QSPR Correlations of the Algistatic Activity of 5-Amino-1-Aryl-1*H*-Tetrazoles

Alan R. Katritzky,[∇]* Ritu Jain,[∇] Ruslan Petrukhin,[∇] Sergey Denisenko,[∇]

and

Thomas Schelenz [•]

[∇] Center for Heterocyclic Compounds, Department of Chemistry, University of Florida P.O. Box 11720, Gainesville, FL 32611- 7200, USA

[•] Martin-Luther-Universitaet Halle-Wittenberg, Institut fuer Physikalische Chemie, Halle, Germany

This article is dedicated to Professor Alexandru T. Balaban on the occasion of his 70th birthday.

RUNNING TITLE

Algistatic Activity of Tetrazoles

ABSTRACT

The algistatic activity of 20 substituted 5-amino-1-aryl-1*H*-tetrazoles was correlated with the second order Kier and Hall index, the XY shadow on the XY rectangle and the total hybridization component of the molecular dipole by the CODESSA program in a three parameter equation with $R^2 = 0.9330$.

KEY WORDS

Tetrazoles, algistatic activity, CODESSA, QSPR correlation.

INTRODUCTION

Tetrazoles [1, 2] are medicinally important heterocycles incorporated in a large number of drugs approved by the FDA. Particularly, 5-amino-1-aryl-1*H*-tetrazole derivatives have been patented for muscle relaxation, anti-inflammatory, anti-arthritic, analgesic, ulcus therapeutic and coccidiostatic properties [3 and references cited therein]. Consequently, studies on the structural aspects and QSPR correlations pertaining to the biological activity of 5-amino-1-aryl-1*H*-tetrazoles are important.

Recently, Schelenz [3] tested a series of 5-amino-1-aryl-1*H*-tetrazoles for growth-inhibiting activity against autotrophic *Chorella vulgaris* cultures. Algae growth was measured photometrically at 680 nm to give activity parameter A defined as the reciprocal value of the isoeffective concentrations C_{50}. The algistatic activity (log A) of the tested heterocycles was shown to depend on their hydrophobicity [3]. A QSAR study of 5-amino-1-aryl-1*H*-tetrazoles used physicochemical and structural parameters such as log P values to correlate the biological activity of tetrazoles [3].

In the present work, we successfully correlated the algistatic activity of 5-amino-1-aryl-1*H*-tetrazoles [3] with three molecular descriptors employing the CODESSA program [4].

METHODOLOGY

The database (Table I) for correlation of the biological activities (using the differential method of measurement) of the 5-amino-1-aryl-1*H*-tetrazoles as reported by Schelenz [3] was created in ISIS/Base [5]. The HyperChem program [6] was applied for preliminary molecular geometry optimization using molecular mechanics MM+ force fields. The final optimization was obtained using the semi-empirical AM1 parameterization method [7] realized in the MOPAC computer program [8]. The output files served for producing constitutional, topological, electrostatic and semi-empirical descriptors by the CODESSA software package [4]. The resulting descriptor matrix was used to search for the multi-linear correlations.

As it is difficult to perform a comprehensive search for the best linear correlation, we used two different methods for the selection of the "best" descriptor subset: the heuristic (HM) and the best multi-linear regression (BMLR) methods. These methods employ different approaches in the discrete local optimization of the correlation coefficient in the descriptor permutation space. As they are local optimization methods, there is no guarantee that the global optimum will be found by them. Although both methods use heuristics for eliminating the descriptors with low probability to be involved in the final multi-parameter correlation, the approaches to this elimination are completely different. The BMLR method commences by correlating the entire function (given property) employing two-parameter regressions with pairs of orthogonal descriptors. An additional descriptor is selected for testing if (i) it is orthogonal to the existing descriptor

set and (ii) its inclusion leads to a maximum increased R^2 value. Completion of the algorithm is achieved when improvement in the value of R^2 becomes insignificant.

The HM for descriptor selection [4], on the other hand, proceeds with a pre-selection of descriptors by eliminating (step by step):

(i) descriptors which do not match any of the following criteria: (a) Fisher F-criterion greater than one unit; (b) R^2 value less than a value defined at the start (usually 0.01); (c) Student's criterion less than that defined (usually 0.1);

(ii) descriptors having a higher squared intercorrelation coefficient than a predetermined level (usually 0.8); retaining descriptors with higher R^2 with reference to the property.

