



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



ISSN : 2231- 3176
CODEN (USA): JCMMDA

Acute toxicity of halogenated phenols: Combining DFT and QSAR studies

A. Ousaa¹, B. Elidrissi¹, M. Ghamali¹, S. Chtita¹, M. Bouachrine² and T. Lakhli^{1*}

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

²MEM, ESTM, University Moulay Ismail, Meknes, Morocco

ABSTRACT

In order to investigate the relationship between activities and structures, a 3D-QSAR study is applied to a set of 43 halogenated phenols compounds. 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. Density functional theory (DFT) and ab-initio molecular orbital calculations have been carried out in order to get insights into the structure, chemical reactivity and property information for the series of study compounds. This study shows that the MRA and MNLN have served also to predict activities, but when compared with the results given by the ANN, we realized that the predictions fulfilled by this latter were more effective. The obtained results suggested that the proposed combination of several calculated parameters could be useful to predict the biological activity of halogenated phenols over *Tetrahymenapyriformis*.

Keywords: 3D-QSAR model, DFT study, Halogenated phenols compounds, *Tetrahymenapyriformis*,

INTRODUCTION

Phenols compounds are basic materials for industry production, which are commonly used in chemical synthesis. They can spread through air and water, with strong carcinogenicity, teratogenicity and mutagenicity [1-2], which will cause great damage to environment, plants, animals and human health. Therefore, it is vital to protect the environment and prevent occupational poisoning by studying the acute toxicity of phenols compounds.

The experiment is a direct way to obtain the toxicity data of organic compounds, which has many deficiencies, such as requirement of myriads of trial organisms, high expense, long time, the difference in measured value between different researchers and so on. Consequently, it would be impossible to gain the toxicity data of all organic compounds by experiment. As new compounds are springing up, other difficulties will follow. So it is necessary to use the theoretical research to make up for disadvantages of the experiment and to predict the toxicity data of compounds quickly and exactly.

QSAR can predict the bioactivity such as toxicity, mutagenicity and carcinogenicity based on structural parameters of compounds and appropriate mathematical models. With the rapid development of computer science and theoretical quantum chemical study, it can speedily and precisely obtain the quantum chemical parameters of compounds by computation.

Moreover, these parameters, which have definite physical meaning, along with the introduction of the QSAR model can increase the interpretability. So quantum chemical theory is extensively applied in establishing QSAR models [3-5].

In this work, we have modeled the toxicity of 43 halogenated phenols compounds to *tetrahymenapyrififormis* using several statistical tools, principal components analysis (PCA), multiple linear regression (MLR), non-linear regression (RNLM) and artificial neural network (ANN) calculations, we accordingly propose a quantitative model, and we try to interpret the activity of these compounds relying on the multivariate statistical analyses.

MATERIALS AND METHODS

Data sources

Acute toxicity data of 43 halogenated phenols to *tetrahymenapyrififormis* were taken from a literature [6]. IC_{50} here means the millimolar concentration causing 50% inhibition of growth about halogenated phenols to *tetrahymenapyrififormis*. The bigger the value of $-\log IC_{50}$ (pIC_{50}), the higher is toxicity of compounds, and vice versa.

The following table shows the studied compounds and the corresponding experimental activities pIC_{50} (Table 1). The experimental toxicity of the studied compounds has been collected from recent work [6]. The range of the toxicity data varies from 0.02 to 2.71 (μM).

Table 1: Halogenated phenols and their observed toxicities against Tetrahymenapyrififormis

