Available online at www.sciencedirect.com
SCIENCE DIRECT®
European Journal of Medicinal Chemistry 41 (2006) 56–62
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://france.elsevier.com/direct/ejmech

Original article

# Radial distribution function descriptors: an alternative for predicting $A_{2\ A}$ adenosine receptors agonists

Maykel Pérez González [a,b,c,*], Carmen Terán [a], Marta Teijeira [a], Aliuska Morales Helguera [c,d]

[a] Department of Organic Chemistry, Vigo University, C.P. 36200, Vigo, Spain
[b] Service Unit, Experimental Sugar Cane Station "Villa Clara-Cienfuegos", Ranchuelo, villa Clara, C. P. 53100, Cuba
[c] Chemical Bioactive Center, Central University of Las Villas, Santa Clara, 54830 villa Clara. Cuba
[d] Department of Chemistry, Faculty of Chemistry and Pharmacy, Central University of Las Villas, Santa Clara, villa Clara, Cuba

Received 29 July 2005; received in revised form 30 August 2005; accepted 31 August 2005
Available online 25 October 2005

**Abstract**

The Radial Distribution Function approach has been applied to the study of the $A_{2\ A}$ adenosine receptors agonist effect of 29 adenosine analogues: $N^6$–arylcarbamoyl, 2-arylalkynyl-$N^6$–arylcarbamoyl, and $N^6$–carboxamido derivatives. A model able to describe around 85% of the variance in the experimental activity was developed with the use of the mentioned approach. In contrast, no one of nine different approaches, including the use of Galvez Topological Charges indexes, BCUT, Geometrical, 2D autocorrelations, Topological, Randić Molecular profile, WHIM, 3D-MORSE and GETAWAY descriptors were able to explain more than 78% of the variance in the mentioned property with the same number of variables in the equation. Finally, the model support that the bulkiness and stereoselectivity play an important role in the affinity for this receptor in this kind of compounds.


*Keywords:* QSAR; $A_{2\ A}$ adenosine receptors agonists; RDF descriptors

## 1. Introduction

Adenosine is the native ligand for adenosine receptors (AR), also referred to as P1 purinergic receptors, which are distinct from P2 purinergic nucleotide receptors. The adenosine receptor class can be further divided into four receptor subtypes which have been identified and cloned from several mammalian tissue types: $A_1$, $A_{2\ A}$, $A_{2B}$, and $A_3$ [1,2]. Agonists and antagonists of these receptors are being explored as potential therapeutic agents for: cardiovascular mediation, ischemia-reperfusion injury, inflammation, Parkinson's disease, and schizophrenia [3,4].

In this connection, we are interested in the impact of adenosine $A_{2\ A}$ receptor agonists, because of the many potential biological implications of stimulating this subclass of AR; much effort has been spent trying to elucidate the structure-activity relationships for adenosine analogues. This has led to the synthesis and evaluation of a great number of $A_{2\ A}$ adenosine analogues [4].

In this sense, the majority of $A_{2\ A}$-selective AR agonists are 2-substituted adenosine derivatives or analogs, frequently bearing an N-alkylcarboxamido modification at the ribose 5′-position. Also, some $N^6$-substitution of adenosine with bulky substituents at $N^6$ have been carried out, but this substitution is detrimental for AR affinity [5].

However, a limited number of models QSAR have been developed about this topic [6–10]. In addition, 3D molecular descriptors have shown to be very useful in QSAR problems in

* Corresponding author. Tel.: (53) (42) 281473; fax: (53)-(42)-281130.
*E-mail address:* mpgonzalez76@yahoo.es (M.P. González).

0223-5234/$ - see front matter © 2005 Elsevier SAS. All rights reserved.
doi:10.1016/j.ejmech.2005.08.004

order to perform a rational analysis of different pharmacological activities [11]. In this connection, a kind of 3D molecular descriptors namely Radial Distribution Function (RDF) have been introduced for modeling physicochemical and biological properties of organic compounds [11].

The successful applications of this theoretical approach to the modeling of affinity of $A_{2B}$ AR [12] have inspired us to perform a more exhaustive study to improve the results obtained with the data set use in this study and in order to test and/or validate the RDF descriptors applicability in this area [10].

## 2. Materials and methods

### 2.1. The radial distribution function approach

These descriptors are based on the distances distribution in the geometrical representation of a molecule and constitute a radial distribution function code [11].

