Journal of Computational Methods in Molecular Design, 2014, 4 (4):121-130

Scholars Research Library
(http://scholarsresearchlibrary.com/archive.html)

ISSN : 2231- 3176
CODEN (USA): JCMMDA

Predicting biological activity of chalcone (1,3-diphenyl-2-propen-1-one) derivatives cytotoxicity against HT-29 human colon adenocarcinoma cell linesby by DFT-QSAR models

Majdouline Larif[1*], Samir Chtita[2], Azeddine Adad[2], Rachid Hmamouchi[2], Mohammed Bouachrine[3] and Tahar Lakhlifi[2]

[1]Laboratory of Separation Process Laboratory, Faculty of Science, University IbnTofail, Kenitra, Morocco
[2]Laboratory of Molecular Chemistry and Natural Substances Laboratory, Faculty of Science, University Moulay Ismail, Meknes, Morocco
[3]ESTM, University Moulay Ismail, Meknes, Morocco

ABSTRACT

Chalcones are 1,3-diphenyl-2-propene-1-one, in which two aromatic rings are linked by a three carbon α,β-unsaturated carbonyl system. These are abundant in edible plants and are considered to be precursors of flavonoids and isoflavonoids. Our objective is to study the relationship between the activities and structure, a 3D-QSAR study is applied to a set of 20 molecules for biological activity prediction derivatives. This study was conducted using the principal component analysis PCA method; the multiple linear regression method MLR and the artificial neural network ANN;The leave-one out cross-validation procedure was used to validate the ANN model for use it to predict the activity of others new compounds. The relevant descriptors obtained from the ANN showed a correlation coefficient of 0.949 models which is a good result. As a result of quantitative structure–activity relationships, we found that the model proposed in this study is constituted of major descriptors used to describe these molecules. The obtained results suggested that the proposed combination of several calculated parameters could be useful to predict the biological activity of derivatives of 1,3-diphenyl-2-propene-1-one.

Keywords: *Biological activity; 3D-QSAR; PCA; MLR; ANN; DFT study.*

INTRODUCTION

Chalcones are 1,3-diphenyl-2-propene-1-one, in which two aromatic rings are linked by a three carbon α, β-unsaturated carbonyl system as,They exhibit diver's biological activities, such as antimicrobial, anticancer, antiprotozoal, antiulcer, anti-inflammatory, among others[1], and thus comprise a class with important therapeutic potential. Benzodiazepines derivatives have long been known for their wide range of therapeutic and pharmacological properties. Nowadays, many members of diazepine family are widely used as anticonvulsant, analgesic, sedative, antidepressive, and hypnotic agents[2]. Benzodiazepine derivatives also find commercial use as dyes for acrylic fibers [3] and [4] and some 2,4-diaryl-7,8-dimethyl-2,3-dihydro-1H-1,5-benzodiazepines have been tested against breast cancer and have shown moderate activity[5].
Cancer is a major public health problem and leading cause of death in many parts of the world. Deaths from cancer worldwide are projected to continue rising, with an estimated 13.1 million deaths in 2030 [6].

Quantitative structure-activity relationship (QSAR), as an important area of chemometrics, has been the subject of a series of investigations [7].

The main aim of QSAR studies is to establish an empirical rule or function relating the structural descriptors of compounds under investigation to bioactivities. This rule or function is then utilized to predict the same bioactivities of the compounds not involved in the training set from their structural descriptors. Whether the bioactivities can be predicted with satisfactory accuracy depends to a great extent on the performance of the applied multivariate data analysis method, provided the property being predicted is related to the descriptors. Many multivariate data analysis methods such as principal components analysis (PCA) and artificial neural network (ANN) have been used in QSAR studies. ANN offers satisfactory accuracy in most cases but tends to over fit the training data. There are a large number of molecular descriptors that can be used in QSAR studies. Once validated, the findings can be used to predict activities of untested compounds. Recently, computer-assisted drug design based on QSAR has been successfully employed to develop new drugs for the treatment of cancer and other diseases [8].

After a QSAR model is built and validated, it can predict the biological activity of novel molecules room their structural properties. A QSAR model can also screen potentially active molecules from a database, as described in the section on applications of the technique. Because the QSAR model can incorporate a wide range of different variables, be it physical, chemical or biological, it can also be utilized in industries apart from drug design [9], such as toxicology [10], food chemistry [11] and other fields.

Within the currently ongoing search for effective anticancer drugs candidates in the present study we have carried out and established a reliable quantitative structure–activity relationship (QSAR) analysis based on 20 chalcones derivatives.

MATERIALS AND METHODS

Experimental data

In our QSAR study, a total of 20chalcones molecules were gathered from the literature [12,13], The log IC_{50} value, concentration of the compound (μM) exhibiting 50% inhibition of cell growth [14], for human colorectal cancer cell line, HT-29, was employed as the dependent variable (Table 1).

