Construction of 3D-QSAR models to predict antiamoebic activities of pyrazoline and dioxazoles derivatives

S. Mbarki [*], K. Dguigui , M. El Hallaoui

LCPPEM, Département de Chimie Faculté des Sciences Dhar El Mahraz, Université Sidi Mohammed ben Abdallah, B.P.1796 ATLAS. Fès. Maroc.

Received in 13 Dec 2010, Revised 22 Jan 2011, Accepted 22 Jan 2011.
** Corresponding author : E-mail : samiramba@hotmail.com; Tel : +212 679 445 114*

Abstract

1-N-substituted thiocarbamoyl-3-phenyl-2-pyrazolines and 3,5-substituted-1,4,2-dioxazoles are potent antiamoebic agents. A 3D-QSAR study is applied to a set of 63 molecules. With The multiple linear regression method (MLR) (r = 0.95), the predicted values of activities are in good agreement with the experimental results. The artificial neural network (ANN) techniques, considering the relevant descriptors obtained from the MLR, showed good results; a correlation coefficient of 0.96 was obtained with an 8-3-1 ANN model. As a result of quantitative structure-activity relationships between 1-N-substituted thiocarbamoyl-3-phenyl-2-pyrazolines and 3,5-substituted-1,4,2-dioxazoles, 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} =0.90).

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

1. Introduction

Amoebiasis is the infection of human gastrointestinal tract by Entamoeba histolytica (E.histolytica), a protozoan parasite capable of invading the intestinal mucosa and may spread to other organs, mainly the liver which usually leads to amoebic liver abscess. This infection remains a significant cause of morbidity and mortality world-wide [1].

This gastrointestinal infection may or may not be symptomatic. It can remain latent in an infected person for several years; amoebiasis is estimated to cause 70,000 deaths per year world wide [2]. Symptoms can range from mild diarrhea to dysentery with blood and mucus in the stool. E. histolytica is usually a commensal organism [3]. Severe amoebiasis infections (known as invasive or *fulminant* amoebiasis) occur in two major forms. Invasion of the intestinal lining causes amoebic dysentery or amoebic colitis. If the parasite reaches the bloodstream it can spread through the body, most frequently ending up in the liver where it causes amoebic liver abscesses. Liver abscesses can occur without previous development of amoebic dysentery. When no symptoms are present, the infected individual is still a carrier, able to spread the parasite to others through poor hygienic practices. While symptoms at onset can be similar to bacillary dysentery, amoebiasis is not bacteriological in origin and treatments differ, although both infections can be prevented by good sanitary practices.

Infection is primarily treated by instituting antiamoebic therapy. Drugs of choice for invasive amoebiasis are tissue active agents, like metronidazole, tinidazole and chloroquine or the more toxic emetine derivatives, including dehydroemetine [4]**.** Recent studies tried to improve the treatment of this infection by developing

antiamoebic therapy [5,6], a set of dioxazoles derivatives showed better activity than the reference drug metronidazole; furthermore, they are non toxic to the human kidney epithelial cells. In the other hand, QSAR studies were reported to identify important structural features responsible for the antiamoebic activity [7].
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 [8,9] are located in industrial research units. So, Classical QSAR and 3D-QSAR are highly active areas of research in drug design [10-11].
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 [12-16]. 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.
In this study, Multiple Linear Regression (MLR) analysis and artificial Neural Network (ANN) calculations are applied to a series of 63 1-N-substituted thiocarbamoyl-3-phenyl-2-pyrazolines and 3,5-substituted-1,4,2-dioxazoles, in order to set up a 3D-QSAR model able to predict antiamoebic activity.

2. Materials

2-1 Experimental data
The experimental IC_{50}(µM) Antiamoebic activities of 1-N-substituted thiocarbamoyl-3-phenyl-2-pyrazolines and 3,5-substituted-1,4,2-dioxazoles are collected from recent publications [5,6]. The observations are converted into minus logarithm scale $logIC_{50}$ and are included in Tables **2-5**.