Formally, the radial distribution function of an ensemble of N atoms can be interpreted as the probability distribution of finding an atom in a spherical volume of radius $r$ [11]. The general form of the radial distribution function code is represented by:

$$g(r) = f \cdot \sum_{i=1}^{N-1} \sum_{j>i}^{N} A_i A_j e^{-B(r-r_{ij})^2}$$

where $f$ is a scaling factor and $N$ is the number of atoms. By including characteristic atomic properties $A$ of the atoms $i$ and $j$, the RDF codes can be used in different tasks to fit the requirements of the information to be represented. These atomic properties enable the discrimination of the atoms of a molecule for almost any property that can be attributed to an atom. The exponential term contains the distance $r_{ij}$ between the atoms $i$ and $j$ and the smoothing parameter $B$, which defines the probability distribution of the individual distances; $B$ can be interpreted as a temperature factor that defines the movement of the atoms. $g(r)$ was calculated for a number of discrete points with defined intervals.

The radial distribution function in this form meets all the requirements for 3D structure descriptors: it is independent of the number of atoms, i. e., the size of a molecule, it is unique regarding the three-dimensional arrangement of the atoms, and it is invariant against translation and rotation of the entire molecule. Additionally, the RDF descriptors can be restricted to specific atom types or distance ranges to represent specific information in a certain three-dimensional structure space, e.g. to describe steric hindrance or structure/activity properties of a molecule.

Finally, the RDF descriptors are interpretable by using simple rules sets, and thus it provides a possibility for conversion of the code back into the corresponding 3D structure. Besides information about interatomic distances in the entire molecule, the RDF descriptors provides further valuable information, e.g. about bond distances, ring types, planar and non-planar systems and atom types.

### 2.2. Data set

Here, a data set of 29 adenosine derivatives for which their activities are reported in the literature by Baraldi et al. [13] was used. Molecular structure, numbering of the substituents and activities of the adenosine derivatives are summarized in Table 1.

The displacement of specified [$^3$H]CGS 21680 binding ($A_{2\ A}$) were measured in rat striatal membranes expressed as $K_i$ in nM ($n$ = 3 - 6).

### 2.3. Molecular Descriptors

The molecular descriptors for the given compounds were calculated using *DRAGON* software [14] on the (x,y,z)-atomic coordinates of the minimal energy conformations determined by the AM1 method in MOPAC 6.0 Package [15]. A total of 511 molcular descriptors of differing types were calculated to describe compound structural diversity. The descriptor typology is as follows: (a) 2D-64 BCUTs, (b) 2D-96 various autocorrelations from the molecular graph, (c) 3D-41 Randić Molecular profile index, (d) 3D-160 3D-MORSE and (e) 3D-150, RDF descriptors.

The list of these molecular descriptors, and their meaning, is provided with literature references by the *DRAGON* package; the calculation procedure is explained in detail, with related literature references, in the Handbook of Molecular Descriptors [11]. However, in the Table 2 we provide the meaning of the descriptors used in the models of the present study.

Descriptors with constant or near constant values inside each group were discarded. For the remaining descriptors pairwise correlation analysis for all kinds of descriptors was performed [10]. This descriptors exclusion method was used to reduce, in a first step, the collinearity and correlation among descriptors.

### 2.4. Orthogonalization of molecular descriptors

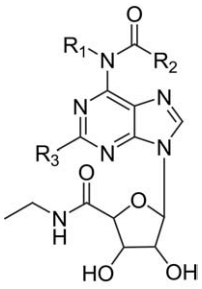
The pairwise correlation analysis is a good method for eliminated the collinearity, but in this case the correlation among the variables persist. For that reason, other method of elimination of the collinearity is necessary.

In order to avoid collinearity, Randić orthogonalization procedure was carried out [16–19]. The main philosophy of this approach is to avoid the exclusion of descriptors on the basis of collinearity 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 impossible to know the relative importance of an index and underestimates the utility of the regression coefficients in a model.

The Randić method of orthogonalization has been described previously [16–19]. Thus, we will give a general overview here. The first step in orthogonalizing the molecular descriptors is to select the appropriated order of orthogonalization, which in this case is the order in which the variables were selected in the forward stepwise search procedure of the linear regression analysis.

