Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters
Bioorganic & Medicinal Chemistry Letters 15 (2005) 3491–3495

BCUT descriptors for predicting affinity toward A_3 adenosine receptors

Maykel Pérez González,[a,b,c,*] Carmen Terán,[c] Marta Teijeira,[c] Pedro Besada[c] and Maria J. González-Moa[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 6 May 2005; revised 27 May 2005; accepted 31 May 2005

Abstract—The BCUT descriptors have been applied to the study of the A_3 adenosine receptor agonist effect of 32 adenosine analogues. A model, able to describe more than 80% of the variance in the experimental activity was developed with the use of the above-mentioned approach. Four different approaches (topological, Galvez topological charges indexes, Randić molecular profiles, and geometrical descriptors) failed to give satisfactory models for this property with the same number of variables in the equation. Although statistically significant models were derived containing descriptors other than BCUT, the best fitted model was still found with these descriptors.

Adenosine regulates many physiological functions through specific cell membrane receptors. All adenosine receptors (ARs) are coupled G-protein, and ubiquitously expressed through the body in different quantities, depending on the tissue class.[1] Four subtypes of ARs have been classified (A_1, A_{2A}, A_{2B}, and A_3), cloned from several species, and pharmacologically characterized. The A_3 AR is the youngest member of the AR family and the last to be cloned.[2] This receptor subtype shows more species differences than other AR subtypes between rodents and humans in amino acid sequence, ligand binding affinity (particularly antagonist), and tissue distribution. In rats, the A_3 receptors are expressed mainly in testis and mast cells, while in humans the highest A_3 receptor densities are found in lung, liver, and cells of the immune system.[3]

Because of the importance of the adenosine receptors, selective and potent ligands for each one of the four receptor subtypes are needed for therapeutic use and as pharmacological tools to study the physiological role of ARs.[4] Previous structural–activity studies of adenosine derivatives as receptor agonists concluded that selectivity is related to specific substitutions on the adenine nucleus. For example, N^6-substituents, such as cyclopentyl, enhance adenosine A_1 receptor selectivity relative to the other subtypes (A_{2A}/A_3).[5] On the other hand, 3-iodobenzyl group induces adenosine A_3 receptor selectivity.[6] Bulky substituents at the 2-position of the adenine moiety, such as aralkylamino,[7] alkylidenehydrazino,[8] or alkynyl,[9] have been described to increase selectivity for the adenosine A_{2A} receptor compared to A_1. For the A_{2B} receptor, however, there are currently no selective agonists. The sugar moiety appears to be essential for agonistic activity and high affinity. Thus, small alkyl 5′-uronamide modification of the ribose often enhances affinity of adenosine derivatives at some subtype.

The synthesis of A_3 ARs specific agonists enables the study of the biological effects and the mechanisms involved upon their activation. Although a large number of adenosine derivatives have been tested at the adenosine A_3 receptor, only a few of them are known as potent and selective agonists.[3] The current standard A_3 agonist is Cl-IB-MECA,[10] an adenosine derivative which bears a 3-iodobenzyl fragment at the 6-amino group and a chlorine atom at C-2, both in the purine ring. In combination, it contains an *N*-methylcarboxamido substituent

Keywords: QSAR; A_3 adenosine receptor agonists; BCUT 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.2005.05.122

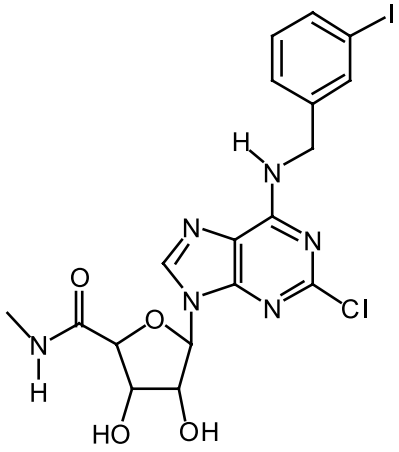


Figure 1. Specific agonist to A_3 adenosine receptor Cl-IB-MECA.

in the 5′ position of the ribose (Fig. 1). This compound appears to exhibit much higher affinity and selectivity for rat than for human ARs.