2-2 Calculation of molecular descriptors
The initial conformations of the compounds are drawn with the "model build" modulus available in ChemOffice 2004. Each molecular structure is firstly pre-optimized with the Molecular Mechanics Force Field (MM+) procedure. The numerical descriptors (see Table 1) for each compound are calculated with Dragon Evaluation version 5.5-2007 which includes several variable types characterizing the 1D, 2D, and 3D structure aspects: constitutional, topological, geometrical, charge….. We have also introduced quantum-chemical descriptors such as HOMO and LUMO energies, and Dipole Length.

3. Methods

3-1 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 $logIC_{50}$. It has served also to select the descriptors used as the input parameters for a back propagation network (ANN).

3-2 Artificial neural network
All the feed-forward ANN used in this paper are three-layer networks, the first (input) layer contains eight 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 [17,18] have proposed a parameter ρ, leading to determine the number of hidden neurons, which plays a major role in determining the best ANN architecture defined as follows:
ρ = (Number of data points in the training set / Sum of the number of connections in the NN).
In order to avoid overfitting or underfitting, it is recommended that $1.8 < \rho < 2.3$ [19].So with three hidden neurones. The output layer represents the calculated activity values log (IC_{50}). The architecture of the ANN used in this work (8-3-1) is depicted in figure 1.
All calculations of NN are done on Matlab 7 using our program written in C language.

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 atomic van der waals volume (Mv) Mean atomic sanders on electonegativity (Me) Mean atomic polarizability (Mp) Mean electropological state (Ms) Number of atoms (nAT) Number of non-H atoms (nSK) Number of bonds(nBT) Number of non-H bonds (nBO)
Geometrical descriptors	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)
Topological descriptors	Balaban distane connectivity index (J) Polarity number (Pol)
Topological charge indices	Global topological charge indix (JGT)
Connectivity indices	Modified Randic connectivity index (XMOD)
Quantum-chemical	Dipole Length(μ) Highest occupied molecular orbital (HOMO) Lowest unoccupied molecular orbital (LUMO)

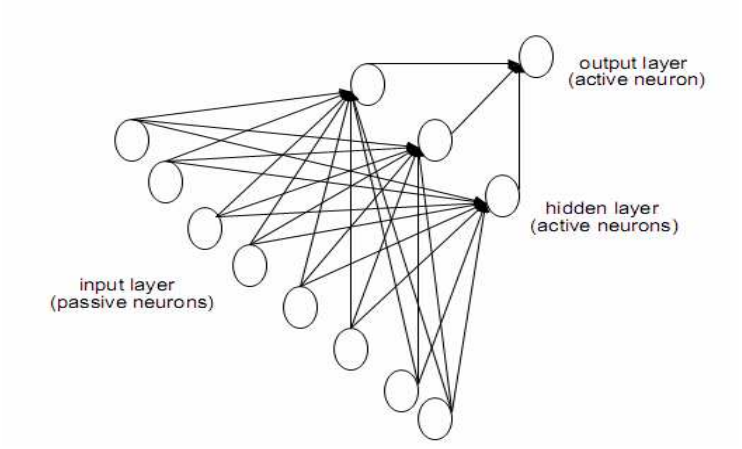
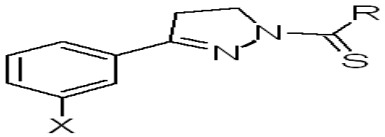


Figure 1: Schematic representation architecture (8-3-1) of the three-layer neural network used in this work.

