J. Mater. Environ. Sci. 6 (1) (2015) 280-288
ISSN : 2028-2508
CODEN: JMESCN

Combining DFT and QSAR results for predicting the cytotoxicity of a series of orthoalkyl substituted 4-X-phenols

M. Ghamali[1], S. Chtita[1], A. Adad[1], R. Hmamouchi[1], M. Bouachrine[2], T. Lakhlifi[1*]

[1] *Molecular Chemistry and Natural Substances Laboratory, Faculty of Science, University Moulay Ismail, Meknes, Morocco*

[2] *ESTM, University Moulay Ismail, Meknes, Morocco*

Received 5 May 2014; Revised 7 November 2014; Accepted 8 November 2014.
** Corresponding Author: E-mail: tahar.lakhlifi@yahoo.fr ; Tel. +212 663853590*

Abstract
In this study we worked on a series of molecules based on phenol against leukemia cell line (L1210). The objective is to find a correlation between the structure and physicochemical properties. These were determined by combining DFT and QSAR results. This study was conducted using the principal component analysis (PCA) method, the multiple linear regression method (MLR), the non-linear regression (RNLM) and the artificial neural network (ANN). We accordingly propose a quantitative model, and we interpret the activity of the compounds relying on the multivariate statistical analysis. This study shows that the prediction results were in excellent agreement with the experimental value.

Keywords: phenols, Leukemia, 3D-QSAR model, DFT study.

1. Introduction
The phenolic hydroxy group has a wide range of cellular activities that have not been clearly investigated. At present there is intense interest in polyphenols which are present in the diet as part of fruits, tea, coffee and wine [1-3] since they have been shown to protect cells from oxidative stress [4]. In addition, these compounds show a wide spectrum of action involving antitumor, antiviral, antibacterial, cardioprotective, prooxidant and antimutagenic activity [1, 5-7]. Recently [8], the Hansch group has examined the cytotoxicity of a set of simple and complex mono-substituted phenols towards a fast-growing murine leukemia cell line (L1210). The objectives of this work are to develop predictive QSAR models for the cytotoxicity of a series of phenols against L1210 leukemia cell line.

Quantitative structure–activity relationship (QSAR) [9, 10] has been widely used for years to provide quantitative analysis of structure and biological activity relationships of compounds. Different QSAR studies were reported to identify important structural features responsible for the biological activity and to develop toxicity models for diverse chemicals by different workers [11, 12]. At present, 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, AIDS, SARS, and other diseases.

In this work, we have modeled the cytotoxicity of phenolic compounds based on ortho alkyl substituted 4-X-phenols (Figure 1) against L1210 leukemia cells using several statistical tools, principal components analysis (PCA), multiple linear regression (MLR), non-linear regression (RNLM) and artificial neural network (ANN) calculations. On the other hand, several quantum chemical methods and quantum-chemistry calculations have been performed in order to study the molecular structure, electronic and topologic properties [13, 14]. The more relevant molecular properties were calculated. These properties are the highest occupied molecular orbital energy E_{HOMO}, the lowest unoccupied molecular orbital energy E_{LUMO}, energy gap ΔE, dipole moment μ, the total energy E_T, the activation energy E_a, the absorption maximum λ_{max}, the Molar Volume (MV), the Molecular Weight (MW), the Molar Refractivity (MR), the Parachor (Pc), the Density (D), the Refractive Index (n), the Surface Tension (γ) and the Polarizability (α).

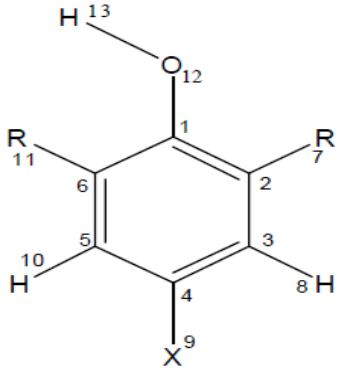

Figure 1: Chemical structure of the series of 2-alkyl and 2,6-dialkyl-4-X-phenols

2. Material and methods

2.1. Material

Previous studies [1] had established a quantitative model of structure activity relationship for a series of phenolic compounds.

Reference [8] provided all cytotoxicity data, in the shape of $-\log IC_{50}$ (pIC_{50}), where C constitutes the molar concentration of X-phenol that induces 50% inhibition of growth in the cell line.

