Available online at www.derpharmachemica.com

ISSN 0975-413X
CODEN (USA): PCHHAX

Der Pharma Chemica, 2016, 8(17):200-212
(http://derpharmachemica.com/archive.html)

Cytotoxicity of compounds based on indole-2-carboxylate: DFT and QSAR study

H. Zaki[1*], Y.Boukarai[2]. M. Benlyas[1], Y. Filali Zegzouti[1], F. Khalil[2] and M. Bouachrine[3*]

[1]*Equipe de Recherche Biologie, Environnement& Santé, Faculty of Science and Technics, Université Moulay Ismail, Meknes, Morocco*
[2]*LAC, Laboratory of Applied Chemistry, Department of Chemistry, Faculty of Science and Technology, University Sidi Mohammed Ben Abdellah, Fez, Morocco*
[3]*Equipe de Recherche Matériaux, Environnement & Modélisation. ESTM, Université Moulay Ismail, Meknes, Morocco*

ABSTRACT

The compounds basedindole-2-carboxylate and derivatives are known for their interesting biological properties such as antiviral and anticancer activities and they have been prepared and extensively studied. In this work we attempt to establish a quantitative structure-activity relationship for cytotoxicity by studying a series of 22 substitutedindole-2-carboxylate and derivatives. We accordingly propose a quantitative model, and we try to interpret the activity of the compounds relying on the multivariate statistical analyses. PCA was served to describe data ; The MLR has served to select the descriptors used as the input parameters the ANN. This method MRA have served also to predict activities, but when compared with the results given by the ANN, were alized that the predictions fulfilled by this latter were more effective. The DFT-B3LYP method, with the basis set 6-31G (d), is employed to calculate some quantum chemical descriptors of the 22 substituted indole-2-carboxylate using Gaussian 03W program, the topological descriptors were computed with chemoffice program.

Keywords: Cytotoxicity, indole-2-carboxylate, DFT, QSAR

INTRODUCTION

Scientific and industrial communities accorded a special attention to indole and derivatives, as there are a major constituent of a large number of compounds occurring in nature. Furthermore, indole derivatives have many applications, being frequently used in the design for the synthesis of new drugs [1-2], also to fix fragrances in the cosmetic industry [3-4]. Another important application is the production of the amino acid tryptophan that is an essential amino acid in the human diet [3]. particularly , 1H-indole-2-carboxylate and 1H-indole-3-carboxylate derivatives have an important role in the pharmaceutical industry due to the important activity in the treatment of various diseases, namely in HIV-1 [5] and hepatitis B [6], where their antiviral action is crucial [7]. Several studies have reported that these compounds also show an anti-proliferative activity oncancer cells [8,9]. Due to their wide applications in industrial and pharmaceutical processes, theknowledge of relationship between structure and the cytotoxicity of 1H-indole -2-carboxylate derivatives is essential to predict the cytotoxicity necessary to identify their harmful effects on humans and It is also one of the main steps in drug design and developments of chemotherapy.

By the application of QSAR study the time and the price of drug discovery can be reduced greatly. The aim of the present study is to develop a significant QSAR model and propose pharmacophore hypothesis to develop the activities of the molecules.

The computational analysis and the statistical study by using the Principal component analysis (PCA), Partial least squares (PLS), Multiple Non-Linear regression(MNLR) and multiplelinear regression (MLR) were applied to a series of 1H-indole-2-carboxylate and derivatives to develop a QSAR model and predict the activities.

MATERIALS AND METHODS

1-Data set

The experimental cytotoxicity TC50 of a series of indole-2-carboxylate Compounds are collected from the work of Xue et al [7]. The TC50 values in units of molarity (M).The observations are shown in **table1**. The studied series of molecules are composed of 22 derivatives of indole-2-carboxylate **(table 1),** have been studied and analyzed in order to find relationship between theircytotoxicity and 3D structure of molecules.