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



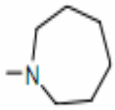
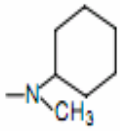
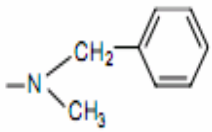
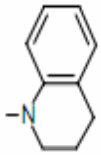
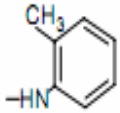
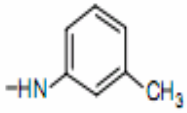
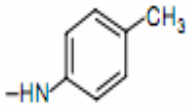
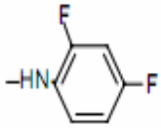
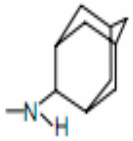
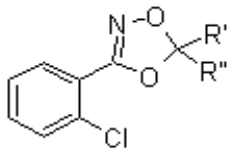
N°	X	R	$logIC_{50(obs)}$	$logIC_{50(MLR)}$	$logIC_{50(ANN)}$	$logIC_{50(CV)}$
1	H		0,572	0,646	0,575	0.443
2	Br		0,450	0,355	0,439	0.445
3	Cl		0,364	0,305	0,401	0.276
4	H		0,642	0,633	0,645	0.419
5	Br		0,037	0,137	0,394	0.458
6	Cl		-0,051	0,041	0,389	0.021
7	H		0,774	0,659	0,798	0.564
8	Br		0,720	0,736	0,638	0.635
9	Cl		0,569	0,634	0,431	0.519
10	H		0,864	0,646	0,860	0.425
11	Br		0,647	0,686	0,696	0.439
12	Cl		0,464	0,582	0,438	0.432
13	H		0,792	0,776	0,788	0.615
14	Br		0,444	0,454	0,444	0.625
15	Cl		0,248	0,250	0,248	-0.288
16	H		0,679	0,564	0,612	0.62
17	Br		0,582	0,419	0,528	0.55
18	Cl		0,225	0,415	0,225	0.438
19	H		0,700	0,698	0,781	0.831
20	Br		0,525	0,536	0,604	0.657
21	Cl		0,449	0,531	0,426	0.294
22	H		0,980	0,808	0,937	0.861
23	Br		0,727	0,654	0,715	0.658
24	Cl		0,380	0,650	0,446	0.276
25	H		0,246	0,118	0,246	0.347
26	Br		-0,174	-0,134	-0,174	-0.328
27	Cl		-0,292	-0,236	-0,292	-0.329
28	H		0,253	0,239	0,253	0.131
29	Br		-0,237	-0,195	-0,282	-0.233
30	Cl		-0,328	-0,354	-0,283	-0.229

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



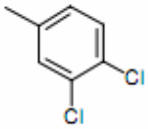
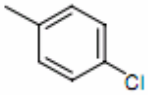
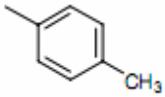
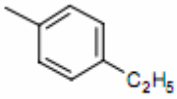
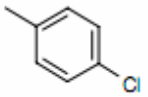
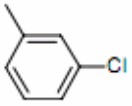
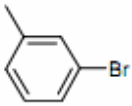
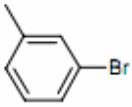
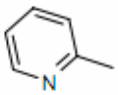
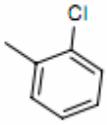
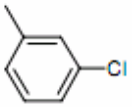
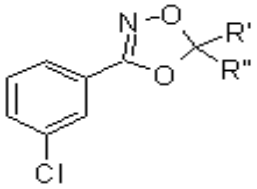
N°	R'	R"	$logIC_{50(obs)}$	$logIC_{50(MLR)}$	$logIC_{50(ANN)}$	$logIC_{50(CV)}$
31	H		-0.092	-0.204	-0.012	-0.163
32	H		-0.292	-0.136	-0.202	-0.187
33	H		0.494	0.409	0.388	0.336
34	H		0.486	0.323	0.388	0.383
35	CH_3		0.461	0.444	0.388	0.391
36	C_2H_5		0.400	0.308	0.388	0.373
37	CH_3		0.364	0.295	0.388	0.331
38	C_2H_5		0.408	0.450	0.388	0.405
39	CH_3		0.210	0.168	0.173	0.443
40	H		-0.387	-0.259	-0.391	-0.418
41	H		-0.143	-0.182	-0.225	-0.242