N°	Name (IUPAC)	pIC_{50}	N°	Name (IUPAC)	pIC_{50}
1	4-fluorophenol	0.017	23	3-bromophenol	1.145
2	2-chlorophenol	0.183	24	4-bromo-2,6-dimethylphenol	1.167
3	2-fluorophenol	0.185	25	2,3,5,6-tetrafluorophenol	1.167
4	2-bromophenol	0.330	26	4-chloro-3,5-dimethylphenol	1.201
5	3-fluorophenol	0.381	27	4-bromo-3,5-dimethylphenol	1.268
6	2-chloro-5-methylphenol	0.393	28	2,3-dichlorophenol	1.276
7	2,6-difluorophenol	0.471	29	4-bromo-6-chloro-2-methylphenol	1.276
8	4-chlorophenol	0.545	30	2,4-dibromophenol	1.398
9	2-bromo-4-methylphenol	0.599	31	3,5-dichlorophenol	1.569
10	2,4-difluorophenol	0.604	32	Pentafluorophenol	1.638
11	4-bromophenol	0.680	33	3,4-dichlorophenol	1.745
12	2-chloro-4,5-dimethylphenol	0.688	34	4-bromo-2,6-dichlorophenol	1.778
13	4-chloro-2-methylphenol	0.701	35	4-chloro-2-isopropyl-5-methylphenol	1.854
14	2,6-dichlorophenol	0.735	36	2,4,6-tribromophenol	2.030
15	4-chloro-3-methylphenol	0.796	37	Pentachlorophenol	2.049
16	2,6-dichloro-4-fluorophenol	0.804	38	2,4,5-trichlorophenol	2.097
17	3-chlorophenol	0.871	39	2,3,5,6-tetrachlorophenol	2.222
18	2,4-dichlorophenol	1.036	40	2,3,5-trichlorophenol	2.373
19	4-chloro-3-ethylphenol	1.081	41	3,4,5,6-tetrabromo-2-methylphenol	2.574
20	2,5-dichlorophenol	1.125	42	Pentabromophenol	2.664
21	3-chloro-4-fluorophenol	1.131	43	2,3,4,5-tetrachlorophenol	2.712
22	2,4,6-trichlorophenol	1.410	-	-	-

Molecular descriptors

All computations were performed by using Gaussian 03W program [7]. The geometries of all 43 theoretically possible halogenated phenols were optimized with DFT method at the B3LYP/6-31G (d) level and frequency calculations were performed at the same level for all of the possible geometries to ensure they are minimal on the potential energy surface. Then choose some related structural parameters from the results of quantum computation: 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 factor of oscillation $f_{(SO)}$.

As well as the ChemSketch program (Demo version 10.0) [8] was employed to calculate the others molecular descriptors such as: 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).

Statistical analysis

The structures of 43 halogenated phenols to *tetrahymenapyrififormis* were studied by statistical methods based on the principal component analysis (PCA) [9] using the software XLSTAT version 2013 [10]. PCA is a statistical technique useful for summarizing all the information encoded in the structures of the compounds. It is also very helpful for understanding the distribution of the compounds [11]. This is an essentially descriptive statistical method which aims to present, in graphic form, the maximum of information contained in the data table 1 and table 2.

The multiple linear regression (MLR) analysis with descendent selection and elimination of variables was employed to model the structure activity relationships. It is a mathematic technique that minimizes differences between actual

and predicted values. It has served also to select the descriptors used as the input parameters in the Multiples nonlinear regression (MNL) and artificial neural network (ANN).

The (MLR), the (MNL) were generated using the software XLSTAT version 2013 [10], to predict cytotoxic effects IC₅₀ activities. Equations were justified by the correlation coefficient (R), mean squared error (MSE) [10]. ANN is artificial systems simulating the function of the human brain. Three components constitute a neural network: the processing elements or nodes, the topology of the connections between the nodes, and the learning rule by which new information is encoded in the network. While there are a number of different ANN models, the most frequently used type of ANN in QSAR is the three-layered feed forward network [12]. In this type of networks, the neurons are arranged in layers (an input layer, one hidden layer and an output layer). Each neuron in any layer is fully connected with the neurons of a succeeding layer and no connections are between neurons belonging to the same layer.

According to the supervised learning adopted, the networks are taught by giving them examples of input patterns and the corresponding target outputs. Through an iterative process, the connection weights are modified until the network gives the desired results for the training set of data. A back propagation algorithm is used to minimize the error function. This algorithm has been described previously with a simple example of application [13] and a detail of this algorithm is given elsewhere [14].

The ANNs analysis was performed with the use of Matlab software version 2009a Neural Fitting tool (nftool) toolbox [15].