The following table shows the chemical structures of the studied compounds and the corresponding experimental activities pIC_{50}. The experimental toxicity of the studied compounds has been collected from recent work [1] (Table 1). The range of the toxicity data varies from 3.02 to 4.9 (M).

Table1: Observed toxicity for the series of 2-alkyl and 2,6-dialkyl-4-X-phenols

N°	Substituent	pIC_{50}
1	2,6-di-Me	3.02
2	2,6-di-OMe	3.86
3	2,4,6-tri-Me	3.20
4	2,6-di-CMe_3	3.85
5	2,6-di-CMe_3-4-Me	4.04
6	2,6-di-Et	3.26
7	2,6-di-$CHMe_2$	3.25
8	2,4,6-tri-CMe_3	3.90
9	2-CMe_3-6-Me	3.73
10	2,6-di-CMe_3,4-NO_2	4.90
11	2,6-di-CMe_3-4-Et	3.91
12	2,6-di-CMe_3-4-Br	4.11
13	2,4-di-CMe_3	4.24
14	2-CMe_3-4-Me	3.80
15	2,4-di-Me	3.04
16	2-Me-4-F	3.09
17	2-Me-4-NO_2	3.49
18	2-Me-4-Br	3.46
19	2-Me-4-OMe	3.39
20	2-Me-4-COMe	3.14
21	2-CMe_3-4-Et	3.80

2.2. Methods

2.2.1 Principal Components Analysis (ACP)

The phenolic compounds (1 to 21) were studied by statistical methods based on the principal component analysis (PCA) [15] using the software XLSTAT version 2013. This is an essentially descriptive statistical method which aims present, in graphic form, the maximum of information contained in a 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.

2.2.2 Multiple linear regressions

The multiple linear regression statistic technique 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 XLSTAT, version 2013, to predict pIC_{50}. It has served also to select the descriptors used as the input parameters for a back propagation network (ANN).

2.2.3 Artificial Neural Networks (ANNs)

The ANNs analysis was performed with the use of Matlab software version 2009a Neural Fitting tool (nftool) toobox on a data set of 2-alkyl and 2,6-dialkyl-4-X-phenols [16].

A number of individual models of ANN were designed built up and trained. Generally the network was built for tree layers; one input layer, one hidden layer and one output layer were considered [17]. The input layer was consisted of fifteen artificial neurons of linear activation function. 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 2.

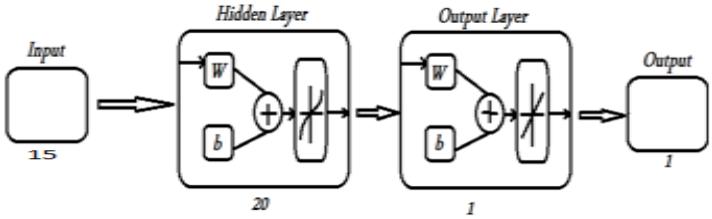

Figure 2: The ANNs architecture

The data subjected to ANN analysis was 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 to 1 range.

The set of phenolic compounds of the cytotoxicity were subjected to the ANN analysis. First, for the learning set of compounds, i.e., 21 ortho alkyl substituted 4-X-phenols were used. ANN models were designed, built and trained. The learning set of data is used in ANNs to recognize the relationship between the input and output data. Then for the revision of ANN model designed and selected, the validation set of three compounds was used. Testing set with three 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 [18]. 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 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 Root Mean Square Error (RMSE). Formula as follows:

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

Where n = number of compounds, p_{exp} = experimental value, p_{pred} = predicted value and summation is over all patterns in the analyzed data set [19,20]. The scripts were run on a personal PC.

2.2.4 DFT calculations

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 [21-23] using the B3LYP functional [24,25] 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.

3. Results and discussion

A QSAR study was carried for a series of 21 ortho alkyl substituted 4-X-phenols, in order to determine a quantitative relationship between structure and cytotoxicity.

Table 2 shows the values of the calculated parameters obtained by DFT/B3LYP 6-31G (d) calculations and by ACD/ ChemSketch program from the fully optimized structures of ortho alkyl substituted 4-X-phenols.

