Journal of Computational Methods in Molecular Design, 2014, 4 (3):28-37

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

ISSN : 2231- 3176
CODEN (USA): JCMMDA

Toxicity *in-vivo* of nitro-aromatic compounds: DFT and QSAR results

B. Elidrissi[a], A. Ousaa[a], M. Ghamali[a], S. Chtita[a], M. A. Ajana[a], M. Bouachrine[b] and T. Lakhlifi[a*]

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

ABSTRACT

Quantitative Structure–Activity Relationship (QSAR) model is presented for the estimation of the toxicity of 28 nitro-aromatic compounds including some well-known explosives. This work was conducted using the principal component analysis (PCA) method, the multiple linear regression method (MLR), the multiple non-linear regressions (MNLR) and the artificial neural network (ANN). The predicted results of various nitro-aromatic compounds afford reliable prediction of LD_{50} with respect to experimental data. Density functional theory (DFT) 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 MLR and ANN have served also to predict activities, but when compared with the results given by the RNLM, we realized that the predictions fulfilled by this latter were more effective.

Keywords: Toxicity, nitroaromatic, 3D-QSAR model, DFT study.

INTRODUCTION

Nitro-aromatic compounds are relatively rare in nature and have been introduced into the environment mainly by human activities; they are widely used in medicine, industry and agriculture. This important class of industrial chemicals is widely used in the synthesis of many diverse products, including dyes, polymers, pesticides, and explosives. Unfortunately, their extensive use has led to environmental contamination of soil and ground water [1]. The nitro group, which provides chemical and functional diversity in these molecules, also contributes to the recalcitrance of these compounds to biodegradation. Recalcitrance is further compounded by their acute toxicity, mutagenicity, and easy reduction into carcinogenic aromatic amines. Nitro-aromatic compounds are hazardous to human health and are registered on the U.S. Environmental Protection Agency's list of priority pollutants for environmental remediation. Although the majority of these compounds are synthetic in nature, microorganisms in contaminated environments have rapidly adapted to their presence by evolving new biodegradation pathways that take advantage of them as sources of carbon, nitrogen, and energy [2,3]. This review provides an overview of the synthesis of both man-made and biogenic nitro-aromatic compounds, the bacteria that have been identified to grow on and completely mineralize nitro-aromatic compounds, and the pathways that are present in these strains. The possible evolutionary origins of the newly evolved pathways are also discussed.

Since its introduction more than forty-five years ago [4], structure-activity relationships have been developed for various areas of applications, e.g. estimating of the different substance characteristics as well as their toxicity levels. Quantitative structure activity relationships (QSAR) are widely used to predict toxicity from chemical structure and corresponding physicochemical properties. The development and application of QSAR techniques started with the prediction of toxicity caused by baseline toxicants [5]. The numerous QSAR studies have been carried out in order to explain or predict toxic influence of nitro-compounds on different living systems [6,7]. In the recent papers [8,9] the QSAR analysis of oral toxicity on rats has been developed on 28 selected nitro-aromatic molecules. In this work we describe structure-toxicity relationship for the above compounds using simpler descriptor. Therefore, the aim of

the present study is the analysis of possibility of preliminary virtual screening of toxicity of nitro-aromatics by QSAR models on the base of their molecular composition, and 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.
In this study, we have modeled the toxicity of several organic compounds based on nitro-aromatic (Figure 1) using several statistical tools, principal components analysis (PCA), multiple linear regression (MLR), multiple non-linear regression (MNLR) and artificial neural network (ANN) calculations. The objectives of this work are to develop predictive QSAR models for the toxicity of our studied molecules. On the other hand, several quantum chemical methods and Quantum-chemistry calculations have been performed in order to study the molecular structure and electronic properties [10,11]. The geometry as well as the nature of their molecular orbital, HOMO (highest occupied molecular orbital) and LUMO (lowest unoccupied molecular orbital) are involved in the properties of biological activity of organic compounds. The more relevant molecular properties were calculated, these properties are the highest occupied molecular orbital energy E_{HOMO}, the lowest unoccupied molecular orbital energy $\mathbf{E_{LUMO}}$, energy gap $\mathbf{\Delta E}$, dipole moment $\boldsymbol{\mu}$, the total energy $\mathbf{E_T}$, the activation energy $\mathbf{E_a}$ and the absorption maximum $\boldsymbol{\lambda_{max}}$.

