Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 13 (2005) 601–608
Bioorganic & Medicinal Chemistry

A radial distribution function approach to predict A_{2B} agonist effect of adenosine analogues

Maykel Pérez González,[a,b,*] Carmen Terán,[c] Yagamare Fall,[c] Marta Teijeira[c] and Pedro Besada[c]

[a]*Unit of Services, Department of Drug Design, Experimental Sugar Cane Station 'Villa Clara-Cienfuegos', Ranchuelo, Cuba*
[b]*Department of Drug Design, Chemical Bioactives Center, Central University of Las Villas, Santa Clara, Villa Clara, Cuba*
[c]*Department of Organic Chemistry, Vigo University, CP 36200, Vigo, Spain*

Received 24 September 2004; revised 27 October 2004; accepted 30 October 2004
Available online 10 December 2004

Abstract—The radial distribution function (RDF) approach has been applied to the study of the A_{2B} agonist effect of a set of 89 adenosine analogues reported with this activity. A model able to describe more than 70% of the variance in the experimental activity was developed with the use of the mentioned approach. In contrast, none of the eleven different approaches including the use of Constitutional, Topological, Molecular walk count, BCUT, Galvez topological charge indices, 2D autocorrelations, Randić molecular profiles, Geometrical, 3D Morse, WHIM and GETAWAY descriptors was able to explain more than 47% of the variance in the mentioned property with the same number of descriptors.

1. Introduction

Adenosine regulates a great variety of physiological functions in nervous, cardiovascular, renal, immune and other systems of organism through specific cell membrane receptors.[1] Four different adenosine receptors subtypes (A_1, A_{2A}, A_{2B} and A_3) have been defined on the basis of biological experiments and receptor cloning and all may be coupled to adenylate cyclase via G proteins.[2] The A_1 and A_3 subtypes inhibit the adenylate cyclase activity diminishing the production of second messenger cyclic AMP.[3,4] On the contrary, A_{2A} and A_{2B} subtypes stimulate adenylate cyclase to produce cyclic AMP.

A_1 and A_{2A} receptors are activated by nanomolar concentrations of adenosine. Whereas A_{2B} and A_3, which are low affinity receptors, become activated only when adenosine levels increase into the micromolar range during periods of inflammation, hipoxia or ischemia.[5–7]

Quantitative determination of the tissue distribution of A_{2B} adenosine receptors is not possible at this time due to the lack of potent radioligands with subtype selectivity. In many systems the presence of this adenosine receptors subtype was confirmed by stimulation of adenylate cyclase activity.[8] The widespread occurrence of the A_{2B} adenosine receptor has opened up a huge number of therapeutical applications, the treatment of asthma being the most prominent one. Several reports point to important A_{2B}-mediated actions in cardiovascular effects of adenosine,[9,10] and a future therapeutic purpose for adenosine A_{2B} receptor agonist could thus be the treatment of septic shock.[8]

Although a large number of adenosine derivatives have been tested at the adenosine A_{2B} receptor, no potent and selective agonists are known so far. 5′-N-ethylcarboxamidoadenosine (NECA, **1**, Fig. 1), a high affinity

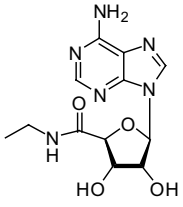


Figure 1. The molecular structure of 5′-N-ethylcarboxamidoadenosine (NECA).

Keywords: Dragon; A_{2B} adenosine receptors; QSAR; RDF descriptors.
* Corresponding author. Tel.: +53 4228 1473; fax: +53 4228 1130; e-mail: mpgonzalez76@yahoo.es

0968-0896/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2004.10.063

ligand at all adenosine receptors, is one of more potent agonist on this subtype with an EC_{50} in the low micromolar range.[11,12]

In this research, we tried to describe the quantitative structure–activity relationship (QSAR) study of A_{2B} agonist effect of different families of adenosine analogues. The QSAR studies are a powerful method for the design of bioactive compounds and the prediction of corresponding activity with physical and chemical properties.[13–16]

In the context of in silico methods for modeling physicochemical and biological properties of chemicals the radial distribution function (RDF) approach has been introduced.[17]

The successful application of this theoretical approach to deriving the 3D structure of organic molecules from their infrared spectra[18,19] have inspired us to test and/or validate the RDF descriptors applicability in assessing discoveries of new drugs.