All the compoundswere evaluated for their cytotoxic activity by the 3-(4,5 dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay based on mitochondrial reduction of yellow MTT tetrazolium dye to a highly colored blue formazan product [15].The following figure 1 represents the basic structure chalcones.

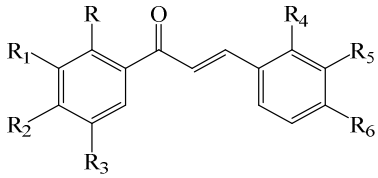

Figure 1:Basic structure ofchalcones

Table 1: selected training set of chalcone derivatives and their respective pI_{50}

Compound	R	R1	R2	R3	R4	R5	R6	pI_{50}
1	H	H	NH_2	H	Cl	H	Cl	1,602
2	H	H	OCH_3	H	Cl	H	Cl	1,276
3	OH	H	H	H	Cl	H	Cl	1,627
4	H	H	NH_2	H	Cl	H	H	1,469
5	H	H	H	H	H	H	OCH_3	1,384
6	H	H	NH_2	H	H	H	OCH_3	1,746
7	H	NH_2	H	H	H	H	OCH_3	1,631
8	H	H	OCH_3	H	H	H	OCH_3	1,691
9	H	H	H	H	H	H	OCH_3	1,775
10	H	H	H	F	H	H	OCH_3	2,005
11	H	H	NH_2	H	F	H	H	1,68
12	H	H	OCH_3	H	F	H	H	1,276
13	H	H	H	H	H	H	OH	1,391
14	H	H	OCH_3	H	H	H	OH	1,269
15	H	H	H	H	OCH_3	H	OCH_3	2,116
16	H	H	F	F	H	OCH_3	H	1,726
17	H	H	NH_2	H	OCH_3	OCH_3	H	1,843
18	H	NH_2	H	H	OCH_3	OCH_3	H	1,641
19	H	H	OCH_3	H	OCH_3	OCH_3	H	1,621
20	H	H	H	H	CF_3	H	H	1,632

Computational methods

DFT (density functional theory) methods were used in this study. These methods have become very popular in recent years because they can reach similar precision to other methods in less time and less cost from the computational point of view. In agreement with the DFT results, energy of the fundamental state of a polyelectronic system can be expressed through the total electronic density, and in fact, the use of electronic density instead of wave function for calculating the energy constitutes the fundamental base of DFT [16-19], using the B3LYP functional [19,20] and a 6-31G* basis set. The B3LYP, a version of DFT method, uses Becke's three-parameter functional (B3) and includes a mixture of HF with DFT exchange terms associated with the gradient corrected correlation functional of Lee, Yang and Parr (LYP). The geometry of all species under investigation was determined by optimizing all geometrical variables without any symmetry constraints.

Calculation of molecular descriptors

Calculation of molecular descriptors using Gaussian 03W

From the results of the DFT calculations, the quantum chemical descriptors were obtained for the model building as follows: the total energy (**Et** (u.a.)), the highest occupied molecular orbital energy **(E_{HOMO}** (eV)), the lowest unoccupied molecular orbital energy **(E_{LUMO}** (eV)), the energy difference between the LUMO and the HOMO energy (**Gap** (eV)), absorption maximum **λ_{max}**, the total dipole moment of the molecule (**μ** (Debye)), absolute hardness **η**), absolute electron negativity **(χ)** and reactivity index (**ω**) [21]. (**η**),(**χ**) and **(ω**) were determined by the following equations:

$$\eta = \frac{(E_{LUMO} - E_{HOMO})}{2} \quad ; \quad \chi = -\frac{(E_{LUMO} + E_{HOMO})}{2} \quad ; \quad \omega = \frac{\chi^2}{2\eta}$$

Principal components analysis

Twenty molecules were studied by statistical methods based on the principal component analysis (PCA) [22,23] using the software XLSTAT 2009.

Essentially a descriptive statistical method which aims to present, in graphic form, the maximum of information contained in the data (Table 1).

PCA is a statistical technique useful for summarizing all the information encoded in the structures of compounds. It is also very helpful for understanding the distribution of the compounds.

Multiple linear and nonlinear regressions (MLR and MNLR)

The multiple linear and nonlinear regression statistics techniques are used to study therelation between one dependent variable and several independent variables. The multiple linear and non linear regression models (MLR and MNLR) are generated using the software XLSTAT, version 2009, to predict pIC_{50}.

The optimal number of components (N) is employed to do validation MLR and MNLR analysis to get the final model parameters such as correlation coefficient R^2, standard deviation (S) and Fischer test value (F) [24].