MATERIALS AND METHODS

Material

Previous studies [12] had established a quantitative model of structure activity(QSAR) relationship for series molecular structures approach have been applied to predict the oral rat toxicity of selected nitro-aromatic compounds. The activity under investigation is the toxicity lethal dose 50 (LD_{50}, acute, oral rat) expressed in mol/kg body weight and logarithmic form ($-\log LD_{50}$). Is logic to be considered, the minimal values of LD_{50}, when they are given in an interval. The following table shows the chemical compounds studied and the corresponding experimental activities $-\log LD_{50}$.

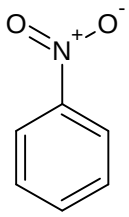

Figure 1: The structural template of nitrobenzene

Table 1: Observed values of investigated nitro-aromatics toxicity [12,13]

N°	Compound	$-\log DL_{50}$ (obs.)
1	Benzene	**-1,860**
2	Toluene	**-1,780**
3	Nitrobenzene	**-0,690**
4	1-Methyl-2-nitrobenzene	**-0,810**
5	1-Methyl-4-nitrobenzene	**-1,190**
6	2-Nitrophenol	**-0,380**
7	3-Nitrophenol	**-0,370**
8	4-Nitrophenol	**-0,160**
9	1-Chloro-2-nitrobenzene	**-0,230**
10	3-Chloronitrobenzene	**-0,390**
11	4-Chloronitrobenzene	**-0,430**
12	3-Nitrobenzoic acid	**-0,610**
13	4-Nitrobenzoic acid	**-1,070**
14	1-Chloromethyl-4-nitrobenzene	**-1,020**
15	1,3-Dimethyl-2-nitrobenzene	**-1,120**
16	1,4-Dimethyl-2-nitrobenzene	**-1,210**
17	1,4-Dichloro-2-nitrobenzene	**-1,320**
18	1,2-Dichloro-4-nitrobenzene	**-0,520**
19	1,3-Dinitrobenzene	**0,310**
20	2-Methyl-1,3-dinitrobenzene	**-0,140**
21	1-Methyl-2,4-Dinitrobenzene	**-0,170**
22	2,4-Dinitrophenol	**0,410**
23	1-Fluoro-2,4-dinitrobenzene	**0,570**
24	1,3,5-Trinitrobenzene	**-0,110**
25	1,2,4-Trichloro-5-nitrobenzene	**-0,670**
26	2-Methyl-4,6-dinitrophenol	**0,520**
27	2-Methyl-1,3,5-trinitrobenzene	**-0,490**
28	1,2,3,4,5-Pentachloro-6 nitrobenzene	**-0,570**

Calculation of molecular descriptors

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 poly-electronic 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 [16-18] using the B3LYP functional [19,20] and a 6-31G(d) 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.

The 3D structures of the molecules were generated using the Gauss View 3.0, and then, all calculations were performed using Gaussian 03W program series, Geometry optimization of twenty-eight compounds was carried out by B3LYP method employing 6–31G (d) basis set.

ChemSketch program (Demo version 10.0) [21] was employed to calculate the others molecular descriptors, Molar Volume MV (cm^3), Molecular Weight MW, Molar Refractivity MR (cm^3), Parachor Pc (cm^3), Density D (g/cm^3), Refractive Index n, Surface Tension γ(dyne/cm) and Polarizability α (cm^3) [22].

Statistical analysis

Principal Components Analysis (ACP)

The molecules of benzene, toluene, nitro-aromatic derivatives (1 to 28) were studied by statistical methods based on the principal component analysis (PCA) [21,22] using the software XLSTAT 2009.

This is an essentially a descriptive statistical method which aims to present, in graphic form, the maximum 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 Regressions (RLM)

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) [22] was generated using the software XLSTAT 2009, to predict antifungal activities $-\log LD_{50}$. It has served also to select the descriptors used as the input parameters for a back propagation network (ANN).

Artificial Neural Networks (ANNs)

The ANN analysis was performed with the use of Mathlab software v 2008a Neural Fitting tool (nftool) toolbox on a data set of nitrobenzene derivatives herbicide activity [23,24].

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 [25]. The input layer consisted of fifteen 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.

