ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

3D-QSAR Models to Predict Antiamoebic Activities of the Cyclised pyrazolines and 2-(quinolin-8-yloxy) acetohydrazones.

Samira Mbarki[1]*, Menana ELHALLAOUI [1]

[1]LIMME, Département de Chimie Faculté des Sciences Dhar El Mahraz, Université Sidi Mohammed Ben Abdallah, B.P.1796 ATLAS. Fès. Maroc.

ABSTRACT

The present article is an attempt to the 3D QSAR studies for 39 molecules of 1-N-substituted cyclised pyrazoline analogues of thiosemicarbazones **1-21** and 2-(quinolin-8-yloxy)acetohydrazones and their cyclised products **22-39** by using The multiple linear regression method (MLR) and artificial neural network (ANN) techniques , considering the relevant descriptors obtained from the MLR; a correlation coefficient of 0.974 was obtained with an 4-3-1 ANN model. As a result of quantitative structure-activity relationship of 2-(quinolin-8-yloxy) acetohydrazones and their cyclised products(1;2;3-thiadiazole and 1;2;2selenadiazole) derivatives, we found that the model proposed in this study is constituted of major descriptors used to describe these molecules. This model is statistically significant and shows very good stability towards data variation in leave-one-out (LOO) cross-validation (r_{cv} =0.942).

Keywords: Antiamoebic activity; 3D-QSAR model; MLR; ANN; LOO

Corresponding author

ISSN: 0975-8585

INTRODUCTION

Amoebiasis is the second leading cause of death from parasitic Disease worldwide [1]. More than 50 million people worldwide are infected and up to 110,000 of these die every year [2]. Infection is primarily treated by instituting antiamoebic therapy. Antiamoebic drugs such as metronidazole, tinidazole, ornidazole, emetine kill amoeba in host tissue and organs (tissue amoebicides) whereas drugs like iodoquinol, diloxanide furoate, paromomycin act on large intestine (luminal amoebicides) are used for treatment. Particularly metronidazole is the most preferred treatment choice as 90% of patients respond to the therapy[3]. Also in the last several years a large number of new compounds have been isolated and/or synthesized of which a few have shown in vitro activity against E. histolytica. However, resistances to metronidazole in many pathogenic bacteria and protozoa as well as several side effects are also well documented [4]. Therefore, the development of new alternative antiamoebic drugs devoid of side effects is still needed.To find the structural requirements for more active amoebicidal agents, comparative QSAR studies had been performed in an antecedent work on Construction of 3D-QSAR models to predict antiamoebic activities of pyrazoline and dioxazoles derivatives [5] and 3D-QSAR for α-GLUCOSIDASE inhibitory activity of N-(PHENOXYALKYL) PHTHALIMIDE derivatives [6]. In this paper we achieve 3D-QSAR studies on some 1-N-substituted thiocarbamoyl-3-phenyl-2-pyrazolines for the same goal.

The quantitative structure-activity relationships (QSAR) are certainly a major factor in contemporary drug design. Thus, it is quite clear why a large number of users of QSAR [7-8] are located in industrial research units. So, Classical QSAR and 3D-QSAR are highly active areas of research in drug design [9-10].

The basis for various quantitative structure–activity relationship (QSAR) methods is the 'description' of the molecular structures by means of numbers. At present, there are a large number of molecular descriptors that can be used in QSAR studies [11-15]. For instance, computer programs such as Dragon5.5 compute up to 3224 descriptors, which may have very different complexity but can be classified according to their 'dimensionality' in: zero dimensional 0D, 1D, 2D, and 3D molecular descriptors.

To establish the relation between structural characteristics of molecule and its properties the mathematical methods can be used. Multiple Linear Regression (MLR) analysis and artificial Neural Network (ANN) are of the mathematical methods which have an extent application, in this work calculations are applied to a series of 39 molecules 1-N-substituted cyclised pyrazolines analogues of thiosemicarbazones 1-21 and 2-(quinolin-8-yloxy)acetohydrazones and their cyclised products 22-39 in order to set up a 3D-QSAR model able to predict antiamoebic activity.

