its ability to accurately predict the activity for new chemicals. Typically, a QSAR model will be able to make predictions for compounds that are within an applicability domain (AD) defined by the molecular descriptors space used to derive the model. This space has been defined in various ways to assess the AD in QSAR models [54]. The method used in this work is the most common and it is based on the leverage values (h) for each compound [47,48,55–58]. The AD was set inside a squared area within $\pm x$ standard deviations and a leverage threshold, h*. The x values are often fixed in $x = 2$ or 3 and the h* value is generally established at $3\kappa/n$, where n is the number of training chemicals and κ the number of model parameters plus one. Only those chemicals which are within the AD of the model should be used to predict antagonist Adenosine A_{2B} activity [55].

Structural interpretation

The structural interpretation was carried out in two ways: (i) more generally, based on the coefficients of descriptors included in the ensemble model and (ii) more particularly, by computing the contributions to the activity of some substructures based on a work recently published by Polishchuk et al. [59] adapted to a classification problem.

The interpretation workflow consisted of the following steps: (i) remove the chosen fragment from the relevant

molecule, (ii) calculate the descriptors relevant to the rest of the molecule, (iii) obtain the activity value predicted by the model (taking into account AD), (iv) estimate the fragment's contribution as the difference between predicted value of the molecule and predicted values of the fragments obtained in step (iii) (Eq. 3), and (v) repeat steps (i)–(iv) for all the molecules containing the chosen fragment and for all the models included in the ensemble.

$$P_{AB}(A) = P(AB) - P(B), \qquad (3)$$

where $P_{AB}(A)$ is the local contribution of moiety A to the activity value of the molecule AB, $P(AB)$ is the predicted value of the molecule AB containing the fragment A, $P(B)$ is the predicted activity value for molecular part B of the compound AB remaining after removal of the fragment A. Once the contribution for each fragment inside the applicability domain of the corresponding model had been established, we classified their contribution as positive or negative if their relative frequency was higher than 50 %. Instant JChem version 6.2.1 [60] was used for molecular decomposition and removal of fragments.

Results and discussion

QSAR model

Following the procedure outlined above, a series of models was developed. As a consequence, several models were available for each family of descriptors. To compare models, the CMD was calculated between pairs of models based on their molecular descriptors [51]. Based on the CMD, the most different models were chosen in such a way that they preserved the most information and diversity. Table 1 shows the best (based on ROCFIT values) and the most diverse models derived for each family of descriptors. The descriptors used in each model are included in Table S3 of Online Resource 1.

It can be observed that DRAGON descriptors provide models with a higher predictive accuracy than those provided by MOE.

Moreover, the results of the Y-randomization and cross-validation test (Table S2 in Online Resource 1) suggest that DRAGON models provide a robust quantitative structure-activity classification (ROCFIT values below 2.5 for cross validation and high ROCFIT values for y-randomization).

In drug discovery, one of the principal applications of QSAR models would be to predict the receptor affinity of ligands. To assess the predictive accuracy of the QSAR model, an additional test was conducted using the external test set. This resulted in an activity classification of around 75–80 % (Table 1), underlining the predictive accuracy of the model.

An evaluation of the individual models shows model D1 to be better than the others, both as regards the goodness of fit and goodness of prediction for the test (80 %) and external (82.2 %) sets, respectively. Equation 4 represents this model, along with the standardized descriptors used.

$$\begin{aligned} A = {} & 1.25 \cdot \text{SpPosA\_X} - 2.55 \cdot \text{J\_D/Dt} \\ & + 4.19 \cdot \text{SpDiam\_B(v)} - 5.78 \cdot \text{SpDiam\_B(e)} \\ & - 0.82 \cdot \text{MATS4e} - 0.97 \cdot \text{GATS2m} + 3.31 \cdot \text{GGI6} \\ & - 2.43 \cdot \text{GGI9} + 5.85 \cdot \text{SM(05)}_{\text{EA(bo)}} \\ & - 4.14 \cdot \text{SM(04)}_{\text{EA(ri)}} \end{aligned} \qquad (4)$$

Both, the equation and the variables have good statistical significance. The statistical significance of the equation is above 0.05 %, while all the variables show a statistical significance of above 0.02 %.

Ensemble model predictions of Adenosine A$_{2B}$ receptor affinity

When a multidimensional scaling (MDS) was applied (Fig. 4) to the distance matrix containing the CMD index values for all the model pairs, two distinct groups could be identified for each family of descriptors (MOE or DRAGON), which is an indication that each family of descriptors modelled different characteristics of the activity under study. This led us to improve the accuracy of the predictions obtained by the individual models through multi-classifiers using three methods: majority vote, weighted majority vote and Naive Bayes.

The results of these combinations are shown in Table 2. To evaluate the prediction accuracy of these combinatorial QSAR approaches, the test set and an external set were used.

Out of all the combinations, the accuracy values obtained using the unanimity majority vote approach were very low, as might be expected since all the predictions obtained when very different models are combined cannot be expected to coincide.

For other methods, the results were better than those obtained for the best individual model, D1. In order to assess Adenosine A$_{2B}$ receptor affinity, we chose the combination of models that had the best accuracy in most methodologies, namely models M8, D1 and D4.