Table 2: the values of the fifteen chemical descriptors

	MW	MR (cm^3)	MV (cm^3)	Pc (cm^3)	n	γ (dyne/cm)	D (g/cm^3)	α (cm^3)	E_T (Ua)	E_{HOMO} (eV)	E_{LUMO} (eV)	ΔE (eV)	μ debye	Ea (eV)	λ_{max} (nm)
1	122.16	37.78	120.40	297.50	1.54	37.20	1.01	14.97	-10513.54	-5.72	0.29	6.01	1.49	5.27	235.33
2	154.16	41.49	135.80	335.60	1.52	37.20	1.13	16.44	-14609.04	-5.26	0.57	5.83	2.22	5.12	242.36
3	136.19	42.60	136.60	335.20	1.54	36.10	1.00	16.89	-11584.15	-5.53	0.31	5.84	1.43	5.16	240.45
4	206.32	64.90	221.20	518.30	1.50	30.10	0.93	25.73	-16936.12	-5.70	0.24	5.94	1.94	5.23	237.23
5	220.35	69.73	237.50	556.00	1.50	30.00	0.93	27.64	-18006.73	-5.52	0.25	5.78	1.87	5.12	242.41
6	150.22	47.23	153.40	375.40	1.53	35.80	1.00	18.72	-12654.45	-5.65	0.37	6.02	1.59	5.27	235.28
7	178.27	56.50	188.00	452.30	1.51	33.50	0.95	22.40	-14795.39	-5.74	0.22	5.96	1.81	5.23	236.98
8	262.43	83.29	287.90	666.40	1.49	28.60	0.91	33.02	-21218.15	-5.54	0.25	5.80	1.86	5.11	242.86
9	164.24	51.34	170.80	407.90	1.51	32.50	0.96	20.35	-13724.87	-5.75	0.23	6.00	1.71	5.22	237.59
10	251.32	71.45	233.00	573.80	1.53	36.70	1.08	28.32	-22504.83	-6.53	-2.03	4.50	5.96	4.78	259.60
11	234.38	74.45	254.00	594.90	1.50	30.00	0.92	29.51	-19077.24	-5.54	0.25	5.80	1.89	5.11	242.50
12	285.22	72.59	237.40	568.80	1.52	32.90	1.20	28.78	-86947.33	-5.78	-0.14	5.65	3.10	5.02	247.01
13	206.32	64.90	221.20	518.30	1.50	30.10	0.93	25.73	-16936.30	-5.69	0.16	5.85	1.32	5.87	211.26
14	164.24	51.34	170.80	407.90	1.51	32.50	0.96	20.35	-13724.87	-5.67	0.16	5.83	1.30	5.12	242.03
15	122.16	37.78	120.40	297.50	1.54	37.20	1.01	14.97	-10513.53	-5.63	0.21	5.84	1.07	5.14	241.47
16	126.13	32.95	108.30	267.00	1.52	36.90	1.16	13.06	-12145.00	-5.82	-0.15	5.67	2.03	5.09	243.41
17	153.14	39.50	115.90	315.40	1.60	54.70	1.32	15.66	-15011.58	-6.75	-2.14	4.61	5.61	4.80	258.93
18	187.03	40.64	120.30	310.40	1.59	44.30	1.55	16.11	-79454.10	-5.93	-0.24	5.68	2.56	4.98	249.07
19	138.16	39.63	128.10	316.60	1.53	37.20	1.08	15.71	-12561.30	-5.27	0.16	5.43	2.23	4.98	248.95
20	150.18	42.98	135.60	345.10	1.55	41.80	1.11	17.04	-13599.56	-6.23	-1.17	5.06	4.30	4.52	247.36
21	178.27	56.07	187.30	446.90	1.51	32.30	0.95	22.22	-14795.38	-5.68	0.15	5.83	1.34	5.12	242.11

3.1. Principal component analysis

The set of descriptors encoding the 21 ortho alkyl substituted 4-X-phenols, topologic, electronic and energetic parameters are submitted to PCA analysis [26]. The first three principal axes are sufficient to describe the information provided by the data matrix. Indeed, the percentages of variance are 46.26%; 33.57% and 9.70% for the axes F1, F2 and F3, respectively. The total information is estimated to a percentage of 89.53%.