Thereby, the aim of this work is to find rationality in the search of novel A_{2B} adenosine receptors agonist compounds using RDF approach. Secondly, to continue the validation of this method in describing biological activity of heterogeneous series of compounds and its comparison with other methodologies in order to demonstrate its value as a good QSAR model.

2. The radial distribution function approach

The 3D coordinates of the atoms of molecules can be transformed into a structure code that has a fixed number of descriptors irrespective of the size of a molecule. This task is performed by a structure coding technique referred to as radial distribution function code (RDF code).[17,18] In general, there are some prerequisites for a structure code:

- independence from the number of atoms, that is, the size of a molecule,
- unambiguity regarding the three-dimensional arrangement of the atoms, and
- invariance against translation and rotation of the entire molecule.

Formally, the radial distribution function of an ensemble of N atoms can be interpreted as the probability distribution to find an atom in a spherical volume of radius r.[19] The equation represents the radial distribution function code as it is used in this investigation:

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

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. The exponential term contains the distance r_{ij} between the atoms i and j and the smoothing parameter B that defines the probability distribution of the individual distances. $g(r)$ was calculated at a number of discrete points with defined intervals.

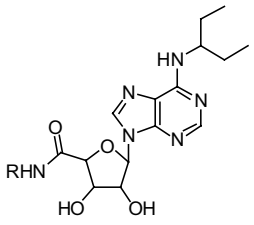
The atomic properties A_i and A_j used in this equation enable the discrimination of the atoms of a molecule for almost any property that can be attributed to an atom. Such distribution function provides, besides information about interatomic distances in a whole molecule, the opportunity to gain access to other valuable information, for example, bond distance, ring types, planar and nonplanar systems and atoms types. This fact is a most valuable consideration for a computer-assisted code elucidation. The radial distribution function in this form meets the entire requirement mentioned above, especially invariance against linear translations.

3. Data sets and computational strategies

In the present study was used a data set of 89 adenosine analogues for which their activities are reported in the literature.[12] The parameter studied was the cyclic AMP (cAMP) production, in Chinese Hamster Ovary cells (CHO) expressing human A_{2B} receptors, shown as percentage of the production by 100 μM NECA. The structures of analyzed compounds as well as the calculated and experimental values of their activity are shown in Tables 1–8 and in Figures 2, 3.

DRAGON[20] computer software was employed to calculate the molecular descriptors. In this way we carry out geometry optimization calculations for each compound of this study using the quantum chemical semi-empirical method AM1[21] included in MOPAC 6.0.[22] Twelve models were developed using the computer software Dragon,[20] calculating the Constitutional, Topological, Molecular walks counts, BCUT, Galvez topological

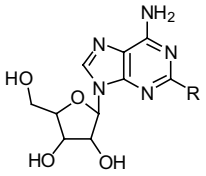
Table 1. Activities of N^6-(3-pentyl), 5′-*N*-disubstituted carboxamidoadenosides on cAMP production in CHO cells expressing human A_{2B} receptors



Compounds	R	% Stimulation (obs)	% Stimulation (pred.)
1	Methyl	0.63	0.27
2	Allyl	0.41	0.33
3	Isopropyl	0.69	0.40
4	Cyclopropyl	0.78	0.22
5	3-Pentyl	0.06	0.25
6	Benzyl	0	0.10
7	2-NH_2-ethyl	0.01	0.05

Compounds were tested at 100 μM. Activities are given as percentage of stimulation with 100 μM NECA.

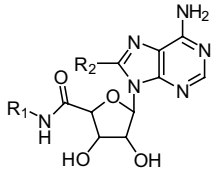
Table 2. Activities of C2-substituted adenosines on cAMP production in CHO cells expressing human A_{2B} receptors



Compounds	R	% Stimulation (obs)	% Stimulation (pred)
8	SH	0.01	0.02
9	SCH_2CH_2CN	0	0.05
10	S-cyclohexyl	0.03	0.005
11	S-hexadecyl	0.05	0.07
12	S-Bn	0.03	0.05
13	NH–Ph	0.10	0.02
14	NH–PhEt	0.03	0.01
15	F	0.23	0.12
16	I	0.02	0.04

Compounds were tested at 100 μM. Activities are given as percentage of stimulation with 100 μM NECA.

