Pergamon

Bioorganic & Medicinal Chemistry 9 (2001) 2787–2792

BIOORGANIC & MEDICINAL CHEMISTRY

QSAR Studies on Acylated Histamine Derivatives

Vijay K. Agrawal[a] and Padmakar V. Khadikar[b,*]

[a]*Department of Chemistry, A.P.S. University, Rewa-486 003, India*
[b]*Research Division, Laxmi Pest and Fumigation Pvt. Ltd., 3 Khatipura, Indore 452 007, India*

Received 4 January 2001; accepted 9 April 2001

Abstract—H_3-receptor antagonists activity in terms of $-\log K_i$ for a series of acylated histamine derivatives was modeled using topological indices, namely negentropy (N), molecular redundancy (MRI), and valence connectivity index ($^mx^v$) indices. Excellent results were obtained in multiple regression analysis upon the introduction of a dummy parameter (indicator parameter). Consistant increase in R^2_A value indicated that inspite of observed collinearity the proposed models are significant. © 2001 Elsevier Science Ltd. All rights reserved.

Introduction

The existence of a third histamine receptor subtype was suggested in 1983. This presynaptically located autoreceptor was named the histamine H_3-receptor.[1] The H_3-receptors function as heteroreceptors on serotoninergic,[2] cholinergic,[3] noradrenergic,[4] dopaminergic,[5] and peptidergic[6] neurons. It was also found that H_3-receptor antagonists influence cerebral functions like microcirculation and vigilance[7] by modulating the release of histamine as well as other neurotransmitters.

In the antagonists field the discovery of thio-peramide,[8] related acyl derivatives,[9] and recently developed isothioureas[10] made it possible to evaluate cerebral H_3-receptors in vivo and in animal in vivo.

Stark et al.[11] have pointed out that the simplest acyl derivatives of the endogenous ligand, N^{α}-acylated histamine shows moderate H_3-receptors antagonist activity and that variation of acyl substituents of the primary amino group of histamines or different methylated hexamine derivatives should increase the antagonists activity without increasing toxicity. Consequently, they have prepared N^{α}-acylated histamines (Fig. 1, Table 1) and investigated their H_3-receptors antagonist in vivo activity. They observed that these compounds possess moderate to pronounced H_3-receptors antagonists activity.

In addition, Stark et al.[11] also observed that there is no difference in activity between the acetyl derivatives (compound **1**, Table 1) and the phenylacetyl derivative (compound **2**, Table 1). Also, that exchange of 1-methylene group by oxygen or sulfur has no advantage on H_3-receptors activity (**4**–**7**). The optimal distance between the polar amide function and the hydrophobic ring substituent seems to be of special importance (**2**,**3**,**10**,**13**). However, quantitative structure–activity (QSAR) relationship study is yet to be made.

In view of the above, we have undertaken the present study in that we have adopted H_3-receptor antagonists activity (expressed as $-\log K_i$) from the work of Stark et al.[11] and modelled it topologically using negentropy[12] (N), molecular redundancy[13] (MRI), and valence connectivity indices[14,15] ($^mx^v$). We observed that excellent results are obtained in multivariate correlations upon the introduction of indicator parameters (Ip_1 and Ip_2). The results are discussed below.

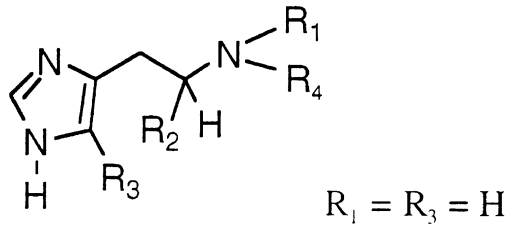
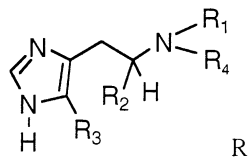


Figure 1. Acylated histamine derivatives used in the present study.