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



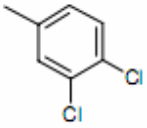
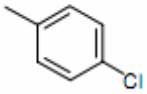
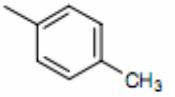
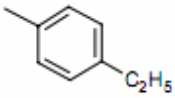
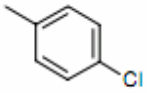
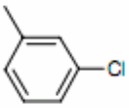
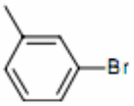
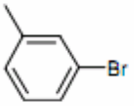
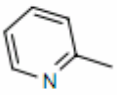
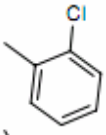
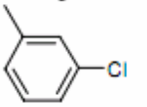
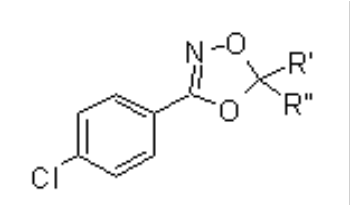
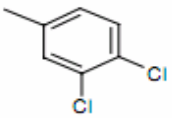
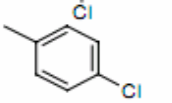
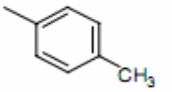
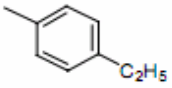
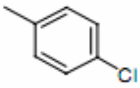
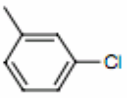
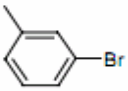
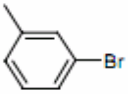
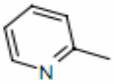
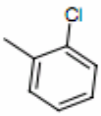
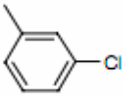
N°	R'	R"	$logIC_{50(obs)}$	$logIC_{50(MLR)}$	$logIC_{50(ANN)}$	$logIC_{50(CV)}$
42	**H**		**0.083**	**-0.106**	**0.078**	**0.06**
43	**H**		**-0.125**	**-0.030**	**-0.060**	**-0.087**
44	**H**		**0.452**	**0.512**	**0.388**	**0.422**
45	**H**		**0.444**	**0.496**	**0.388**	**0.385**
46	CH_3		**0.468**	**0.562**	**0.388**	**0.437**
47	C_2H_5		**0.441**	**0.409**	**0.388**	**0.437**
48	CH_3		**0.433**	**0.385**	**0.388**	**0.428**
49	C_2H_5		**0.367**	**0.543**	**0.388**	**0.412**
50	CH_3		**0.238**	**0.335**	**0.262**	**0.217**
51	H		**-0.208**	**-0.133**	**-0.175**	**-0.219**
52	H		**-0.041**	**-0.094**	**-0.111**	**-0.136**

Table 5 Studied compounds and their observed antiamoebic activities $logIC_{50(obs)}$, and calculated $logIC_{50}$ with MLR; ANN and CV methods.



N°	R'	R"	$logIC_{50(obs)}$	$logIC_{50(MLR)}$	$logIC_{50(ANN)}$	$logIC_{50(CV)}$
53	H		0.053	-0.215	-0.021	-0.032
54	H		-0.276	-0.149	-0.252	-0.114
55	H		0.433	0.519	0.388	0.385
56	H		0.373	0.420	0.388	0.392
57	CH_3		0.433	0.426	0.388	0.385
58	C_2H_5		0.389	0.294	0.388	0.458
59	CH_3		0.417	0.335	0.388	0.464
60	C_2H_5		0.403	0.443	0.388	0.395
61	CH_3		0.199	0.255	0.223	0.229
62	H		-0.319	-0.193	-0.287	-0.306
63	H		-0.066	-0.131	-0.157	-0.214

3-3 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" [20-22]. 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.