Table 3. Activities of C8-substituted NECA and MECA analogues on cAMP production in CHO cells expressing human A_{2B} receptors



Compounds	R_1	R_2	% Stimulation (obs)	% Stimulation (pred)
17	Et	Br	0	0.04
18	Et	NH_2	0.09	0.09
19	Et	EtNH	0	0.02
20	Et	Me_2N	0	0.002
21	Me	MeNH	0	0.05
22	Me	Me_2N	0	0.02

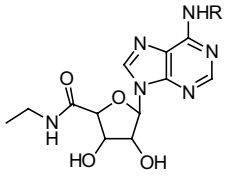
Compounds were tested at 100 μM. Activities are given as percentage of stimulation with 100 μM NECA.

charge indices, 2D autocorrelations, Randić molecular profiles, Geometrical, RDF, 3D-MORSE, WHIM and GETAWAY descriptors.[23] All statistical analysis and data exploration was carryied out using the Statistic 6.0.[24] The most significant parameters were identified from the data set using Forward stepwise regression methods,[25] where the independent variables are individually added or deleted from the model at each step of the regression depending on the Fisher ratio values selected to enter and to remove until the 'best' model is obtained.

Examining the regression coefficient, the standard deviation, the significance and the number of variables in the equation determined the quality of the model.

In addition, the regression models obtained were validated by calculating q^2 values. The q^2 are obtained from 'leave-one-out' (LOO) and leave-group-out (LGO) testing, also known as cross-validation. A data point 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 has

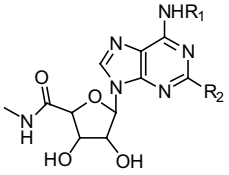
Table 4. Activities of N^6-substituted NECA analogues on cAMP production in CHO cells expressing human A_{2B} receptors



Compounds	R	% Stimulation (obs)	% Stimulation (pred)
23	3-CH_3O–phenyl	0.63	0.65
24	3-CH_3–benzyl	0.16	0.33
25	4-CH_3O–benzyl	0.14	0.28
26	3-CH_3O–benzyl	0.91	0.56
27	3-F-benzyl	0.80	0.49
28	3-Cl–benzyl	0.84	0.29
29	3-Br–benzyl	0.95	0.89
30	3-I–benzyl	0.71	0.57
31	2-NO_2–benzyl	0.87	0.42
32	3-NO_2–benzyl	0.57	0.03
33	4-NO_2–benzyl	0.41	0.38
34	Benzyl	0.24	0.36

Compounds were tested at 100 μM. Activities are given as percentage of stimulation with 100 μM NECA.

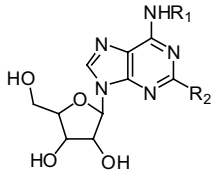
Table 5. Activities of N^6-substituted MECA analogues on cAMP production in CHO cells expressing human A_{2B} receptors



Compounds	R_1	R_2	% Stimulation (obs)	% Stimulation (pred)
35	3-CH_3-benzyl	H	0.13	0.05
36	3-CF_3-benzyl	H	0	−0.13
37	4-Cl-benzyl	H	0.01	0.06
38	3-Br-benzyl	H	0.16	0.17
39	4-Br-benzyl	H	0.03	0.07
40	3-I-benzyl	H	0.10	0.20
41	3-I-benzyl	Cl	0.03	0.11
42	3-I-benzyl	S-Me	0	−0.007
43	3-I-benzyl	NH-Me	0	−0.03
44	4-SO_3H-benzyl	H	0	0.06
45	3-I-4-NH_2-benzyl	H	0.18	0.36
46	3-NO_2-benzyl	H	0.04	0.10
47	(*R*)-1-Phenyl-ethyl	H	0.23	0.37
48	(*S*)-1-Phenyl-ethyl	H	0.15	0.34
49	2-Methyl-furan	H	0	0.17
50	Benzyl	H	0.15	−0.03

Compounds were tested at 100 μM. Activities are given as percentage of stimulation with 100 μM NECA.