Artificial neural networks (ANN)

The ANN analysis was performed with the use of Matlab software version 7.0 using a program written in C language Neural toolbox on a data set of structures of 20 chalcones derivatives [25,26]. A number of individual models of ANN were designed built up and trained. Generally the network was built for three layers; one input layer, one hidden layer and one output layer were considered [27]. The input layer was consisted of eight artificial neurons of linear activation function (Figure 2). The number of artificial neural in the hidden layer was adjusted experimentally. The hidden layer consisted of 20 artificial neural. One neuron formed the output layer of sigmoid function activation. The architecture of the applied ANN models is presented in (Figure 3).

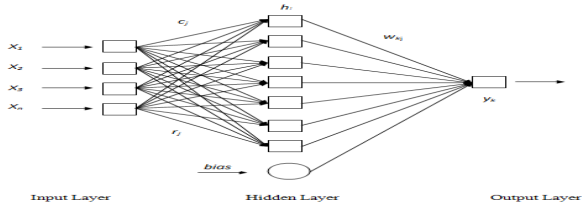

Figure 2: Neuron Layout of ANN

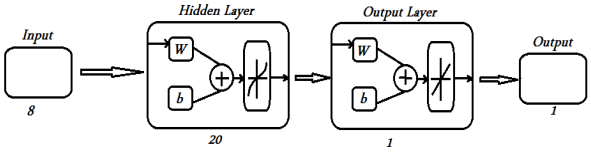

Figure 3: The ANN architecture.

The data subjected to ANN analysis were randomly divided into three sets: a learning set, a validation set and a testing set. Prior to that, the whole data set was scaled within the 0-1 range.

The set of structures of 20 chalcones derivatives [28] was subjected to the ANN analysis. First, for the learning set of compounds, i.e., selected training set of 20 chalcone derivatives. The learning set of data is used in ANN to recognize the relationship between the input and output data. Then for the revision of the ANN model designed and selected, the validation set of 20 compounds was used. Testing set with eight compounds was provided to be an independent evaluation of the ANN model performance for the finally applied network. In this study, we selected the sigmoid as a basis function [29].

The operation of the output layer is linear, which is given as below:

$$y_k(X) = \sum_{j=1}^{n_k} w_{kj} h_j(X) + b_k$$

Where y_k is the k^{th} output layer unit for the input vector X, w_{kj} is the weight connection between the k^{th} output unit and the j^{th} hidden layer unit and b_k is the bias that allows a transfer function "non-zero" given by the following equation:

$$\text{Bias} = \sum (\bar{y} - y)$$

Where y is the measured value and$\bar{y}$is the value predicted by the model.

The accuracy of the model was mainly evaluated by the root mean square error (RMSE). Formula is given as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} . \sum_{i=1}^{n} (p_{exp} - p_{pred})^2}$$

Where n = number of compounds, p_{exp} = experimental value, p_{pred} = predicted value and summation is of overall patterns in the analyzed data set [30,31]. The scripts were run on a personal PC.

RESULTS

Data for analysis

A QSAR study was carried for a series of 20 chalcone derivatives, in order to determine a quantitative relationship between structure and toxicity.

Table 2 shows the values of the calculated parameters obtained by DFT/B3LYP 6-31G* optimization of the studied Selected training set of 20 chalcones derivatives.

Table 2: Values of the twelve chemical descriptors

Molec.	pI_{50}	Et	E_{HOMO}	E_{HUMO}	Gap	μ	χ	η	ω	Ea	λ_{max}	$f_{(SO)}$
1	1,602	-44346	-5,806	-1,978	3,8280	3,8349	3,8917	1,9140	3,9565	4,2510	291,66	0,5404
2	1,276	-45422	-3,924	0,087	4,0107	2,3570	1,9182	2,0054	0,9174	3,7181	333,46	0,0606
3	1,627	-44887	-6,332	-2,185	4,1469	3,3212	4,2589	2,0734	4,374	4,0957	302,72	0,5792
4	1,469	-33879	-5,626	-1,763	3,8626	4,6616	3,6947	1,9313	3,5341	3,4537	358,99	0,0553
5	1,384	-20928	-5,856	-1,785	4,0709	4,2239	3,8206	2,0354	3,5858	3,8474	322,25	0,6661
6	1,746	-22435	-5,612	-1,527	4,0848	3,8558	3,5697	2,0424	3,1196	4,026	307,96	0,4213
7	1,631	-22435	-5,560	-1,678	3,8825	2,5075	3,6191	1,9412	3,3737	3,8781	319,71	0,6879
8	1,691	-24046	-5,770	-1,646	4,1240	2,9293	3,7078	2,0620	3,3336	3,7892	327,20	0,6445
9	1,775	-22976	-5,802	-1,657	4,1450	3,8229	3,7300	2,0725	3,3565	4,0080	309,34	0,6724
10	2,005	-26332	-6,009	-2,041	3,9680	5,2198	4,0251	1,9840	4,0831	3,7441	331,14	0,6124
11	1,680	-22019	-5,709	-1,716	3,9933	5,2873	3,7127	1,9966	3,4518	4,3396	285,70	0,4666
12	1,276	-23630	-6,109	-1,840	4,2697	4,3550	3,9745	2,1348	3,6997	4,3234	286,78	0,3799
13	1,391	-19858	-5,939	-1,816	4,1234	3,6236	3,8777	2,0617	3,6466	3,9149	316,70	0,6220
14	1,269	-22976	-5,849	-1,674	4,1749	2,445	3,7618	2,0875	3,3896	3,8213	324,46	0,5374
15	2,116	-24046	-5,595	-1,545	4,0502	5,4096	3,5704	2,0251	3,1474	3,7831	327,73	0,5717
16	1,726	-26332	-6,202	-2,135	4,0676	2,4557	4,1686	2,0338	4,2722	4,1078	301,82	0,3407
17	1,843	-25553	-5,651	-1,589	4,0619	4,1110	3,6198	2,0309	3,2259	3,6408	340,54	0,1999
18	1,641	-25553	-5,583	-1,750	3,8329	3,5899	3,6668	1,9164	3,5079	4,0830	303,66	0,1365
19	1,621	-27164	-6,037	-1,707	4,3301	2,8592	3,8724	2,1651	3,4630	3,8771	319,79	0,2886
20	1,632	-26987	-6,574	-2,086	4,4875	5,6475	4,3298	2,2438	4,1777	4,2796	289,71	0,4016