Structural interpretation of the Adenosine A$_{2B}$ receptor affinity

In a more general interpretation, we used the descriptors of the three models included in the ensemble model, focusing our interpretation on the sign of the standardized coefficients of the descriptors that produce a variation in the Wilks'

Table 1 QSAR models selected based on CMD and ROCFIT[a]

Name	MD	Size	TRAINING						TEST								
			Se	Sp	Ac	λ	AUC	D^2	Se	Sp	Ac	Se^b	Sp^b	Ac^b	E	ROCFIT	AUC
M1	MOE	8	90.77	75.38	83.08	0.58	0.88	2.87	76.86	80.95	77.46	76.86	80.95	77.46	0	1.67	0.80
M2	MOE	8	87.69	78.46	83.08	0.55	0.89	3.21	71.07	85.71	73.24	71.19	85.71	73.38	3	1.68	0.81
M3	MOE	6	87.69	73.85	80.77	0.59	0.87	2.74	76.03	76.19	76.06	77.97	76.19	77.70	3	1.71	0.79
M4	MOE	4	81.54	73.85	77.69	0.64	0.85	2.21	74.38	80.95	75.35	76.27	80.95	76.98	3	1.75	0.82
M5	MOE	12	90.77	81.54	86.15	0.49	0.92	4.13	75.21	80.95	76.06	74.79	80.95	75.71	2	1.76	0.81
M6	MOE	7	81.54	76.92	79.23	0.57	0.88	3.00	75.21	76.19	75.35	76.27	76.19	76.26	3	1.77	0.79
M7	MOE	5	86.15	70.77	78.46	0.63	0.84	2.30	74.38	80.95	75.35	74.38	80.95	75.35	0	1.82	0.79
M8	MOE	6	86.15	73.85	80.00	0.63	0.86	2.32	71.07	85.71	73.24	70.83	90.00	73.57	2	1.97	0.82
D1	DRAGON	10	89.23	83.08	86.15	0.46	0.93	4.67	78.51	90.48	80.28	77.78	90.48	79.71	4	1.02	0.86
D2	DRAGON	10	86.15	83.08	84.62	0.48	0.93	4.24	78.51	90.48	80.28	78.15	90.48	80.00	2	1.09	0.86
D3	DRAGON	10	83.08	84.62	83.85	0.48	0.92	4.34	78.51	90.48	80.28	77.97	90.48	79.86	3	1.10	0.87
D4	DRAGON	8	84.62	80.00	82.31	0.55	0.89	3.28	79.34	85.71	80.28	79.17	85.71	80.14	1	1.24	0.84
D5	DRAGON	5	86.15	84.62	85.38	0.58	0.88	2.86	75.21	80.95	76.06	75.21	80.95	76.06	0	1.30	0.82
D6	DRAGON	7	80.00	86.15	83.08	0.53	0.90	3.50	76.86	85.71	78.17	75.65	85.71	77.21	6	1.30	0.87

Name	MD	Size	EXTERNAL								
			Se	Sp	Ac	Se^b	Sp^b	Ac^b	E	ROCFIT	AUC
M1	MOE	8	75.00	76.19	75.18	75.42	80.00	76.09	3	1.99	0.81
M2	MOE	8	72.50	80.95	73.76	73.45	80.00	74.44	8	1.99	0.82
M3	MOE	6	75.83	71.43	75.18	77.39	70.00	76.30	6	2.22	0.79
M4	MOE	4	81.67	71.43	80.14	82.05	70.00	80.29	4	2.19	0.83
M5	MOE	12	75.00	61.90	73.05	75.00	65.00	73.53	5	2.80	0.74
M6	MOE	7	75.00	80.95	75.89	76.52	80.00	77.04	6	1.81	0.82
M7	MOE	5	80.83	66.67	78.72	81.03	66.67	78.83	4	2.31	0.74
M8	MOE	6	79.17	71.43	78.01	80.51	71.43	79.14	2	2.13	0.81
D1	DRAGON	10	83.33	76.19	82.27	84.35	76.19	83.09	5	1.19	0.76
D2	DRAGON	10	75.00	66.67	73.76	75.21	65.00	73.72	4	2.24	0.75
D3	DRAGON	10	82.50	76.19	81.56	82.35	75.00	81.29	2	1.48	0.76
D4	DRAGON	8	80.83	80.95	80.85	80.51	80.95	80.58	2	1.36	0.84
D5	DRAGON	5	75.83	71.43	75.18	76.47	71.43	75.71	1	1.60	0.78
D6	DRAGON	7	80.83	76.19	80.14	81.42	80.00	81.20	8	1.38	0.77

[a] MD: the program used to calculate the Molecular Descriptors, ROCFIT: Receiver Operating Characteristic graph Euclidean Distance corrected with FIT (λ), λ: Wilk's lambda, AUC: area under the ROC curve Se: sensitivity (true positive rate), Sp: specificity (true negative rate), Ac: accuracy, E: number of chemical outside Applicability Domain
[b] Refers to the statistics but considering the applicability domain of the model

Fig. 4 Multidimensional scaling (MDS) scatter plot of the 14 final models from Table 1
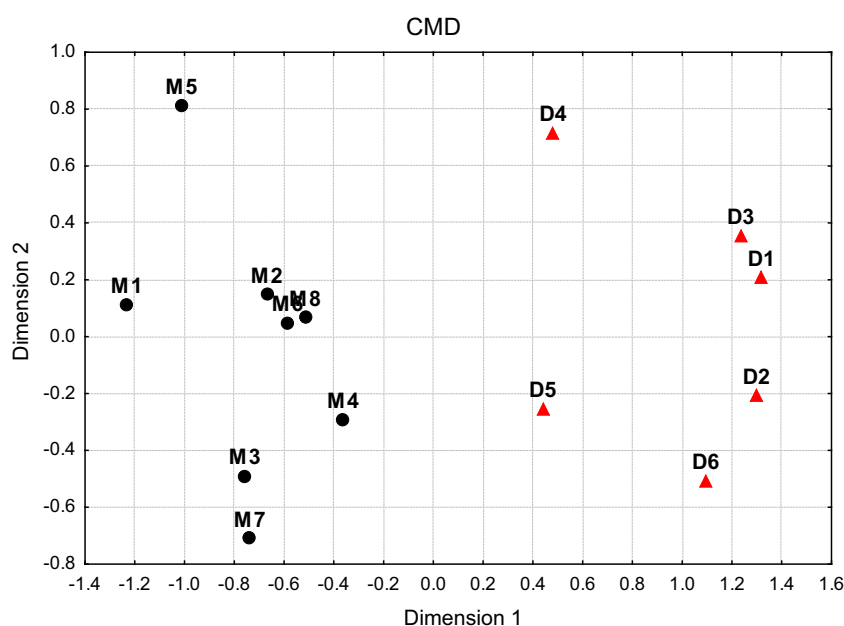