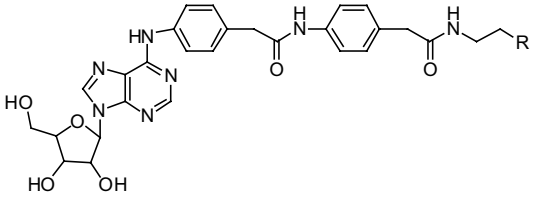
Table 6. Activities of N^6-substituted adenosines on cAMP production in CHO cells expressing human A_{2B} receptors



Compounds	R_1	R_2	% Stimulation (obs)	% Stimulation (pred)
51		H	0.03	0.10
52	4-NH_2-benzyl	H	0.12	0.21
53	2-CH_3-benzyl	H	0.21	0.11
54	2-Cl-benzyl	H	0.35	0.07
55	3-I-benzyl	H	0.18	0.02
56	3-I-benzyl	Cl	0.06	0.003
57	3-I-benzyl	NH_2	0.06	0.05
58	4-SO_3H-phenylpropyl	H	0.01	0.23
59	4-SO_3H-phenylbutyl	H	0.02	0.01
60	4-SO_3H-phenyldecyl	H	0.04	0.09
61	*n*-Butyl	H	0.23	0.08
62	*n*-Decyl	H	0.03	0.15
63	4-NH_2-butyl	H	0.10	−0.12
64	10-NH_2-decyl	H	0	0.07
65	*N*,*N*-Dimethyl	H	0	0.03
66	*N*,*N*-Dipropyl	H	0	0.07
67	NH_2	H	0	0.26
68	(Adenosine-N^6-yl)butyl	H	0.15	0.08
69	(Adenosine-N^6-yl)decyl	H	0.13	0.11

Compounds were tested at 100 μM. Activities are given as percentage of stimulation with 100 μM NECA.

Table 7. Activities of N^6-functionalized congeners on cAMP production in CHO cells expressing human A_{2B} receptors



Compounds	R_1	% Stimulation (obs)	% Stimulation (pred)
70	NHAc	0.13	0.78
71	$N(CH_3)_3$	0.17	0.09
72	OH	0.19	0.04

Compounds were tested at 100 μM. Activities are given as percentage of stimulation with 100 μM NECA.

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,[26] and is therefore a more meaningful summary statistic for QSAR models. Analysis of residuals from the regression equations was used to identify outliers, which were removed to aid analysis.[27]

Table 8. Activities of deazaadenosine analogues on cAMP production in CHO cells expressing human A_{2B} receptors

Compounds	Analogue	% Stimulation (obs)	% Stimulation (pred)
73	1-Deazaadenosine	0.10	0.04
74	2-Chloro-1-deazaadenosine	0.20	0.34
75	3-Deazaadenosine	0.01	0.05
76	1,3-Dideazaadenosine	0.01	0.09
77	7-Deazaadenosine	0	0.06
78	1,7-Dideazaadenosine	0	0.06

Compounds were tested at 100 μM. Activities are given as percentage of stimulation with 100 μM NECA.

4. Result and discussion

In this work, the model selection was subjected to the principle of parsimony.[26] Then, we chose a function with high statistical signification but having very few parameters (descriptors) as possible.

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

$$A = 0.095 + 0.051 \cdot RDF130u - 0.131 \cdot RDF140v + 0.149 \cdot RDF050v - 0.146 \cdot RDF050p + 0.317 \cdot RDF140p - 0.127 \cdot RDF155p \quad (1)$$

$$N = 89 \quad S = 0.131 \quad R^2 = 0.653 \quad F = 25.229$$
$$p < 0.0001 \quad q^2 = 0.582 \quad S_{cv} = 0.152$$

where A is the studied property, in our case, the activity on cAMP production in CHO cells expressing human A_{2B} receptors, 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 the correlation coefficient of the cross-validation, p is the significance of the variables in the model and S_{cv} is the standard deviation of the cross-validation. The meaning of the variables included in the model appears in Table 10.