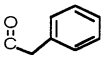
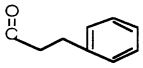
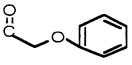
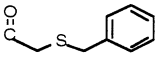
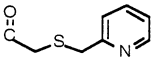
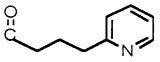
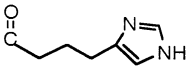
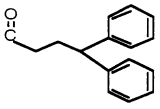
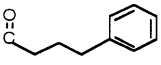
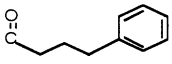
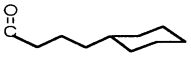
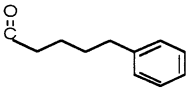
*Corresponding author. E-mail: vijay-agrawal@lycos.com

0968-0896/01/$ - see front matter
PII: S0968-0896(01)00147-X

Table 1. Structural details, indicator parameters and antogonists activity of the compounds used



$R_1 = R_3 = H$

Compound	R_2	R_4	Ip_1	Ip_2	$-\log K_i$
1	H		0	0	5.90
2	H		1	0	6.00
3	H		1	0	6.20
4	H		1	0	6.10
5	H		1	0	6.20
6	H		0	1	5.70
7	H		0	1	6.00
8	H		0	1	6.00
9	H		0	0	6.00
10	H		1	0	7.10
11	CH_3		1	0	6.00
12	H		0	0	7.30
13	H		1	0	6.70

Results and Discussion

Table 1 records the structural details and values of Ip_1 and Ip_2. This table also records the values of $-\log K_i$. The calculated molecular descriptors (N, MRI, $^m\chi^v$) are given in Table 2.

The correlation matrix for the parameters used is demonstrated in Table 3, while the regression parameters and the quality of statistically significant correlations are given in Table 4.

Table 5 records the estimated values of $-\log K_i$ from the most appropriate models. The observed values of $-\log K_i$ and the residue, that is difference between the observed and estimated $-\log K_i$ are also summarized in Table 5 for comparison.

Finally, Figure 2 demonstrates the correlation between observed and estimated $-\log K_i$.

A perusal of Table 1 shows that the H_3-receptor antagonists activity ($-\log K_i$) for the set of compounds used is as follows:

$$12 > 10 > 13 > 3 = 5 > 4 > 2 = 7 = 8 = 9 = 11 > 1 > 6$$

The aforementioned sequence of activity demonstrates very high degeneracy in the observed H_3-receptor antagonists activity ($-\log K_i$). However, the data presented in Table 2 shows that no degeneracy is present in any of the topological descriptors (N, MRI, $^m\chi^v$) used.

Table 2. Calculated topological indices for the compounds used

Compound	N	MRI	$^0\chi^v$	$^1\chi^v$	$^2\chi^v$
1	26.8972	0.0892	5.9244	3.5090	1.9313
2	39.3696	0.1826	9.5185	5.6268	3.0800
3	42.8365	0.2074	10.2256	6.1268	3.4031
4	41.3160	0.1755	9.9267	6.0242	3.1803
5	46.6272	0.1678	10.6338	5.9154	3.3958
6	46.8160	0.1337	10.1036	6.0640	3.2772
7	46.5567	0.1971	10.8025	6.4866	3.6381
8	42.6020	0.2117	10.1475	6.0640	3.3314
9	55.8768	0.3076	13.1900	8.6696	4.6637
10	46.4170	0.2268	10.9324	6.1727	3.6439
11	51.6846	0.2184	11.8553	7.0476	3.9599
12	52.1928	0.2782	11.6586	7.3373	4.6909
13	49.4214	0.2526	11.6395	7.7269	3.9975

Table 3. Correlation matrix for the inter-correlation of structural descriptors and their correlation with the activity

	N	MRI	$^0\chi^v$	$^1\chi^v$	$^2\chi^v$	Ip_1	Ip_2	$-\log K_i$
N	1.0000							
MRI	−0.8142	1.0000						
$^0\chi^v$	0.9788	0.8734	1.0000					
$^1\chi^v$	0.9525	0.9069	0.9724	1.0000				
$^2\chi^v$	0.9514	0.9281	0.9477	0.9605	1.0000			
Ip_1	0.0158	0.0137	−0.1135	−0.0281	−0.0480	1.0000		
Ip_2	0.0043	−0.2258	−0.0515	−0.0562	−0.1104	−0.5916	1.0000	
$-\log K_i$	0.3321	0.5363	0.3313	0.2838	0.4943	0.1921	−0.4091	1.0000

Table 4. Regression parameters and quality of correlation for multivariate correlations carried out in the present study