In this sense, we used RDF075p = $^1\Omega$RDF075p as the first orthogonal variable. Afterwards, the successive residuals of the step–by–step regressions between each variable selected

Table 1
Structures and affinities in radioligand binding assays at rat brain $A_{2\ A}$ adenosine receptors used in the current work



| Comp | $R_1$ | $R_2$ | $R_3$ | $\log(K_i(A_{2\ A}))$[a] |
|---|---|---|---|---|
| 1 | H | 4-biphenyl | H | 3.554 |
| 2 | H | 2,4-Cl-Ph-$CH_2$ | H | 1.425 |
| 3 | H | 4-$CH_3O$-Ph | H | 2.375 |
| 4 | H | 2-Cl-Ph | H | 3.292 |
| 5 | H | Ph | H | 2.826 |
| 6 | H | Ph$CH_2$NH | H | 3.254 |
| 7 | H | 4-$SO_2NH_2$PhNH | H | 3.072 |
| 8 | H | 4-$CH_3$CO-PhNH | H | 3.021 |
| 9 | H | (R)-α-phenylethyl-NH | H | 2.446 |
| 10 | H | (S)-α-phenylethyl-NH | H | 3.473 |
| 11 | H | 5-Me-isoxazol-3-yl-NH | H | 2.946 |
| 12 | H | 1,3,4-thiadiazol-2-yl-NH | H | 2.962 |
| 13 | H | 4-n-$C_3H_7O$-PhNH | H | 2.407 |
| 14 | H | Ph-$CH_2CH_2$NH | H | 3.423 |
| 15 | H | 3,4-MeO-Ph-$CH_2CH_2$NH | H | 3.196 |
| 16 | H | Fur-2-yl-$CH_2$NH | H | 3.270 |
| 17 | H | 4-(pyridin-2-yl-$NHSO_2$)PhNH | H | 2.869 |
| 18 | H | 4-(5-Me-isoxazol-3-yl-$NHSO_2$) PhNH | H | 2.990 |
| 19 | H | 4-(pyrimidin-2-yl-$NHSO_2$)PhNH | H | 1.225 |
| 20 | 4-$NO_2$-Ph-NH-CO | 4-$NO_2$-Ph-NH | H | 3.403 |
| 21 | 5-Cl-pyridin-2-yl-NH-CO | 5-Cl-pyridin-2-yl-NH | H | 2.903 |
| 22 | H | 3-Cl-Ph-NH | Cl | 2.746 |
| 23 | H | 4-MeO-Ph-NH | Cl | 1.771 |
| 24 | H | 3-Cl-Ph-NH | I | 4.025 |
| 25 | H | 4-MeO-Ph-NH | I | 2.415 |
| 26 | H | 3-Cl-Ph-NH | n-$C_4H_9$-C≡C | 3.853 |
| 27 | H | 3-Cl-Ph-NH | Ph-C≡C | 4.688 |
| 28 | H | 4-MeO-Ph-NH | n-$C_4H_9$-C≡C | 3.367 |
| 29 | H | 4-MeO-Ph-NH | Ph$(CH_2)_3$-C≡C | 3.223 |

[a] Displacement of specified [$^3$H]CGS 21680 binding ($A_{2\ A}$) in rat striatal membranes expressed as $K_i$ in nM.

in the model and the others in order of statistical significance were calculated. All these residuals were used as the remnant orthogonal variables in the Eq. 3. In this analysis the least squares method selected all orthogonal analogs of collinear variables. It ensured us that, in spite of variables collinearity, each variable have an amount of information do not encoded in the others.

### 2.5. Statistical methods

#### 2.5.1. Genetic algorithm (GA) analysis

Five models were developed using the descriptors obtained by Dragon computer software. All statistical analysis and data exploration was carrying out using the Statistic 6.0 [20]. The most significance parameters were identified from the data set using Genetic Algorithm Analysis (GA) [21,22].

GA is a class of methods based on biological evolution rules. The first step is to create a population of linear regression models. These regression models mate with each other, mutate, crossover, reproduce, and then evolve through successive generations toward an optimum solution. The GA simulation conditions were 10 000 generations and 300 populations. The models were linear combinations of 5 descriptors. The GA procedure was repeated n-times to confirm that the selected descriptors are the most optimal descriptor set for describing the modelled property.