From a mechanistic point of view, inclusions of a variable into a QSAR suggest that the variable is related directly to how the chemistry of a molecule influences biological activity. Whilst biological activity is certainly a multivariate process, there are not an infinite number of controlling factors and a QSAR would be expected to have no more variables than factors controlling biological activity. In this sense, a model with seven variables was carried out.

$$A = 0.075 + 0.050 \cdot RDF130u - 0.124 \cdot RDF140v + 0.169 \cdot RDF050v - 0.169 \cdot RDF050p + 0.289 \cdot RDF140p - 0.114 \cdot RDF155p + 0.006 \cdot RDF065m \quad (2)$$

$$N = 89 \quad S = 0.129 \quad R^2 = 0.666$$
$$F = 23.104 \quad p < 0.0001 \quad q^2(\mathrm{LOO}) = 0.585$$
$$q^2(\mathrm{LGO}) = 0.558 \quad S_{cv} = 0.153$$

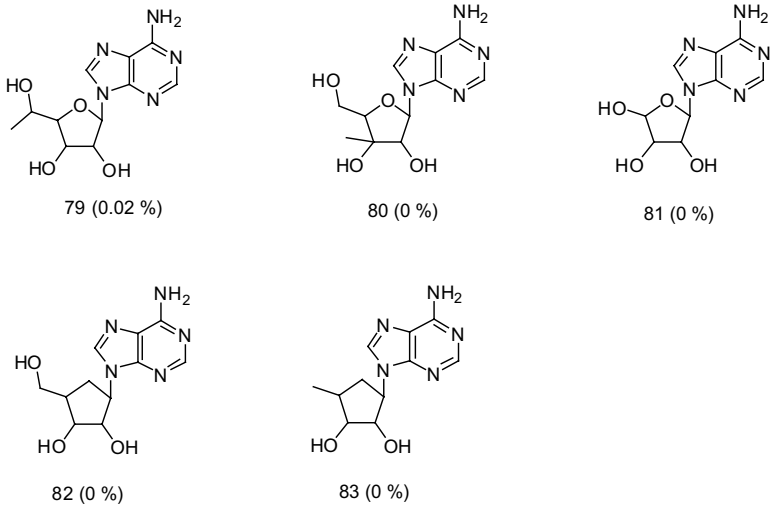


Figure 2. Activities of ribose modified analogues on cAMP production in CHO cells expressing human A_{2B} receptors. Compounds were tested at 100 μM. Activities are given as percentage of stimulation with 100 μM NECA.

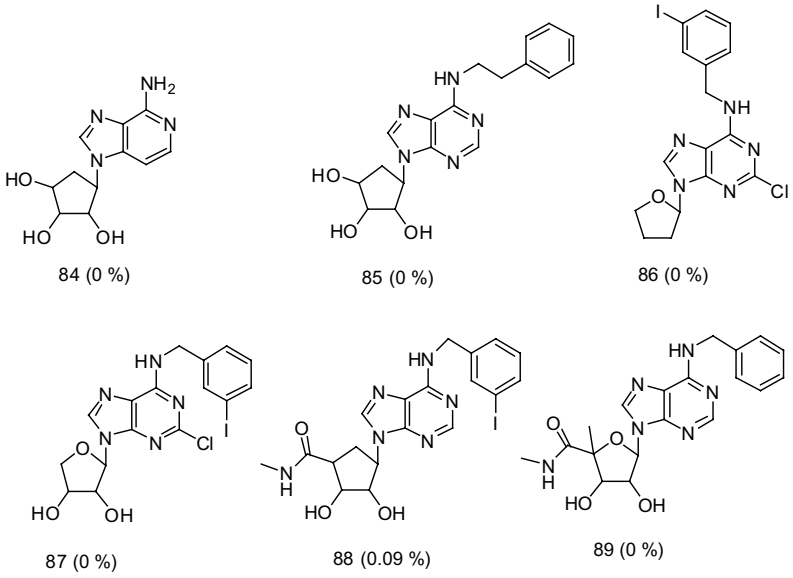


Figure 3. Activities of variously modified analogues on cAMP production in CHO cells expressing human A_{2B} receptors. Compounds were tested at 100 μM. Activities are given as percentage of stimulation with 100 μM NECA.

Considering the evolution of R^2 and the q^2 coefficients, Eq. 1 was taken as the optimal QSAR model for this kind of descriptors. The model of Eq. 2 presents similar predictive capabilities when adding more parameters, therefore did not justify the increase of equation complexity.