MATERIALS

Experimental data

The experimental IC_{50}(µM) antiamoebic activities of cyclised pyrazoline and acetohydrazones are collected from recent publications[16-17]. The observations are

ISSN: 0975-8585

converted into minus logarithm scale logIC_{50} and are included in Tables **2-4**.

Calculation of molecular descriptors

The initial conformations of the compounds are drawn with "model build" modulus available in ChemOffice 2004. Each molecular structure is firstly pre-optimized with the Molecular Mechanics Force Field (MM+) procedure. For each compound, the numerical descriptors (Table 1) are calculated with Dragon version 5.5-2007 which includes several variable types characterizing the 1D, 2D, and 3D structure aspects: constitutional, topological, geometrical, charge…..

Table 1: Descriptors chosen for the QSAR model, and used in this study.

Category of descriptors	Name of the descriptors
Constitutional descriptors	Molecular Weight (MW) Sum of atomic van der waals volumes (Sv) Sum of atomic polarizabilities (Sp) Mean electropological state (Ms) Number of non-H atoms (nSK) Number of non-H bonds (nBO) Number of Sulfur atoms (nS) Number of Carbon atoms (nC)
Topological descriptors	Balaban distane connectivity index (J) Harary H index(Har) Wiener W index (W)
Geometrical descriptors	Harmonic oscillator model of aromaticity index(HOMA) Harmonic oscillator model of aromaticity index total (HOMT) 3D-Wiener index (W3D) 3D-Balaban index (J3D) 3D-Harary index (H3D)
Molecular properties	Hydrophilic factor (Hy) Ghose-grippen molar refractivity (AMR) Moriguchi octanol-water partition coeff(logP)(MlogP) Ghose-Crippen octanol-water partition coeff(logP)(AlogP)

METHODS

Multiple linear regressions (MLR)

The statistic technique multiple linear regression is used to study the relation between one dependent variable and several independent variables. It is a mathematic technique that minimizes differences between actual and predicted values. The multiple linear regression model (MLR) was generated using the software SYSTAT, version 12, to

predict antiamoebic activities logIC50. It has served also to select the descriptors used as the input parameters for a back propagation network (ANN).

Artificial neural network

All the feed-forward ANN used in this paper are three-layer networks, the first (input) layer contains five neurones, representing the relevant descriptors obtained in MLR technique. Although there are neither theoretical nor empirical rules to determinate the number of hidden layers or the number of neurone layers, one hidden layer seems to be sufficient in the most chemical application of ANN. Some authors [18-19] have proposed a parameter ρ, leading to determine the number of hidden neurons, which plays a major role in determining the best ANN architecture. It is defined as follows:

ρ = (Number of data points in the training set / Sum of the number of connections in the NN).

Therefore, in order to avoid overfitting or underfitting, it is recommended to take into account the ρ value; $1.8 < \rho < 2.3$ [20]. Thus, the ANN used in this work is formed by three hidden neurones, and the output layer represents the calculated activities values log (IC50). So, the final ANN architecture is (4-3-1), it is depicted in figure 1.
All calculations of NN are done on Matlab 7 using our program written in C language.