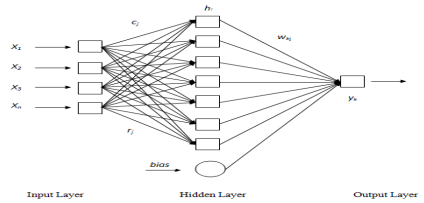

Figure 2: Neuron Layout of ANN

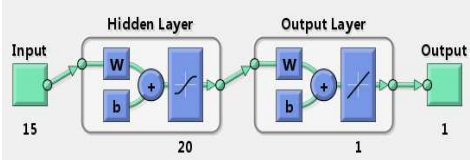

Figure 3: The ANN architecture

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 nitro-aromatics derivatives of herbicide activity [26] were subjected to the ANN analysis. First, for the learning set of compounds, i.e., 28 nitro-aromatic derivatives were used. ANN models were designed, built and trained. 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 four compounds was used. Testing set with four 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 [27]. The operation of the output layer is linear, which is given as below:

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

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

$$\text{Bias} = \sum(\bar{y} - y) \quad \textbf{(2)}$$

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:

$$\text{RMSE} = \sqrt{\frac{1}{n}.\sum_{i=1}^{n}(p_{exp} - p_{pred})^2} \quad \textbf{(3)}$$

where n = number of compounds, p_{exp} = experimental value, p_{pred} = predicted value and summation is of overall patterns in the analyzed data set [28,29]. The scripts were run on a personal PC.

RESULTS

This study was carried for a series of 28 our compounds: benzene, toluene and nitro-aromatic 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 and ACD/ChemSketch program of the studied nitro-aromatics.

Principal component analyses (PCA)

In this part, PCA was applied to select a training set from among 28 compounds studied.

The set of descriptors encoding the 28 nitro-aromatic compounds, electronic, energetic and topologic parameters are submitted to PCA analysis [30]. The first three principal axes are sufficient to describe the information provided by the data matrix. Indeed, the percentages of variance are 61.95%; 18.49% and 7.15% for the axes F1, F2 and F3, respectively. The total information is estimated to a percentage of 87.59%. The principal component analysis (PCA) [31] 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 fifteen descriptors are shown in table 3 as a correlation matrix and in figure 4 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 1: values of the calculated parameters obtained by DFT/B3LYP 6-31G* optimization and ACD/ChemSketch program of the studied nitro-aromatics