Quantitative structure–activity relationship (QSAR) studies are a powerful method for the design of bioactive compounds and the prediction of activity according to physical and chemical properties. [11–17] For that reason, in this research we have tried to describe the QSAR between the A_3 receptor and series of 5′-*N*-ethylcarboxamido analogues of adenosine (adenosine-5′-ethyluronamide derivatives) structurally related to Cl-IB-MECA.

The BCUT metrics of Pearlman approach has been introduced[18] in the context of in silico method for modeling physicochemical and biological properties of chemicals. The successful application of this theoretical approach has inspired us to test and/or validate the BCUT descriptors applicability in assessing discoveries of new adenosine analogues. Thereby, the aim of this work was to find rationality in the search of novel A_3 adenosine receptor agonist compounds using BCUT approach and to continue the validation of this method in describing the biological activity of series of compounds and its comparison with other methodologies to demonstrate its value as a good QSAR model.

The BCUT metrics are extensions of parameters originally developed by Burden.[19] The Burden parameters are based on a combination of the atomic number for each atom and a description of the nominal bond-type for adjacent and nonadjacent atoms. They incorporate connectivity information and atomic properties (e.g., atomic charge, polarizability, and hydrogen bond abilities) that are relevant to intermolecular interactions.

The BCUT metrics expand the number and types of atomic features that can be considered, and also provide a greater variety of proximity measures and weighting schemes. Many BCUT metrics can be generated, depending on the choices of connectivity and atomic information, and on the scaling factors controlling the relative balance of these two kinds of information. The BCUT metrics can capture sufficient structural features of molecules to yield useful measurement of molecular diversity. The result is a new whole-molecule descriptor that has shown significant utility in the measurement of molecular diversity and related tasks.

In this study, we have selected a data set of 32 adenosine analogues whose affinity at A_3 receptors was reported.[20] The affinity of these compounds was measured by displacement of specified [^{125}I]AB-MECA binding at rat A_3 receptors expressed in CHO cells, given as K_i in nM. The experimental values of this property are collected in Table 1.

Before calculating the molecular descriptors, we carried out geometry optimization calculations for each compound of this study using the quantum chemical semi-empirical method AM1[21] included in Mopac 6.0 computer software.[22]

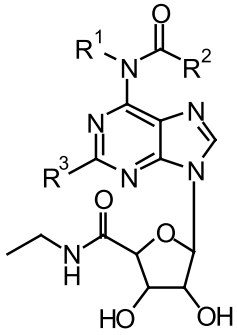
Dragon[23] computer software was employed to calculate the molecular descriptors. Five kinds of molecular descriptors including different groups, topological indices, BCUT descriptors, Galvez topological charge indices, Randić molecular profiles, and geometrical descriptors[24], were calculated for this set of compounds. Descriptors with constant values inside each group of descriptors were discarded. For the remaining descriptors, pairwise correlation analysis for all kinds of descriptors was performed. The following descriptor exclusion methods were used to reduce, in a first step, the collinearity and correlation between descriptors.

This procedure consists of the elimination of the one descriptor from each pair with the modulus of the correlation coefficients higher than a predefined value R_{max} (0.90). The procedure must be carried out with care. Indeed, we let $R_{ij} = R(d_i,d_j)$ be the correlation coefficient between descriptors d_i and d_j. Then, from $R_{ij} > R_{max}$ and $R_{jk} > R_{max}$ it does not follow that $R_{ik} > R_{max}$. So, in this case if d_j is eliminated, d_k must be retained.

The most significant parameters were identified from the data set using genetic algorithm (GA) analysis of the five kinds of descriptors obtained by Dragon computer software. All statistical analysis and data exploration were carried out using the Statistic 6.0.[25] The 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, cross-over, 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 six descriptors. The GA procedure was repeated n times to confirm that the selected descriptors are the optimal descriptor set for describing the modeled property. Examining the regression coefficients, the standard deviations, the significances, and the number of variables in the equation we determined the quality of the models.