4. Results and discussion

4-1 Multiple linear regressions

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

$$LogIC_{50} = -68.413 - 1.272(HOMO) + 10.172(S_V) + 2.387(J) + 6.326(JGT) - 1.487(MlogP) + 22.414(MS) - 1.423(nAT) - 7.022(nSk)$$

n= 63 r =0.95 s= 0.133 F-ratio = 15.61

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; nSK; nAT), Topological charge indices (JGT) , Topological descriptors (J) , Molecular properties (MlogP) and Quantum-chemical (HOMO) are the most important in the establishment of the QSAR model for pyrazoline and dioxazoles derivatives.

The correlation of the observed activities with the RLM calculated ones are illustrated in figure 2.

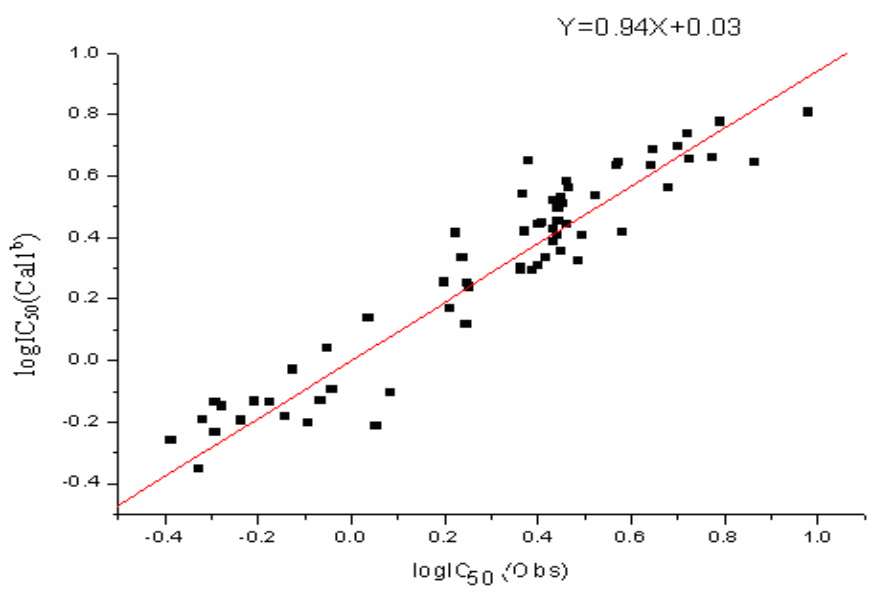


Figure. 2. Predicted antiamoebic activities by (MLR) in comparison with experimental values

4-2 Artificial Neural networks

Neural networks (ANN) can be used to generate predictive models of quantitative structure–activity relationships (QSAR) between a set of molecular descriptors obtained from the MLR and observed activity. The correlation of the observed activities with the ANN calculated ones are illustrated in Figure 3. The correlation coefficient **r = 0.96** and Standard Error of Estimate **s = 0.138**, obtained with the Neural network, show that the selected descriptors by LMR are pertinent and that the model proposed to predict activity is relevant.