Table 2: the values of the sixteen chemical descriptors

N°	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)	f (so)
1	112.10	28.12	92.00	229.40	1.52	38.50	1.22	11.15	-11074.34	-5.95	-0.31	5.64	1.96	5.06	244.81	0.047
2	128.56	33.02	99.80	258.10	1.58	44.70	1.29	13.09	-20887.12	-6.25	-0.35	5.90	0.93	5.14	241.37	0.032
3	112.10	28.12	92.00	229.40	1.52	38.50	1.22	11.15	-11074.38	-6.13	-0.10	6.02	0.79	5.25	236.07	0.022
4	173.01	35.82	104.00	272.70	1.60	47.20	1.66	14.20	-78383.56	-6.22	-0.37	5.85	0.94	5.07	244.41	0.001
5	112.10	28.12	92.00	229.40	1.52	38.50	1.22	11.15	-11074.40	-6.17	-0.09	6.08	0.80	5.27	235.11	0.023
6	142.58	37.85	116.00	295.80	1.57	42.10	1.23	15.00	-21957.75	-6.15	-0.32	5.83	1.23	5.09	243.57	0.041
7	130.09	28.12	96.20	236.50	1.50	36.40	1.35	11.14	-13776.32	-6.30	-0.10	6.21	1.91	5.27	235.27	0.007
8	128.56	33.02	99.80	258.10	1.58	44.70	1.29	13.09	-20887.05	-6.09	-0.40	5.69	2.49	5.03	246.57	0.029
9	187.03	40.64	120.30	310.40	1.59	44.30	1.55	16.11	-79454.17	-6.02	-0.34	5.68	1.33	4.95	250.28	0.001
10	130.09	28.12	96.20	236.50	1.50	36.40	1.35	11.14	-13776.43	-6.13	-0.42	5.72	0.66	5.10	243.25	0.042
11	173.01	35.82	104.00	272.70	1.60	47.20	1.66	14.20	-78383.44	-6.04	-0.40	5.64	2.41	4.98	249.13	0.026
12	156.61	42.67	132.30	333.40	1.56	40.20	1.18	16.91	-23028.36	-5.94	-0.16	5.78	1.60	4.99	248.54	0.049
13	142.58	37.85	116.00	295.80	1.57	42.10	1.23	15.00	-21957.69	-5.97	-0.01	5.96	2.84	5.06	245.01	0.029
14	163.00	37.92	111.70	294.00	1.59	47.80	1.46	15.03	-33401.79	-6.49	-0.69	5.80	2.08	5.02	247.16	0.033
15	142.58	37.85	116.00	295.80	1.57	42.10	1.23	15.00	-21957.70	-5.98	-0.24	5.74	2.46	5.07	244.56	0.029
16	180.99	37.91	115.90	301.10	1.57	45.40	1.56	15.03	-36103.80	-6.48	-1.00	5.48	1.37	4.87	254.58	0.066
17	128.56	33.02	99.80	258.10	1.58	44.70	1.29	13.09	-20887.07	-6.29	-0.37	5.92	1.11	5.15	240.97	0.028
18	163.00	37.92	111.70	294.00	1.59	47.80	1.46	15.03	-33401.86	-6.36	-0.75	5.60	1.07	4.92	251.97	0.038
19	156.61	42.57	132.60	334.70	1.56	42.57	1.18	16.87	-23028.16	-5.94	-0.19	5.75	2.51	5.07	244.52	0.031
20	163.00	37.92	111.70	294.00	1.59	47.80	1.46	15.03	-33401.88	-6.52	-0.73	5.79	1.46	5.02	247.03	0.037
21	146.55	33.02	104.00	265.20	1.55	42.20	1.41	13.09	-23589.04	-6.24	-0.66	5.58	2.24	4.98	249.11	0.056