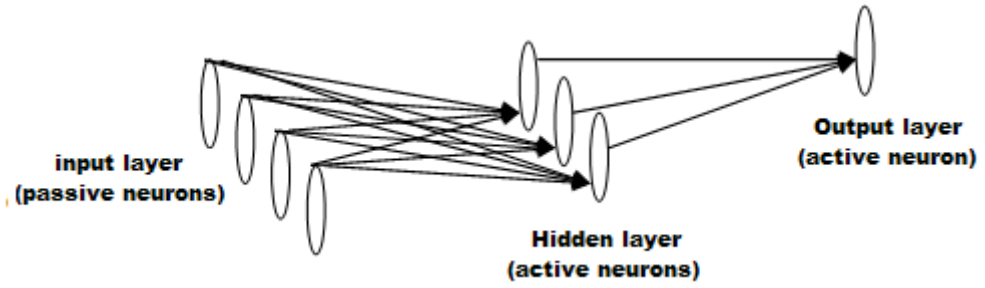


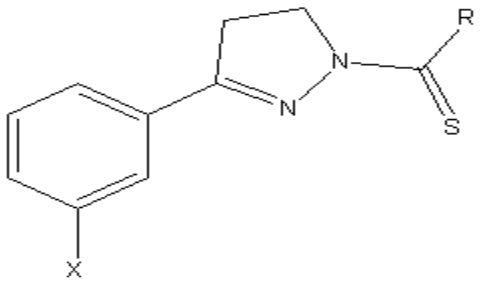
Figure 1. Schematic representation architecture (4-3-1) of the three-layer neural network used in this work.

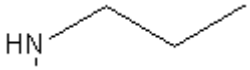
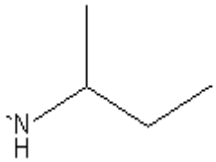
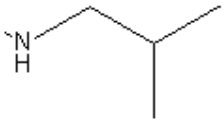
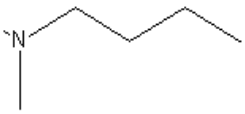
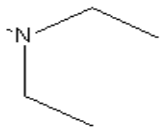
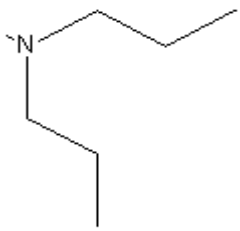
Cross-validation technique

Cross-validation is a popular technique used to explore the reliability of statistical models. Based on this technique, a number of modified data sets are created by deleting in each case one or a small group of molecules, these procedures are named respectively “leave-one-out” and “leave-some-out” [21-23]. For each data set, an input-output model is developed. The model is evaluated by measuring its accuracy in predicting the responses of the remaining data (the ones that have not been used in the development of the model). In this study we used, the leave-one-out (LOO) procedure.

ISSN: 0975-8585

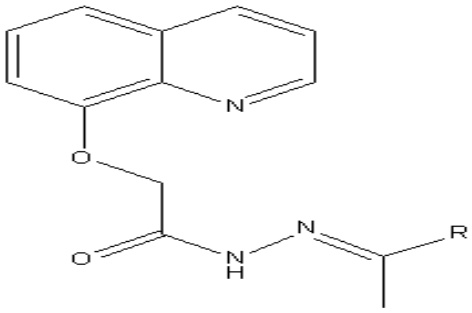
Table 2: Studied compounds and their observed antiamoebic activities $logIC_{50}$ (Obs), and calculated $logIC_{50}$ with MLR; ANN and CV methods.



N°	X	R	$logIC_{50}$(Obs)	$logIC_{50}(Cal_{MLR})$	$logIC_{50}(Cal_{ANN})$	$logIC_{50}(prd_{LOO})$
1	H		1.140	1.332	1.213	1.206
2	Br		0.929	0.914	1.057	1.092
3	Cl		0.903	0.875	0.972	0.978
4	H		1.362	1.435	1.151	1.258
5	Br		1.182	1.101	1.030	1.080
6	Cl		1.086	0.968	0.960	1.213
7	H		1.367	1.263	1.313	1.212
8	Br		1.152	0.995	1.115	1.151
9	Cl		1.090	0.834	0.996	0.967
10	H		1.040	1.252	1.185	1.097
11	Br		0.785	0.986	0.785	0.958
12	Cl		0.690	0.788	0.974	0.609
13	H		0.756	0.559	0.563	0.330
14	Br		0.380	0,408	0.183	0.160
15	Cl		-0.155	0.144	-0.315	-0.162
16	H		0.623	0.400	0.580	0.540
17	Br		0.079	0.079	0.209	0.233
18	Cl		0.000	0.059	0.001	-0.158
19	H					
20	Br		0.301	0.557	0.730	0.374
21	Cl		-0.097	-0.176	0.101	-0.019
			-0.222	-0.289	-0.303	-0.395

