Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 15 (2005) 2641–2645
Bioorganic & Medicinal Chemistry Letters

Geometry, topology, and atom-weights assembly descriptors to predicting A_1 adenosine receptors agonists

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

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

Received 7 January 2005; revised 2 March 2005; accepted 7 March 2005
Available online 9 April 2005

Abstract—The GEometry, Topology, and Atom-Weights AssemblY (GETAWAY) approach has been applied to the study of the A_1 adenosine receptors agonist effect of 32 adenosine analogues: N^6-arylcarbamoyl, 2-arylalkynyl-N^6-arylcarbamoyl, and N^6-carboxamido derivatives. A model, able to describe more than 77% of the variance in the experimental activity, was developed with the use of the above mentioned approach. Five different approaches (Topological, Galvez Topological Charges indexes, Randić Molecular Profiles, Geometrical, and WHIM descriptors) failed to give satisfactory models ($R^2 = 0.70$) for this property with the same number of variables in the equation. Although statistically significant models were derived containing descriptors other than GETAWAY, the best fitted out model was still found with these descriptors.

1. Introduction

Extracellular adenosine acts as a local modulator at four receptor subtypes: A_1, A_{2A}, A_{2B}, and A_3. These receptor subtypes are G protein coupled and are able to modulate, either positively or negatively, the adenylate cyclase activity. While A_1 and A_3 adenosine receptor subtypes are coupled to adenylate cyclase in an inhibitory manner, A_{2A}, and A_{2B} stimulate this enzyme. Adenosine receptors are ubiquitously expressed in the body, many cells express several subtypes, but in different densities.[1] A_1 and A_{2A} receptors play important roles in cardiovascular and central nervous systems (CNS). On the other hand the A_{2B} subtype may also have important functions in the regulation of vascular tone and functions of mast cells. Expression levels for A_3 adenosine receptors are generally low and are highly species dependent and this fact is a limit in the knowledge about their physiological role.[2] Ligands for the A_3 receptor seem to be useful for treatment of inflammation, neurodegenerative diseases, asthma, and cardiac ischaemia.[3–5]

All receptor subtypes have been cloned, thus providing tools for studying each receptor subtype solely by using artificially constructed test systems.[6]

In this sense, the A_1 receptor is the best-characterized member of the adenosine receptor family. Its medicinal chemistry is particularly well developed with many agonists, partial agonist, and antagonist reported.[7–11] Also molecular biology aspects of the A_1 receptor have been extensively described, it has been cloned from different species, including humans.[1]

Nevertheless, the search of new A_1 agonists may be fruitfully facilitated by a quantitative structure–activity relationship (QSAR) study, which investigates molecular properties that can lead to strong binding with the A_1 adenosine receptors (i.e. hydrophobicity, polarizability, polar surface, etc.).[12–15] In this connection, the main objective of this research is to demonstrate, for the first time, how the GEometry, Topology, and Atom-Weights AssemblY (GETAWAY) descriptors are able to model this biological property and the higher quality of these descriptors to study this biological property if compared with other approaches by means of the statistical parameters of the obtained models.

Keywords: QSAR; A_1 adenosine receptors agonists; GETAWAY descriptors.
* Corresponding author. Tel.: +53 42 281473; fax: +53 42 281130; e-mail: mpgonzalez76@yahoo.es

0960-894X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2005.03.028

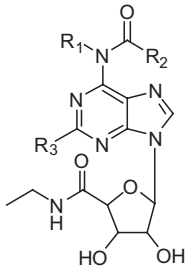
2. Data sets and computational strategies

A data set of 32 adenosine agonists for which the affinities of A_1 adenosine receptors were reported in the literature was selected to carry out the present study.[16] The parameter studied is the log (K_i) where K_i is the displacement of a specified [^{3}H]R-PIA binding (A_1) in rat brain membranes, expressed in nanomolar concentration. The compounds as well as the experimental values of K_i are shown in Table 1.

The DRAGON[17] computer software was employed to calculate the molecular descriptors.