22	197.45	42.81	123.70	329.80	1.61	50.50	1.60	16.97	-45916.49	-6.57	-1.05	5.52	1.42	4.82	257.22	0.040
23	173.01	35.82	104.00	272.70	1.60	47.20	1.66	14.20	-78383.46	-6.26	-0.39	5.87	1.02	5.11	242.57	0.031
24	201.06	45.47	136.50	348.00	1.58	42.10	1.47	18.02	-80524.73	-5.82	-0.13	5.70	2.87	5.05	245.56	0.021
25	166.07	28.10	104.70	250.70	1.45	32.90	1.59	11.14	-19180.07	-6.76	-0.44	6.32	1.38	5.25	235.98	0.001
26	156.61	42.67	132.30	333.40	1.56	40.20	1.18	16.91	-23028.33	-5.89	-0.09	5.80	1.97	5.25	235.98	0.001
27	201.06	45.47	136.50	348.00	1.58	42.10	1.47	18.02	-80524.77	-5.86	-0.09	5.77	1.92	5.09	243.57	0.017
28	163.00	37.92	111.70	294.00	1.59	47.80	1.46	15.03	-33401.77	-6.49	-0.67	5.82	1.32	5.05	245.43	0.026
29	221.48	45.54	132.20	346.20	1.60	46.90	1.67	18.05	-91968.76	-6.16	-0.59	5.57	3.56	4.91	252.53	0.028
30	251.90	43.51	120.20	323.20	1.64	43.51	1.64	17.25	-148394.69	-6.27	-0.76	5.51	0.93	4.84	256.31	0.038
31	163.00	37.92	111.70	294.00	1.59	47.80	1.46	15.03	-33401.82	-6.63	-0.73	5.90	2.38	5.05	245.44	0.019
32	184.06	28.10	108.90	257.90	1.43	31.40	1.69	11.14	-21775.11	-6.56	-1.85	4.72	1.98	6.28	197.44	0.002
33	163.00	37.92	111.70	294.00	1.59	47.80	1.46	15.03	-33401.71	-6.33	-0.71	5.63	2.61	4.94	251.19	0.033
34	241.90	45.61	127.90	344.50	1.63	52.50	1.89	18.08	-103412.89	-6.51	-1.04	5.47	1.38	4.66	266.22	0.001
35	184.66	52.03	166.10	410.80	1.54	37.30	1.11	20.63	-25169.29	-5.84	-0.11	5.72	2.43	5.07	244.78	0.034
36	330.80	51.20	136.40	373.70	1.67	56.30	2.42	20.29	-218405.74	-6.45	-1.18	5.27	1.49	4.71	263.28	0.044
37	266.34	52.60	147.60	401.60	1.63	54.70	1.80	20.85	-70945.45	-6.82	-1.45	5.37	1.91	4.68	264.94	0.032
38	197.45	42.81	123.70	329.80	1.61	50.50	1.60	16.97	-45916.48	-6.56	-1.03	5.53	2.05	4.82	257.16	0.045
39	231.89	47.71	135.60	365.70	1.62	52.80	1.71	18.91	-58431.02	-6.85	-1.23	5.63	1.40	4.75	260.89	0.001
40	197.45	42.81	123.70	329.80	1.61	50.50	1.60	16.97	-45916.50	-6.76	-1.00	5.75	1.03	4.95	250.70	0.023
41	423.72	63.72	168.80	461.90	1.68	55.90	2.51	25.26	-289487.11	-6.38	-1.45	4.94	3.49	4.48	276.89	0.002
42	488.59	66.58	168.70	474.70	1.72	62.50	2.89	26.39	-358427.65	-6.61	-1.97	4.64	1.72	4.15	298.82	0.001
43	231.89	47.71	135.60	365.70	1.62	52.80	1.71	18.91	-58430.96	-6.71	-1.24	5.47	2.08	4.77	260.03	0.035