Table 2 Multi-classifier accuracy results using different subsets of single classifiers and different output combinations

Nos.	Combination	Single classifier results		Ensemble QSAR model									
				Majority vote						Weighted majority vote		Naive Bayes	
		Max results[a]		Simply		Plurality		Unanimity					
		Test	External	Test	External	Test	External	Test	External	Test	External	Test	External
1	M5, M8, D1, D4, D5	80.3	82.3	83.8	84.4	83.8	84.4	46.5	47.5	83.8	85.1	83.8	85.1
2	M2, M7, D1, D3, D4	80.3	82.3	83.1	84.4	83.1	85.8	52.8	50.4	83.1	85.8	83.1	86.5
3	M2, M5, D1, D4, D5	80.3	82.3	83.1	85.1	83.1	85.1	47.2	46.8	83.1	85.8	83.1	85.8
4	M7, M8, D1, D3, D4	80.3	82.3	83.1	84.4	83.1	85.8	52.1	53.2	83.1	85.8	83.1	86.5
5	M3, M8, D1, D3, D4	80.3	82.3	83.1	84.4	83.1	85.8	52.1	51.1	83.1	85.8	83.1	85.8
6	M3, M4, M5, D3, D4	80.3	81.6	83.1	83.7	84.5	84.4	48.6	48.2	84.5	83.7	84.5	84.4
7	M8, D1, D4	80.3	82.3	83.1	84.4	83.1	84.4	57.7	59.6	83.1	85.8	83.1	87.2

[a] It is the best accuracy obtained from the single classifiers that combined to form the multi-classifier

lambda higher than 30 % (Table 3). Through the sign of the coefficients in the ensemble model, descriptors such as SMR_VSA2, balabanJ and J_D/Dt were found to have a negative coefficient in the Adenosine A_{2B} receptor affinity model, while descriptors SM05_EA(bo), SpDiam_B(v), GGI6, SM04_EA(bo) and ChiA_X had a positive coefficient. The most significant descriptors—SM04_EA (bo), SM05_EA (bo), Chi_X, balabanJ and J_D/Dt—are related with topological factors such as shape, volume, the bonding order, the degree of branching and many other factors. If we relate the variation in the descriptor values with the structures of the training set, several things are evident. In the first place, an increase in number of cycles and length of the molecule increases the Adenosine A_{2B} receptor affinity of substances due to an increase in the descriptor values of

SM04_EA (bo), SM05_EA (bo) and Chi_X and a decrease in the values of balabanJ and J_D/Dt, which contribute negatively to the modelled activity. In the second place, the last two descriptor values (balabanJ and J_D/Dt) are also related to the degree of branching of the molecules, which, in turn, is related with an increase in volume. Probably, this increase in volume produces a steric hindrance between the molecule and the protein pocket.

As can be seen, the model related an increase in GGI6 values with an increase in the Adenosine A_{2B} receptor affinity due to its positive coefficient. This descriptor reflects the charge distribution of a molecule and could be related to the electrostatic interaction between ligand and receptor. This kind of descriptor has been mentioned by other authors in drug design, when it was related with the electronic fac-

Table 3 List of descriptors with the greatest influence on the values of Wilks' lambda

	M8			D1			D4			Definition
	Coeff. Std.	$\Delta\lambda^a$	$\%\lambda^b$	Coeff. Std.	$\Delta\lambda^a$	$\%\lambda^b$	Coeff.	$\Delta\lambda^a$	$\%\lambda^b$	
SMR_VSA2	−1.03	0.034	9.3							The sum of the accessible van der Waals surface area (in Å2) Over all atoms i such that Molar refractivity is in (0.26, 0.35]
balabanJ	−1.04	0.260	70.1							Balaban's connectivity topological index.
SM05_EA(bo)				5.85	0.012	2.2				Spectral moment of order 5 from augmented edge adjacency mat. weighted by bond order
SpDiam_B(v)				4.19	0.102	18.8				Spectral diameter from Burden matrix weighted by van der Waals volume
GGI6				3.31	0.093	17.1				Topological charge index of order 6
J_D/Dt				−2.55	0.167	30.7				Balaban-like index from distance/detour matrix
SM04_EA(bo)							3.21	0.319	70.2	Spectral moment of order 4 from augmented edge adjacency mat. weighted by bond order
ChiA_X							0.58	0.024	5.4	Average Randic-like index from chi matrix

[a] Variation in lambda value due to the introduction of descriptor in the model
[b] Percentage of the variation in lambda

tors that driving the binding interactions with the receptor [61,62]. For the other two adenosine subtypes, A$_1$ and A$_3$, it appears that electronic factors are an important property in the ligand-receptor interaction [63–66]. For the A$_{2B}$ subtype, electronic factors do not seem so important judging by the minor influence of the GGI6 descriptor (Table 3). In addition, the results obtained by other authors for the A$_2$ subtype show a strong dependence on hydrophobic rather than electronic factors [67]. In contrast, pharmacophoric models proposed by others authors using 140 diverse A$_{2B}$ antagonists identified four important regions: a hydrophobic-aliphatic, a hydrogen bond acceptor and two aromatics regions [68]. Since in our model all molecules possess a xanthine ring, electronic interactions driven by aromatic ring are not a distinguishing feature, for which reason the model probably does not sufficiently reflect this kind of interaction.