This is followed by listing the remaining descriptors in decreasing order of the correlation coefficients when used in global search for 2-parameter correlations. Each significant 2-parameter correlation by F-criteria is recursively expanded to an n-parameter correlation till the normalized F-criterion remains greater than the startup value. The top N correlations by R^2 as well as F-criteria are saved. The HM usually produces correlations 2-5 times faster than BMLR, with quality comparable to that obtained with the latter. The rapidity of calculations from the HM render it the first method of choice in practical research.

RESULTS AND DISCUSSION

Schelenz [3] has related the algistatic activity of tetrazoles with the experimentally determined logP (octanol-water partition coefficient). LogP is widely accepted as an

indication of the distribution of analytes into biological membranes [9]. Schelenz [3] correlated logA with logP having $R^2 = 0.8064$. To investigate the contribution of various molecular descriptors to the biological activity of tetrazoles, we have now applied the CODESSA program to the logA values. The correlation of the biological data with the molecular descriptors of 20 5-amino-1-aryl-1*H*-tetrazoles employing CODESSA program afforded two- and three-parameter correlations, which are given in Table II (HM) and in Table III (BMLR) respectively.

Treatment of the descriptors scales by the HM provides two-parameter (Eq. **1**) and three-parameter (Eq. **2**) correlations.

$$\log A = (1.2265 \pm 0.1407)\ {}^{2}\chi + (8.4762 \pm 1.5198) S_{R,XY} - (6.6553 \pm 1.2540)$$

$$n = 20,\ R^2 = 0.8394,\ F = 44.42,\ s = 0.221 \qquad (\mathbf{1})$$

$$\log A = (1.0888 \pm 0.1028)\ {}^{2}\chi + (7.4533 \pm 1.0827) S_{R,XY} - (1.6068 \pm 0.3673)\mu_h - (2.9416 \pm 1.2172)$$

$$n = 20,\ R^2 = 0.9269,\ F = 67.58,\ s = 0.154 \qquad (\mathbf{2})$$

Using the BMLR method, we obtained the same two-parameter correlation (Eq. **1**), but a better three-parameter correlation (Eq. **3**).

$$\log A = (1.200 \pm 0.1037)\ {}^{2}\chi - (1.723 \pm 0.348)\mu_h + (6.935 \pm 0.877) V_{XYZ} - (0.204 \pm 0.878)$$

$$n = 20,\ R^2 = 0.9330,\ F = 74.23,\ s = 0.0216 \qquad (\mathbf{3})$$

In the equations above, ${}^{2}\chi$ denotes the second order Kier and Hall index, $S_{R,XY}$ is the ration of XY shadow on XY rectangle, μ_h is the total hybridization component of the

molecular dipole, and V_{XYZ} is a molecular volume. Plots of the best two- and three-parameter correlations (Eq.**1**, **3**) are depicted in Figure 1 and Figure 2.

The Kier and Hall indexes are traditionally connected with molecular shape as are the shadow descriptors. Involvement of the shape descriptors can be explained as the influence of steric factors on probability of forming a binding complex. The molecular volume can be a measure of size and also depicts the influence of the steric factor. The total hybridization component of the molecular dipole is a specific measure of the charge distribution in the molecule, thus reflecting (i) an electrostatic component of inter molecular interaction, and (ii) the probability and energy of a hydrogen bond forming. All the descriptors are easily understandable in terms of the classical biochemical approach and clearly indicate the importance of the shape, size, and hybridization of a molecule in explaining the biological activity of 5-amino-1-aryl-1*H*-tetrazoles.