ISSN: 0975-8585

Table 3: Studied compounds and their observed antiamoebic activities $logIC_{50}$(Obs), and calculated $logIC_{50}$ with MLR; ANN and CV methods.



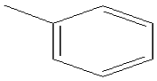
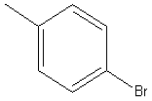
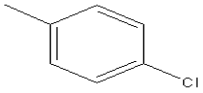
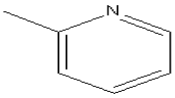
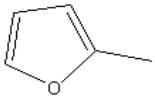
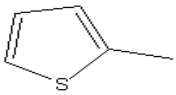
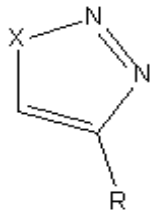
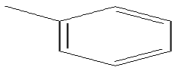
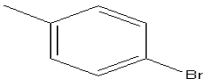
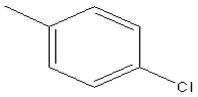
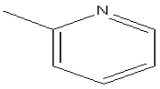
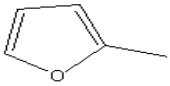
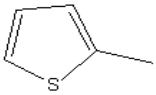
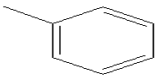
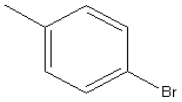
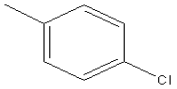
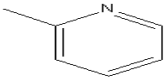
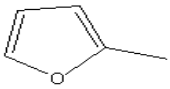
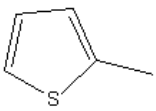
N°	R	$logIC_{50}$(Obs)	$logIC_{50}(Cal_{MLR})$	$logIC_{50}(Cal_{ANN})$	$logIC_{50}(prd_{LOO})$
22		-1.523	-1.542	-1.334	-1.534
23		-1.398	-1.344	-1.335	-1.377
24		-1.398	-1.427	-1.336	-1.274
25		-1.301	-1.271	-1.335	-1.351
26		-1.301	-1.494	-1.335	-1.369
27		-1.097	-0.964	-1.335	-1.373

Table 4: Studied compounds and their observed antiamoebic activities $logIC_{50}$(Obs), and calculated $logIC_{50}$ with MLR; ANN and CV methods.



ISSN: 0975-8585

N°	X	R	$logIC_{50}$(Obs)	$logIC_{50}(Cal_{MLR})$	$logIC_{50}(Cal_{ANN})$	$logIC_{50}(prd_{LOO})$
28	**S**		**0.873**	**0.471**	**0.488**	**0.449**
29	**S**		**-0.620**	**-0.266**	**0.002**	**-0.036**
30	**S**		**-0.409**	**-0.289**	**-0.408**	**-0.540**
31	**S**		**0.737**	**0.231**	**0.737**	**0.217**
32	S		**-0.638**	**-0.083**	**-0.638**	**0.443**
33	S		**0.352**	**0.310**	**0.214**	**0.148**
34	Se		**0.659**	**0.676**	**0.706**	**0.647**
35	Se		**0.470**	**0.137**	**0.262**	**0.114**
36	Se		**-0.237**	**-0.080**	**-0.438**	**-0.349**
37	Se		**-0.337**	**0.511**	**-0.007**	**-0.291**
38	Se		**0.757**	**0.183**	**0.757**	**0.016**
39	Se		0.723	0.605	0.567	0.689