Through these results, it can be concluded that the affinity of xanthine moieties for the adenosine A$_{2B}$ receptor is related with topological factors (probably related with the hydrophobicity), steric hindrance and, to a lesser extent, electronic factors (hydrogen bond, aromatic interactions, etc). In summary, in this model, long and narrow molecules positively contribute to the affinity of the A$_{2B}$ receptor, which suggests that the pocket of the protein will share the same characteristics.

The binding contribution studied by the ensemble model can be seen in Tables 4 and 5. For every fragment, the values (positive or negative) whose frequency exceeds 50 % of the total data (in Tables 4 and 5) with a confidence level greater than 95 % in the hypothesis testing are presented in bold. With the family of descriptors used, it was impossible to calculate the descriptors for molecules containing more than one isolated fragment and only the contributions of the fragments at the end of the molecules could be calculated. For this reason, it was not possible to obtain contributions for internal zones of the molecules.

In the xanthine-based antagonist, the presence of alkyl groups in position 3, for example propyl, produced a negative effect on Adenosine A$_{2B}$ affinity, while in position 1 the binding affinity was improved (Table 4). The same negative effect was observed for a deazaxanthine-based antagonist. This result agrees with the effect indicated by Kalla et al. [16] in a differentiation study of the substitution at the N-1 and N-3 positions of the xanthine core. Similar results for subtype A$_2$ were also obtained by Doichinova et al. [67] for a group of substituted xanthines. Probably, the formation of hydrogen bonds in position 3 of the xanthine ring with the amino acid residues of the protein pocket is an important feature in the binding affinity. Furthermore, in position 1, hydrophobic interactions appear to be more important.

In their study of 1,3-dialkyl-8-(hetero)aryl-9-OH-9-deazaxanthines, Stefanachi et al. [30] observed that deazaxanthines with hydroxyl groups in position 9 were more potent than those unsubstituted in position 7 and 9, which, in turn, were more potent than deazaxanthines with hydroxyl groups in position 7. In Table 4, a clear decrease in the affinity was observed when these hydroxyl groups were in position 7, which agrees with the relationship obtained by the above authors. This hydroxyl group produces a decrease in hydrophobicity and probably a decrease in the interactions with the protein pocket. However, there are no conclusive results regarding substitution in position 9 by –OH groups.

Although other researchers [30] observed a dramatic decrease in activity due to methylation at the N7 position in a deazaxanthine-based antagonist, the ensemble model is inconclusive in this respect. Rather, the same moiety, methyl group, in xanthine-based antagonists produced an improved affinity. The same was observed by other authors with xanthine derivatives [35,36]. The presence of methyl groups in position 7 promotes receptor affinity due to hydrophobic interactions.

Furthermore, the chloride group in position 9 (Table 4) produced a decrease in the affinity of the deazaxanthine-based antagonist. This same result was reported by other authors [33], who observed a decrease in activity of between 0.1 and 1.2 log units, independently of the type of halogen used.

These two effects in the N7 and C9 position may be due to the loss of aromaticity and hydrophobicity caused by the substituents preventing aromatic pi-pi or hydrophobic interactions that occur between the pyrrol ring and the amino acid residues of the protein pocket.
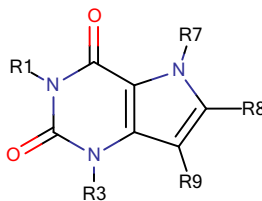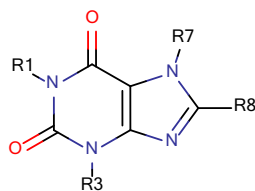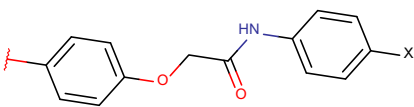
Moreover, as Table 5 shows, the introduction of furan-2-ylmethyl in position 3 of xanthine-based derivatives is beneficial for A$_{2B}$ affinity, as seen in the results obtained by Rodriguez-Borges et al. [36]. In this position, aliphatic groups have an opposite effect, presumably because the interactions are preferably aromatic in this zone of the protein pocket.

In general terms, it was seen that the presence of aromatic groups in position 8 in both xanthines as in deazaxanthines leads to an increase in affinity. Due to the small number of fragments studied, the significance of this observation cannot be assessed. Although the dataset contains a high number of aryl groups in this position, very few of them are in the optimal condition to carry out the correct fragmentation with the descriptors used.

The introduction of a methoxy group in either the meta- or ortho-position of the 8-phenyl ring (Table 5) in deazaxanthines yielded a lower affinity, as was also observed by Carotti et al. [33]. Such a decrease in activity may be due to the methoxy moiety in this position having a steric effect in the narrow cavity of A$_{2B}$.

The same authors also observed an increase in affinity in the presence of electron-withdrawing groups in the para