N°	-lgLD$_{50}$	MW	MR (cm^3)	MV (cm^3)	Pc (cm^3)	n	γ (dyne/ cm)	D (g/ cm^3)	A (cm^3)	E_T (ua)	E_{HOMO} (ev)	E_{LUMO} (ev)	ΔE (ev)	μ (Debay)	Ea (ev)	λ_{max} (nm)
1	-1,860	78,112	26,250	89,400	207,200	1,498	28,800	0,873	10,400	-6324,131	-6,706	0,099	6,806	0,0001	7,394	167,680
2	-1,780	92,138	31,070	105,700	244,900	1,499	28,800	0,871	12,320	-7394,760	-6,409	0,145	6,554	0,3195	7,233	171,400
3	-0,690	193,109	32,790	101,200	262,700	1,561	45,300	1,215	13,000	-11892,718	-7,597	-2,431	5,166	4,5800	4,790	258,840
4	-0,810	137,136	37,620	117,500	300,400	1,553	42,600	1,166	14,910	-12963,063	-7,060	-1,965	5,096	3,7131	4,762	260,340
5	-1,190	137,136	37,620	117,500	300,400	1,553	42,600	1,177	14,910	-12961,480	-7,369	-2,320	5,049	5,2069	4,789	258,920
6	-0,380	139,109	34,670	99,700	277,700	1,612	60,200	1,395	13,740	-13940,441	-6,666	-1,807	4,858	5,1141	4,809	257,830
7	-0,370	139,109	34,670	99,700	277,700	1,612	60,200	1,395	13,740	-13940,831	-6,784	-2,398	4,386	5,8264	4,805	258,060
8	-0,160	139,109	34,670	99,700	277,700	1,612	60,200	1,395	13,740	-13940,916	-6,925	-2,223	4,703	5,3388	4,795	258,560
9	-0,230	157,554	37,690	113,200	298,600	1,580	48,300	1,398	14,940	-24407,028	-7,306	-2,132	5,174	4,7097	4,800	258,310
10	-0,390	157,554	37,690	113,200	298,600	1,580	48,300	1,391	14,940	-24407,118	-7,352	-2,202	5,149	3,3962	4,791	258,800
11	-0,430	157,554	37,690	113,200	298,600	1,580	48,300	1,391	14,940	-24407,129	-7,227	-2,184	5,043	2,2289	4,796	258,530
12	-0,610	167,119	39,720	113,800	324,800	1,615	66,400	1,468	15,740	-17026,919	-7,887	-2,654	5,234	5,2759	5,031	246,470
13	-1,070	167,119	39,720	113,800	324,800	1,615	66,400	1,468	15,740	-17027,464	-7,878	-2,922	4,956	3,6701	4,791	258,810
14	-1,020	171,581	42,560	128,900	338,000	1,574	47,100	1,330	16,870	-25478,022	-7,761	-2,684	5,077	3,8672	4,792	258,750
15	-1,120	151,163	42,440	133,800	338,000	1,547	40,700	1,129	16,820	-14033,728	-6,958	-1,962	4,996	3,3753	4,735	261,860
16	-1,210	151,163	42,440	133,800	338,000	1,547	40,700	1,129	16,820	-14033,894	-6,991	-2,245	4,746	4,5946	4,761	260,410
17	-1,320	192,000	42,580	125,100	334,400	1,595	50,900	1,533	16,880	-36921,819	-7,364	-2,739	4,626	3,9591	4,799	258,380
18	-0,520	192,000	42,580	125,100	334,400	1,595	50,900	1,533	16,880	-36921,725	-7,359	-2,358	5,002	2,1203	4,792	258,750
19	0,310	168,107	39,340	113,100	318,200	1,612	62,600	1,486	15,590	-17460,273	-7,743	-2,260	5,482	3,8713	3,961	313,030
20	-0,140	182,134	44,160	129,300	355,800	1,598	57,200	1,407	17,500	-18531,435	-7,897	-2,852	5,045	2,9266	3,910	317,080
21	-0,170	182,134	44,160	129,300	355,300	1,598	57,200	1,407	17,500	-18531,699	-8,119	-2,979	5,140	4,8467	3,905	317,490
22	0,410	184,106	41,220	111,500	333,200	1,660	79,600	1,650	16,340	-19300,887	-6,780	-4,146	2,634	6,1586	3,190	388,710
23	0,570	186,097	39,330	117,300	325,300	1,585	59,100	1,586	15,590	-20162,437	-7,760	-2,646	5,114	3,2228	3,914	316,740
24	-0,110	213,105	45,880	124,900	373,700	1,655	80,000	1,705	18,190	-23028,484	-8,367	-2,954	5,413	0,0126	3,971	312,230
25	-0,670	226,445	47,480	137,100	370,300	1,609	53,200	1,651	18,820	-49436,417	-7,526	-2,903	4,623	2,7867	4,798	258,400
26	0,520	198,133	46,050	127,800	370,800	1,639	70,800	1,550	18,250	-20579,728	-7,478	-2,724	4,754	6,7207	3,816	324,940
27	-0,490	227,131	50,710	141,200	411,300	1,637	71,900	1,608	20,100	-24099,753	-8,485	-3,481	5,004	1,5349	3,922	316,130
28	-0,570	295,335	57,270	161,000	442,100	1,629	56,800	1,834	22,700	-74465,202	-7,612	-2,647	4,964	2,3915	3,908	317,260

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

	-loqgLD$_{50}$	MW	MR	MV	Pc	n	γ	D	α	E_T	E_{HOMO}	E_{LUMO}	ΔE	μ	E_a	λ_{max}
-logLD$_{50}$	**1**															
MW	0,427	**1**														
MR	0,311	0,891	**1**													
MV	0,083	0,774	0,944	**1**												
Pc	0,362	0,884	**0,994**	0,923	**1**											
n	0,724	0,683	0,605	0,314	0,643	**1**										
γ	**0,737**	0,550	0,483	0,185	0,547	0,959	**1**									
D	0,673	0,852	0,737	0,503	0,753	0,910	0,812	**1**								
α	0,311	0,892	**1,000**	**0,944**	**0,994**	0,605	0,483	0,737	**1**							
E_T	-0,116	-0,810	-0,748	-0,713	-0,688	-0,400	-0,178	-0,684	-0,748	**1**						
E_{HOMO}	-0,337	-0,643	-0,605	-0,511	-0,655	-0,495	-0,520	-0,566	-0,605	0,287	**1**					
E_{LUMO}	-0,607	-0,710	-0,663	-0,473	-0,704	-0,819	-0,777	-0,791	-0,663	0,353	0,596	**1**				
ΔE	-0,500	-0,396	-0,367	-0,202	-0,380	-0,645	-0,573	-0,556	-0,367	0,222	-0,016	0,793	**1**			
μ	0,408	-0,001	-0,025	-0,129	-0,005	0,320	0,291	0,147	-0,025	0,158	0,141	-0,429	-0,642	**1**		
E_a	-0,788	-0,702	-0,676	-0,499	-0,718	-0,795	-0,753	-0,775	-0,676	0,330	0,542	0,902	0,713	-0,439	**1**	
λ_{max}	**0,801**	0,661	0,642	0,456	0,686	0,780	0,774	0,741	0,642	-0,287	-0,478	-0,846	-0,691	0,356	**-0,955**	**1**