On the other hand, 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.[27] 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[28] it is noted that it is not acceptable to remove the outlier repeatedly from a QSAR analysis simply to improve a correlation. In the current work, the compounds **70** and **82** present large residuals and should be considered as outliers. At removal of these compounds from the training set the next equation is obtained:

$$A = 0.095 + 0.051 \cdot RDF130u - 0.131 \cdot RDF140v + 0.149 \cdot RDF050v - 0.146 \cdot RDF050p + 0.317 \cdot RDF140p - 0.127 \cdot RDF155p \quad (3)$$

$$N = 87 \quad S = 0.124 \quad R^2 = 0.701 \quad F = 29.702$$
$$p < 0.0001 \quad q^2 = 0.648 \quad S_{cv} = 0.132$$

As can be seen the R^2 and q^2 increase on deleting these compounds, therefore should be considered as potential outliers. In addition, a result of LGO demonstrated that the q^2 of this test ($q^2 = 0.625$) improved the below results. This behavior may be due to the presence of a large substituent at N^6 in these compounds the consequence of which being a rather different mechanism of action of these compounds, hence changing their usual activity.[10] However, before making the interpretation of the model we need to orthogonalize the molecular descriptors included in such models due to the high intercorrelation existing between some of them as shown in Table 9.

4.1. Orthogonalization of descriptors

The orthogonalization process of molecular descriptors was introduced by Randić 10 years ago as a way of improving the statistical interpretation of the model built by using interrelated indices.[29–33] The main philosophy of this approach is to avoid the exclusion of descriptors on the basis of its collinearity with other variables previously included in the model. The acceptable level of collinearity to avoid is a more subjective issue. In view of the authors, the collinearity of the variables should be as low as possible because the interrelatedness among the different descriptors can result in highly unstable regression coefficient, which makes impossible to know the relative importance of an index and underestimates the utility of the regression coefficient model.

The Randić method of orthogonalization has been described in detail in several publications.[29–32] Thus, we will give a general overview here. The first step in orthogonalizing the molecular descriptors in model (3) is to select the appropriate 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. The first variable (RDF130u) is taken as the first orthogonal descriptors ^1O(RDF130u), and the second one is orthogonalized with respect to it by taking the residual of its correlation with ^1O(RDF130u). The process is repeated until all the variables are completely orthogonalized, and the orthogonal variables are then used to obtain the new model. In Table 10 we resume the results of the orthogonalization of molecular descriptors included in model (3).

4.2. Interpretation and comparison with other approach

As we previously pointed out, 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 eleven models using the same data set that was included in the RDF QSAR model. The results obtained with the use of Constitutional, Topological, Molecular walks counts, BCUT, Galvez topological charge indices, 2D autocorrelations, Randić molecular profiles, Geometrical, RDF, 3D-MORSE, WHIM and GETAWAY descriptors are given in Table 11.

As can be seen there are remarkable differences concerning the explanation of the experimental variance given

Table 11. The statistical parameters of the lineal regressions models obtained for the twelve kinds of descriptors

Descriptors	Variables	S	R^2	F	$p<$
Constitutional	5	0.235	0.155	3.050	0.014
Topological	10	0.176	0.552	9.626	0.0001
Molecular walk count	10	0.226	0.263	2.788	0.0001
BCUT	10	0.167	0.598	11.624	0.0001
Galvez topological charge indices	6	0.231	0.193	3.288	0.005
2D autocorrelations	12	0.179	0.548	7.684	0.0001
Randić molecular profiles	11	0.231	0.241	2.225	0.021
Geometrical	8	0.240	0.309	4.474	0.0001
RDF	**6**	**0.124**	**0.701**	**29.702**	**0.0001**
3D-MORSE	12	0.188	0.505	6.461	0.0001
WHIM	11	0.207	0.391	4.505	0.0001
GETAWAY	12	0.186	0.515	6.728	0.0001

Table 9. Correlation coefficients between the six most significant variables

	RDF130u	RDF140u	RDF050v	RDF050p	RDF140p	RDF155p
RDF130u	1.00	0.81	0.62	0.62	0.77	0.64
RDF140u		1.00	0.45	0.44	0.98	0.86
RDF050v			1.00	0.99	0.38	0.22
RDF050p				1.00	0.38	0.22
RDF140p					1.00	0.91
RDF155p						1.00

Table 10. Regression coefficients in model (3) for orthogonal molecular descriptors