The linear models obtained were validated by calculating q^2 values, which are calculated from leave-one-out (LOO) test, also known as cross-validation. Thus, a data point is removed from the set, the regression

Table 1. Structures and affinities of compounds in radioligand binding assays at rat A_3 adenosine receptors expressed in CHO cells used in the current work



Compound	R^1	R^2	R^3	K_i (A_3)[a] (nM)
1	H	4-Biphenyl	H	979
2	H	2,4-Cl–Ph–CH_2	H	167
3	H	4-CH_3O–Ph	H	837
4	H	2-Cl–Ph	H	754
5	H	Ph	H	824
6	H	$PhCH_2NH$	H	38.3
7	H	4-SO_2NH_2PhNH	H	9.73
8	H	4-CH_3CO–PhNH	H	20.9
9	H	(*R*)-α-Phenylethyl–NH	H	16.3
10	H	(*S*)-α-Phenylethyl–NH	H	319
11	H	5-Me–isoxazol-3-yl–NH	H	532
12	H	1,3,4-Thiadiazol-2-yl–NH	H	5550
13	H	4-*n*-C_3H_7O–PhNH	H	107
14	H	Ph–CH_2CH_2NH	H	149
15	H	3,4-MeO–Ph–CH_2CH_2NH	H	411
16	H	Fur-2-yl–CH_2NH	H	713
17	H	4-(Pyridin-2-yl–$NHSO_2$)PhNH	H	54.1
18	H	4-(5-Me–isoxazol-3-yl–$NHSO_2$)PhNH	H	155
19	H	4-(Pyrimidin-2-yl–$NHSO_2$)PhNH	H	405
20	4-NO_2–Ph–NH–CO	4-NO_2–Ph–NH	H	168
21	5-Cl–pyridin-2-yl–NH–CO	5-Cl–pyridin-2-yl–NH	H	5700
22	H	3-Cl–Ph–NH	Cl	77.6
23	H	4-MeO–Ph–NH	Cl	17.1
24	H	3-Cl–Ph–NH	I	315
25	H	4-MeO–Ph–NH	I	251
26	H	3-Cl–Ph–NH	*n*-C_4H_9–C≡C	581
27	H	3-Cl–Ph–NH	Ph–C≡C	611
28	H	3-Cl–Ph–NH	PhCH(OH)–C≡C	696
29	H	4-MeO–Ph–NH	*n*-C_4H_9–C≡C	211
30	H	4-MeO–Ph–NH	Ph–C≡C	154
31	H	4-MeO–Ph–NH	PhCH(OH)–C≡C	324
32	H	4-MeO–Ph–NH	$Ph(CH_2)_3$–C≡C	89.5

[a] Displacement of specified [^{125}I]AB-MECA binding at rat A_3 receptors expressed in CHO cells, expressed as K_i in nM (n = 3–6).

recalculated, and the predicted value for that point is then compared to its actual value. This process is repeated until each datum has been omitted once. The sum of squares of these deletion residuals can be then used to calculate q^2, a statistic equivalent to R^2 that 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. Therefore, it is a more meaningful summary statistic for QSAR models.

In this connection, the best QSAR model obtained with the BCUT descriptors is given below together with the statistical parameters of the regression:

$$\log(K_i) = 11.619 + 4.204 \cdot BELm7 + 11.721 \cdot BELv2 - 7.415 \cdot BEHe7 + 5.610 \cdot BEHp3 + 13.131 \cdot BELp5 - 7.010 \cdot BELp6 \quad (1)$$

$N = 32$; $R = 0.899$; $S = 0.322$; $F_{exp} = 17.535$; $p < 10^{-5}$; $q^2 = 0.724$; $S_{cv} = 0.385$, where N is the number of compounds used, R is the correlation coefficient, S is the standard deviation of the regression, F_{exp} is the Fisher ratio at the 95% confidence level, p is the significance of the variables in the model, q^2 is the square of the correlation coefficient of the leave-one-out cross-validation, and S_{cv} is the standard deviation of the leave-one-out cross-validation. The variables in the model (Eq. 1) are essentially independent, their intercorrelation being very low ($R^2 = 0.409$).