Correlation matrix

Table 2 shows the correlation matrix (Pearson (n)) therefor obtained between different descriptors.

Table 2: Correlation matrix (Pearson (n)) between different obtained descriptors

	Et	E_{HOMO}	E_{LUMO}	Gap	μ	χ	η	ω	E_a	λ_{max}	$f_{(SO)}$	pI_{50}
Et	**1**											
E_{HOMO}	0,206	**1**										
E_{LUMO}	**0,576**	**0,768**	**1**									
Gap	0,331	**-0,688**	-0,064	**1**								
μ	0,031	-0,074	-0,043	0,066	**1**							
χ	-0,384	**-0,958**	**-0,919**	**0,451**	0,065	**1**						
η	0,331	**-0,688**	-0,064	**1,000**	0,066	**0,451**	**1**					
ω	**-0,575**	**-0,781**	**-1,000**	0,084	0,040	**0,927**	0,084	**1**				
E_a	-0,235	-0,408	-0,426	0,153	0,091	0,441	0,153	0,427	**1**			
λ_{max}	0,213	0,408	0,416	-0,165	-0,055	-0,437	-0,165	-0,418	**-0,997**	**1**		
$f_{(SO)}$	0,080	-0,070	-0,053	0,049	-0,101	0,067	0,049	0,057	0,144	-0,183	**1**	
pI_{50}	0,159	0,153	0,165	-0,051	0,323	-0,168	-0,051	-0,156	-0,248	0,233	0,034	**1**

Bold values are different from 0 at a level significant for p < 0.05
At a very significant for p < 0,01; At a highly significant to p < 0,001

Correlation circle

Principal component analysis (PCA) was also performed to detect the connection between the different variables. The principal component analysis revealed from the correlation circle (Figure 4) shows that the F1 axis (40.46%) presents the energy of the variance while the axis F2 (20.11%) of the variance is located by the other parameters of energy [21].

χ and $\mathbf{E_{HOMO}}$ are perfectly correlated (r = 0,958), both variables are redundant.
$\boldsymbol{\lambda_{max}}$ and $\mathbf{E_a}$ are strongly negatively correlated (r = -0,997).

$\mathbf{E_{LUMO}}$ and **ω** are strongly negatively correlated (r = -1).

η and **Gap**are perfectly correlated (r = 1).

The following variables then removed are:**Gap, λ_{max},χ**and **ω.**

On the other hand, the correlation circle (Figure 4) indicates the correlation between electronic descriptors.

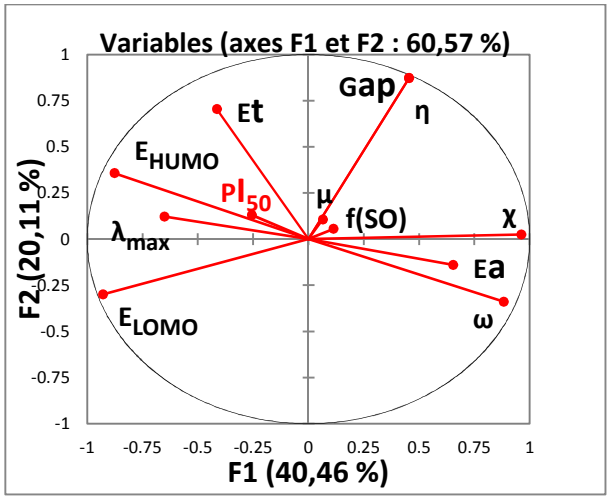

Figure 4:Circle correlation between descriptors

Cartesian diagram

Projection on the plan 1×2

The Cartesian diagram analyses of projections according to the plane F1–F2 (60.57%) of the total variance of the studied molecules.