Table 2
BCUT, 2D autocorrelations, Randić Molecular profile, 3D-MORSE, and RDF descriptors of the QSAR regression reported in this study

| Symbol | Meaning |
|---|---|
| BELv7 | Highest eigenvalue n. 7 of Burden matrix / weighted by atomic van der Waals volumes |
| BEHe7 | Highest eigenvalue n. 7 of Burden matrix / weighted by atomic Sanderson electronegativities |
| ATS6v | Broto-Moreau autocorrelation of a topological structure – lag 6 / weighted by atomic van der Waals volumes |
| ATS1e | Broto-Moreau autocorrelation of a topological structure – lag 1 / weighted by atomic Sanderson electronegativities |
| ATS7e | Broto-Moreau autocorrelation of a topological structure – lag 7 / weighted by atomic Sanderson electronegativities |
| MATS3e | Moran autocorrelation – lag 3 / weighted by atomic Sanderson electronegativities |
| GATS7e | Geary autocorrelation – lag 7 / weighted by atomic Sanderson electronegativities |
| DP04 | Molecular profile no. 04 |
| RDF100 m | Radial distribution function – 10.0 / weighted by atomic masses |
| RDF135 m | Radial distribution function – 13.5 / weighted by atomic masses |
| RDF140 m | Radial distribution function – 14.0 / weighted by atomic masses |
| RDF130v | Radial distribution function – 13.0 / weighted by atomic van der Waals volumes |
| RDF075p | Radial distribution function – 7.5 / weighted by atomic polarizabilities |
| Mor31u | 3D-Morse – signal 31 / unweighted |
| Mor05 m | 3D-Morse – signal 05 / weighted by atomic masses |
| Mor14 m | 3D-Morse – signal 14 / weighted by atomic masses |
| Mor02v | 3D-Morse – signal 02 / weighted by atomic van der Waals volumes |
| Mor10p | 3D-Morse – signal 10 / weighted by atomic polarizabilities |
| BEHm3 | Highest eigenvalue n.3 of burden matrix / weighted by atomic masses |
| BEHm6 | Highest eigenvalue n.6 of burden matrix / weighted by atomic masses |
| BELp7 | Lowest eigenvalue n.7 of burden matrix / weighted by atomic polarizabilities |
| DP01 | Molecular profile no. 01 |
| DP02 | Molecular profile no. 02 |
| DP05 | Molecular profile no. 05 |
| DP07 | Molecular profile no. 07 |

Examining the regression coefficients, the standard deviations, the significances and the number of variables in the equation determined the quality of the models.

### 2.6. Validation of the models

Lineal models were validated by calculating $q^2$ values. The $q^2$ values are calculated from "leave-one-out" (LOO) and leave–group–out (LGO) testiness, also known as cross–validation. One, or a group of compounds (in this case the 25% of the data set) is removed from the set and the regression recalculated; the predicted value for that point is then compared to its actual value. This is repeated until each datum or data group has been omitted once; the sum of squares of these deletion residuals can then be used to calculate $q^2$, an equivalent statistic to $R^2$. The $q^2$ values can be considered a measure of the predictive power of a regression equation: whereas $R^2$ can always be increased artificially by adding more parameters (descriptors), $q^2$ decreases if a model is overparameterized [23], and is therefore a more meaningful summary statistic for QSAR models.

## 3. Results and discussion

### 3.1. Quantitative structure activity relationships models

The model selection was subjected to the principle of parsimony. The five-dimensional models are characterized by the best compromise between predictive power and model complexity. The addition of another variable does not lead to such an increase in predictive power such that the complexity increase is counterbalanced. In addition, the relation between cases and variables should be 5:1, reason why the addition of new variable violates this principle.

The best QSAR model obtained with the RDF descriptors is given below together with the statistical parameters of the regression.

$$-\log(Ki) = 1.372 + 0.209 \cdot RDF075p - 0.156 \cdot RDF135m - 0.293 \cdot RDF130v + 0.215 \cdot RDF100m - 0.121 \cdot RDF140m \quad (1)$$

$N = 29$ $S = 0.351$ $R^2 = 0.806$ $F = 19.073$
$q^2$ (LOO) $= 0.731$ $S_{LOO} = 0.411$ $q^2$ (LGO) $= 0.705$ $S_{LGO} = 0.431$

where N is the number of compounds included in the model, $R^2$ is the correlation coefficient, S the standard deviation of the regression, F the Fisher ratio, $q^2$(LOO); $S_{LOO}$ and $q^2$(LGO); $S_{LGO}$ the correlation coefficient and the standard deviation of the cross – validation leave one out and leave group out respectively.

The meaning of the variables included in the model and the whole of the other descriptors used in the current work appear in Table 2.