In this way, we carried out geometry optimization calculations for each compound of this study using the quantum chemical semi-empirical method AM1[18] included in MOPAC 6.0.[19] Six models were developed using the DRAGON computer software, the Topological, Galvez Topological Charges indexes, Randić Molecular Profiles, Geometrical, WHIM, and GETAWAY descriptors were calculated.[20] All statistical analyses and data exploration were carried out using the STATISTIC 6.0 software.[21] The statistical processing to obtain the QSAR models was carried out by using the genetic algorithm method; all the parameters such as population size mutation probabilities, cross-over probabilities, and smoothing were fixed at their default values. Analysis of residuals and deleted residuals from the regression equations was used to identify outliers. The statistical significance of the models was determined by examining the regression coefficient, the standard deviation, the number of variables, the

Table 1. Structures and affinities in radioligand binding assays at rat brain A_1 adenosine receptors of compounds used in the current work



Compd	R_1	R_2	R_3	$K_i(A_1)$[a]/nM
1	H	4-Biphenyl	H	5.92
2	H	2,4-Cl–Ph–CH_2	H	16
3	H	4-CH_3O–Ph	H	86.4
4	H	2-Cl–Ph	H	1170
5	H	Ph	H	252
6	H	PhCH_2NH	H	171
7	H	4-SO_2NH_2PhNH	H	453
8	H	4-CH_3CO–PhNH	H	72.7
9	H	(*R*)-α-Phenylethyl–NH	H	433
10	H	(*S*)-α-Phenylethyl–NH	H	537
11	H	5-Me–isoxazol-3-yl–NH	H	146
12	H	1,3,4-Thiadiazol-2-yl–NH	H	208
13	H	4-*n*-C_3H_7O–PhNH	H	247
14	H	Ph–CH_2CH_2NH	H	129
15	H	3,4-MeO–Ph–CH_2CH_2NH	H	1770
16	H	Fur-2-yl–CH_2NH	H	419
17	H	4-(Pyridin-2-yl-$NHSO_2$)PhNH	H	292
18	H	4-(5-Me-isoxazol-3-yl-$NHSO_2$)PhNH	H	901
19	H	4-(Pyrimidin-2-yl-$NHSO_2$)PhNH	H	725
20	4-NO_2–Ph–NH–CO	4-NO_2–Ph–NH	H	89.1
21	5-Cl–pyridin-2-yl–NH–CO	5-Cl–pyridin-2-yl–NH	H	761
22	H	3-Cl–Ph–NH	Cl	90
23	H	4-MeO–Ph–NH	Cl	22
24	H	3-Cl–Ph–NH	I	103
25	H	4-MeO–Ph–NH	I	29.4
26	H	3-Cl–Ph–NH	*n*-C_4H_9–C≡C	2040
27	H	3-Cl–Ph–NH	Ph–C≡C	1330
28	H	3-Cl–Ph–NH	PhCH(OH)–C≡C	814
29	H	4-MeO–Ph–NH	*n*-C_4H_9–C≡C	428
30	H	4-MeO–Ph–NH	Ph–C≡C	4790
31	H	4-MeO–Ph–NH	PhCH(OH)–C≡C	75.1
32	H	4-MeO–Ph–NH	Ph$(CH_2)_3$–C≡C	14,400

[a] Displacement of specified [^{3}H]R-PIA binding (A_1) in rat brain membranes, expressed as K_i in nM (n = 3–6).

cross-validation leave-one-out statistics, and the proportion between the cases and variables in the equation.

3. Quantitative structure–activity relations

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

$$\begin{aligned}\log(K_i) &= 23.886 - 8.143 \cdot (\mathrm{H8v}) - 45.062 \cdot (\mathrm{REIG}) \\ &\quad - 10.686 \cdot (\mathrm{R2u^+}) + 91.678 \cdot (\mathrm{R7u^+}) \\ &\quad - 11.937 \cdot (\mathrm{R5v}) + 29.425 \cdot (\mathrm{R1v^+}) \\ N &= 32, \quad S = 0.383, \quad R_2 = 0.773, \\ F &= 14.188, \quad p < 0.0001, \quad S_{cv} = 0.501, \\ q_2 &= 0.664 \end{aligned} \tag{1}$$

where N is the number of compounds included in the model, S is the standard deviation of the regression, R^2 is the squared correlation coefficient, F is the Fischer ratio, p is the significance of the variables in the model, S_{cv} is the standard deviation of the cross-validation, and q^2 is the square of the correlation coefficient of the cross-validation.

One of the most important variables in this equation is REIG. This descriptor has a negative influence in the studied property and, for this reason, increases the affinity of the adenosine analogues for the A_1 receptors.

The REIG descriptor is defined as the first eigenvalue of the influence/distance matrix of the magnitude in study. This descriptor has larger values for branched molecules, for this reason according to this behavior; adenosine analogues with branched groups might have a high affinity for these kinds of receptors in special N^6-position.

On the other hand, a group of two variables $R2u^+$ and R5v decreases the $\log(K_i)$ and for this reason increases the affinity for A_1 adenosine receptor subtypes. Especially the term R5v is expected to have a high dependence on conformational changes as encoding information on pairs of atoms near each other. For that reason, according to our model and the previous report, the stereoselectivity plays an important role for the affinity at A_1 adenosine receptors.