RDF130u	RDF140u	RDF140p	RDF155p	RDF050p	RDF050v	Intercept	R^2	S
0.6860						0.1421	0.3507	0.1753
0.6860	8.3682					0.1421	0.5769	0.1423
0.6860	8.3682	−88.9281				0.1421	0.6205	0.1356
0.6860	8.3682	−88.9281	−4.2691			0.1421	0.6587	0.1293
0.6860	8.3682	−88.9281	−4.2691	−0.0087		0.1421	0.6652	0.1289
0.6860	8.3682	−88.9281	−4.2691	−0.0087	−6.4163	0.1421	0.7010	0.1241

Table 12. The statistical parameters of the lineal regressions models obtained with six variables for the twelve kinds of descriptors

Descriptors	Variables	S	R^2	F	q^2
Constitutional	nCIC, RBF, nDB, nS, nR05, nR06	0.228	0.212	3.689	0.115
Topological	X4A, X5A, Lop, IC3, IC5, T(N..S)	0.192	0.440	10.763	0.314
Molecular walk count	MWC05, MWC08, MWC10, SRW05, SRW07, SRW10	0.231	0.188	3.164	0.056
BCUT	BEHm8, BELm1, BELm2, BEHe7, BEHp5, BEHp7	0.188	0.462	11.751	0.235
Galvez topological charge indices	GGI4, GGI6, GGI7, JGI1, JGI8, JGI9	0.231	0.193	3.288	0.117
2D autocorrelations	MATS3m, MATS1e, MATS2e, MATS1p, GATS8v, GATS8v	0.187	0.471	12.211	0.321
Randić molecular profiles	DP04, DP06, SP01, SP02, SP03, SHP2	0.233	0.178	2.966	0.098
Geometrical	J3D, MAXDN, MAXDP, MEcc, G(N..N), G(O..I)	0.237	0.289	5.557	0.203
RDF	**RDF130u**, **RDF140v**, **RDF050v**, **RDF050p**, **RDF140p**, **RDF155p**	**0.124**	**0.701**	**29.702**	**0.648**
3D-MORSE	Mor16m, Mor18m, Mor20m, Mor08e, Mor12e, Mor10p	0.307	0.355	7.526	0.236
WHIM	E1v, L2e, G3m, L2p, E1s, Gm	0.254	0.305	6.016	0.205
GETAWAY	R3u+, R1m+, R3v+, R3e+, R5e+, R7e+	0.208	0.347	7.266	0.257

by these models compared to the RDF one. While the RDF QSAR model explains more than 65% of activities the rest of the models are unable to explain more than 55% of such variance. The model obtained using the Topological descriptors explain the 55.2% of the data variance, but this model need 10 variables for a relation cases/variables of 8.9. In the case of RDF methodology only six variables are needed for relation cases/variables of 14.83, thus showing the great superiority of this model over those generated with the Dragon software. In order to compare the models statistic parameters, we carried out the models using for all of them the same number of variables. As can be seen, in Table 12, the results obtained with the model using the RDF descriptors are much better than those of other methodologies, which are unable of explain more than 40% of the data variance. 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.

The RDF model not only overtakes the other eleven 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 the structural variability of the compounds in the data set 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 Table 12 the value of the determination coefficient of leave-one-out cross validation for the model obtained with the RDF ($q^2 = 0.582$) was the highest for 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.

5. Concluding remarks

We have shown that the RDF approach is able to describe the A_{2B} agonist activity of different family of adenosine analogues. In fact, we have developed a model for predicting this effect of a data set of 89 compounds, which is both statistically and chemically sounded. This model explains more than 70% of the variance in the experimental activity with an acceptable predictive power. These features are significantly better than that obtained from eleven other different methodologies.

Acknowledgements

Maykel Pérez acknowledges Professors Gloria Cristalli, Sauro Vittori and Rosaria Volpini for sending him valuable information for the development of this paper. In this sense, we acknowledge the comments and contribution of Prof. Dr. Johann Gasteiger for his contribution to increase the quality of the present article.