Although, the model presents good statistical parameters, the outliers removed from a QSAR are essential. An outlier to a QSAR is identified normally by having a large standard residual and can indicate the limits of applicability of QSAR

models. There are several reasons for their occurrence in QSAR studies, for example, chemicals might be acting by a mechanism different from that of 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 the large data sets. Although it is acceptable to remove a small number of outliers from QSAR it is noted that it is not acceptable to remove the outlier repeatedly from a QSAR analysis simply to improve a correlation. From this data, after the application several statistical techniques for the detection of outliers, the compound 23 present large residual and should be considered as outlier. The outliers' number were extracted considering that a number of outliers lower than 10% of the general data are classically accepted in the literature as threshold limit value for outliers' extraction. In our case the outlier number represented only a 3.5% of the whole data.

Analysis of the residuals and deleted residual for equation 1 identified to compound 23 as significant outlier. Table 3

Table 3
Observed and predicted activity, residual and deleted residual of the compounds used in this study according to the equation 1

| Compounds | Observed activity | Predicted activity | Residual | Deleted Residual |
|---|---|---|---|---|
| 1 | 3.554 | 3.415 | 0.139 | 0.149 |
| 2 | 1.425 | 1.661 | –0.237 | –0.331 |
| 3 | 2.375 | 2.911 | –0.536 | –0.585 |
| 4 | 3.292 | 3.245 | 0.047 | 0.054 |
| 5 | 2.826 | 2.675 | 0.151 | 0.166 |
| 6 | 3.254 | 3.224 | 0.029 | 0.035 |
| 7 | 3.072 | 2.694 | 0.378 | 0.514 |
| 8 | 3.021 | 3.463 | –0.441 | –0.482 |
| 9 | 2.446 | 2.017 | 0.428 | 0.522 |
| 10 | 3.473 | 3.408 | 0.064 | 0.071 |
| 11 | 2.946 | 2.736 | 0.210 | 0.239 |
| 12 | 2.962 | 3.146 | –0.184 | –0.211 |
| 13 | 2.407 | 2.649 | –0.242 | –0.293 |
| 14 | 3.423 | 3.491 | –0.068 | –0.074 |
| 15 | 3.196 | 2.668 | 0.528 | 0.589 |
| 16 | 3.270 | 2.984 | 0.286 | 0.309 |
| 17 | 2.869 | 2.948 | –0.078 | –0.088 |
| 18 | 2.990 | 2.834 | 0.156 | 0.176 |
| 19 | 1.225 | 1.536 | –0.310 | –0.523 |
| 20 | 3.403 | 3.475 | –0.071 | –0.109 |
| 21 | 2.903 | 2.625 | 0.278 | 0.402 |
| 22 | 2.746 | 3.055 | –0.309 | –0.341 |
| 23 | 1.771 | 2.615 | –0.844 | –1.026 |
| 24 | 4.025 | 3.495 | 0.530 | 0.609 |
| 25 | 2.415 | 2.394 | 0.021 | 0.052 |
| 26 | 3.853 | 3.695 | 0.158 | 0.176 |
| 27 | 4.688 | 4.890 | –0.203 | –0.395 |
| 28 | 3.367 | 3.466 | –0.099 | –0.146 |
| 29 | 3.223 | 3.005 | 0.218 | 0.692 |

Removal of these compounds and subsequence re-analysis of the dataset produced a following QSAR.

$$-\log(Ki) = 1.509 + 0.198 \cdot RDF075p - 0.141 \cdot RDF135m - 0.285 \cdot RDF130v + 0.219 \cdot RDF100m - 0.151 \cdot RDF140m \quad (2)$$

N = 28 S = 0.299 $R^2$ = 0.849 F = 24.812
$q^2$ (LOO) = 0.786 $S_{LOO}$ = 0.368 $q^2$ (LGO) = 0.749 $S_{LGO}$ = 0.407

In this case, the improved of statistical parameters of the equation 2 (standard deviation, the regression coefficient and the $q^2$ of the cross validation LOO and LGO) justified broadly from statistical point of view the remove of this compound.

On the other hand, we carry out a Pairwise Correlation Analysis of the variables to eliminating the correlation among the descriptors, however this persists. For that reason, before making the interpretation of the model we need to orthogonalize the molecular descriptors included in such models due to the intercorrelation existing among some of them as shown in the following Table 4 for the RDF descriptors.