Table 4 Contributions of fragments in the final ensemble QSAR model

		n^a	P^b	N^c	$p$value^d	n^a	P^b	N^c	$p$value^d
R1	$-CH_2CH_2(CH_3)_2$	–	–	–	–	5	2	3	–
	$-CH_3$	121	66	55	–	–	–	–	–
	$-nC_3H_7$	**109**	**71**	**38**	**<0.05**	70	35	35	–
	$-CH_2CH_2OCH_3$	3	2	1	–	–	–	–	–
	$-nC_2H_5$	–	–	–	–	23	13	10	–
R3	$-CH_3$	**122**	**10**	**112**	**<0.05**	–	–	–	–
	$-nC_3H_7$	**104**	**22**	**82**	**<0.05**	64	8	56	<0.05
	$-CH_2CH_2OCH_3$	3	1	2	–	–	–	–	–
	$-CH_2CH_2CH_2OCH_3$	9	0	9	–	**14**	**2**	**12**	**<0.05**
	$-CH_2CH_2(CH_3)_2$	–	–	–	–	5	0	5	–
	Furan-2-ylmethyl	–	–	–	–	**28**	**21**	**7**	**<0.05**
R7	$-CH_3$	6	3	3	–	**21**	**13**	**8**	**<0.05**
	$-OH$	**14**	**0**	**14**	**0.05**	–	–	–	–
R8	Furan-2-ylmethyl	–	–	–	–	12	8	4	–
	Thiophen-2-ylmethyl	–	–	–	–	9	4	5	–
	$-CH_2Ph$	–	–	–	–	9	5	4	–
	Thiophen-2–yl	2	2	0	–	–	–	–	–
	Thiophen-3-yl	1	1	0	–	–	–	–	–
	Phenyl	7	5	2	–	–	–	–	–
	Furan-2-yl	1	1	0	–	2	2	0	–
	4-Aminophenyl	3	2	1	–	–	–	–	–
	4-Methylphenyl	5	4	1	–	–	–	–	–
	[structure: X=Halogen]	**18**	**16**	**2**	**<0.05**	–	–	–	–
	4-(Carboxymethoxy)phenyl	6	5	1	–	–	-	–	–
R9	$-OH$	13	8	5	0.416	–	–	–	–
	$-Cl$	**15**	**3**	**12**	**<0.05**	–	–	–	–
	$-Br$	3	2	1	–	–	–	–	–

[a] The total number of predicted values for fragments
[b] The number of fragments classified as positive
[c] The number of fragments classified as negative
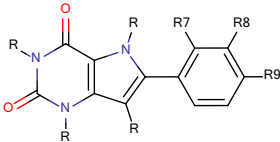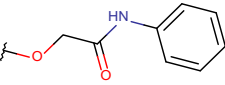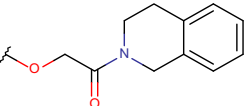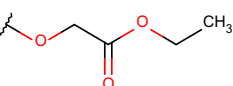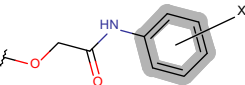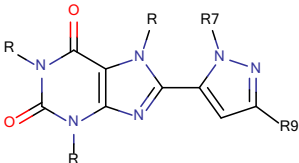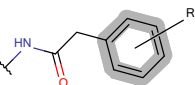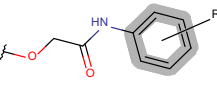[d] p-value of hypothesis test

position of the anilide phenyl ring, which they attributed to increased lipophilicity. We observed the same relationship of the fragment contribution in Table 5.

No conclusive results were observed in the case of the fragments studied for 8-(pyrazol-5-yl) xanthine derivatives.

The greater subfragmentation contributions obtained in the ensemble model agree with the affinities mentioned in the literature. All of these findings point to the adequate representation of the Adenosine A$_{2B}$ receptor affinity on the part of the developed model, not only because of the precision obtained in the predictions, but also due to the contribution values obtained as a result of subfragmentation.

Table 5 Contributions of fragments in the final ensemble QSAR model

		n^a	P^b	N^c	p value^d
R7	–F	2	0	2	–
	–OCH$_3$	8	0	8	–
R8	–OCH$_3$	**21**	**0**	**21**	**<0.05**
		2	2	0	–
R9	–NH$_2$	6	1	5	–
	–CH$_3$	9	0	9	–
	–O–CH$_2$COOH	**17**	**5**	**12**	**<0.05**
		14	9	5	–
		9	3	6	–
	 R: Halogen	**48**	**33**	**5**	**<0.05**

		n^a	P^b	N^c	p value^d
R7	–CH$_3$	46	20	26	0.452
R9	 R: Any	23	9	14	0.374
	 R: Any	9	6	3	–

[a] The total number of fragments
[b] The number of fragments classified as positive
[c] The number of fragments classified as negative
[d] p value of hypothesis test

Conclusion

The adenosine A$_{2B}$ receptor affinity of a series of 413 ligands was studied using QSAR methodology. The ensemble QSAR model obtained good classifications for both the test and external sets. In addition, most of the contributions to the adenosine A$_{2B}$ receptor affinity derived by subfragmentation of the molecules in the training set agreed with the relationships observed by other authors. These two factors mean the ensemble QSAR model could be used for predicting future

adenosine A_{2B} antagonist candidates. In the subfragmentation study of this activity, several fragments with a significant effect on the activity can be identified. Among these effects, we highlight the finding that long hydrophobic substituents in position 1 make the molecule a more potent adenosine A_{2B} antagonist, although the same moiety has the opposite effect in position 3.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Mohamed T, Osman W, Tin G, Rao P (2013) Selective inhibition of human acetylcholinesterase by xanthine derivatives: in vitro inhibition and molecular modeling investigations. Bioorg Med Chem Lett 23:4336–4341. doi:10.1016/j.bmcl.2013.05.092

2. Salhiyyah K, Senanayake E, Abdel-Hadi M, Booth A, Michaels JA (2012) Pentoxifylline for intermittent claudication. Cochrane Database Syst Rev CD005262. doi:10.1002/14651858.CD005262.pub2

3. Czechtizky W, Dedio J, Desai B, Dixon K, Farrant E, Feng QX, Morgan T, Parry D (2013) Integrated synthesis and testing of substituted xanthine based DPP4 inhibitors: application to drug discovery. ACS Med Chem Lett 4:768–772. doi:10.1021/ml400171b

4. Cazzola M, Page C, Calzetta L, Matera M (2012) Pharmacology and therapeutics of bronchodilators. Pharmacol Rev 64:450–504. doi:10.1124/pr.111.004580