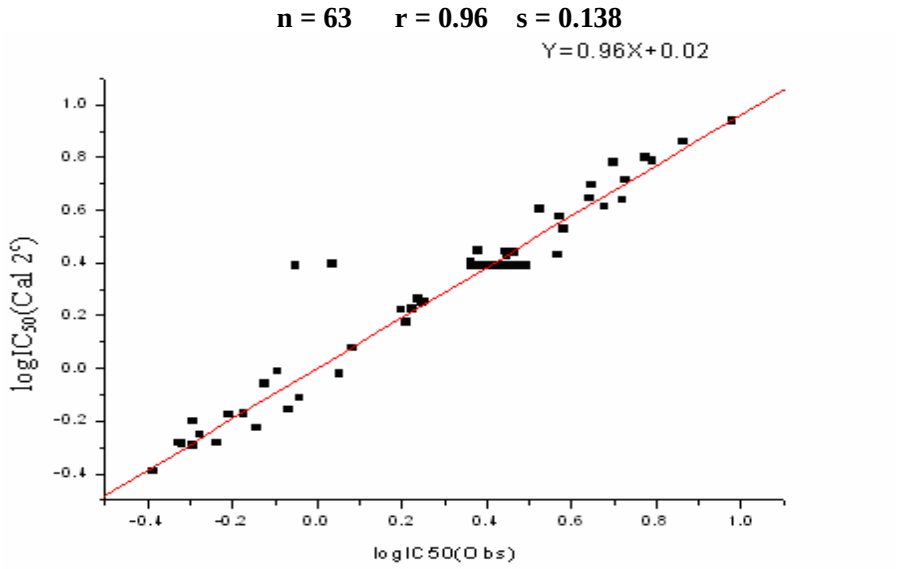


Figure. 3. Predicted antiamoebic activities by (ANN) correlated to experimental values

4-3 Validation

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

n = 63 **r_{cv} = 0.9** **s = 0.130**

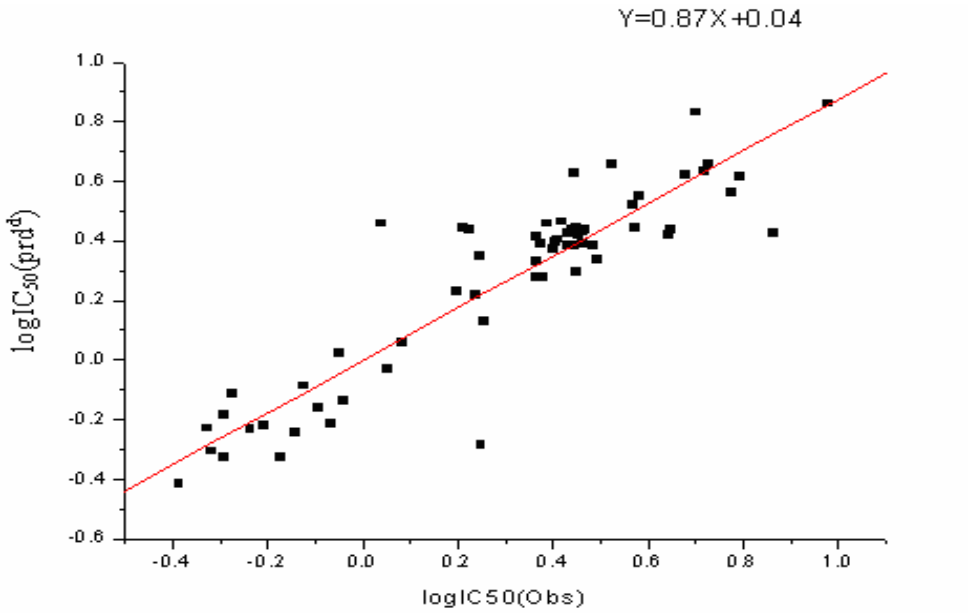


Figure. 4. Predicted antiamoebic activities by (CV) in comparison with experimental values

A good correlation was obtained with cross validation r_{cv}=0.9. 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 (rcv = 0.9).

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.

5. Conclusion

In this study, we investigated the best linear QSAR regression equations established in this study. Based on this result, a comparison of the quality of de MLR and ANN models shows that the ANN models have substantially better predictive capability because the ANN approach gives better results than MLR. ANN was able to establish a satisfactory relationship between the molecular descriptors and the antiamoebic activity.