Correlation circle

The principal component analysis (PCA) was also performed to detect the connection between the different variables. The principal component analysis revealed the correlation circle (Figure 4) shows that the F1 axis (61.95% of the variance) appears to represent the density **D**, and the F2 axis (18.49% of the variance) seems to represent the dipole moment **μ**.

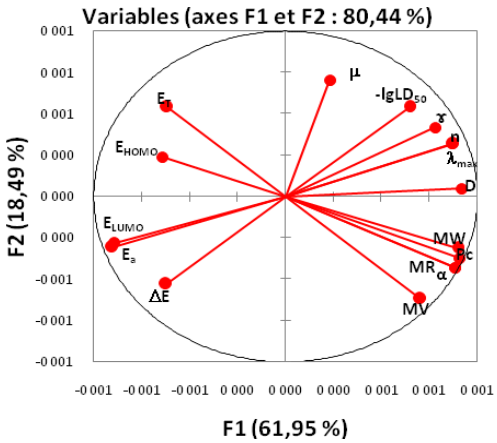

Figure 4: Correlation circle

Analysis of projections according to the planes F1–F2 and F1-F3 (80.44% and 69.10% of the total variance respectively) of the studied molecules (Figure 5) is showing in figure 5:

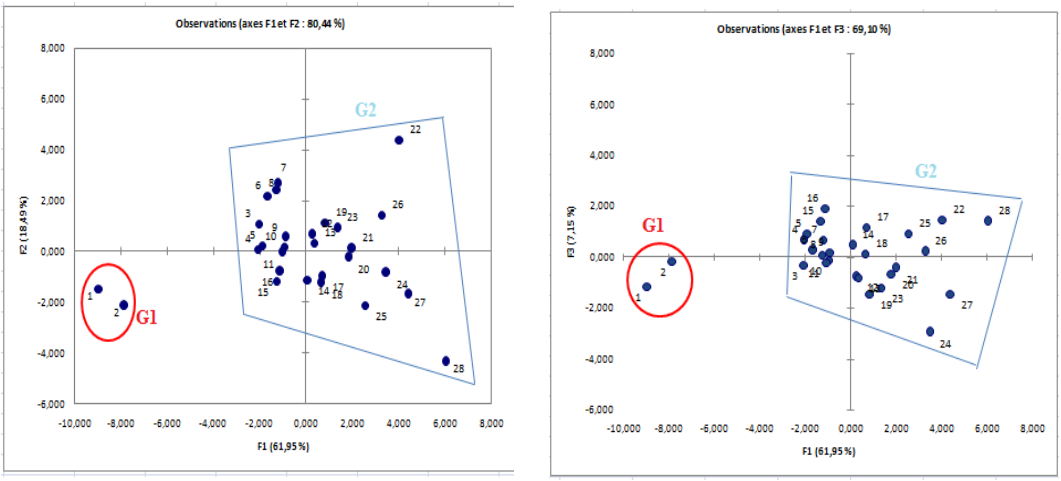

Figure 5: Cartesian diagram according to F1-F2 and F1-F3

Multiple linear regressions (MLR)

To establish quantitative relationships between toxicity **-logLD$_{50}$** and selected descriptors, our array data were subjected to a multiple linear and nonlinear regression. 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 -logLD$_{50}$, but the best relationship obtained by this method is only one corresponding to the linear combination of several descriptors: the molecular weight **MW**, the refractive index **n**, the surface tension **γ**, the density **D**, the total energy $\mathbf{E_T}$, the energy $\mathbf{E_{LUMO}}$, the dipole moment **μ** and the absorption maximum $\boldsymbol{\lambda_{max}}$.