In addition, the model establishes a strong relationship among GETAWAY descriptors and the property under study characterized by a good squared regression coefficient (0.773), which explains around 77% of the variance of data. In our opinion this fact is a determining factor at the time of selecting the best model to be used later; besides it presents the greatest F of Fischer ($F = 14.188$) and the lowest standard deviation for the data ($S = 0.383$), which confirms the former selection.

Additionally, a low susceptibility to over-fitting, due to excess in variable input, is ensured by a highly adjusted squared regression coefficient (0.758), which decreases only 1.94 % with respect to the squared regression coefficient.

On the other hand, one of the objectives of the current work is to compare the reliability of the GETAWAY descriptors to describe the property under study as compared with other different descriptors and methods. Consequently, we have developed other five models using the same data set that was used in the GETAWAY QSAR model. The results obtained with the use of Topological, Galvez Topological Charges indexes, Randić Molecular Profiles, Geometrical, and WHIM descriptors and the meaning of these are given in Tables 2 and 3.

As can be seen, there are remarkable differences concerning the explanation of the experimental variance given by these models compared to the GETAWAY one. While the GETAWAY QSAR model explains more than 77% of activity, the rest of the models are unable to explain more than 70% of such variance; all these models also have important statistic parameters of a lower quality compared to the GETAWAY approach, such as the Fischer ratio (F) and the standard deviation (S).

The GETAWAY model not only overtakes the other five models in the statistical parameters of the regression but its most important feature is also related to the stability to the inclusion–exclusion of compounds as measured by the correlation coefficient and standard deviation of the cross-validation. 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 2, the value of the determination coefficient of leave-one-out cross-validation for the model obtained with the GETAWAY descriptors ($q^2 = 0.669$) was the highest for the whole set of models, proving the predictive power of this approach and the stability of the model. In addition this model presents the lowest value of the standard deviation in the cross-validation ($S_{cv} = 0.501$); therefore the previous criteria are reaffirmed.

Table 2. The statistical parameters of the lineal regressions models obtained for the six kinds of descriptors

Descriptors	Variables	S	R^2	F	q^2	S_{cv}
Topological	MSD, CIC1, VRA1, MPC09, piPC09, T(N..S)	0.439	0.701	9.750	0.491	0.573
Galvez Topological Charges indexes	GGI2, GGI3, GGI8, GGI9, GGI10, JGI5	0.464	0.666	8.302	0.427	0.608
Randić Molecular Profiles	DP01, SP03, SP04, SP07, SP12, SP13	0.482	0.640	7.387	0.403	0.620
Geometrical	W3D, AGDD, DDI, ADDD, MAXDP, FDI	0.493	0.623	6.878	0.366	0.640
WHIM	E3u, P2m, G3m, L2s, E2s, Gu	0.481	0.642	7.471	0.371	0.637
GETAWAY	**H8v, REIG, R2u$^+$, R7u$^+$, R5v, R1v$^+$**	**0.383**	**0.773**	**14.188**	**0.669**	**0.501**

Table 3. Topological, Galvez Topological Charges indexes, Randić Molecular Profiles, Geometrical, WHIM, GETAWAY descriptors of the QSAR regression reported in this study

Symbol	Definition
CIC1	Complementary information content (neighborhood symmetry of 1-order)
SP12	Shape profile no. 12
SP03	Shape profile no. 03
SP04	Shape profile no. 04
SP12	Shape profile no. 12
FDI	Folding degree index
H8v	H autocorrelation of lag 8/weighted by atomic van der Waals volumes
REIG	First eigenvalue of the R matrix
$R2u^+$	R maximal autocorrelation of lag 2/unweighted
$R7u^+$	R maximal autocorrelation of lag 7/unweighted
R5v	R autocorrelation of lag 5/weighted by atomic van der Waals volumes
$R1v^+$	R maximal autocorrelation of lag 1/weighted by atomic van der Waals volumes
T(N..S)	Sum of topological distances between N..S
piPC09	Molecular multiple path count of order 09
MPC09	Molecular path count of order
VRA1	Randic-type eigenvector-based index from adjacency matrix
MSD	Mean square distance index (Balaban)
DP01	Molecular profile no. 01
SP07	Shape profile no. 07
SP13	Shape profile no. 13
JGI5	Mean topological charge index of order 5
GGI10	Topological charge index of order 10
GGI9	Topological charge index of order 9
GGI8	Topological charge index of order 8
GGI3	Topological charge index of order 3
GGI2	Topological charge index of order 2
W3D	3D-Wiener index
AGDD	Average geometric distance degree
DDI	D/D index
ADDD	Average distance/distance degree
MAXDP	Maximal electro topological positive variation
Gu	G total symmetry index/unweighted
E2s	Second component accessibility directional WHIM index/weighted by atomic electro topological states
L2s	Second component size directional WHIM index/weighted by atomic electro topological states
G3m	Third component symmetry directional WHIM index/weighted by atomic masses
P2m	Second component shape directional WHIM index/weighted by atomic masses
E3u	Third component accessibility directional WHIM index/unweighted