The principal component analysis (PCA) [27] was conducted to identify the link between the different variables. Correlations between the sixteen descriptors are shown in Table 3 as a correlation matrix and in Figure 3 these descriptors are represented in a correlation circle.

The Pearson correlation coefficients are summarized in the following table 3. The obtained matrix provides information on the negative or positive correlation between variables.

* The Polarizability (α) is perfectly correlated with the Molar Refractivity (MR) for r = 1. Both variables are redundant. Taking into account these observations, we removed the polarizability (α) order not to distort the rest of calculation.
* The Parachor (Pc) is strongly correlated with the Molar Refractivity (MR) (r=0.999) and the Polarizability (α) (r=0.999).
* The Parachor (Pc) is strongly correlated with the Molar Volume (MV) for r=0.997.
* The dipole moment (μ) is highly negatively correlated with the gap energy (ΔE) (r= -0.951) and the energy E_{LUMO} (r= -0.948).

Table 3: Correlation matrix (Pearson (n)) between different obtained descriptors

	pIC_{50}	MW	MR	MV	Pc	n	γ	D	α	E_T	E_{HOMO}	E_{LUMO}	ΔE	μ	E_a	λ_{max}
pIC_{50}	**1**															
MW	**0,796**	**1**														
MR	**0.759**	0.930	**1**													
MV	0.739	0.903	**0.996**	**1**												
Pc	0.761	0.922	**0.999**	**0.997**	**1**											
n	-0.404	-0.412	**-0.648**	**-0.707**	-0.661	**1**										
γ	-0.370	-0.443	**-0.646**	**-0.699**	-0.649	**0.949**	**1**									
D	-0.148	-0.097	-0.439	-0.499	-0.456	**0.843**	0.769	**1**								
α	0.759	0.930	**1.000**	**0.996**	**0.999**	-0.647	-0.646	-0.439	**1**							
E_T	-0.239	-0.529	-0.223	-0.168	-0.193	-0.295	-0.100	-0.632	-0.223	**1**						
E_{HOMO}	-0.146	-0.086	0.055	0.116	0.054	-0.562	-0.647	-0.443	0.055	0.117	**1**					
E_{LUMO}	-0.241	-0.119	0.038	0.100	0.032	-0.548	**-0.668**	-0.461	0.038	0.086	0.940	**1**				
ΔE	-0.298	-0.135	0.020	0.077	0.011	-0.485	**-0.624**	-0.433	0.020	0.051	0.803	0.958	**1**			
μ	0.329	0.239	0.048	-0.017	0.054	0.492	0.619	0.491	0.048	-0.187	**-0.842**	**-0.948**	**-0.951**	**1**		
E_a	0.083	0.007	0.165	0.211	0.163	-0.485	-0.547	-0.454	0.165	0.149	0.509	0.640	0.692	-0.702	**1**	
λ_{max}	0.065	0.085	-0.103	-0.156	-0.104	0.503	0.552	0.517	-0.103	-0.226	-0.496	-0.642	-0.705	0.706	-0.904	**1**

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 3) shows that the F1 axis (46.26% of the variance) is mainly due to the Surface Tension (γ), while the axis F2 (33.57% of the variance) is located by the other parameters topologic.

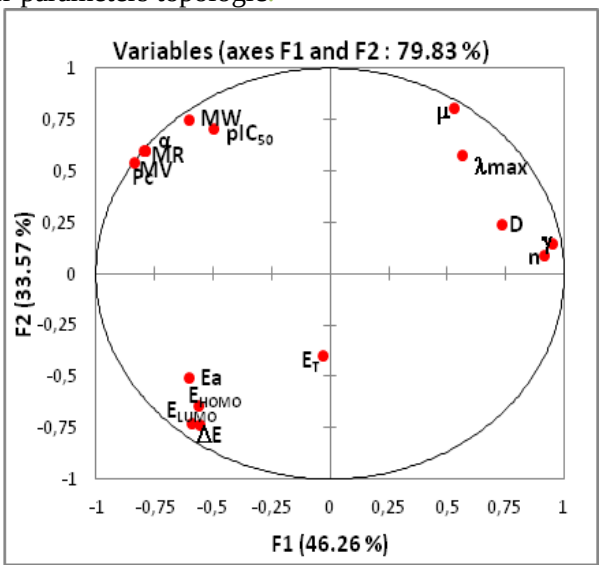

Figure 3: Correlation circle

On the other hand, the projection according to the plan F1-F2 (79.83% of the total variance) of the studied molecules (Figure 4) shows that we can discern two groups of molecules: The group 1 (G1) containing the compounds with $pIC_{50} \leq 3.80$, the group 2 (G2) containing the compounds with $pIC_{50} > 3.80$.
In this representation, the compound 2 that should be in group 2 (high value of pIC_{50}), but that an exception because they contain group which it's not similar to those of other compounds of this series.