Model no.	Parameters used	A_i $i=1, 2, 3, 4, 5$	Intercept (B)	Standard deviation (SD)	Correlation coefficient (R)	F-ratio	R_A^2	Q
1	$^1\chi^v$	$A_1=-1.0126$ (±0.2486)	5.7430	0.2817	0.8460	2.591	0.6589	3.003
	$^2\chi^v$	$A_2=1.9428$ (±0.4110)						
2	$^0\chi^v$	$A_1=-0.6368$ (±0.1950)	6.2721	0.3173	0.8219	6.248	0.5675	2.590
	$^2\chi^v$	$A_2=1.7946$ (±0.4629)						
	Ip_1	$A_3=0.5332$ (±0.2039)						
3	$^1\chi^v$	$A_1=-1.0376$ (±0.2290)	5.5986	0.2589	0.8854	10.886	0.7119	3.420
	$^2\chi^v$	$A_2=1.9911$ (±0.3788)						
	Ip_1	$A_3=0.2436$ (±0.1495)						
4	$^1\chi^v$	$A_1=-0.9486$ (±0.2387)	5.8255	0.2660	0.8786	10.152	0.6959	3.303
	$^2\chi^v$	$A_2=1.8231$ (±0.3963)						
	Ip_2	$A_3=-0.2666$ (±0.1791)						
5	MRI	$A_1=7.1138$ (±3.2528)	6.2488	0.2400	0.9024	13.163	0.6987	3.760
	$^1\chi^v$	$A_2=-1.0824$ (±0.2142)						
	$^2\chi^v$	$A_3=1.5169$ (±0.4006)						
6	MRI	$A_1=6.3745$ (±3.0150)						
	$^1\chi^v$	$A_2=-1.0962$ (±0.1967)						
	$^2\chi^v$	$A_3=1.6017$ (±0.3708)						
	Ip_1	$A_4=0.2045$ (±0.1241)	6.0750	0.2199	0.9281	12.430	0.7921	0.4221

This is obvious as they belong to second generation topological indices proposed by Balaban.[16,17] According to Balaban, second generation topological indices are best suited for QSAR modelling.

An inspection of Table 3 shows that all the molecular descriptors are linearly correlated. Note that all the terms in the correlation matrix with value >0.4 are susceptible for collinearity.[18] The term value indicates that: (i) MRI and $^o\chi^v$, (ii) MRI and $^1\chi^v$, and (iii) MRI and $^2\chi^v$ are highly inter-correlated. However, the inter-correlation of MRI and N is comparatively poor. On the other hand, the inter-correlation amongst $^0\chi^v$, $^1\chi^v$, $^2\chi^v$ is significantly high. None of the molecular descriptors singly correlates with H_3-receptor antagonists activity indicating, thereby, absence of any univariate correlations in modelling the activity.

In order to investigate occurrence of statistically significant QSAR models we have subjected the data to step-wise regression analyses and the correlations so obtained are recorded in Table 4.

The regression analyses indicated that no mono-parametric correlations which are statistically significant are possible. These are, therefore, not shown in Table 4.

Only one biparametric correlation containing $^1\chi^v$ and $^2\chi^v$ is observed. However, there are as many as four statistically significant tri-parametric correlations (Table 4). Finally, stepwise regression has resulted into a tetra-parametric correlation. All these significant correlations are discussed below.

As stated earlier only one bi-parametric correlation has a high statistical significance. This correlation as shown below (model-1) gives a very good account for the influence of first and second order branching on the high activity.

$$-\log K_i = -1.0126\ (\pm0.2486)\,{}^1\chi^v + 1.9428 \times (\pm0.2486)\,{}^2\chi^v + 5.7430 \quad (5)$$

Table 5. Estimated values of $-\log K_i$ from eqs. (6) and (7) and their comparison with observed values

		Estimated-log K_i			
		Eq 6		Eq 7	
Compound	$-\log K_i$ (Residue)	Est.	Residue	Est.	Residue
1	5.90	6.01	−0.11	5.89	0.01
2	6.00	6.13	−0.13	6.21	−0.21
3	6.20	6.25	−0.05	6.34	−0.14
4	6.10	5.80	0.30	5.89	0.21
5	6.20	6.19	0.01	6.30	−0.10
6	5.70	5.60	0.10	5.53	0.17
7	6.00	6.14	−0.14	6.04	−0.04
8	6.00	6.24	−0.24	6.11	−0.11
9	6.00	6.13	−0.13	6.00	0.00
10	7.10	6.71	0.39	6.80	0.30
11	6.00	6.18	−0.18	6.29	−0.29
12	7.30	7.40	−0.10	7.23	0.07
13	6.70	6.39	0.31	6.48	0.22