The Cartesian diagram does not provide information on the distribution of molecules because the molecular structure of chalcone derivatives used in this study is very similar, which does not distinguish specific groups of molecules.

Multiple linear regressions (MLR)

In order to propose a mathematical model and to evaluate quantitatively the substituent's physicochemical effects on the activity $\mathbf{pI_{50}}$ of the totality of the set of these 20 molecules, we submitted the data matrix constituted obviously from the 11 physicochemical variables corresponding to the 20 molecules, to a progressive multiple regression analysis. This method used the coefficients R, R^2, and the F-values to select the best regression performance. Where R is the correlation coefficient; R^2 is the coefficient of determination; **MSE** is the mean squared error; F is the Fisher F-statistic. Treatment with multiple linear regressions is more accurate because it allows you to connect the structural descriptors for each activity of 20 molecules to quantitatively evaluate the effect of substituent.

$$\mathbf{pI_{50}} = 41{,}05 + 1{,}04\ 10^{-5}\ \mathbf{Et} + 1{,}02\ \mathbf{E_{HOMO}} + 28{,}66\ \mathbf{E_{LUMO}} + 0{,}15\ \boldsymbol{\mu} + 15{,}66\ \boldsymbol{\omega} - 5{,}01\ \mathbf{E_a} - 0{,}06\ \boldsymbol{\lambda_{max}} - 0{,}21\ \mathbf{f_{(SO)}} \quad \text{(Equation 1)}$$

N = 20 R^2 = 0, 641 R=0,801 RMCE = 0,180

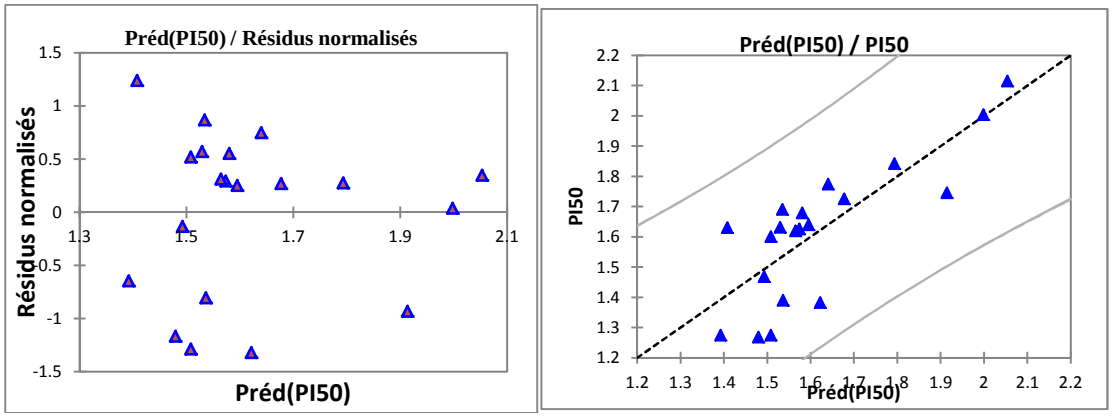

Figure 5: Relationship between the estimated values of pI_{50}, their predictions and their residues established by MLR

As a remark (Table 3), the model the values are different from 0 at a significant level p <0.05 for Pr<0,001 with F(8,11) = 2,40. The figure 5 shows a very regular distribution of toxicity values depending on the experimental values [21].

Table 3: Analyses of variance

Source	DDL	Sum of squares	Mean square	F	Pr > F
Model	8	0,639	0,080	2,460	< 0,0001
Error	11	0,357	0,032	-	-
Total corrected	19	0,996	-	-	-

Multiple non linearregression (MNLR)

We have used also the technique of nonlinear regression model to improve the structure-activity relationship to quantitatively evaluate the effect of substituent. It takes into account several parameters. This is the most common tool for the study of multidimensional data. We have applied to the data matrix constituted obviously from the descriptors proposed by MLR corresponding to the 20 molecules. The coefficients R, R^2, and the F-values are used to select the best regression performance.We used a pre-programmed function of XLSTAT following:

$$\mathbf{Y = a + (b\ X1 + c\ X2 + d\ X3 + e\ X4 \ldots) + (f\ X12 + g\ X22 + h\ X32 + i\ X42\ldots)}$$