RESULTS AND DISCUSSION

Multiple Linear Regressions

The QSAR model built using multiple linear regression (MLR) method is represented by the following equation:

ISSN: 0975-8585

$$\mathbf{LogIC_{50}} = 8.634 - 0.139\mathbf{(Sv)} - 4.150\mathbf{(Ms)} + 1.960\mathbf{(J3D)} + 2.958\mathbf{(Hy)}$$

n= 39 r =0.939 s= 0.315 F-ratio = 63.171

Where n is the number of compounds, r is the correlation coefficient, s is the standard deviation, F is the Fisher F-statistic.

We can notice that the descriptors related to the Constitutional descriptors (Ms; Sv), Geometrical descriptor (J3D), and Molecular propertie (Hy) are the most important descriptors in the establishment of the QSAR model for the cyclised pyrazolines and acetohydrazones.

The correlation of the observed activities with the MLR calculated ones is illustrated in figure 2.

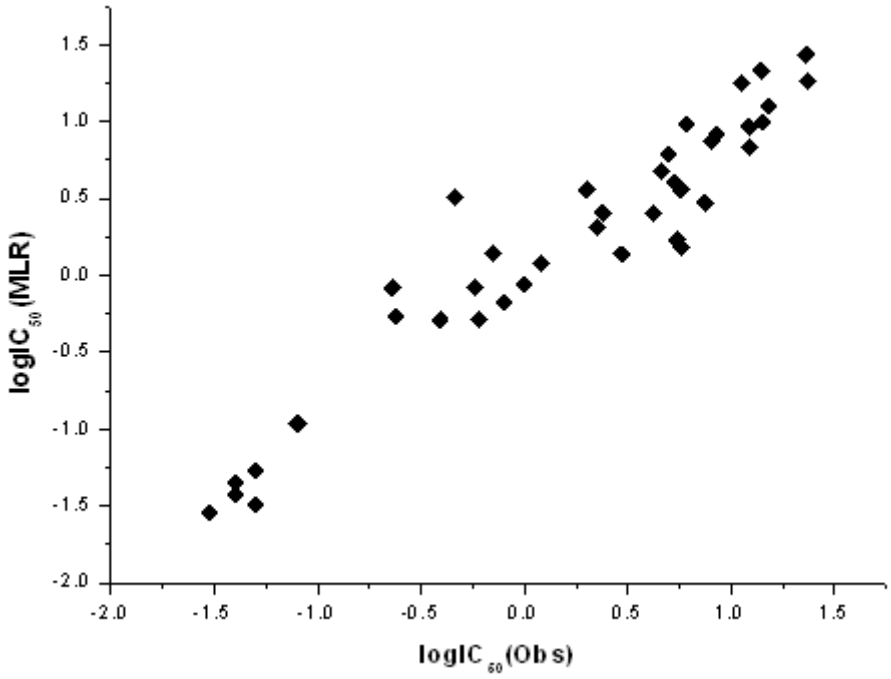


Figure 2. Plot of observed versus predicted antiamoebic activities by (MLR)

Neural networks

Neural networks (NN) can be used to generate predictive model of quantitative structure–activity relationship (QSAR) between a set of molecular descriptors obtained from the MLR and Observed activities.

The correlation of the observed activities with the ANN calculated ones are illustrated in Figure 3.