As a result of the orthogonalization process of the equation 2 we obtain:

$$-\log(Ki) = 0.226 + 0.07 \cdot {}^{1}\Omega RDF075p - 0.11 \cdot {}^{2}\Omega RDF135m - 0.098 \cdot {}^{3}\Omega RDF130v + 0.184 \cdot {}^{4}\Omega RDF100m - 0.151 \cdot {}^{5}\Omega RDF140m \quad (3)$$

N = 28 S = 0.299 $R^2$ = 0.849 F = 24.812
$q^2$ (LOO) = 0.786 $S_{LOO}$ = 0.368 $q^2$ (LGO) = 0.749 $S_{LGO}$ = 0.407

The most important variable in this equation are weight with the atomic masses, van der Waals volumes and atomic polarizabilities. In this model, the negative contributions to the log (K*i*), reason way decrease the affinity of the adenosine analogues for the $A_{2\ A}$ receptors, of the indexes that conform the model are weight with the atomic masses and radius of van der Waals, but the most significant behavior is that these descriptors are correspond to a radius of 13.0 to 14.0 Å. Formally, the RDF of a molecule of *A* atoms can be interpreted as the probability distribution of finding an atom in a spherical volume of radius R. In this sense, according to our model a spherical molecular volume with these dimensions could have certain restrictions to the addition of bulky substituents. This interpretation suggests that the adenosine analogues not be able to accommodate large substituents in the $N^6$ and C2 positions

Table 4
Correlation Matrix of the variables in the model 2

| | RDF075p | RDF135 m | RDF130v | RDF100 m | RDF140 m |
|---|---|---|---|---|---|
| **RDF075p** | 1.00 | 0.61 | 0.66 | 0.62 | 0.48 |
| **RDF135 m** | | 1.00 | 0.44 | 0.54 | 0.42 |
| **RDF130v** | | | 1.00 | 0.83 | 0.44 |
| **RDF100 m** | | | | 1.00 | 0.51 |
| **RDF140 m** | | | | | 1.00 |

at the same time for example. This observation agrees with the explanation reported by Müller et al. [24] some years ago.

On the other hand, the variable $^{1}\Omega 1RDF075p$ has a positive influence in the studied property, reason way increase the affinity of the adenosine analogues for the $A_{2\ A}$ receptors. This descriptor is weight with atomic polarizability and suggests that adenosine analogues might accommodate atoms or atoms group in different positions such as Iode and benzene that possess certain dimensional restrictions, because very bulky substituents do not be tolerate. This behavior was observed by Erica W. van Tilburg et al. [25] in a set of 2, 5'-disubstituted adenosine derivatives. These authors reported that the adenosine $A_{2\ A}$ receptor accommodate only 2-substituents with a restrained spacer. In this sense, Cristalli et al. [26] in a set of 2-aralkynyl and 2-heteroalkynyl derivates of adenosine-5'-N-ethyluronamide, reported that the affinity was reduced by introducing the bulkier naphthyl ring in position C2 and Viziano et al. [27], at the same position, reported that the presence of a bulky chain as the diphenyl derivates instead of benzene derivates is detrimental for $A_{2\ A}$ affinity. A similar report was introduce by Gungor et al. in a set of $N^6$-Substituted AR Agonists where noticed that the 2-position substituent's size may influence the activity on $A_2$ receptors [28].

Therefore, according to we model, the previous report and the criterion of Homma et al. one of the major determinants for the affinity at $A_{2\ A}$ ARs would be bulkiness of substituents attached at the different positions and the combinations of these in the adenosine derivatives [29].

### 3.2. Comparison with other approach

As we previously explain, one of the objectives of the current work is to compare the reliability of the RDF descriptors to describe the property under study as compared with other different descriptors and methods. Consequently, we have developed other four models using the same data set that was included in the RDF QSAR model. The results obtained with the use of BCUT, 2D autocorrelations, Randić Molecular profile, 3D-MORSE and RDF descriptors and models reported previously for us are given in Table 5.

As can be seen there are remarkable differences concerning the explanation of the experimental variance given by these models compared to the RDF one. While the RDF QSAR model explains around the 85% of the experimental activity the rest of the models are unable to explain more than 80% of such variance activity. Moreover, important statistic parameters such as the Fischer ratio (F) and the standard deviation (S) are of higher quality in the case of RDF model.