References and notes

1. Jacobson, K. A.; van Galen, P. J.; Williams, M. *J. Med. Chem.* **1992**, *35*, 407.
2. Klinger, M.; Freissmuth, M.; Nanoff, C. *Cell. Signal.* **2002**, *14*, 99.
3. Libert, F.; Schiffman, S. N.; Lefort, A.; Parmentier, M.; Gerard, C.; Dumont, J. E.; Vanderhaeghen, J.; Vassart, G. *EMBO J.* **1991**, *10*, 1677.
4. Zhou, Q. Y.; Li, C. Y.; Olah, M. E.; Johnson, R. A.; Stiles, G. L.; Civelli, O. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7432.
5. Jacobson, K. A. *Drugs of future* **1995**, *20*, 689.
6. Olah, M. E.; Stiles, G. L. *Annu. Rev. Pharmacol. Toxicol.* **1995**, *35*, 581.
7. Maenhaut, C.; Van sande, J.; Liebert, F.; Abramowicz, M.; Parmentier, M.; Vanderhaeghen, J. J.; Dumont, J.; Vassart, G.; Schiffmann, S. *Biochem. Biophys. Res. Commun.* **1990**, *173*, 1169.
8. Volpini, R.; Costanzi, S.; Vittori, S.; Cristalli, G.; Klotz, K.-N. *Curr. Top. Med. Chem.* **2003**, *3*, 427.
9. Dubey, R. K.; Gillespie, D.; Mi, Z. C.; Jackson, E. K. *Circulation* **1997**, *96*, 2656.
10. Morrison, R. R.; Talukder, M. A. H.; Ledent, C.; Mustafa, S. J. *Am. J. Physiol.* **2002**, *282*, H437.
11. Zwart, M.; Link, R.; Kunzel, J. K. V.; Cristalli, G.; Jacobson, K. A.; Townsend-Nicholson, A. P.; Ijzerman, A. *Nucleos. Nucleot.* **1998**, *17*, 969.
12. Poulsen, S.; Quinn, R. *Bioorg. Med. Chem.* **1998**, *6*, 619.
13. González, M. P.; Morales, A. H.; González, H. D. *Polymer* **2004**, *45*, 2073.
14. García-Domenech, R.; Gálvez, J.; Moliner, R.; García-March, F. *Drug. Invest.* **1991**, *3*, 344.

15. González, M. P.; González, H. D.; Molina, R. R.; Cabrera, M. A.; Ramos de Armas, R. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1192.
16. González, M. P.; Morales, A. H. *J. Comput. Aided Mol. Des.* **2003**, *10*, 665.
17. Gasteiger, J.; Sadowski, J.; Schuur, J.; Selzer, P.; Steinhauer, L.; Steinhauer, V. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1030.
18. Gasteiger, J.; Schuur, J.; Selzer, P.; Steinhauer, L.; Steinhauer, V.; Fresenius, J. *Anal. Chem.* **1997**, *359*, 50.
19. Hemmer, M. C.; Steinhauer, V.; Gasteiger, J. *Vib. Spectrosc.* **1999**, *19*, 151.
20. Todeschini, R.; Consonni, V.; Pavan, M. (**2002**) Dragon. Software version 2.1.
21. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
22. Stewar, J. J. P. *MOPAC Manual*, 6th ed.; Seiler Research Laboratory: US Air Force academy, Colorado Springs, CO, 1990; p 189.
23. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, Germany, 2000.
24. STATISTICA version. 6.0 (**2002**) Statsoft.
25. Kowalski, R. B.; Wold, S. Pattern Recognition in Chemistry. In *Handbook of Statistics*; Krishnaiah, P. R., Kanal, L. N., Eds.; North Holland: Amsterdam, 1982; pp 673–697.
26. Hawkins, D. M. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1.
27. Lipnick, R. L. *Sci. Total Environ.* **1991**, *109*, 131.
28. Devillers, J.; Lipnick, R. L. Practical applications of regression analysis in environmental QSAR studies. In *Practical Applications of Quantitative Structure–Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, K., Devillers, J., Eds.; Kluwer: Dordrecht, 1990; pp 129–143.
29. Randić, M. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 311.
30. Randić, M. *New. J. Chem.* **1991**, *15*, 517.
31. Randić, M. *J. Mol. Struct. (Theochem)* **1991**, *233*, 45.
32. Lucic, B.; Nikolic, S.; Trinajstic, N.; Juric, D. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 532.
33. Klein, D. J.; Randić, M.; Babic, D.; Lucic, B.; Nikolic, S.; Trinajstic, N. *Int. J. Quantum Chem.* **1997**, *63*, 215.