4. Concluding remarks

We have shown that the GETAWAY approach is able to explain the A_1 agonist effect for this set of compounds. In fact, we have developed a model for predicting the agonist activity of a data set of 32 adenosine analogues: N^6-arylcarbamoyl, 2-arylalkynyl-N^6-arylcarbamoyl, and N^6-carboxamido derivatives. This model is statistically and chemically sound and explains more than 77% of the variance in the experimental activity with a good predictive power. These features are significantly better than others obtained from five different methodologies contained in the Dragon Software. For these reasons, we can assert that the GETAWAY approach should be used as an efficient alternative to develop new analogues of the A_1 adenosine receptor agonists.

Acknowledgments

Maykel Pérez González acknowledges Professors Christa E. Muller and Gloria Cristalli for sending him valuable information for the development of this paper. In this sense, we thank the Universidad de Vigo and the Xunta de Galicia (PGIDT01PX130114PR) for financial support. Marta Teijeira thanks the Xunta de Galicia for the Parga Pondal contract.

References and notes

1. Fredholm, B. B.; Ijzerman, A. P.; Jacobson, K. A.; Klotz, K. N.; Linden, J. *Pharmacol. Rev.* **2001**, *53*, 527.
2. Dalpiaz, A.; Manfredini, S. *Curr. Med. Chem.* **2002**, *9*(21), 1923.
3. Baraldi, P. G.; Cacciari, B.; Romagnoli, R.; Spalluto, G. *Exp. Opin. Ther. Patents* **1999**, *9*, 515.
4. Hannon, J. P.; Pfaunnkuche, H. J.; Fozard, J. R. *Br. J. Pharmacol.* **1995**, *115*, 945.
5. Lubitz, D. K. J. V.; Lin, R. C. S.; Popik, P. *Eur. J. Pharmacol.* **1995**, *275*, 23.
6. Olah, M. E.; Stiles, G. L. *Annu. Rev. Pharmacol. Toxicol.* **1995**, *35*, 581.
7. Scammells, P. J.; Baker, S. P.; Belardinelli, L.; Olsson, R. A.; Russell, R. A.; Knevitt, S. A. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 811.
8. Hutchinson, S. A.; Baker, S. P.; Scammells, P. J. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 933.
9. Hutchinson, S. A.; Baker, S. P.; Linden, J.; Scammells, P. J. *Bioorg. Med. Chem.* **2004**, *12*, 4877.

10. Müller, C. *Il farmaco* **2001**, *56*, 77.
11. Soudijn, W.; van Wijngaarden, I.; IJzerman, A. P. *Curr. Top. Med. Chem.* **2003**, *3*, 355.
12. González, M. P.; Terán, C.; Fall, Y.; Teijeira, M.; Besada, P. *Bioorg. Med. Chem.* **2005**, *13*, 601.
13. González, M. P.; Terán, C. *Bioorg. Med. Chem.* **2004**, *12*, 2985.
14. González, M. P.; Terán, C. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3077.
15. González, M. P.; Terán, C. *Bull. Math. Biol.* **2004**, *66*, 907.
16. Baraldi, P. G.; Cacciari, B.; Pineda de las Infantas, M. J.; Romagnoli, R.; Spalluto, G.; Volpini, R.; Constanza, S.; Vittori, S.; Cristalli, G.; Merman, N.; Kyung-Sung, P.; Xiao-duo, J.; Jacobson, K. A. *J. Med. Chem.* **1998**, *41*, 3174.
17. Todeschini, R.; Consonni, V.; Pavan, M. DRAGON Software version 2.1, 2002.
18. Michael, J. S.; Dewar, E.; Zoebisch, G.; Eamonn, F.; Stewart, J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
19. Stewar, J. J. P. MOPAC Manual, 6th ed.; Frank J. Seiler Research Laboratory, U.S. Air Force Academy, Colorado Springs, CO, 1990; p 189.
20. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, Germany, 2000.
21. StatSoft, STATISTICA 6.0. Copyright_1984–2002.