According the previous result is important highlight that the kinds of 3D descriptors explain better the property under study than the 2D ones, therefore this behavior suppose that the stereochemistry of the adenosine analogues play an important role in the affinity for these receptors. Jacobson et al. [30] prepared a pair of methanocarba-adenosine analogues for exploring the role of sugar puckering in ligand recognition. The (S)-methanocarba analogue of adenosine was only weakly active in binding to $A_{2\ A}$ adenosine receptors, presumably because of an unfavorable conformation that decreases receptor binding. In contrast, the methanocarba analogues constrained in the (N)-conformation, displayed high receptor affinity. PHPNECA possesses an asymmetrical carbon in the side chain in 2-position. For that reason Klotz et al. [31] investigated the stereoselectivity of the respective diastereomers. These authors demonstrated that the *S*-diastereomer is about 20-fold more potent at $A_{2\ A}$-receptor than the *R*-form. Reason why, confirm we conclusion about the importance of the stereoselectivity for the affinity at $A_{2\ A}$ adenosine receptors.

Finally, the RDF model not only overtakes the other models in the statistical parameters of the regression but more importantly in the stability to the inclusion–exclusion of compounds as measured by the correlation coefficient and standard deviation of the cross-validation. Because of that these statistics of the leave-one-out cross validation might be considered as a good measurement of the predictability of the models. As can be seen in the Table 5 the value of the determination coefficient of leave-one-out cross validation for the model obtained with the RDF ($q^2 = 0.786$) was the highest for the all analyses model proving the predict power of this approach and the stability of the model. These results have shown that the RDF descriptors not only explain the experimental data, but seem to be the best one in doing so.

## 4. Concluding remarks

We have shown that the RDF approach is able to describe the $A_{2\ A}$ agonist activity of different $N^6$–arylcarbamoyl, 2-arylalkynyl-$N^6$–arylcarbamoyl, and $N^6$–carboxamido of adenosine derivatives. In fact, we have developed a model for pre-

Table 5
The statistical parameters of the lineal regressions models obtained for the five kinds of descriptors and others reported in the literature for the same training set

| Descriptors | Variables | S | $R^2$ | F | $q^2$ | $S_{cv}$ |
|---|---|---|---|---|---|---|
| Galvez Topological Charges indexes[a] | GGI4, GGI6, GGI7, JGI4, JGI6 | 0.591 | 0.450 | 3.759 | 0.011 | 0.801 |
| BCUT | BEHm3, BEHm6, BELv7, BEHe7, BELp7 | 0.463 | 0.663 | 9.014 | 0.404 | 0.615 |
| Geometrical[a] | AGDD, SPAN, FDI, G(O..O), G(Cl..Cl) | 0.436 | 0.701 | 10.760 | 0.515 | 0.555 |
| 2D autocorrelations | ATS6v, ATS1e, ATS7e, MATS3e, GATS7e | 0.402 | 0.745 | 13.437 | 0.582 | 0.515 |
| Topological[a] | X5 A, PW5, IC2, IC5, T(N..I) | 0.474 | 0.646 | 8.396 | 0.391 | 0.621 |
| Randić Molecular profile | DP01, DP02, DP04, DP05, DP07 | 0.544 | 0.534 | 5.268 | 0.204 | 0.711 |
| WHIM[a] | G3v, G3e, G1p, Tv, Ds | 0.570 | 0.487 | 4.381 | 0.130 | 0.743 |
| 3D-MORSE | Mor31u, Mor05 m, Mor14 m, Mor02v, Mor10p | 0.385 | 0.767 | 15.128 | 0.665 | 0.461 |
| **RDF** | **RDF100 m, RDF135 m, RDF140 m, RDF130v, RDF075p** | **0.299** | **0.849** | **24.812** | **0.786** | **0.368** |
| GETAWAY[a] | HGM, HATS3u, R6u+, R1v+, R5e+ | 0.375 | 0.778 | 16.149 | 0.681 | 0.464 |

[a] Models obtained in a previous work [10].

dicting this effect of a data set of 29 compounds, which is both statistically and chemically sounded. This model explains more than 84.5% of the variance in the experimental activity with a good predictive power. These features are significantly better than that obtained from other different methodologies. In addition, according to the model interpretation, the major determinants for the affinity at $A_{2\ A}$ adenosine receptors would be bulkiness of substituents in different positions and the stereoselectivity of the adenosine analogues.

## Acknowledgements

We thank the Universidad de Vigo and the Xunta de Galicia (PGIDT01PX130114PR) for financial support. Marta Teijeira thanks the Xunta de Galicia for a Parga Pondal contract.