Where a, b, c, d,..: represent the parameters and X1, X2, X3, X4,...: represent the variables.
The resulting equation was:

$$\mathbf{pI_{50}} = 3772{,}01 + 1{,}71\ 10^{-4}\ \mathbf{Et} + 47{,}20\ \mathbf{E_{HOMO}} - 227{,}77\ \mathbf{E_{LUMO}} + 7{,}03\ 10^{-3}\boldsymbol{\mu} - 96{,}60\ \boldsymbol{\omega} - 715{,}83\ \mathbf{E_a} - 6{,}55\ \boldsymbol{\lambda_{max}} - 6{,}55\ \mathbf{f_{(SO)}} + 2{,}0310^{-9}\mathbf{Et^2} + 3{,}57\ \mathbf{E_{HOMO}^2} - 116{,}17\ \mathbf{E_{LUMO}^2} + 0{,}019\ \boldsymbol{\mu}^2 + 27{,}55\ \boldsymbol{\omega}^2 + 51{,}62\ \mathbf{E_a^2} + 4{,}2710^{-3}\boldsymbol{\lambda_{max}^2} + 7{,}52\ \mathbf{f_{(SO)}^2}$$ (Equation 2)

$$\mathbf{N = 20 \quad R^2 = 0{,}901 \quad R = 0{,}941 \quad RMCE = 0{,}182}$$

With MLNR was obtained significantly better correlation coefficient R = 0,912 (Figure 6) shows a very uniform distribution of the toxicity observed values depending on the experimental values and the correlation between the experimental results and calculated alter them pI_{50}. The residual values tended to zero which is why we did not graph for prediction residuals [21,32].

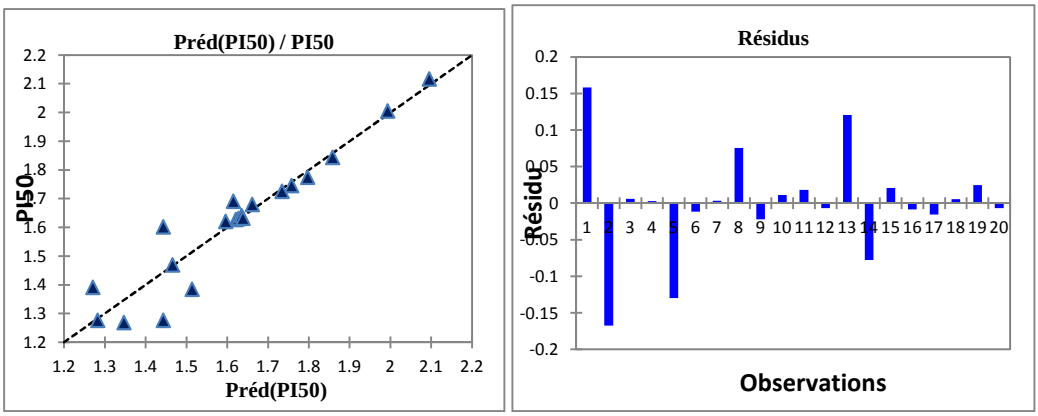

Figure 6: Relationship between the estimated values of pI_{50} and their Predictions established by MNLR

Validation of model - Cross Validation

Artificial neural networks (ANN) are used to generate predictive models of quantitative structure-activity relationships (QSAR) between the 8 molecular descriptors obtained from the MLR and observed activities (input). The ANN calculated activities model were developed using the properties of several studied compounds. For determination of the number of hidden neuronsD. Cherqaoui and D. Villemin [33]have discussed the usefulness of ρ parameter defined as:

ρ = Number of data points in the training set /Sum of the number of connections in the ANN.

The range of 1< ρ<2.2 has been suggested as an empirical guideline of acceptable ρ values. It has been claimed that for ρ<<1.0 the ANN simply memorizes the data. While for ρ>>3.0, the ANN is not able to generalize[34,27].

So we used a number of neurons in the hidden layer allowing to maintain ρ in the 1<ρ<3 range, to avoid these two problems. Two different architectures have been applied (8-2-1; 8-1-1). The two architectureswere trained with the number of cycles limited to 500. The criterion used for the comparison of the two architectures is the correlation coefficient (R).

The architecture (8-2-1) gives better results than the (8-1-1). The output layer represents the calculated activity values IC_{50} calculated using ANN (Table 5).The statistical measures obtained were:

N=20 R^2 = 0,949 R=0, 901 (Architecture 8-2-1)

We see that the ANN approach gives better results than the MLR and MNLR (table 4). This preliminary study allows concluding that the ANN was able to establish a satisfactory relationship between the electronic descriptors and the activity against HT-29 human colon adenocarcinoma cell lines.

Table 4 shows the comparison between the different methods of the correlation coefficient R.

Table 4: Statistical results of comparative all models based on the N = 20 compounds
Statistical result

Statistical result	MLR	MNLR	ANN
R^2	0,641	0,901	0,949
MCE	0,032	0,099	-
RMSE	0,180	0,182	-

R^2: Determination coefficient; S: Standard error of estimated; F: Fischer test value; RMSE: Root mean square error.