5. Diniz C, Borges F, Santana L, Uriarte E, Oliveira J, Goncalves J, Fresco P (2008) Ligands and therapeutic perspectives of adenosine A(2A) receptors. Curr Pharm Design 14:1698–1722. doi:10.2174/138161208784746842

6. Ratheesh A, Jain M, Gude P (2010) Antimetastatic action of pentoxifylline, a methyl xanthine derivative, through its effect on PKC mediated integrin transport in B16F10 melanoma cells. World J Oncol 1:194–203

7. Pacher P, Nivorozhkin A, Szabó C (2006) Therapeutic effects of xanthine oxidase inhibitors: Renaissance half a century after the discovery of allopurinol. Pharmacol Rev 58:87–114. doi:10.1124/pr.58.1.6

8. Francis SH, Sekhar KR, Ke H, Corbin JD (2011) Inhibition of cyclic nucleotide phosphodiesterases by methylxanthines and related compounds. Handb Exp Pharmacol 200:93–133. doi:10.1007/978-3-642-13443-2_4

9. Gu Y, Wang W, Zhu X, Dong K (2014) Molecular dynamic simulations reveal the mechanism of binding between xanthine inhibitors and DPP-4. J Mol Model 20:1–12. doi:10.1007/s00894-014-2075-1

10. Mueller C, Jacobson K (2011) Xanthines as adenosine receptor antagonists. In: Fredholm BB (ed) Methylxanthines, handbook of experimental pharmacology. Springer, Berlin, pp 151–199

11. Fredholm BB, Ijzerman AP, Jacobson KA, Linden J, Muller CE (2011) International union of pharmacology XXV Nomenclature and classification of adenosine receptors. Pharmacol Rev 63:1–34

12. Wilson C, Mustafa S (2009) Adenosine receptors in health and disease. Springer, New York

13. Zhong H, Wu Y, Belardinelli L, Zeng D (2006) A_{2B} adenosine receptors induce IL-19 from bronchial epithelial cells, resulting in TNF-alpha increase. Am J Respir Cell Mol Biol 35:587–592. doi:10.1165/rcmb.2005-04760C

14. Feoktistov I, Ryzhov S, Zhong H, Goldstein A, Matafonov A, Zeng D, Biaggioni I (2004) Hypoxia modulates adenosine receptors in human endothelial and smooth muscle cells toward an A_{2B} angiogenic phenotype. Hypertension 44:649–654. doi:10.1161/01.HYP.0000144800.21037.a5

15. Feoktistov I, Ryzhov S, Goldstein A, Biaggioni I (2003) Mast cell-mediated stimulation of angiogenesis: cooperative interaction between A_{2B} and A3 adenosine receptors. Circ Res 92:485–492. doi:10.1161/01.RES.0000061572.10929.2D

16. Kalla R, Elzein E, Perry T, Li X, Gimbel A, Yang M, Zeng D, Zablocki J (2008) Selective, high affinity A_{2B} adenosine receptor antagonist: N-1 monosubstituted 8-(pyrazol-4-yl)xanthines. Bioorg Med Chem Lett 18:1397–1401. doi:10.1016/j.bmcl.2008.01.008

17. Ryzhov S, Novitskiy SV, Zaynagetdinov R, Goldstein AE, Carbone DP, Biaggioni I, Dikov MM, Feoktistov I (2008) Host A_{2B} adenosine receptors promote carcinoma growth. Neoplasia 10:987–995. doi:10.1593/neo.08478

18. Beukers M, Meurs I, IJzerman A (2006) Structure-afinity relationships of adenosine A_{2B} receptor ligands. Med Res Rev 26:667–698. doi:10.1002/med.20069

19. Moro S, Gao Z-G, Jacobson K, Spalluto G (2006) Progress in the pursuit of therapeutic adenosine receptor antagonists. Med Res Rev 26:131–159. doi:10.1002/med.20048

20. Burnet M, Stohr T, Donsbach M (2014) Patente no 20140142113 A1 20140522 US

21. Teijeira M, Gonzalez MP, Saiz-Urra L, Teran C, Rivero V, Lopez-Romero JM (2011) Topological descriptors for predicting affinity of xanthine derivates to A_{2B} adenosine receptors. In: Putz M (ed) Advances in chemical modeling. Nova Press, Burleigh Heads, pp 267–280

22. Paz O, Brito C, Castilho M (2014) Quantitative insights towards the design of potent deazaxanthine antagonists of adenosine 2B receptors. J Enzym Inhib Med Chem 29:590–598

23. Bonet I, Franco-Montero P, Rivero V, Teijeira M, Borges F (2013) Classifier ensemble based on feature selection and diversity measures for predicting the affinity of A_{2B} adenosine receptor antagonists. J Chem Inf Model 53:3140–3155. doi:10.1021/ci300516w

24. Mansourian M, Fassihi A, Saghaie L, Madadkar-Sobhani A, Mahnam K, Abbasi M (2015) QSAR and docking analysis of A_{2B} adenosine receptor antagonists based on non-xanthine scaffold. Med Chem Res 24:394–407. doi:10.1007/s00044-014-1133-7

25. Joseph TB, Kumar BVSS, Santhosh B, Kriti S, Pramod AB, Ravikumar M, Kishore M (2008) Quantitative structure activity relationship and pharmacophore studies of adenosine receptor A_{2B} inhibitors. Chem Biol Drug Des 72:395–408. doi:10.1111/j.1747-0285.2008.00714.x

26. Song Y, Coupar IM, Iskander MN (2001) Structural predictions of adenosine 2B antagonist affinity using molecular field analysis. Quant Struct Act Relat 20:23–30. doi:10.1002/1521-3838(200105)20:1.;23:AID-QSAR23.;3.0.CO;2-I