RESULTS

QSAR models and analysis

The QSAR analysis was performed using the pIC_{50} of the 43 halogenated phenols to *tetrahymenapyriformis* as reported in [16], the values of the 16 chemical descriptors as shown in table 2.

The principle (for the two studies) is to perform in the first time, a main component analysis (PCA), which allows us to eliminate descriptors that are highly correlated (dependent), then perform a decreasing study of MLR based on the elimination of descriptors (one by one) aberrant until a valid model.

Principal component analysis

The set of descriptors encoding the 43 halogenated phenols, topologic, electronic and energetic parameters are submitted to PCA analysis [17]. The first three principal axes are sufficient to describe the information provided by the data matrix. Indeed, the percentages of variance are 61.48%; 11.35% and 09.51% for the axes F1, F2 and F3, respectively. The total information is estimated to a percentage of 82.34%.

The principal component analysis (PCA) [18] was conducted to identify the link between the different variables. Bold values are different from 0 at a significance level of $p = 0.05$. Correlations between the sixteen descriptors are shown in table 3 as a correlation matrix and in figure 1 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.

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

Variables	pIC_{50}	MW	MR	MV	Pc	n	γ	D	α	E_T	E_{HOMO}	E_{LUMO}	ΔE	μ	Ea	λ_{max}	$f_{(SO)}$
pIC_{50}	1																
MW	0.75	1															
MR	0.76	0.87	1														
MV	0.74	0.78	0.95	1													
Pc	0.78	0.85	0.99	0.98	1												
n	0.54	0.72	0.79	0.57	0.70	1											
γ	0.63	0.72	0.73	0.52	0.66	0.94	1										
D	0.67	0.94	0.68	0.55	0.65	0.67	0.73	1									
α	0.76	0.87	1.00	0.95	0.99	0.79	0.73	0.68	1								
E_T	-0.57	-0.95	-0.78	-0.65	-0.73	-0.74	-0.68	-0.92	-0.78	1							
E_{HOMO}	-0.58	-0.37	-0.18	-0.10	-0.19	-0.26	-0.51	-0.51	-0.18	0.21	1						
E_{LUMO}	-0.75	-0.72	-0.51	-0.44	-0.52	-0.46	-0.63	-0.78	-0.51	0.57	0.77	1					
ΔE	-0.61	-0.76	-0.61	-0.57	-0.62	-0.45	-0.50	-0.73	-0.61	0.68	0.28	0.82	1				
μ	0.26	0.22	0.36	0.42	0.38	0.12	0.09	0.10	0.36	-0.17	0.23	-0.03	-0.26	1			
Ea	-0.47	-0.67	-0.74	-0.56	-0.67	-0.85	-0.82	-0.58	-0.74	0.65	0.24	0.34	0.30	-0.12	1		
λ_{max}	0.54	0.76	0.79	0.62	0.73	0.86	0.85	0.68	0.79	-0.74	-0.29	-0.46	-0.43	0.13	-0.99	1	
$f_{(SO)}$	-0.20	-0.29	-0.16	-0.19	-0.18	-0.01	-0.03	-0.31	-0.16	0.29	0.12	0.12	0.07	-0.05	-0.10	0.04	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 very significant for $p < 0.001$

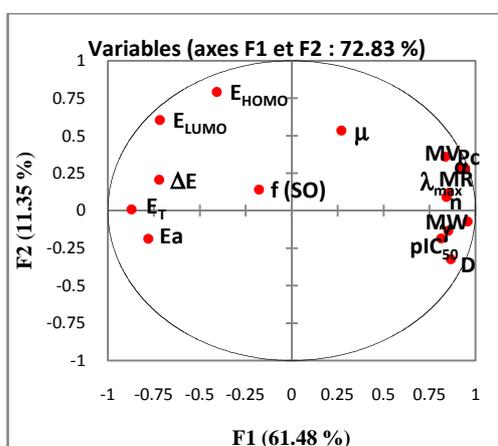


Figure 1: Correlation circle

Analysis of projections according to the plan F1-F2 (72.83% of the total variance) of the studied molecules (Figure 2) is showing in figure 2:

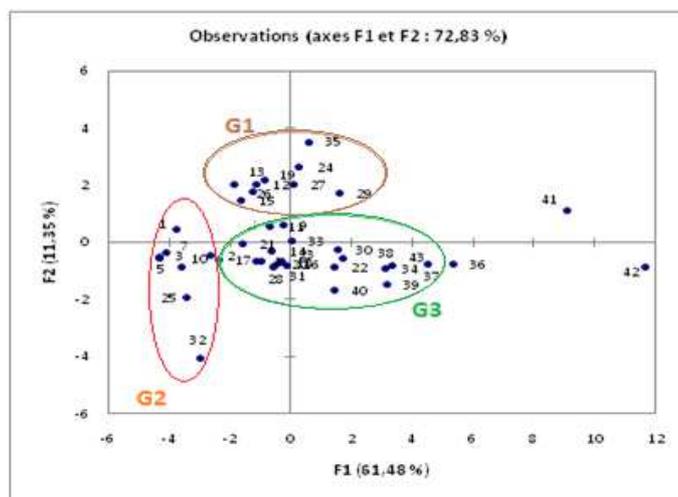


Figure 2: Cartesian diagram according to F1 and F2: Separation between three regions

Multiple Linear Regressions

To establish quantitative relationships between toxicity 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.

Multiple linear regression of the variable toxicity (MLR)

Many attempts have been made to develop a relationship with the indicator variable of toxicity pIC_{50} , but the best relationship obtained by this method is only one corresponding to the linear combination of two descriptors selected: Parachor (Pc) and energy E_{HOMO} .