References

1. Samuel, L., Stanley, J., Sharon, L. and Reed, Initial. Microbes and microbial toxin: paradigms for microbial-mucosal. Interactions VI. Entamoeba histolytica : parasite- host interactions. *Am. J. Phy. Gastrointest. Liver Physiol*. 2806 (2001) 1049.
2. WHO. Life in the 21st Century: a vision for all. The World Health Report 1998. World Health Organization, Geneva, Switzerland. (1998).
3. Haque, R., Mondal, D., Duggal, P., Kabir,M., Roy, S., Farr, B., Sack, B., William, A., Petri, Jr., Entamoeba histolytica infection in children and protection from subsequent amebiasis. Infection & Immunity. *American Society for Microbiology*. 74 (2006) 904.
4. Bansal, D., Sehgal, R., Chawla, Y., Chander Mahajan, R., Malla, N., In vitro activity of antiamoebic drugs against clinical isolates of Entamoeba histolytica and Entamoeba dispar. Annals of Clinical Microbiology and Antimicrobials. *Annals of Clinical Microbiology and Antimicrobials*. 3 (2004) 27.
5. Bhat, AR., Athar, F., Azam, A., New derivatives of 3,5-substituted-1,4,2-dioxazoles: Synthesis and activity against Entamoeba histolytica. Eur J Med Chem.. 44 (2009) 926.
6. Abid, M., Bhat, AR., Athar, F., Azam, A., Synthesis, spectral studies and antiamoebic activity of new 1-N-substituted thiocarbamoyl-3-phenyl-2-pyrazolines. *Eur J Med Chem.* 31 (2007) 426.
7. Adhikari, N., Maiti, MK., tarun, J., Exploring structural requirements of 1-N-substituted thiocarbamoyl-3-phenyl-2-pyrazolines as antiamoebic agents using comparative QSAR modelling. *Bioorganic& medicinal chemistry letters*. 20 (2010) 4021.
8. Trinajstic, N., *Chemical Graph Theory*. (1992) 225-273.
9. a) Bazoui, H., Zahouily, M., Boulaajaj, S., Sebti, S., Zakarya, D., SAR QSAR Environ. Res. 13 (2002) 567. b) Bazoui, H., Zahouily, M., Sebti, S., Boulaajaj, S., Zakarya, D.J., Mol. Model. 8(2002)1-7, c) Agrawala, V. K., Singha, J., Mishraa, K. C., Khadikar, P. V., Jaliwalac Y. A. Arkivoc (2006) 162.
10. Kubinyi, H., QSAR:Hansch analysis and related approaches. In: Mannhold R, Krogsgarrd Larsen P, Timmerman H (eds) Methods and principles in medicinal chemistry. Wiley, Weinheim; 1 (1993).
11. Kubinyi, H., (ed) 3D QSAR in drug design: theory, methods and applications. ESCOM, Leiden (1993).
12. Todeschini, R., Consonni, V., Handbook of Molecular Descriptors; Wiley-VCH: (2000) Germany.
13. Karelson, M., Molecular Descriptors in QSAR/QSPR; John Wiley & Sons: New York, (2000).
14. Diudea, M. V., Ed. QSPR/QSAR Studies by Molecular Descriptors; Nova Science: Huntington, NY, (2001).
15. Balaban, A. T., Ed. From Chemical Graphs to Three-Dimensional Geometry; (1997). New York.
16. Balaban, A. *SAR QSAR Environ. Res.* 8 (1998) 1.
17. So, S., Richards, G. J., *Med. Chem.* 35 (1992) 3207.
18. Andrea, T.A., Kalayeh, H.J., *Med. Chem.* 34 (1991) 2824.
19. Ellhalaoui, M., Modélisatrice moléculaire et étude QSAR d'antagonistes non compétitifs du récepteur NMDA par les méthodes statistiques et le réseau de neurones. Thesis de Doctorat, Fes, Morocco (2002) 106.
20. Efron, B. *J. Am. Stat. Assoc.* 78 (1983) 316.
21. Efroymson, M.A. Multiple regression analysis. In Mathematical Methods for Digital Computers; Ralston, A.;Wilf, H.S., Eds.;Wiley: NewYork (1960).
22. Osten, D.W. *J. Chemom.* 2 (1998) 39.