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

Descriptors	Variables[a]	S	R^2	F	p	q^2
BCUT	BELm7, BELv2, BEHe7, BEHp3, BELp5, BELp6	0.322	0.808	17.535	10^{-5}	0.724
Topological	X2A, PW3, PW5, T(N···N), T(N···Cl), T(N···I)	0.395	0.760	13.282	10^{-5}	0.657
Galvez topological charges indexes	GGI2, GGI7, GGI8, JGI1, JGI10, JGT	0.467	0.594	6.116	10^{-5}	0.353
Randić molecular profiles	DP05, DP07, DP10, DP12, DP17, SP12	0.549	0.441	3.288	10^{-5}	0.135
Geometrical	DELS, SPH, ASP, L/Bw, G(N···O), G(N···Cl)	0.473	0.585	5.878	10^{-5}	0.363

[a] The definition of the terms is largely explained in Ref. 24.

As it can be noticed, in the previous equation three significant variables are related to atomic polarizabilities, where *BEH*p3 is the highest eigenvalue *n*. 3, *BEL*p5 the lowest eigenvalue *n*. 5 and *BEL*p6 the lowest eigenvalue *n*. 6 of the Burden matrix. These variables have different contributions at the property due to their different signs in the equation. This behavior is comprehensible in the base of the existence of different pockets in the A_3 adenosine receptors, and the presence of different sub-structures in the molecule that may interact in different ways with the same receptor. In this sense, *BEH*e7 is the highest eigenvalue *n*. 7 of Burden matrix weighted by atomic Sanderson electronegativities. For this reason, their negative contributions to log (K_i) suggest that an increase of the electronegativity increases the affinity of new analogues for this type of receptor. Similar behavior was reported by Baraldi et al.[20] from the binding assays, where the presence of amide versus urea functionality at the N^6 position was generally detrimental in terms of affinity at rat A_3 receptors. Finally, two variables *BEL*m7, lowest eigenvalue *n*. 7 weighted by atomic masses, and *BEL*v2, lowest eigenvalue *n*. 2 weighted by atomic van der Waals volumes of Burden matrix, contribute to increase the log (K_i) and to decrease the affinity of the ligands. Therefore, this evidences that receptor A_3 does not possess the capability of other receptors in this family that accommodate large sub-structures at different positions.[15]

As we previously pointed out, one of the objectives of the current work is to compare several kinds of molecular descriptors for describing the property under study. Consequently, we have developed four other models using the same data set that was included in our QSAR study with the BCUT descriptors. For this aim, the topological, Galvez topological charges indexes, Randić molecular profiles, and geometrical descriptors were included. The results of these models are shown in Table 2.

The statistical information for the best regressions of affinity for A_3 adenosine receptors with these molecular descriptors show that the BCUT descriptors explain the experimental variance of the data better than the other approximations. This connection also presents the best q^2 (0.724), regarding the better predictive power of the methodologies used. For that reason, we considered that BCUT methodology can be a very useful tool for the prediction of affinity in A_3 adenosine receptors.

The aim of this work was to build a significant comparative study of different sets of molecular descriptors, on the basis of their predictive ability for affinity of different ligands toward A_3 adenosine receptors. Particular attention was center in the BCUT descriptors, because they are used for the first time in this type of studies and because of their behavior from a statistical point of view over other methodologies. From the analysis, we can conclude that the BCUT descriptors have an overall good modeling capability, proving their usefulness in QSAR studies of the affinity of analogues for this receptor subtype.