A cross-validation of the correlations was performed in the following manner. The whole set of 20 compounds was divided into three groups: compounds **1, 4, 7,** etc. form group 1; compounds **2, 5, 8,** etc. form group 2; and compounds **3, 6, 9,** etc. form group 3. Three sub-sets, X, Y and Z were formed from these groups by combining the groups 1 and 2, 2 and 3, and 1 and 3 respectively. The descriptor set previously obtained in the best heuristic three-parameter correlation (Eq.**2**) was retained. Then, using multi-linear regression method we obtained partition coefficients for sub-sets X, Y, and Z. The squared correlation coefficients for “native” sub-sets were 0.9513, 0.9213, and 0.9377, correspondently. Next, the activities for group 3 were predicted using the descriptor partition coefficients from sub-set X (generated from groups 1 and 2), group 1 activities

were predicted using the partition coefficients from sub-set Y and group 2 activities were predicted using the partition coefficients from sub-set Z. The squared correlation coefficients for "non-native" groups were 0.8323, 0.9656, and 0.8874, correspondently. All predicted activities for group 1, 2 and 3 were combined and correlated with their experimental activities of the whole set. This has afforded a correlation coefficient of 0.8641 compared to 0.9269 found in the original three-parameter correlation (Eq.**2**) obtained by the heuristic method. The squared cross-validated correlation coefficient by the leave-one-out method is 0.8834. The R^2cv value reflects the rather small data set.

CONCLUSION

Successful correlation equations emerged upon correlation of the biological activity of 5-amino-1-aryl-1*H*-tetrazoles containing different substituents with molecular descriptors from CODESSA program. The three-parameter correlation equation should enable prediction of the biological activities of unknown or unavailable compounds of this class [10]. The BMLR and HM methods provide comparable results in this correlation study.

REFERENCES

[1] Butler, R.N. (1984). The structure, reactions, synthesis and uses of heterocyclic compounds. *In Comprehensive Heterocyclic Chemistry* (Potts, K.T. Eds.) Pregamon Press, Oxford, **5**, pp.791-838.

[2] Butler, R.N. (1996). The structure, reactions, synthesis and uses of heterocyclic compounds. *In Comprehensive Heterocyclic Chemistry II* (Storr, R.C. Ed.) Pregamon Press, Oxford, **4**, pp. 621-678.

[3] Schelenz, T. (2000). Relationships between algistatic activity and physicochemical properties of 5-amino-1-aryl-1*H*-tetrazoles. *J. Prakt. Chem.*, **342**, 205-210.

[4] Katritzky, A.R., Lobanov, V.S. and Karelson, M. (1996). *CODESSA: Reference Manual*; *Version 2*, University of Florida.

[5] Integrated Scientific Information System Documentation Package, MDL, Inc., (1999).

[6] Hyperchem release for Windows (Version 5.0), Hypercube, Inc. (1996).

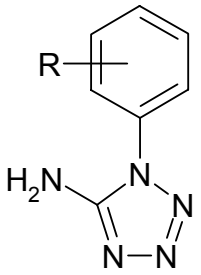
[7] Dewar, M., Zoebisch, E.G., Healy, E.F. and Stewart, J.J.P. (1985) AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.,* **107**, 3902-3909.

[8] Stewart, J.J.P. *et al*. MOPAC 7.0 for UNIX; Quantum Chemistry Program Exchange. Project; 688.

[9] Griffin, S., Wyllie, S.G. and Markham, J. (1999). Determination of octanol - water partition coefficient for terpenoids using reversed - phase high - performance liquid chromatography. *J. Chromatgr. A*, **864**, 221-228.

[10] Zhaoxu, C., Jianfen, F. and Heming, X. (1999). Theoretical study on tetrazole and its derivatives. Part 7: ab initio MO and thermodynamic calculations on azido derivatives of terazole. *J. Mol. Str.(Theochem.)*, **458**, 249-256.

TABLE I Dataset for the structure and the algistatic activity values (logA)



	R	$LogA_{exp}$	$LogA_{calc}$ (eq.**1**)	$LogA_{calc}$ (eq.**3**)
1	H	2.65	2.44	2.56
2	3-F	2.66	2.94	2.71
3	3-Cl	3.21	3.03	3.03
4	3-Me	3.09	3.09	3.03
5	3-NO_2	2.79	2.63	2.76
6	3-OMe	2.92	2.56	3.08
7	4-F	2.49	2.46	2.46
8	4-Me	3.15	3.00	3.28
9	4-NO_2	3.01	3.08	3.11
10	4-OH	2.19	2.63	2.25
11	4-OEt	3.31	3.04	3.19
12	3,4-Cl_2	3.84	3.68	3.63
13	3-Cl-4-Me	3.76	3.91	3.81
14	3,4-$(CH)_4$	4.01	3.81	3.83
15	2-Cl	2.40	2.57	2.46
16	2-Me	2.46	2.47	2.21
17	2-OMe	2.08	2.16	2.14
18	2,5-Cl_2	3.11	3.32	3.27
19	2,5-Me_2	2.98	3.20	3.19
20	2,3-$(CH)_4$	3.07	3.14	3.17