n= 39 r = 0.974 s = 0.385

ISSN: 0975-8585

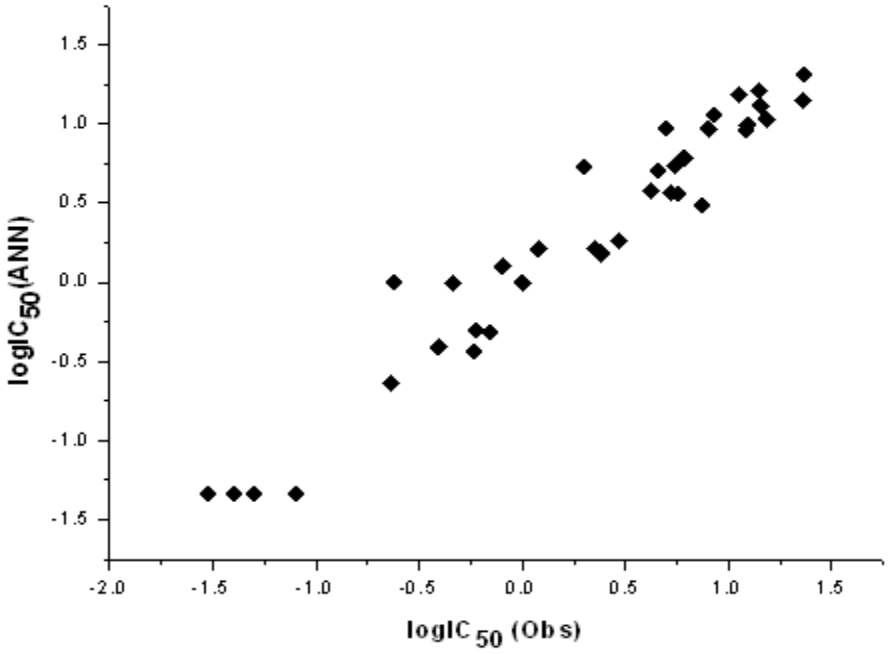


Figure 3. Plot of observed versus predicted antiamoebic activities by (ANN)

The correlation coefficient **r = 0.974** and Standard Error of Estimate **s = 0.385**, obtained with the Neural network, show that the selected descriptors by MLR are pertinent and that the model proposed to predict activity is relevant.

Validation

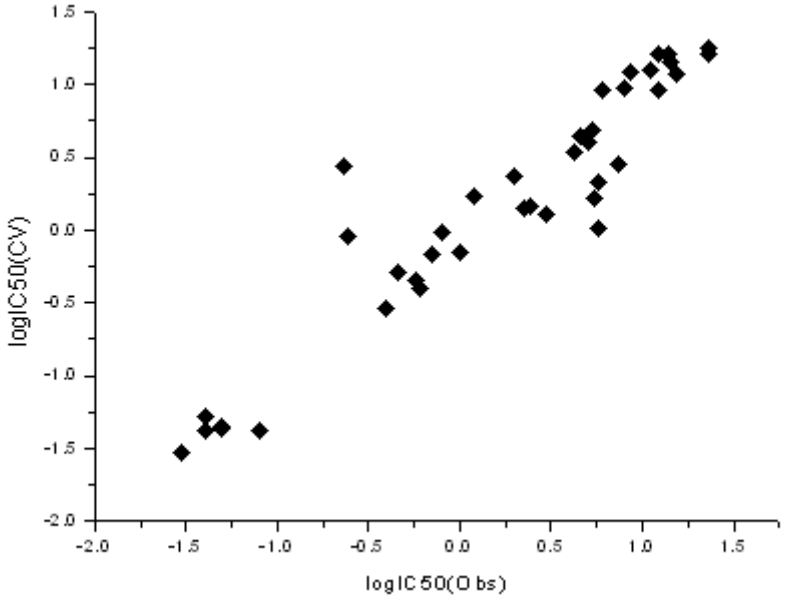


Figure 4. Plot of observed versus predicted antiamoebic activities by (CV)

Before using a QSAR model to predict the activity of new compounds, we should validate it using a validation method. In this paper we validated our model with cross

validation using LOO procedure.

n= 39 r_{cv} =0.942 s = 0.282

The correlation of the observed activities with the CV calculated ones is illustrated in figure 4.

A good correlation was obtained with cross validation r_{cv}=0.942. So the predictive power of this model is very significant.