The resulting equation is:

$$pIC_{50} = -8.331 + 8.532 \cdot 10^{-3} \times Pc - 1.094 \times E_{HOMO} \quad (\text{Equation 1})$$

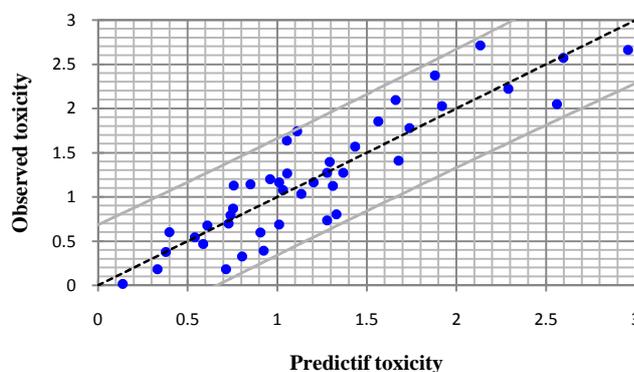


Figure 3: Graphical representation of calculated and observed toxicity by MLR

For our 43 compounds, the correlation between experimental toxicity and calculated one based on this model is quite significant (Figure 3) as indicated by statistical values:

$$N = 43 \quad R = 0.896 \quad R^2 = 0.804 \quad RMSE = 0.322$$

The figure 3 shows a very regular distribution of toxicity values depending on the experimental values.

Multiple nonlinear regression of the variable toxicity (MNLR)

We have also used the technique of nonlinear regression model to improve the structure toxicity 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 43 molecules associated with sixteen variables.

The resulting equation is:

$$\text{pIC}_{50} = -18427.59 + 9.57 \cdot 10^{-02} \times \text{MW} - 2.27 \text{MR} + 2.21 \text{MV} - 0.47 \text{Pc} + 4321.00 \text{n} - 9.68 \gamma + 18.89 \text{D} + 6.43 \cdot 10^{-06} \text{E}_T - 41.88 \text{E}_{\text{HOMO}} + 28.19 \text{E}_{\text{LUMO}} - 0.72 \mu + 1777.84 \text{Ea} + 44.44 \lambda_{\text{max}} + 23.05 \text{f}_{(\text{SO})} + 3.00 \cdot 10^{-04} \text{MW}^2 - 3.02 \cdot 10^{-02} \text{MR}^2 - 1.90 \cdot 10^{-03} \text{MV}^2 + 4.49 \cdot 10^{04} \text{Pc}^2 - 1321.63 \text{n}^2 + 0.11 \gamma^2 - 12.16 \text{D}^2 + 1.70 \cdot 10^{-10} \text{E}_T^2 - 0.90 \text{E}_{\text{HOMO}}^2 - 2.60 \text{E}_{\text{LUMO}}^2 - 2.59 \Delta \text{E}^2 + 0.27 \times \mu^2 - 77.79 \text{Ea}^2 - 4.90 \cdot 10^{-02} \lambda_{\text{max}}^2 - 523.06 \text{f}_{(\text{SO})}^2$$

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

$$\mathbf{N} = 43 \quad \mathbf{R} = 0.958 \quad \mathbf{R}^2 = 0.918 \quad \mathbf{RMSE} = 0.417$$

The toxicity value pIC_{50} predicted by this model is somewhat similar to that observed. The figure 4 shows a very regular distribution of toxicity values based on the observed values.

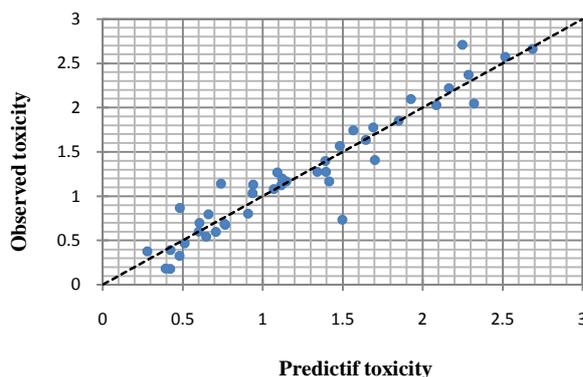


Figure 4: Graphical representation of calculated and observed toxicity by MNL

The obtained coefficient of determination in equation (2) is quite interesting (0.92). To optimize the error standard deviation and a better finish to building our model, we involve in the next part artificial neural networks (ANN).