27. Ivanov AA, Wang B, Klutz AM, Chen VL, Gao ZG, Jacobson KA (2008) Probing distal regions of the A_{2B} adenosine receptor by quantitative structure-activity relationship modeling of known and novel agonists. J Med Chem 51:2088–2099. doi:10.1021/jm701442d

28. Carotti A, Stefanachi A, Ravina E, Sotelo E, Loza MI, Cadavid MI, Centeno NB, Nicolotti O (2004) 8-substituted-9-deazaxanthines as adenosine receptor ligands: design, synthesis and structure-affinity relationships at A(2B). Eur J Med Chem 39:879–887. doi:10.1016/j.ejmech.2004.07.008

29. Stefanachi A, Brea JM, Cadavid MI, Centeno NB, Esteve C, Loza MI, Martinez A, Nieto R, Ravina E, Sanz F, Segarra V, Sotelo E, Vidal B, Carotti A (2008) 1-, 3- and 8-substituted-9-deazaxanthines as potent and selective antagonists at the human A(2B) adenosine receptor. Bioorg Med Chem 16:2852–2869. doi:10.1016/j.bmc.2008.01.002

30. Stefanachi A, Nicolotti O, Leonetti F, Cellamare S, Campagna F, Loza MI, Brea JM, Mazza F, Gavuzzo E, Carotti A (2008) 1,3-Dialkyl-8-(hetero)aryl-9-OH-9-deazaxanthines as potent A_{2B} adenosine receptor antagonists: Design, synthesis, structure-affinity and structure-selectivity relationships. Bioorg Med Chem 16:9780–9789. doi:10.1016/j.bmc.2008.09.067

31. Nieto MI, Balo MC, Brea J, Caamano O, Fernandez F, Garcia-Mera X, Lopez C (2010) Synthesis and pharmacological evaluation of novel substituted 9-deazaxanthines as A(2B) receptor antagonists. Eur J Med Chem 45:2884–2892. doi:10.1016/j.ejmech.2010.03.011

32. Fernandez F, Caamano O, Nieto MI, Lopez C, Garcia-Mera X, Stefanachi A, Nicolotti O, Loza MI, Brea J, Esteve C, Segarra V, Vidal B, Carotti A (2009) 1,3-Dialkyl-8-N-substituted benzyloxycarbonylamino-9-deazaxanthines as potent adenosine receptor ligands: Design, synthesis, structure-affinity and structure-selectivity relationships. Bioorg Med Chem 17:3618. doi:10.1016/j.bmc.2009.03.062

33. Carotti A, Cadavid MI, Centeno NB, Esteve C, Loza MI, Martinez A, Nieto R, Ravina E, Sanz F, Segarra V, Sotelo E, Stefanachi A, Vidal B (2006) Design, synthesis, and structure-activity relationships of 1-,3-,8-, and 9-substituted-9-deazaxanthines at the human A(2B) adenosine receptor. J Med Chem 49:282–299. doi:10.1021/jm0506221

34. Baraldi PG, Tabrizi MA, Preti D, Bovero A, Romagnoli R, Fruttarolo F, Naser Abdel M, Allan R, Varani K, Gessi S, Merighi S, Borea PA (2004) Design, synthesis, and biological evaluation of new 8-heterocyclic xanthine derivatives as highly potent and selective human A_{2B} adenosine receptor antagonists. J Med Chem 47:1434–1447. doi:10.1021/jm0309654

35. Balo MC, Brea J, Caamano O, Fernandez F, Garcia-Mera X, Lopez C, Loza MI, Nieto MI, Rodriguez-Borges JE (2009) Synthesis and pharmacological evaluation of novel 1- and 8- substituted-3-furfutyl xanthines as adenosine receptor antagonists. Bioorg Med Chem 17:6755–6760. doi:10.1016/j.bmc.2009.07.034

36. Rodriguez-Borges JE, Garcia-Mera X, Balo MC, Brea J, Caamano O, Fernandez F, Lopez C, Loza MI, Nieto MI (2010) Synthesis and pharmacological evaluation of novel 1,3,8-and 1,3,7,8-substituted xanthines as adenosine receptor antagonists. Bioorg Med Chem 18:2001–2009. doi:10.1016/j.bmc.2010.01.028

37. Nieto MI, Balo MC, Brea J, Caamano O, Cadavid MI, Fernandez F, Mera XG, Lopez C, Rodriguez-Borges JE (2009) Synthesis of novel 1-alkyl-8-substituted-3-(3-methoxypropyl) xanthines as putative A(2B) receptor antagonists. Bioorg Med Chem 17:3426. doi:10.1016/j.bmc.2009.03.029

38. Talete, srl (2006) Dragon for windows (software for molecular descriptors calculation) version 54. Talete, srl, Milano

39. Chemical Computing Group (2008) Molecular operating environment (MOE). Chemical Computing Group, Montreal

40. Yasri A, Hartsough D (2001) Toward an optimal procedure for variable selection and QSAR model building. J Chem Inf Comput Sci 41:1218–27. doi:10.1021/ci010291a

41. Gore P (2000) Handbook of applied multivariate statistics and mathematical modelling. Academic Press, San Diego

42. Duchowicz PR, Castro EA, Fernndez FM (2006) Alternative algorithm for the search of an optimal set of descriptors in QSAR-QSPR studies. MATCH Commun Math Comput Chem 55:179–192

43. StatSoft Inc. (2002) StatSoft Statistica, version 8.0. StatSoft Inc., Tulsa

44. Pérez-Garrido A, Helguera AM, Borges F, Cordeiro MN, Rivero V, Escudero AG (2011) Two new parameters based on distances in a receiver operating characteristic chart for the selection of classification models. J Chem Inf Model 51:2746–2759. doi:10.1021/ci2003076