Before using a QSAR model to predict the activity of new compounds, we should validate it using a validation method. In this work we validated our model ANN with cross validation using Leave one out LOO procedure (using MATLAB software).

The cross-validation analysis was performed with the leave-one-out (LOO) procedure in which one compound is removed from the dataset and its activity is predicted using the model derived from the rest of the dataset.

The values of predicted activities calculated using Cross Validation is given in Table 5.The statistical measures obtained were:

N = 20 R_{cv}^2 = 0,611 R_{cv}=0, 782

DISCUSSION

On the projection plane 1x2principal component analyses informs us from the circle of correlation descriptors that speak strongly correlated from figure 4.

The Cartesian diagram does not provide information on the distribution of molecules because the molecular structure of chalcone derivatives used in this study is very similar, which does not distinguish specific groups of molecules.

The RNLM takes into account all the descriptors with higher compared with the linear regression RLM.

The cross-validated R^2 value that resulted in optimum number of components N was taken. The robustness of the model was internally evaluated by calculating the R2 from the training set [35]. This study allows concluding that our model was able to establish a satisfactory relationship between the electronic descriptors and the activity studded, and we can used to the activity of new compounds.

Table 5 shows the comparison of observed values with the calculated values of the MLR, MNLR, ANN and CV.

Table 5: Observed values and calculated of pI_{50} according to different methods.

		RLM		RNLM		ANN		CV	
Mol.	**pI_{50} (Obs.)**	**pI_{50} (pred.)**	**Residu**	**pI_{50} (pred.)**	**Residu**	**pI_{50} (pred.)**	**Residu**	**pI_{50} (Pred.)**	**Residu**
1	1,602	1,508	0,094	1,444	0,158	1,439	-0,899	1,577	-0,025
2	1,276	1,508	-0,232	1,444	-0,168	1,439	-0,899	1,602	0,326
3	1,627	1,574	0,053	1,621	0,006	1,610	-1,031	1,439	-0,188
4	1,469	1,493	-0,024	1,466	0,003	1,496	-1,441	1,483	0,014
5	1,384	1,622	-0,238	1,514	-0,130	1,391	-0,725	1,312	-0,072
6	1,746	1,914	-0,168	1,758	-0,012	1,733	-1,312	1,771	0,025
7	1,631	1,408	0,223	1,628	0,003	1,619	-0,931	1,620	-0,012
8	1,691	1,534	0,157	1,616	0,075	1,781	-1,136	1,616	-0,075
9	1,775	1,640	0,135	1,797	-0,022	1,850	-1,178	1,692	-0,083
10	2,005	1,998	0,007	1,994	0,011	2,020	-1,408	2,112	0,107
11	1,680	1,580	0,100	1,662	0,018	1,670	-1,203	1,648	-0,032
12	1,276	1,392	-0,116	1,283	-0,007	1,363	-0,983	1,521	0,245
13	1,391	1,536	-0,145	1,271	0,120	1,339	-0,717	1,362	-0,029
14	1,269	1,479	-0,210	1,347	-0,078	1,224	-0,687	1,567	0,298
15	2,116	2,053	0,063	2,095	0,021	2,013	-1,441	1,920	-0,196
16	1,726	1,677	0,049	1,735	-0,009	1,720	-1,379	1,703	-0,024
17	1,843	1,793	0,050	1,859	-0,016	1,764	-1,564	1,721	-0,122
18	1,641	1,595	0,046	1,635	0,006	1,647	-1,510	1,641	0,000
19	1,621	1,565	0,056	1,596	0,025	1,643	-1,354	1,648	0,027
20	1,632	1,529	0,103	1,639	-0,007	1,616	-1,214	1,530	-0,102

CONCLUSION

In this work, we studied the QSAR regression to predict the toxicity of a series of 20 compounds as R or R^2 different models obtained using different statistical tools and various descriptors was shown in table 5. The study of the quality of ANN, MLR and MNLR models showed that the ANN result has substantially better predictive capability than the other methods. With the ANN approach we have established a relationship between 8 electronic descriptors and toxicity in satisfactory manners, and the resulted models were validated with success by leave-one out cross-validation procedure to check their predictability and robustness.

This study allows concluding that our model was able to establish a satisfactory relationship between the electronic descriptors and the activity against HT-29 human colon adenocarcinoma cell lines, and we can used to the activity of new compounds.

Finally, we can conclude that one of the studied descriptors, which is sufficiently rich in chemical and electronic information to encode the structural features, may be used with other topological descriptors for the development of predictive QSAR models.

Acknowledgment
We are grateful to the “Association Marocaine des ChimistesThéoriciens” (AMCT) for its pertinent help concerning the programs.