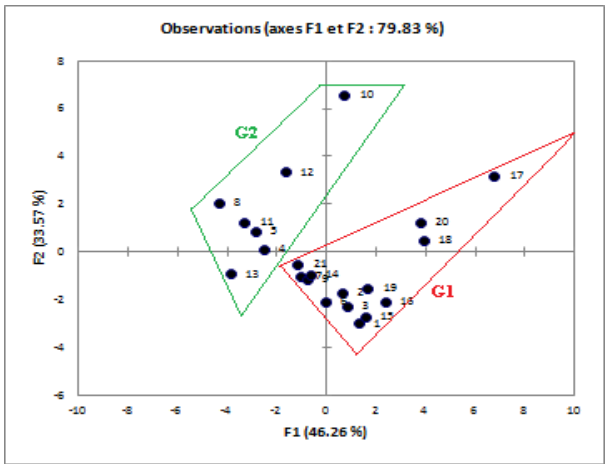

Figure 4: Cartesian diagram according to F1 and F2: Separation between two regions

3.2. Multiple linear regressions

To establish quantitative relationships between cytotoxicity pIC_{50} and selected descriptors, our array data were subjected to a multiple regression linear and were nonlinear. Only variables whose coefficients are significant were retained.

3.2.1. Multiple linear regression of the variable cytotoxicity (MLR)

Many attempts have been made to develop a relationship with the indicator variable of cytotoxicity pIC_{50}, but the best relationship obtained by this method is only one corresponding to the linear combination of several descriptors selected: the molecular weight (MW), the Molar Refractivity (MR)**,** the Molar Volume (MV)**,** the Parachor (Pc)**,** the Refractive Index (n), the Density (D), the total energy E_T and the absorption maximum λ_{max}.
The resulting equation is:

$$\mathbf{pIC_{50}} = 69.460 + 4.060\ 10^{-02}\times\mathbf{MW} + 1.558\times\mathbf{MR} - 0.1325\times\mathbf{MV} - 0.147\times\mathbf{Pc} \quad 46.906\times\mathbf{n} + 5.821\times\mathbf{D} + 7.933\ 10^{-05}\times\mathbf{E_T} - 1.385\ 10^{-02}\times\boldsymbol{\lambda_{max}} \quad (4)$$

For our 21 compounds, the correlation between experimental cytotoxicity and calculated one based on this model is quite significant (Figure 5) as indicated by statistical values:

N = 21 **R = 0.962** **R^2 = 0.926** **RMSE = 0.167**

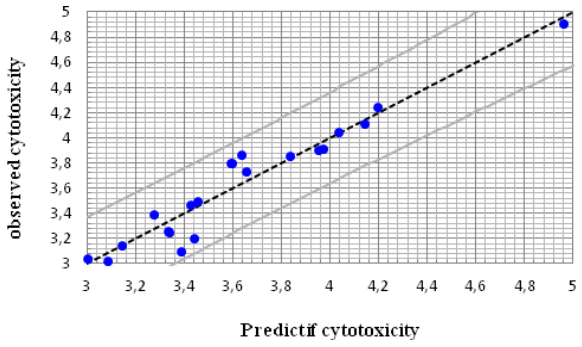

Figure 5: Graphical representation of calculated and observed cytotoxicity by MLR

The figure 5 shows a very regular distribution of cytotoxicity values depending on the experimental values.