45. Olivares-Morales A, Hatley OJD, Turner D, Galetin A, Aarons L, Rostami-Hodjegan A (2014) The use of ROC analysis for the qualitative prediction of human oral bioavailability from animal data. Pharm Res 31:720–730. doi:10.1007/s11095-013-1193-2

46. Nandy A, Kar S, Roy K (2014) Development of classification- and regression-based QSAR models and in silico screening of skin sensitisation potential of diverse organic chemicals. Mol Simul 40:261–274. doi:10.1080/08927022.2013.801076

47. Pérez-Garrido A, Girón-Rodríguez F, Morales Helguera A, Borges F, Combes RD (2014) Topological structural alerts modulations of mammalian cell mutagenicity for halogenated derivatives. SAR QSAR Environ Res 25:17–33. doi:10.1080/1062936X.2013.820791

48. Morales Helguera A, Pérez-Garrido A, Gaspar A, Cagide F, Vina D, Cordeiro MNDS, Borges F (2013) Combining QSAR classification models for predictive modeling of human monoamine oxidase inhibitors. Eur J Med Chem 59:75–90. doi:10.1016/j.ejmech.2012.10.035

49. Wold S, Eriksson L, Clementi S (1995) Statistical validation of QSAR results. In: van de Waterbeemd H (ed) Chemometric methods in molecular design. Wiley-VCH, Berlin, pp 309–338. doi:10.1002/9783527615452.ch5

50. de Cerqueira Lima P, Golbraikh A, Oloff S, Xiao Y, Tropsha A (2006) Combinatorial QSAR modeling of P-glycoprotein substrates. J Chem Inf Model 46:1245–1254. doi:10.1021/ci0504317

51. Todeschini R, Ballabio D, Consonni V, Manganaro W, Mauri A (2009) Canonical measure of correlation (CMC) and canonical measure of distance (CMD) between sets of data part 1 theory and simple chemometric applications. Anal Chim Acta 648:45–51. doi:10.1016/j.aca.2009.06.032

52. Krzanowski WJ (1988) Principles of multivariate analysis. Oxford University Press, New York

53. Kuncheva LI (2004) Combining pattern classifiers methods and algorithms. Wiley, New Jersey

54. Netzeva TI, Worth AP, Aldenberg T, Benigni R, Cronin MTD, Gramatica P, Jaworska JS, Kahn S, Klopman P, Marchant CA, Myatt G, Patlewicz GY, Perkins R, Roberts DW, Schultz TW, Stanton DT (2005) Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. ATLA 33:155–173

55. Gramatica P (2007) Principles of QSAR models validation: internal and external. QSAR Comb Sci 26:694–701. doi:10.1002/qsar.200610151

56. Pérez-Garrido A, Morales Helguera A, Caravaca G, Cordeiro MNDS, scudero AG (2010) A topological substructural molecular design approach for predicting mutagenesis end-points of α, β-unsaturated carbonyl compounds. Toxicology 268:64–77. doi:10.1016/j.tox.2009.11.023

57. Pérez-Garrido A, Morales Helguera A, Ruiz JMM, Rentero PZ (2012) Topological sub-structural molecular design approach: radical scavenging activity. Eur J Med Chem 49:86–94. doi:10.1016/j.ejmech.2011.12.030

58. Tropsha A, Gramatica P, Gombar V (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR Models. QSAR Comb Sci 22:69–77. doi:10.1002/qsar.200390007

59. Polishchuk PG, Kuz'min VE, Artemenko AG, Muratov EN (2013) Universal approach for structural interpretation of QSAR/QSPR models. Mol Inf 32:843–853. doi:10.1002/minf.201300029

60. Instant Jchem version 6.2.1 (2014) Chemaxon Ltd, Budapest, https://www.chemaxon.com

61. Liu H, Gramatica P (2007) QSAR study of selective ligands for the thyroid hormone receptor beta. Bioorg Med Chem 15:5251–5261. doi:10.1016/j.bmc.2007.05.016

62. Helguera A, Combes R, González M, Cordeiro M (2008) Aplications of 2D descriptors in drug design: a DRAGON tale. Curr Top Med Chem 8:1628–1655. doi:10.2174/156802608786786598

63. Borghini A, Pietra D, Domenichellia P, Bianuccia A (2005) QSAR study on thiazole and thiadiazole analogues as antagonists for the adenosine A1 and A3 receptors. Bioorg Med Chem 13:5330–5337. doi:10.1016/j.bmc.2005.05.041

64. Bhattacharya P, Leonard J, Roy K (2005) Exploring QSAR of thiazole and thiadiazole derivatives as potent and selective human adenosine A3 receptor antagonists using FA and GFA techniques. Bioorg Med Chem 13:1159–1165. doi:10.1016/j.bmc.2004.11.022

65. Roy K, Leonard J, Sengupt C (2004) QSAR of adenosine receptor antagonists part 3: exploring physicochemical requirements for selective binding of 1,2,4-triazolo5,1-i.purine derivatives with human adenosine A3 receptor subtype. Bioorg Med Chem Lett 14:3705–3709. doi:10.1016/j.bmcl.2004.05.007

66. Bhattacharya P, Roy K (2005) QSAR of adenosine A3 receptor antagonist 1,2,4-triazolo4,3-a.quinoxalin-1-one derivatives using chemometric tools. Bioorg Med Chem Lett 15:3737–3743. doi:10.1016/j.bmcl.2005.05.051

67. Doichinova I, Natcheva R, Mihailova D (1994) QSAR studies of S-substituted xanthines as adenosine receptor antagonists. Eur J Med Chem 29:133–138. doi:10.1016/0223-5234(94)90210-0

68. Cheng F, Xu Z, Liu G, Tang Y (2010) Insights into binding modes of adenosine A_{2B} antagonists with ligand-based and receptor-based methods. Eur J Med Chem 45:3459–3471. doi:10.1016/j.ejmech.2010.04.039