The resulting equation is:

$$-\ \mathbf{logLD_{50}} = -\ 22.858 + 2.802.10^{-03}\times\mathbf{MW} + 10.887\times\mathbf{n} - 7.151.10^{-02}\times\boldsymbol{\gamma} + 5.755\times\mathbf{D} + 6.274.10^{-05}\times\mathbf{E_T} + 0.622\times\mathbf{E_{LUMO}} + 6.215.10^{-02}\times\boldsymbol{\mu} + 1.091.10^{-02}\times\boldsymbol{\lambda_{max}} \qquad \text{(Equation 4)}$$

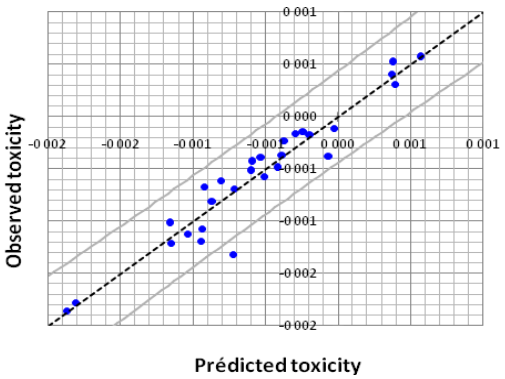

Figure 6: Graphical representation of calculated and observed toxicity by MLR

For our 28 compounds, the correlation between experimental toxicity and calculated one based on this model are quite significant (Figure 6) as indicated by statistical values:

N = 28 R = 0.961 R^2 = 0.923 RMSE = 0.207

The figure 6 shows a very regular distribution of toxicity values depending on the experimental values.

Multiple nonlinear regression of the variable toxicity (MNLR)

We have used also the technique of nonlinear regression model to improve the structure–toxicity in a quantitative way. It takes into account several parameters. This is the most common tool for the study of multidimensional data. We have applied it to table 2 containing 28 molecules associated with fifteen variables.

The resulting equation is:

$$-\mathbf{logLD_{50}} = 11635{,}317 + 7{,}144 \times \mathbf{MW} - 70{,}084 \times \mathbf{MV} - 50{,}952 \times \mathbf{Pc} - 117352{,}688 \times \mathbf{n} + 17{,}878 \times \boldsymbol{\gamma} - 2084{,}566 \times \mathbf{D} + 1352{,}531 \times \boldsymbol{\alpha} - 1{,}396.10^{-02} \times \mathbf{E_T} - 271{,}157 \times \mathbf{E_{HOMO}} + 210{,}306 \times \mathbf{E_{LUMO}} - 21{,}281 \times \boldsymbol{\mu} + 11793{,}651 \times \mathbf{E_a} + 204{,}634 \times \boldsymbol{\lambda_{max}} - 2{,}399.10^{-02} \times \mathbf{MW^2} + 0{,}231 \times \mathbf{MV^2} + 0{,}113 \times \mathbf{Pc^2} + 37295{,}048 \times \mathbf{n^2} - 0{,}203 \times \boldsymbol{\gamma}^2 + 377{,}229 \times \mathbf{D^2} - 53{,}756 \times \boldsymbol{\alpha}^2 + 1{,}7316.10^{-07} \times \mathbf{E_T}^2 - 12{,}787 \times \mathbf{E_{HOMO}}^2 + 29{,}867 \times \mathbf{E_{LUMO}}^2 - 6{,}179 \times \boldsymbol{\Delta E}^2 + 3{,}381 \times \boldsymbol{\mu}^2 - 626{,}155 \times \mathbf{E_a}^2 - 0{,}186 \times \boldsymbol{\lambda_{max}}^2$$ (Equation 5)

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

N = 28 R = 0.999 R^2 = 0.999

The toxicity value – $logLD_{50}$ predicted by this model is somewhat similar to that observed. The figure 7 shows a very regular distribution of toxicity values based on the observed values.

With MLNR was obtained significantly better correlation coefficient R = 0, 999. Figure 7 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 –$logLD_{50}$. The residual values tended to zero which is why we did not graph for prediction residuals.

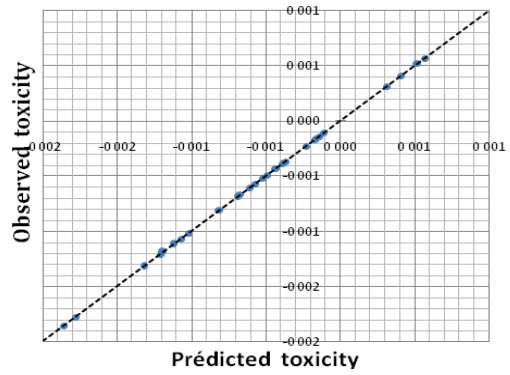

Figure 7: Graphical representation of calculated and observed toxicity

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.