3.2.2. Multiple nonlinear regression of the variable cytotoxicity (MNLR)

We have also used the technique of nonlinear regression model to improve the structure cytotoxicity in a quantitative way, taking into account several parameters. This is the most common tool for the study of multidimensional data. We have applied it to Table 2 containing 21 molecules associated with fifteen variables.
The resulting equation is:

$$\mathbf{pIC_{50}} = 33314{,}370 + 2{,}218\times\mathbf{MW} - 28{,}170\times\mathbf{MR} - 5{,}282\times\mathbf{MV} + 4{,}584\times\mathbf{P} - 43122{,}751\times\mathbf{n} + 0{,}235\times\boldsymbol{\gamma} - 296{,}882\times\mathbf{D} + 6{,}026\ 10^{-5}\mathbf{E_T} + 4{,}380\times\mathbf{E_{HOMO}} - 1{,}609\times\mathbf{E_{LUMO}} + 0{,}616\times\boldsymbol{\mu} - 5{,}155\times\mathbf{Ea} - 0{,}150\times\boldsymbol{\lambda_{max}} - 2{,}586\ 10^{-3}\times\mathbf{MW^2} + 0{,}253\times\mathbf{MR^2} + 2{,}052\ 10^{-3}\times\mathbf{MV^2} - 3{,}652\ 10^{-3}\times\mathbf{P^2} + 14155{,}278\times\mathbf{n^2} - 6{,}282\ 10^{-2}\times\boldsymbol{\gamma^2} + 29{,}759\times\mathbf{D^2} \quad (5)$$

The obtained parameters describing the topological and the electronic aspects of the studied molecules are:

N = 21 R = 0.999 R^2 = 0.999

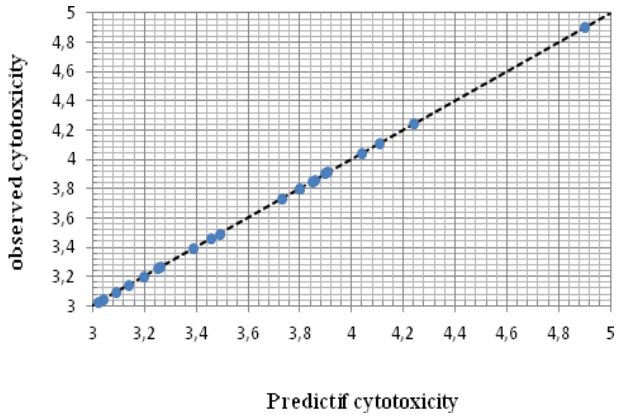

Figure 6: Graphical representation of calculated and observed cytotoxicity by MNLR

With MLNR was obtained significantly better correlation coefficient R = 0, 999
Figure 6 shows a very uniform distribution of the cytotoxicity observed values depending on the experimental values and the correlation between the experimental results and calculated alter them pIC_{50}. The residual values tended to zero which is why we did not graph for prediction residuals.
As part of this conclusion, we can say that the cytotoxicity values obtained from nonlinear regression are highly correlated to that of the observed cytotoxicity comparing to results obtained by MLR method.

3.3. Artificial neural networks ANN

In order to increase the probability of good characterization of studied compounds, 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 ANN calculated cytotoxicity model was developed using the properties of several studied compounds. The correlation between ANN calculated and experimental cytotoxicity values are very significant as illustrated in figure 7 and as indicated by R and R^2 values.

N = 21 R = 0.956 R^2 = 0.914 RMSE = 0.0536

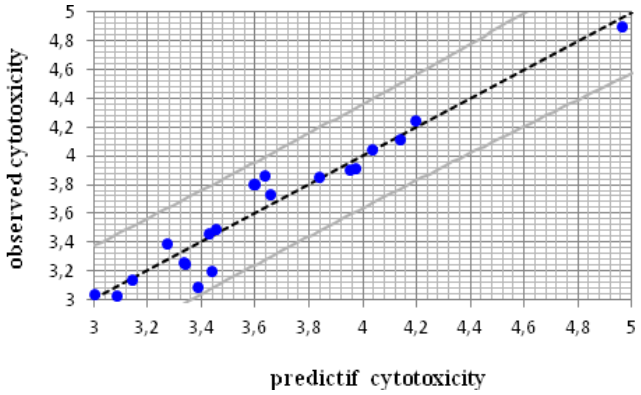

Figure 7: Correlations of observed and predicted activities calculated using ANN

The statistic of the three steps of the calculation by the ANNs: training, validation and test are illustrated in table 4.