REFERENCES

[1] RSaxena, A S Srivastava, S Chandra,*Bioorg. Med. Chem. Lett.*, **2000**, 10, 2159.
[2]S Suvarna; K Krishna;S Kumar; K Hatti;D Latha; GChandrasekara Reddy,*Eur. J. Med. Chem.***, 2013**, 62, 435-442.
[3] K Chandrabose; V S Raja;L Hoyun; PTrivedi, *Biomed. &Preve.Nutr.*, **2013,** 3,325-330.
[4] A ANabila; S AAmal; M S Makarem, *J. of Saud. Chem. Soc.*, **2012**,16, 45-53.
[5] P Singh; AAnand; V Kumar;*European J. of Med. Chem.*,**2014**, 85,758-777.
[6] X Wu; HZeng; X Zhu; Q Ma; Y Hou; X Wu,*Eur.J. Pharm. Sci.*,**2013**, 50, 323-334.
[7] E X Esposito; A J Hopfinger; J D Madura,*Methods Mol. Biol*.,**2004**, 275, 131-214.
[8] R Perkins; H Fang; W Tong; W Welsh, *Environ. Toxicol. Chem*., **2003**, 22, 1666-1679.
[9] Q S Du; R B Huang; K C Chou,*Curr. ProteinPept. Sci*.,**2008**, 9, 248-260.
[10] S P Bradbury, *Toxicol. Lett*.**1995**, 79, 229-237.
[11] K Martinez-Mayorga; J L Medina-Franco,*Adv. Food. Nutr. Res*., **2009**, 58, 33-56.
[12] S Syam; S IAbdelwahab; M A Al-Mamary; S Mohan, *Molecules*, **2012**, 17, 6179-6195.
[13] J Wu; C Wang; Y Cai; J Peng; D Liang; Y Zhao; S Yang; X Li; X Wu; G Liang, *Med. Chem. Res*.,**2012**, 21, 444-452.
[14] L Xie; X Zhai; L Ren; H Meng; C Liu; W Zhu; Y Zhao, *Chem. Pharm.Bull*.,**2011**, 59, 984-990.

[15] TMosmann,*J. Immunol. Methods*, **1983**, 65, 55-63.
[16.C Adamo; V Barone, *Chem. Phys. Lett.*, **2000**, 330, 152-160.
[17] M Parac; S Grimme, *J. Phys. Chem.*, **2003**, A106, 6844-6850.
[18] U Sakar; R Parthasarathi; V Subramanian and P K Chattaraji, *Internet Electronic J. Mol. Design*, **2004**, 1-24.
[19].A D Becke, *J. Chem. Phys.*, **1993**, 98, 1372.
[20] C Lee; W Yang; R G Parr, *Phys. Rev.*, **1988**, B37, 785-789.
[21] M Larif; S Chtita; A Adad; R Hmamouchi; M Bouachrine; T Lakhlifi,*IJARCSSE*, **2013**, 3 (4), 1-6.
[22] J N Hogarh; N Seike; Y Kobara; A Habib; J JNamd;J S Lee; L Qilu; X Liu; L Jun; G Zhang; S Masunaga,*Chemosphere*, **2012**, 86, 718-726.
[23] A M Taurino; D Dello Monaco; S Capone; M Epifani; R Rella; P Siciliano; L Ferrara; GMaglione; A Basso; D Balzarano,*Sensors and Actuators*, **2003**, B 95, 123-131.
[24] C Rücker; G J Rücker, *Chem. Inf. Model*.,**2007**, 47, 2345-2357.
[25] H Demuth; M Hugan;M Beal, Neural Network Toolbox. For use with MATHLAB, User Guid's, Version 9, **2011**.
[26] M Larif; A Adad; R Hmamouchi; A I Taghki; A Soulaymani; A Elmidaoui; M Bouachrine; T Lakhlifi, *Arab. J. Chem.*, **2013**, in press.
[27] G Hea; L Fenga; H Chena, *International Symposium on Safety Science and Engineering in China*, *Proc. Engin*, **2012**, 43, 204-209.
[28] R Chimizou; H Iwamura; T Fujita, *Agric. Food Chem.*,**1988**, 36, 1276.
[29] N Turkkan, *Revue de l'Université de Moncton*, **1993**, 26 (1), 205-221.
[30] P Y Lee,*J. Hazard. Mater*.,**2009**, 165, 156-161.
[31] G Jing; Z Zhou; J Zhuo, *Chemosphere*, **2012**, 86, 76-82.
[32] R Hmamouchi; I A Taghki;M Larif;A Adad; A Abdellaoui; M Bouachrine; T Lakhlifi, *JOCPR*,**2013**, 5(9), 198-209.
[33] D Cherquaoui; D Villemin, *J. Chem. Soc. Far. Trans.*, **1994**, 90, 97-102.
[34]T A Andrea; H Kalayeh, *J. Med. Chem*. **1991**, 34, 2824-2836.
[35] F Giraud; C Loge; M Le Borgne; FY M Pagniez, *SAR and QSAR in Environ.Res.*,**2006**, 17(3), 299-309.