The most important result of this investigation is that in vitro antiamoebic activity could be predicted using QSAR methods. So, the model proposed in this study shows high predictive power (r_{cv} = 0.942).

One of the most important observations that can be drawn from this study is that different descriptors representing the majority of classes of descriptors proposed to build a QSAR model were selected. Therefore, we conclude that the antiamoebic activity is related to the Constitutional, Topological, Molecular and Quantum-chemical descriptors.

CONCLUSION

In this study, QSAR model includes some Molecular descriptors, regression quality indicates that these descriptors provide valuable information and have significant role in the assessment of the activity of cyclised the pyrazoline and 2-(quinolin-8-yloxy) acetohydrazones. The artificial neural network (ANN) techniques, considering the relevant descriptors obtained from the MLR, showed good agreement between the observed and the predicted values was excellent.

REFERENCES

[1] SL Stanley Jr. Lancet 2003;361:1025-1034
[2] Schuster H, Chiodini PL. Curr Opin Infect Dis 2001;14:587–591.
[3] A Azam, SM Agarwal. Curr Bioact Compd 2007;3:121-133.
[4] C Wassmann, A Hellberg, E Tannich, I Bruchhaus. J Biol Chem 1999;274:26051-26056.
[5] S Mbarki, K Dguigui, M El Hallaoui. J Mater Environ Sci 2011;2:61-70.
[6] S Mbarki, M El Hallaoui, K Dguigui. IJRRAS 2012;11:395-401.
[7] Trinajstic N. Chemical Graph Theory, 2nd Edn.; Boca Raton, F.L. Eds.; CRC Press, 1992; 20.
[8] a) Bazoui H, Zahouily M, Boulaajaj S, Sebti S, Zakarya D. Environ Res 2002 ;13 :567, b) Bazoui H, Zahouily M, Sebti S, Boulaajaj S, Zakarya D. J Mol Model 2002;8: 1-7, c) Agrawala VK, Singha J, Mishra KC, Khadikar PV, Jaliwalac YA. Arkivoc 2006;ii:162.
[9] Kubinyi H. QSAR: Hansch analysis and related approaches. In: Mannhold R, Krogsgarrd Larsen P, Timmerman H (eds) Methods and principles in medicinal chemistry, vol 1. Wiley, Weinheim; 1993.
[10] Kubinyi H (ed) 3D QSAR in drug design: theory, methods and applications. ESCOM, Leiden 1993.
[11] Todeschini R, Consonni V. Handbook of Molecular Descriptors; Wiley-VCH: 2000,

Germany.
[12] Karelson M. Molecular Descriptors in QSAR/QSPR; John Wiley & Sons: New York, 2000.
[13] Diudea MV, Ed. QSPR/QSAR Studies by Molecular Descriptors; Nova Science: Huntington, NY, 2001.
[14] Balaban AT, Ed. From Chemical Graphs to Three-Dimensional Geometry; 1997. New York.
[15] Balaban A. Environ Res 1998;8:1.
[16] Mohammad Abid, Amir Azam. Bioorg Med Chem 2005;13:2213-2220.
[17] Faisal Hayat, Attar Salahuddin, Jamil Zargan, Amir Azam. European J Med Chem 2010;45:6127-6134.
[18] So S, Richards G. J Med Chem 1992;35:3207.
[19] Andrea TA, Kalayeh H. J Med Chem 1991;34:2824.
[20] M Ellhalaoui, Modélisatrice moléculaire et étude QSAR d'antagonistes non compétitifs du récepteur NMDA par les méthodes statistiques et le réseau de neurones; Thèse de doctorat, 2002, 106.
[21] Efron B. J Am Stat Assoc 1983;78:316.
[22] Efroymson MA. Multiple regression analysis. In Mathematical Methods for Digital Computers ;Ralston A, Wilf HS. Eds. Wiley: New York, 1960.
[23] Osten DW. J Chemom 1998;2:39.