Table 4: Values obtained by ANNs

	Samples	**RMSE**	**R**	**R^2**
Training	15	0.0535931	0.956284	0.914479
Validation	3	0.0418975	0.955253	0.912508
Test	3	0.212149	0.976446	0.953446

R: correlation coefficient; R^2: determination coefficient; RMSE: root mean square error.

The obtained squared correlation coefficient (R^2) value is 0.914 for this data set of 2-alkyl and 2,6-dialkyl-4-X-phenols. It confirms that the multiple nonlinear regression (MNLR) results were the best to build the quantitative structure activity relationship models.
In this part, we investigated the best linear QSAR regression equations established in this study. Based on this result, a comparison of the quality of the CPA, MLR and ANN models shows that the MNLR models have substantially better predictive capability because the MNLR approach gives better results than MLR and ANN. MNLR was able to establish a satisfactory relationship between the molecular descriptors and the activity of the studied compounds.

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

N°	pIC_{50} (obs.)	pIC_{50} (calc.)		
		MLR	**NMLR**	**ANN**
1	3.020	3.086	3.020	2.8091
2	3.860	3.634	3.860	3.363
3	3.200	3.440	3.200	2.890
4	3.850	3.838	3.850	4.132
5	4.040	4.035	4.040	4.019
6	3.260	3.334	3.260	3.105
7	3.250	3.341	3.250	3.700
8	3.900	3.954	3.900	3.775
9	3.730	3.656	3.730	3.591
10	4.900	4.965	4.900	4.820
11	3.910	3.975	3.910	3.972
12	4.110	4.142	4.110	4.161
13	4.240	4.198	4.240	4.130
14	3.800	3.594	3.800	3.514
15	3.040	3.001	3.040	2.813
16	3.090	3.387	3.090	3.391
17	3.490	3.454	3.490	3.203
18	3.460	3.428	3.460	3.669
19	3.390	3.275	3.390	3.096
20	3.140	3.143	3.140	3.735
21	3.800	3.599	3.800	3.655

Conclusion

In this work we have investigated the QSAR regression to predict the cytotoxicity of phenolic compounds.
Comparison of key statistical terms like R or R^2 of different models obtained by using different statistical tools and different descriptors has been shown in table 5.
The study of the quality of the MLR, MNLR and ANN models show that the MNLR result has substantially better predictive capability than the other methods. With MNLR approach, we have established a relationship between several descriptors and inhibition values pIC_{50} of several organic compounds based on ortho alkyl substituted 4-X-phenols in satisfactory manners.
Finally, we can conclude that studied descriptors, which are sufficiently rich in chemical, electronic and topological information to encode the structural feature may be used with other descriptors for the development of predictive QSAR models.

Acknowledgment-WE ARE GRATEFUL TO THE "ASSOCIATION MAROCAINE DES CHIMISTES THÉORICIENS" (AMCT) FOR ITS PERTINENT HELP CONCERNING THE PROGRAMS.