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 toxicity model was developed using the properties of several studied compounds. The correlation between ANN calculated and experimental toxicity values are very significant as illustrated in figure 8 and as indicated by R and R^2 values.

$$\mathbf{N} = 43 \quad \mathbf{R} = 0.998 \quad \mathbf{R}^2 = 0.996 \quad \mathbf{RMSE} = 0.003$$

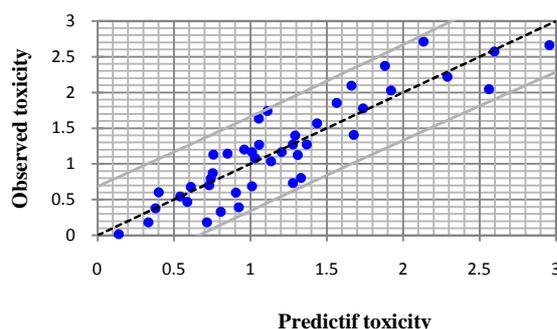


Figure 5: Correlations of observed and predicted activities calculated using ANN

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	31	$2.58 \cdot 10^{-03}$	0.998	0.996
Validation	6	0.109	0.921	0.848
Test	6	0.725	0.905	0.819

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

DISCUSSION

Principal component analysis

* The toxicity (pIC_{50}) is well correlated with the Parachor (Pc) ($r=0.782$ and $p<0.05$) and the molecular weight (MW) ($r=0.748$ and $p <0.05$) and the Molar Refractivity (MR) ($r=0.756$ and $p <0.05$) and The Polarizability (α) ($r=0.756$ and $p <0.05$) at a significant level.

* The Polarizability (α) is positively correlated with the Molar Volume (MV) ($r=0.954$ and $p<0.05$) and the Parachor (Pc) ($r=0.991$ and $p<0.05$) at a significant level.

* The energy of activation E_a is negatively correlated with maximum of absorption λ_{max} for $r=0.987$ and $p<0.05$ at a significant level.

* The Polarizability (α) is strongly correlated with the Molar Refractivity (MR) for $r=1$ and $p<0.001$ at a high level. Both variables are thus redundant. Taking into account these observations, we removed the polarizability (α) order not to distort the rest of calculation.

The principal component analysis revealed from the correlation circle (Figure 1) shows that the F1 axis (61.48% of the variance) is clearly connected to the molecular weight (MW), while the axis F2 (11.35% of the variance) is located by the other parameters of energy.

Analysis of projections according to the plan F1-F2 (72.83% of the total variance) of the studied molecules (Figure 2) shows that the molecules are dispersed, according to the of halogenated of halogenated phenols, in three classes of compounds belonging to three groups: the group 1 (G1) containing the phenol substituted by halogenated and the hydrocarbons, the group 2 (G2) containing the phenol substituted by fluorine "donors by mesomeric effect" , and the group 3 (G3) containing the phenol substituted by chlorine "donors by mesomeric effect".

In this representation, the compounds 41, 42, 36 (donor by mesomeric effect) with $pIC_{50} >2,03$, is an exception because they are phenols substituted by bromine.

Statistical Analysis

As part of this conclusion, we can say that the toxicity values obtained from nonlinear regression are highly correlated to those of the observed toxicity comparing to results obtained by MLR method.

The obtained squared correlation coefficient (R^2) value is 0.998 for this data set of halogenated phenols. It confirms that the artificial neural network results were the best to build the quantitative structure activity relationship models. 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 the CPA, 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 activity of the studied compounds.

CONCLUSION

In this work we have investigated the QSAR regression to predict the toxicity of several compounds based on halogenated phenols.

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 and ANN models show that the ANN result has substantially better predictive capability than the other methods. With ANN approach, we have established a relationship between several descriptors and inhibition values pIC_{50} of halogenated phenols.

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.