TABLE II Correlations of the biological activity of tetrazoles (logA) by Heuristic Method

# P	R^2	F	S^2		B	ΔB	t	Name of descriptor
1	0.5455	21.60	0.1303	0	-0.017	0.645	-0.027	Intercept
				1	1.037	0.223	4.648	Kier & Hall index (order 2)
2	0.8394	44.42	0.0488	0	-6.655	1.254	-5.307	Intercept
				1	1.227	0.141	8.719	Kier & Hall index (order 2)
				2	8.476	1.520	5.577	XY Shadow on XY rectangle
3	0.9269	67.58	0.0236	0	-2.942	1.217	-2.416	Intercept
				1	1.089	0.1028	10.591	Kier & Hall Index (Order 2)
				2	7.453	1.083	6.884	XY shadow on XY Rectangle
				3	1.607	0.367	-4.374	Total hybridization component of the molecular dipole

TABLE III Correlations of the biological activity of tetrazoles (logA) by the Best Multi-Linear Regression Method

# P	R^2	F	S^2		B	ΔB	t	Name of descriptor
2	0.8394	44.42	0.0488	0	-6.655	1.254	-5.307	Intercept
				1	1.227	0.141	8.719	Kier & Hall index (order 2)
				2	8.476	1.520	5.577	XY Shadow on XY rectangle
3	0.9330	74.23	0.0216	0	-0.204	0.878	-0.232	Intercept
				1	1.200	0.1037	11.573	Kier & Hall Index (Order 2)
				2	-1.723	0.348	-4.947	Total hybridization component of the molecular dipole
				3	6.935	0.877	7.292	Molecular volume/ XYZ Box

FIGURE 1 Plot of calculated vs. experimental log A based on equation **1.**

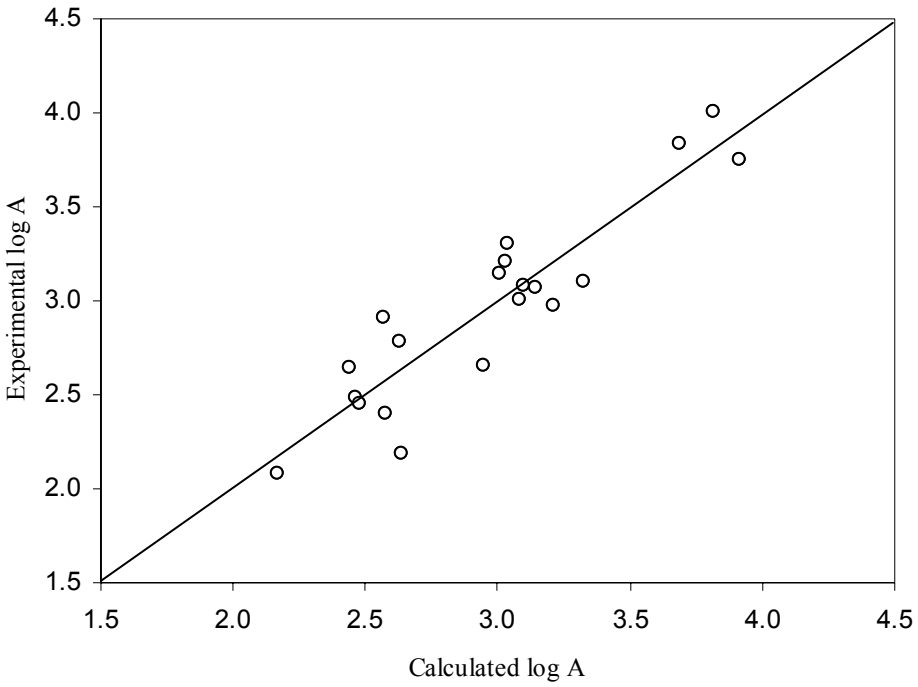


FIGURE 2 Plot of calculated vs. experimental log A based on equation **3.**