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

Table 4: Values obtained by ANN

ANN	Samples	MSE	R	R^2
Training	20	0.002	0.997	0.994
Validation	4	0.625	0.963	0.926
Test	4	0.092	0.995	0.991

N = 28 R = 0.997 R^2 = 0.994 RMSE = 0.002

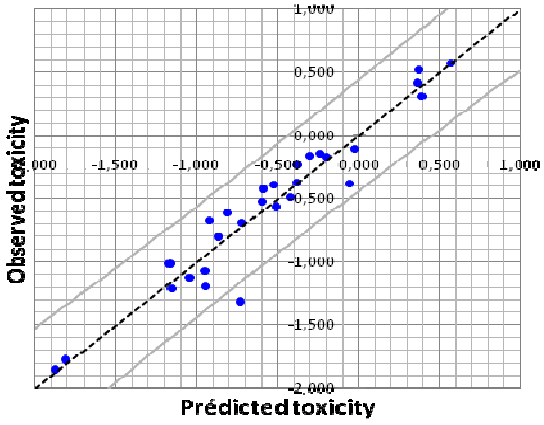

Figure 8: Graphical representation of calculated and observed toxicity -logLD_{50}

DISCUSSION

Principal component analysis

* The toxicity is well correlated with the maximum of absorption λ_{max} (r= 0.801 and p <0.05) and the surface tension γ (r=0.737 and p<0.05) at a significant level.
* The polarizability α is positively correlated with the parachor Pc (r =0.994 and p < 0.05) and the molar volume MV (r= 0.944 and p <0.05) at a significant level.

* The energy of activation E_a is negatively correlated with the maximum of absorption λ_{max} (r= 0.995 and p <0.05) at a significant level.

* The polarizability α is strongly correlated with the molar refractivity MR (r =1 and p< 0.001) at a high level.

Analysis of projections according to the planes F1–F2 and F1-F3 (80.44% and 69.10% of the total variance respectively) of the studied molecules (Figure 5) shows that the molecules are dispersed, according to the structure of the R group of benzene, in two classes of compounds belonging to two groups: The group 1 don't containing a nitro group (The low toxicity -logLD_{50}<-1.76) and group 2 containing a nitrogen (belongs to nitro group), oxygen (belongs to nitro group, hydroxyl and carboxyl), fluorine and chlorine atoms promote toxicity increase (–logLD_{50}>-1.31).

Statistical Analysis

The obtained multiple nonlinear regression correlation coefficient R value is 0.999 for this data set of nitrobenzene derivatives. 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, MNLR 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.

We have investigated the QSAR regression to predict the toxicity (LD_{50} values) of nitro-aromatic 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.

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

N°	Obs (-logLD_{50})	Pred (-logLD_{50})		
		MLR	MNLR	ANN
1	-1,860	-1,871	-1,860	-2,43
2	-1,780	-1,811	-1,780	-2,51
3	-0,690	-0,719	-0,690	-0,64
4	-0,810	-0,866	-0,810	-0,83
5	-1,190	-0,947	-1,190	-1,13
6	-0,380	-0,063	-0,380	-0,42
7	-0,370	-0,384	-0,370	-0,44
8	-0,160	-0,299	-0,160	-0,24
9	-0,230	-0,370	-0,230	-0,30
10	-0,390	-0,530	-0,390	-0,46
11	-0,430	-0,594	-0,430	-0,25
12	-0,610	-0,809	-0,610	-0,72
13	-1,070	-0,941	-1,070	-1,08
14	-1,020	-1,160	-1,020	-2,12
15	-1,120	-1,039	-1,120	-1,10
16	-1,210	-1,156	-1,210	-1,19
17	-1,320	-0,727	-1,320	-1,38
18	-0,520	-0,601	-0,520	-0,57
19	0,310	0,393	0,310	0,31
20	-0,140	-0,239	-0,140	-0,56
21	-0,170	-0,194	-0,170	-0,93
22	0,410	0,367	0,410	0,42
23	0,570	0,566	0,570	0,56
24	-0,110	-0,026	-0,110	-0,11
25	-0,670	-0,924	-0,670	-0,78
26	0,520	0,376	0,520	0,50
27	-0,490	-0,420	-0,490	-0,50
28	-0,570	-0,511	-0,570	-0,57

CONCLUSION

In this work, the study of the quality of the MLR, MNLR and ANN models shows 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 toxicity in satisfactory manners.