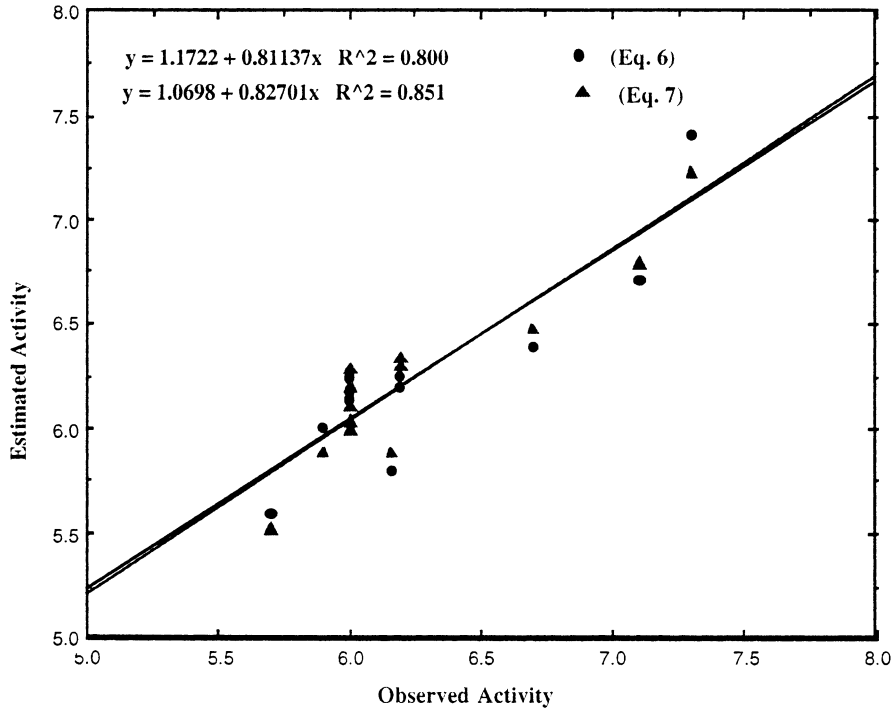


Figure 2. Correlation of observed versus estimated $-\log K_i$ using eqs 6 and 7.

This indicates that the first and second order branching as well as presence of hetero-atom has dominant role in exhibiting high activity. However, this model accounts for only 84.60% correlation. In view of this we have attempted still higher order multiple correlations.

Successive correlation analyses gave four tri-parametric models (models 2–5, Table 4). Out of these models, the model-2 is rejected on the ground that it has R-value much smaller than the model-1 discussed above.

Now the question arises which out of the remaining three tri-parametric models is the most appropriate model in modelling H_3-receptor antagonists activity ($-\log K_i$). We have, therefore, used quality factor Q,[19] which is defined as the ratio of correlation coefficient (R) to the standard error of estimation (Se), (Q = R /Se) for this purpose. Highest value of Q was observed for a tri-parametric correlation involving MRI, $^1\chi^v$ and $^2\chi^v$. This correlation is found as follows:

$$-\log K_i = 7.1138(\pm 3.2528)\ \mathrm{MRI} - 1.0824 \times (\pm 0.2142)\ {}^1\chi^v + 1.5169(\pm 0.4006)\,{}^2\chi^v + 6.2488 \quad (6)$$

This eq 6 again suggests that first and second order branching have predominant role in modelling the H_3-receptor antagonists activity ($-\log K_i$). The positive sign associated with $^2\chi^v$ indicates favorable effects of spatial demand on the periphery of the antagonists. Furthermore, the positive sign associated with MRI in eq 6 suggest that molecular redundancy affects $-\log K_i$.

Finally, step-wise regression ended into an excellent model containing MRI, $^1\chi^v$, $^2\chi^v$ and Ip_1 (model-6, Table 1). This model is shown below:

$$-\log K_i = 6.3745(\pm 3.0150)\mathrm{MRI} - 1.0962 \times (\pm 0.1965)\,{}^1\chi^v + 1.6017(\pm 0.3708)\,{}^2\chi^v + 0.2045(\pm 0.1241)\ Ip_{1+} 6.0750 \quad (7)$$

The Q-value (0.4221) for eq 7 also supports it (eq 7) to be the best QSAR model for modelling $-\log K_i$. This improvement in the model is due to the occurrence of Ip_1 term in eq 7. Thus, except for compound **12**, the presence of aromatic ring in the substituent R is essential for exhibiting high activity.