References

1. Moridani M.Y., Siraki A., O'Brien P.J., Quantitative structure toxicity relationships for phenols in isolated rat hepatocytes, J. Chemico-Biological Interactions 145 (2003) 213-223.
2. Buren J.V., Vos L.D., Pilnik W., Measurement of chlorogenic acid and flavonol glycosides in apple juice by a chromatographic-fluorometric method, J. Food Sci. 38 (1973) 656-658.
3. Challis B.C., Barlett C.D., Possible carcinogenic effects of coffee constituents, Nature 254 (1975) 532-533.
4. Skaper S.D., Fabris M., Ferrari V., Dalle Carbonare M., Leon A., Quercetin protects cutaneous tissue-associated cell types including sensory neurons from oxidative stress induced by glutathione depletion: cooperative effects of ascorbic acid, Free Radic. Biol. Med. 22 (1997) 669-678.
5. Rice-Evans C.A., Miller N.J., Paganga G., Structure- antioxidant activity relationships of flavonoids and phenolic acids, Free Radic. Biol. Med. 20 (1996) 933-956.
6. Tanaka T., Kojima T., Kawamori T., Wang A., Suzui M., Okamoto K., Mori H., Inhibition of 4-nitroquinoline-1-oxide-induced rat tongue carcinogenesis by the naturally occurring plant phenolics caffeic, ellagic, chlorogenic and ferulic acids, Carcinogenesis 14 (1993) 1321-1325.
7. Tanaka T., Kojima T., Kawamori T., Yoshimi N., Mori H., Chemoprevention of diethylnitrosamine-induced hepatocarcinogenesis by a simple phenolic acid protocatechuic acid in rats, Cancer Res. 15 (1993) 2775-2779.
8. Selassie C. D., Verma R. P., Kapur S., Shusterman A. J., Hansch C., J. Chem. Soc., Perkin 2 (2002) 1112.
9. Hansch C., Muir R.M., Fujita T., Maloney P.P, Geiger F., Streich M., J. Am. Chem. Soc. 85 (1963) 2817–2825.
10. Bodor N., Current Medicinal Chemistry 5 (1988) 353-380. From book: Biochemistry of Redox Reactions, by Bernard Testa, editor: London [u, a], Acad. Press (1995).
11. Elhallaoui M., Elasri M., Ouazzani F., Mechqrane A., Lakhlifi T., Quantitative structure-activity relationships of noncompetitive antagonists of the NMDA receptor: a study of a series of MK801 derivative molecules using statistical methods and neural network. Int. J. Mol. Sci. 4 (2003) 249–262.
12. Jing G., Zhou Z., Zhuo J., Quantitative structure-activity relationship (QSAR) study of toxicity of quaternary ammonium compounds on Chlorella pyrenoidosa and Scenedesmus quadricauda. Chemosphere 86 (2012) 76–82.
13. Ghamali M., Chtita S., Adad A., Hmamouchi R., Bouachrine M., Lakhlifi T., Biological activity of molecules based on benzylpiperidine inhibitors of human acetylcholinesterase (HuAChE). Predicting by Combining DFT and QSAR calculations. International Journal of Advanced Research in Computer Science and Software Engineering, volume 4, Issue 1 (2014) 536-546.
14. Chtita S., Larif M., Ghamali M., Adad A., Hmamouchi R., Bouachrine M., Lakhlifi T., Studies of two different cancer cell lines activities (MDAMB-231 and SK-N-SH) of imidazo[1,2-a]pyrazine derivatives by combining DFT and QSAR results. International Journal of Innovative Research in Science, Engineering and Technolog, 2 (11), (2013) 6586-6601.
15. Larif M., Adad A., Hmamouchi R., Taghki A. I., Soulaymani A., Elmidaoui A., Bouachrine M., Lakhlifi T., Biological activities of triazine derivatives. Combining DFT and QSAR results, article in press in Arabian Journal of Chemistry (2013).
16. Adad A., Hmamouchi R., Taghki A. I., Abdellaoui A., Bouachrine M., Lakhlifi T., Atmospheric half-lives of persistent organic pollutants (POPs) study combining DFT and QSPR results, J. of Chemical and Pharmaceutical Research, Vol. 5, Issue 7 (2013) 28-41.
17. Zupan J., Gasteiger J., Neural Networks for Chemistry and Drug Design: An Introduction, second ed., VCH, Weinheim (1999).
18. Turkkan N., Revue de l'Université de Moncton 26 (1), (1993) 205-221.
19. Lee P.Y., Chen C.Y.J., Hazard. Mater. 165 (2012) 156-161.
20. Jing G., Zhou Z., Zhuo J., Chemosphere 86 (2012) 76-82.
21. Adamo C., Barone V., Chem. Phys. Lett. 330 (2000) 152-160.
22. Parac M., Grimme S., All calculations were done by GAUSSIAN 03 W software, J. Phys. Chem. A 106 (2003) 6844-6850.
23. Gaussian 03, Revision B.01, M. J. Frisch, and al., Gaussian Inc., Pittsburgh, PA (2003).
24. Becke A.D., J. Chem. Phys. 98 (1993) 1372.
25. Lee C., Yang W., Parr R.G., Phys. Rev. B. 37 (1988) 785-789.
26. STATITCF Software, Technical Institute of cereals and fodder (1987), Paris, France.
27. Jonathan N., Nobuyasu H., Yuso S., Ahsan K., Jae-Jak H., Jong-Sik N.N., Xiang L.L.Q., Jun L.L., Gan Z., Shigeki M., Chemosphere 86 (2012) 718–726.