We can conclude that one 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] M H Keshavarz; M Jaafari, *Propellants, Explos. Pyrotechn*., **2006**, 31, 216.
[2] J V Dilley; C A Tyson; R J Spanggord; D P Sasmore; B W Nexwell; J C Dacre, *J. Toxicol. Environ*., **1982**, 9, 587–610.
[3] L S Djerassi; L Vitany, *Br. J. Ind. Med*., **1975**,32, 54–58.
[4] C Hansch; P P Maloney; T Fujita; R M Muir, *Nature*, **1962**,194, 178–180.
[5] B Escher; R P Schwarzenbach, *Aquat. Sci*., **2002**, 64, 20–35.
[6] W K Agrawal; P V Khadikar, *Bioorg. Med. Chem*., **2001**, 9, 3035–3040.
[7] M T D Cronin; B W Gregory; T W Schultz, *Chem. Res. Toxicol*., **1998**, 11, 902–908.
[8] O Isayev; B Rasulev; L Gorb; J Leszczynski, *Mol. Diver*, **2006**, 10, 233–245.

[9]V E Kuz'min; E N Muratov; A G Artemenko; L G Gorb; M Qasim; J Leszczynski, *J. Comp. Aid. Mol. Des*., **2008**, 6, 9179.
[10] K Laarej ; M Bouachrine; S Radi; S Kertit; B Hammouti, *E-J. Chem*, **2010**, 7 (2), 419–424.
[11] S Chtita; M Larif; M Ghamali; A Adad; R Hmamouchi; M Bouachrine; T Lakhlifi, *Int. J. Inn. Res. Sci., Engin. Technol.,* **2013**, 2 (11), 6586-6601.
[12] V E Kuz'min; E N Muratov; A G Artemenko; L Gorb; M Qasim; J Leszczynski, *J. Comp. Aid. Mol. Des*., **2008**, 72, 1373–1380.
[13] H R Pouretedal; M H Keshavarz, *Propellants, Explos. Pyrotechn*., **2010**, 8 (1), 78-89.
[14] C Adamo; V Barone, *Chem. Phys. Lett.*, **2000**, 330, 152–160.
[15] M Parac; S Grimme, *J. Phys. Chem.*, **2003**, A 106, 6844–6850.
[16] Gaussian 03, Revision B.01, M. J. Frisch, M. J. and al., Gaussian, Inc., Pittsburgh, PA, **2003**.
[17] A D Becke, *J. Chem. Phys.*, **1993**,98, 1372.
[18] C Lee; W Yang; R G Parr, *Phys. Rev*., **1988**, B 37, 785-789.
[19] Advanced Chemistry Development Ino., Canada,www.acdlabs.com/resources/freeware/chemsketch/, **2009**.
[20] S Chtita; M Larif; M Ghamali and all, *IJIRSET*, **2013**, 2 (11), 6586-6601.
[21] J N Hogarh; N Seike; Y Kobara; A Habib; J J Namd; J S Lee; Qilu Li; X Liu; Jun Li; G Zhang; S Masunaga, *Chemosphere*, **2012**, 86, 718–726.
[22] M Ghamali; S Chtita; R Hmamouchi; A Addad; M Bouachrine; T Lakhlifi, *IJARCSSE*, **2014**, 4 (1), 536–546.
[23] A Adad; R Hmamouchi; A I Taghki; A Abdellaoui; M Bouachrine; T Lakhlifi, *J. Chem. Pharm. Res.*, **2013**, 5 (7), 28-41.
[24] D Zakarya; A Boulaamail; E M Larfaoui; T Lakhlifi, *SAR QSAR Environ. Res.*, **1997**, 6, 183–203.
[25] J Zupan; J Gasteiger, *Neural Networks for Chemistry and Drug Design*, *second ed. VCH*, *Weinheim*, **1999**.
[26] R Chimizou; H Iwamura; T Fujita, *Agric. Food Chem.*, **1988**, 36, 1276.
[27] N Turkkan, *Revue de l'Université de Moncton*, **1993**, 26 (1), 205–221.
[28] G Jing; Z Zhou; J Zhuo, *Chemosphere*, **2012**, 86, 76-82.
[29] P Y Lee; C Y Chen, *J Hazard. Mater.*, **2009**, 165, 156-161.
[30] STATITCF Software, *Technical Institute of cereals and fodder*, Paris, France, **1987**.
[31] 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.