It is worthy to mention that in all the above correlations the signs of $^1\chi^v$ and $^2\chi^v$ terms remain the same, meaning thereby, the effect due to the first and second order connectivity on the activity ($-\log K_i$) is identical. The numerical values of these coefficients accounts for their role in proposing statistically significant models.

The aforementioned relationships needs further explanation. It contains molecular valence connectivity index and MRI which are highly inter-correlated. However, for these indices the corresponding coefficients are significantly higher than their standard error of estimation. This indicates that inspite of high collinearity eq 7 can be considered as statistically significant. Furthermore, adjusted R^2_A for eq 7 is significantly higher than eq 6. R^2_A is particularly important when the number of independent variables is large relative to the sample size.[18] It

is measure of the percent explained variation in the dependent variable that takes into account the relationship between the number of cases and number of independent variables in the regression model. All these points are in favour of eq 7.

In the present case, both R^2 and R^2_A increases as we pass from biparameteric model to tetra-parameteric model. This increase in R^2_A indicates that inspite of collinearity defect all the proposed models are significant. The statistical significance of the proposed models is further established from the fact that in all the cases the coefficients of molecular descriptors involved in the model are significantly larger than their respective standard deviations.

Finally, it is worth mentioning that the most significant models-5 and -6 (Table 4) contain MRI as one of the correlating parameters. This indicates that MRI is a dominating parameter for the exhibition of the activity.

From the aforementioned results and discussion, we conclude that inspite of existence of collinearity among MRI, $^1\chi^v$, $^2\chi^v$, all the models discussed above are found to be statistically significant. This can be explained on the basis of R^2_A-values which takes into account the adjustment of R^2.

In order to confirm our findings we have calculated $-\log K_i$ using models-5 and -6 (eq 6 and 7) and compared them with the observed values of $-\log K_i$. Such a comparison is shown in Table 5 and demonstrated in Fig. 2. The predictive correlation coefficient (0.851) for the model-6 indicates it to be the most appropriate model for modelling $-\log K_i$. At this stage, it is worth mentioning that though initially we have used seven molecular descriptors (N, MRI, $^0\chi^v$, $^1\chi^v$, $^2\chi^v$, Ip_1, Ip_2) the correlation analyses shows that only four (MRI, $^1\chi^v$, $^2\chi^v$, Ip_1) molecular descriptors are useful for proposing statistically significant models for modelling $-\log K_i$.

Conclusions

From the results and discussion made so far, we conclude that-receptor antagonists activity in terms of $-\log K_i$ can be modelled using MRI and connectivity index upon introduction of dummy parameters and that adjusted R^2_A accounts for the statistical significance of a model in that collinearity exists and that the number of independent variables are comparatively large relative to sample size.

Experimental

Material and methods

Calculation of topological indices. All the topological indices referred above are well defined in the literature and thus we will not discuss them in details but will give the appropriate expressions for their calculations.

Negentropy (N). A drug molecule is considered to be an information source with an information content available to respective tissue. In non-specific interactions, much of the information content has quality as judged by the receptor. Quantitation of the information content using Shannon's equation gives the molecular negentropy. The negentropy is thus the negative of information entropy. This index is shown to rank molecules according to symmetry and to encode structural characteristics influencing physico-chemical properties and biological activity in certain cases.

The negentropy per atom[12] (i) is calculated using the following expression and subsequently by multiplying it with number of atoms present in the molecule gives molecular negentropy (N):

$$i = -K\sum_j P_j \log P_j \tag{1}$$

where K is a constant depending on the logarithmic base, j is the set, and P_j is the complete array of probabilities for the sets.

Molecular redundancy index (MRI). MRI[13] is a molecular symmetry descriptor and indicates the capacity and symmetry of a molecule this can be computed using the following expression:

$$\mathrm{MRI} = \frac{\sum n_i \log n_i}{N \log N} \tag{2}$$

where n is the number of atoms of the same kind in the ith atom set, i is the number of different atoms sets and $N = \Sigma n_i$ is the total number of atoms in the molecule.

Molecular valence connectivity index ($^m\chi^v$). The connectivity index χ (G) = χ of a molecular graph G is defined[14] as:

$$\chi = \sum_{\text{edges}} [\mathrm{d}(i)\mathrm{d}(j)]^{-0.5}$$

where the sum is taken over all edges of G, while d(i) and d(j) are valencies of vertices i and j making up the edge i–j.