Table 5: Observed values and calculated values of pIC₅₀ according to different methods

N°	pIC ₅₀ (obs.)		pIC ₅₀ (calc.)	
	MLR	NMLR	ANN	
1	0.017	0.137	0.054	0.053
2	0.183	0.714	0.394	0.409
3	0.185	0.330	0.407	0.273
4	0.330	0.804	0.505	0.461
5	0.381	0.378	0.264	0.426
6	0.393	0.923	0.453	0.449
7	0.471	0.586	0.525	0.431
8	0.545	0.540	0.669	0.558
9	0.599	0.905	0.650	0.708
10	0.604	0.398	0.610	0.584
11	0.680	0.609	0.772	0.693
12	0.688	1.010	0.791	0.726
13	0.701	0.728	0.571	0.756
14	0.735	1.277	1.489	0.769
15	0.796	0.739	0.643	0.837
16	0.804	1.329	0.929	0.802
17	0.871	0.753	0.485	0.377
18	1.036	1.133	1.011	1.060
19	1.081	1.031	1.076	1.309
20	1.125	1.309	1.086	0.811
21	1.131	0.756	0.916	1.111
22	1.410	1.676	1.672	2.277
23	1.145	0.849	0.728	1.218
24	1.167	1.010	1.444	0.951
25	1.167	1.203	1.143	1.084
26	1.201	0.960	1.106	1.237
27	1.268	1.055	1.108	1.314
28	1.276	1.278	1.376	1.323
29	1.276	1.367	1.327	1.271
30	1.398	1.291	1.392	1.413
31	1.569	1.434	1.533	1.570
32	1.638	1.053	1.645	1.611
33	1.745	1.110	1.603	1.607
34	1.778	1.737	1.682	1.771
35	1.854	1.563	1.840	1.267
36	2.030	1.919	2.090	2.029
37	2.049	2.560	2.328	3.580
38	2.097	1.660	1.870	2.145
39	2.222	2.289	2.192	2.272
40	2.373	1.879	2.241	2.042
41	2.574	2.595	2.511	2.524
42	2.664	2.956	2.686	2.625
43	2.712	2.133	2.256	3.685

Acknowledgment

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

REFERENCES

- [1] Y B Zang, *Chin. Agric. Sci. Bull.* **2012**, 28, 282-285.
- [2] P R Zhan; H T Wang; Z X Chen, *J. Agro-Environ. Sci.* **2008**, 27, 801-804.
- [3] J W Chen; W J Peijnenburg; X Quan, *Chemosphere* **2000**, 40, 1319-1326.
- [4] E Zvinanash; T T Du; T Griff, *Chemosphere*, **2009**, 75, 1531-1538.
- [5] M J Zhu; F Ge; R L Zhu, *Chemosphere*, **2010**, 80, 46-52.
- [6] T D Mark; T Cronin; W Schultz, *Chemosphere*, **1996**, 32, 1453-1468.
- [7] M Parac; S Grimme, All calculations were done by GAUSSIAN 03W software, *J. Phys. Chem. A*, **2003**, 106, 6844-6850.
- [8] Advanced Chemistry Development Inc., Toronto, Canada **2009**.
www.acdlabs.com/resources/freeware/chemsketch/.
- [9] M Larif; A Adad; R Hmamouchi; A I Taghki; A Soulaymani; A Elmidaoui; M Bouachrine; T Lakhlifi, *Arab. J. Chem.*, in press, **2013**.
- [10] XLSTAT **2013** software (XLSTAT Company). <http://www.xlstat.com>.
- [11] R Hmamouchi; A I Taghki; M Larif; A Adad; A Abdellaoui; M Bouachrine and T Lakhlifi, *J. Chem. Pharm. Research*, **2013**, 5(9), 198-209.

- [12] J Zupan; J Gasteiger, *Neural Networks for Chemists: an Introduction*. VCH Publishers Weinheim, **1993**.
- [13] D Cherquaoui; D Villemin, *J. Chem. Soc. Far. Trans.*, **1994**, 90, 97-102.
- [14] J A Freeman; D M Skapura, *J. Operat. Res. Soc.*, **1991**, 43(11), 1106.
- [15] M Ghamali; S Chtita; A Adad; R Hmamouchi; M Bouachrine; T Lakhlifi, *IJARCSSE*, **2014**, 4, 536-546.
- [16] G Hea; L Fenga; H Chena, *International Symposium on Safety Science and Engineering in China, Procedia Engineering*, **2012**, 43, 204 – 209.
- [17] STATITCF Software, *Technical Institute of cereals and fodder*, Paris, France **1987**.
- [18] N Jonathan; H Nobuyasu; S Yuso; K Ahsan; H Jae-Jak; N N Jong-Sik; L L Q Xiang; L L Jun; Z Gan; M Shigeki, *Chemosphere*, **2012**, 86, 718–726.