Acknowledgments

Maykel Pérez González acknowledges Professors Kenneth A. Jacobson and Christa E. Müller for sending him valuable information for the development of this paper. We acknowledge 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. Olah, M. E.; Stiles, G. L. *Pharmacol. Ther.* **2000**, *85*, 55.
2. Olah, M. E.; Gallo-Rodriguez, C.; Jacobson, K. A.; Stiles, G. L. *Mol. Pharmacol.* **1994**, *45*, 978.
3. Müller, C. *Curr. Top. Med. Chem.* **2003**, *3*, 445.
4. Fishman, P.; Bar-Yehuda, S. *Curr. Top. Med. Chem.* **2003**, *3*, 463.
5. Van der Wenden, E. M.; Carnielli, M.; Roelen, H. C. P. F.; Lorenzen, A.; von Frijtag Drabbe Kunzel, J. K.; IJzerman, A. P. *J. Med. Chem.* **1998**, *41*, 102.
6. Van Tilburg, E. W.; Von Frijtag Drabbe Kunzel, J.; Groote, M.; Vollinga, R. C.; Lorenzen, A.; IJzerman, A. P. *J. Med. Chem.* **1999**, *42*, 1393.
7. Hutchison, A. J.; Williams, M.; de Jesus, R.; Yokoyama, R.; Oei, H. H.; Ghai, G. R.; Webb, R. L.; Zoganas, H. C.; Stone, G. A.; Jarvis, M. F. *J. Med. Chem.* **1990**, *33*, 1919.
8. Niiya, K.; Olsson, R. A.; Thompson, R. D.; Silvia, S. K.; Ueeda, M. *J. Med. Chem.* **1992**, *35*, 4557.
9. Cristalli, G.; Eleuteri, A.; Vittori, S.; Volpini, R.; Lohse, M. J.; Klotz, K. N. *J. Med. Chem.* **1992**, *35*, 2363.
10. Kim, H. O.; Ji, X.; Siddiqi, S. M.; Olah, M. E.; Stiles, G. L.; Jacobson, K. A. *J. Med. Chem.* **1994**, *37*, 3614.
11. González, M. P.; González, H. D.; Molina, R. R.; Cabrera, M. A.; Ramos de Armas, R. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1192.
12. Rios-Santamaría, I.; García-Doménech, R.; Gálvez, J.; Cortijo, J.; Santamaría, P.; Morcillo, E. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 477.
13. Gupta, M. K.; Sagar, R.; Shaw, A. K.; Prabhakar, Y. S *Bioorg. Med. Chem.* **2005**, *13*, 343.
14. González, M. P.; Terán, C.; Fall, Y.; Teijeira, M.; Besada, P. *Bioorg. Med. Chem.* **2005**, *13*, 601.
15. González, M. P.; Terán, C. *Bioorg. Med. Chem.* **2004**, *12*, 2985.

16. González, M. P.; Terán, C.. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3077.
17. González, M. P.; Terán, C. *Bull. Math. Biol.* **2004**, *66*, 907.
18. Pearlman, R. S.; Smith, K. M. In *3D-QSAR and Drug Design: Recent Advances*; Kubinyi, H., Martin, Y., Folkers, G., Eds.; Kluwer Academic: Dordrecht, Netherlands, 1997, pp 339–353.
19. Burden, F. R. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 225.
20. Baraldi, P. G.; Cacciari, B.; Pineda de las Infantas, M. J.; Romagnoli, R.; Spalluto, G.; Volpini, R.; Costanzi, S.; Vittori, S.; Cristalli, G.; Merman, N.; Kyung-Sung, P.; Xiao-duo, J.; Jacobson, K. A. *J. Med. Chem.* **1998**, *41*, 3174.
21. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
22. MOPAC version 6.0. Frank J. Seiler Research Laboratory, US Air Force Academy, Colorado Springs, CO, 1993.
23. Todeschini, R.; Consonni, V.; Pavan, M. Dragon. Software version 2.1, 2002.
24. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim: Germany, 2000.
25. STATISTICA version 6.0. Statsoft, Inc, 2002.