In the case of hetero-systems, the connectivity index is given in terms of valency delta δ^v.

The valence connectivity index[14,15] is calculated using the following expression:

$$^m\chi^v = {}^m\chi^v(\mathrm{G}) = \sum_{ij} [\delta_i^v \delta_j^v \ldots\ldots \delta_m^v]^{-0.5} \tag{3}$$

where the sum is taken over all bonds i–j of the molecule. Valence delta values are given by

$$\delta_i^v = \frac{Z_i^v - H_i}{Z_i - Z_j - 1} \tag{4}$$

where Z_i is the atomic number of atom i, Z_i^v is the number of valence electron of the atom i and H_i is the number of hydrogen atoms attached to atom i.

Indicator parameters (Ip_1 and Ip_2). Indicator parameters are dummy parameters accounting for structural characteristics not covered in the definition of topological indices used. In the present case we have used two such dummy parameters, namely Ip_1 and Ip_2, such that Ip_1 is taken as unity if one benzene ring is present in the substituent R_4, otherwise it is zero. Similarly, Ip_2 is an indicator parameter which assumed the value of unity when nitrogen is present in the substituent R_4, otherwise its value is taken as zero. These indicator parameters, however, cannot differentiate among compounds **1**, **9**, and **12**. Looking to the size of the samples, we cannot use more indicator parameters.

Statistical analysis. We have used the maximum R^2 improvement method to identify prediction models. This method finds the 'best' one variable model, the 'best' two variable model and so forth for the prediction of property/activity. Several models (combinations of variables) were examined to identify combinations of variables with good prediction capabilities. In all regression models developed we have examined a variety of statistics associated with residues, that is the Wilks–Shapiro test for normality and Cooks D-statistics for outliers, to obtain the most reliable results.[18]

Acknowledgements

The authors wish to thank Professor Istvan Lukovits, Hungarian Academy of Sciences, Budapest, Hungary for providing Regres-1 program for carrying out multiple regression analyses for correlating antagonists activity.

References and Notes

1. Arrang, J. M.; Garbarg, M.; Schwartz, J. C. *Nature, (London)* **1983**, *302*, 832.
2. Schlicker, E.; Betz, R.; Gothert, M. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **1988**, *337*, 588.
3. Claphan, J.; Kilpatrick, G. J. *Br. J. Pharmacol.* **1992**, *107*, 919.
4. Molderings, G. J.; Weibenborn, G.; Schlicker, E.; Likungu, J.; Gothert, M. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **1992**, *346*, 46.
5. Schlicker, E.; Fink, K.; Detzner, M.; Gothert, M. *J. Neural. Transm.* **1993**, *93*, 1.
6. Garbarg, M.; Arrang, J. M.; Lloren-Cortes, C. In *Presynaptic Receptors and Neuronal Transporters*; Lager, S. Z., Galzin, M., Consteritin, J., Eds; Pergamon: Oxford, 1991.
7. Lin, J. S.; Sakai, K.; Vanni-Mercier, G. *Brain. Res.* **1990**, *523*, 325.
8. Lipp, R.; Stark, H.; Schunack, W. *Ser. Receptor Biochemistry and methodology* **1992**, *16*, 57.
9. Arrang, J. M.; Garbarg, M.; Lancelot, J. C. *Nature, (London)* **1987**, *327*, 117.
10. Lipp, R.; Arrang, J. M.; Garbarg, M.; Luger, P.; Schwartz, J. C.; Schunack, W. *J. Med. Chem.* **1992**, *35*, 4434.
11. Stark, H.; Lipp, R.; Arrang, J. M.; Garbarg, M.; Schwartz, J. C.; Luger, P.; Schunack, W. *Eur. J. Med. Chem.* **1994**, *29*, 695.
12. Shanon, C. A.; Weaver, W. *The Mathematical Theory of Communication*; University of Illinois Press: Urbana Ill, 1949.
13. Kier, L. B. *J . Pharm. Sci.* **1980**, *69*, 807.
14. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Relationship*; Wiley, New York, 1986.
15. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
16. Balaban, A. T. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 23.
17. Balaban, A. T. *Math. Chem.* **1986**, *21*, 115.
18. Chaterjee, S.; Hadi, A. S.; Price, B. *Regression Analysis by Examples*, 3rd Ed.; Wiley: New York, 2000.
19. Pogliani, L. *Amino Acids* **1994**, *6*, 141.