## References

[1] C.E. Müller, T. Scior, Pharm. Acta Helv. 68 (1993) 77–111.
[2] V. Ralevic, G. Burnstock, Pharmacol. Rev. 50 (1998) 413–492.
[3] M.P. DeNinno, Annu. Rep. Med. Chem. 33 (1998) 111–121.
[4] G. Cristalli, C. Lambertucci, S. Taffi, S. Vittori, R. Volpini, Curr. Top. Med. Chem. 3 (2003) 387–401.
[5] B. Cacciari, G. Pastorin, G. Spalluto, Curr. Top. Med. Chem. 3 (2003) 403–411.
[6] M.P. González, C. Teran, Bioorg. Med. Chem. Lett. 14 (2004) 3077–3079.
[7] M.P. González, C. Teran, Bull. Math. Biol. 66 (2004) 907–920.
[8] M.P. González, C. Teran, Bioorg. Med. Chem. 12 (2004) 2985–2993.
[9] M.P. González, C. Terán, M. Teijeira, P. Besada, Bioorg. Med. Chem. Lett. 10 (2005) 2641–2645.
[10] M.P. González, C. Terán, M. Teijeira, M.J. Gonzalez-Moa, Eur. J. Med. Chem. (in press).
[11] R. Todeschini, V. Consonni, Handbook of molecular descriptors, Wiley-VCH, Weinheim, Germany, 2000.
[12] M.P. González, C. Terán, Y. Fall, M. Teijeira, P. Besada, Bioorg. Med. Chem. 13 (2005) 601–608.
[13] P.G. Baraldi, B. Cacciari, M.J. Pineda de las Infantas, R. Romagnoli, G. Spalluto, R. Volpini, S. Constanza, S. Vittori, G. Cristalli, N. Merman, P. Kyung-Sung, J. Xiao-duo, K.A. Jacobson, J. Med. Chem. 41 (1998) 3174–3185.
[14] R. Todeschini, V. Consonni, M. Pavan, Dragon. Software version 2.1. 2002.
[15] J.J.P. Stewar, (1990) MOPAC manual, 6th ed., p 189, Frank J. Seiler Research Laboratory, U. S. Air Force academy, Colorado Springs, CO.
[16] M. Randić, J. Chem. Inf. Comput. Sci. 31 (1991) 311–320.
[17] M. Randić, N. J. Chem. 15 (1991) 517–525.
[18] M. Randić, J. Mol, Struct. (Teochem) 233 (1991) 45–59.
[19] B. Lučić, S. Nikolić, N. Trinajstić, D. Jurić, J. Chem. Inf. Comput. Sci. 35 (1995) 532–538.
[20] StatSoft, Statistica 6.0. Copyright _1984–2002.
[21] D.E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, Massachusetts, MA, 1989.
[22] R. Leardi, R. Boggia, M. Terrile, J. Chemom. 6 (1992) 267-281.
[23] D.M. Hawkins, J. Chem. Inf. Comput. Sci. 44 (2004) 1–12.
[24] C.E. Müller, T. Scior, Pharm Act. Helv. 68 (1993) 77–111.
[25] E.W. van Tilburg, J. von Frijtag Drabbe, M. de Groote, A.P. IJzerman, J. Med. Chem. 45 (2002) 420–429.
[26] G. Cristalli, E. Camaioni, S. Vittori, P.A. Borea, A. Conti, S. Dionisotti, E. Ongini, A. Monopoli, J. Med. Chem. 38 (1995) 1462–1472.
[27] M. Viziano, E. Ongini, A. Conti, C. Zocchi, M. Seminati, D. Pocar, J. Med. Chem. 38 (1995) 3581–3585.
[28] T. Gungor, P. Malabre, J.M. Teulon, F. Camborde, J. Meignen, F. Hertz, A. Virone-Oddos, F. Caussade, A. Cloarect, J. Med. Chem. 37 (1994) 4307–4316.
[29] H. Homma, Y. Watanabe, T. Abiru, T. Murayama, Y. Nomura, A. Matsuda, J. Med. Chem. 35 (1992) 2881–2890.
[30] K.A. Jacobson, J. Xiao-duo, L. An-Hu, N. Melman, M.A. Siddiqui, S. Kye-Jung, V.E. Marquez, R.G. Ravi, J. Med. Chem. 43 (2000) 2196–2203.
[31] K.N. Klotz, E. Camaioni, R. Volpini, S. Kachler, S. Vittori, G. Cristalli, Naunyn Schmiedebergs Arch. Pharmacol. 360 (1999) 103–108.