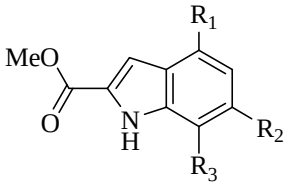
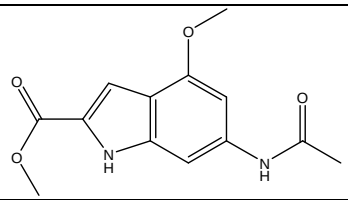
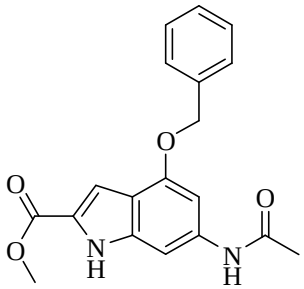
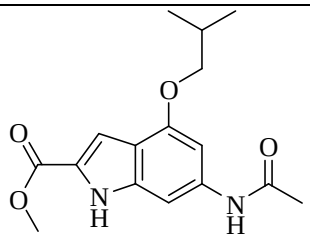
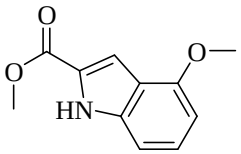
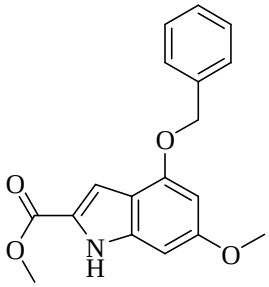
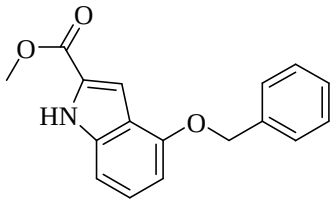
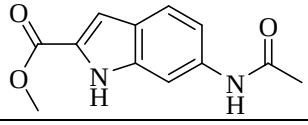
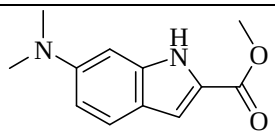
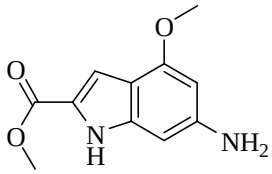
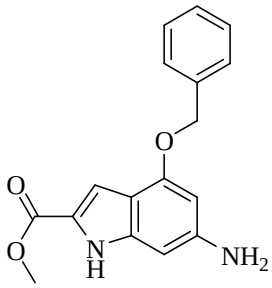
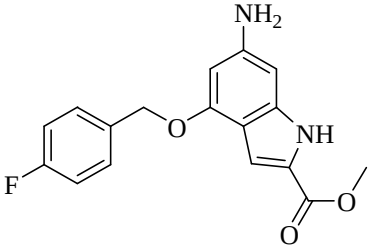
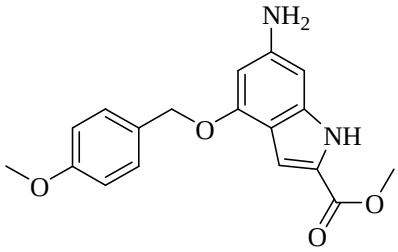
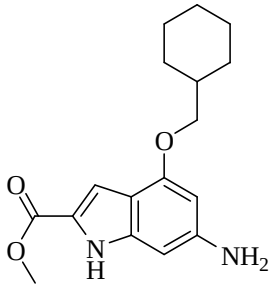
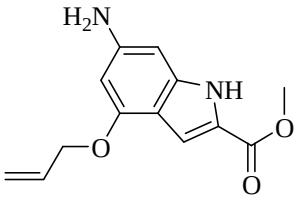
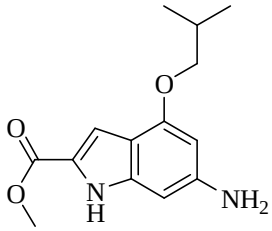
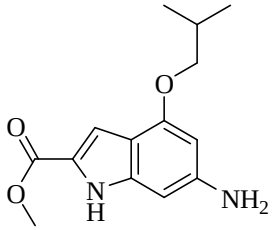
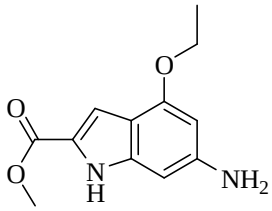
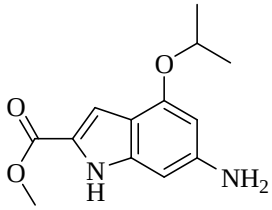
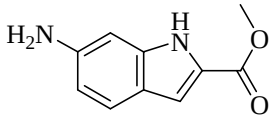
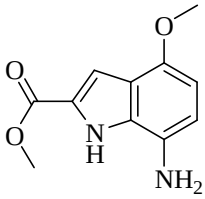
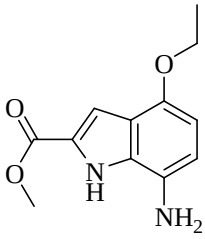


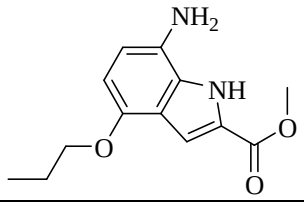
Fig. 1: Studied compounds (Table 1)

Table 1: Chemical structure and observed activities of studied compounds

Molecule	Structure	**TC50 $_1$**	**TC50 $_2$**	**TC50 $_3$**
M1		**763,36**	**763,36**	**763,36**
M2		**45,59**	**45,59**	**341,63**
M3		**50,69**	**50,69**	**24,38**
M4		**225,46**	**225,46**	**225,46**
M5		**49,55**	**49,55**	**92,06**

M6		**54,84**	**54,84**	**136,98**
M7		**597,71**	**597,71**	**414,44**
M8		**212,02**	**212,02**	**711,97**
M9		**174,95**	**174,95**	**174,95**
M10		**14,51**	**14,51**	**108,64**
M11		**16,37**	**16,37**	**273,57**
M12		**52,91**	**52,91**	**354,2**
M13		**11,75**	**11,75**	**14,13**

M14		**210,33**	**210,33**	**130,28**
M15		**177,77**	**177,77**	**177,77**
M16		**122,33**	**122,33**	**40,76**
M17		**94,96**	**94,96**	**65,85**
M18		**625,85**	**625,85**	**465,6**
M19		**243,26**	**243,26**	**90,79**
M20		**145,68**	**145,68**	**11,23**
M21		**18,29**	**31,67**	**31,67**

M22		**115,44**	**115,44**	**5,77**

2- Computational details

DFT (Density Functional Theory) methods were used in this study. These methods have become very popular in recent years because they can reach exactitude similar to other methods in less time and are less expensive 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 [10,11]. All calculations were done by GAUSSIAN 03 W software [12] using the B3LYP functional [13] and a 6-31G* basis set [14]. 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. Frontier molecular orbital's (HOMO and LUMO), the absolute electronegativity, the absolute hardness, the softness, Gap Energy (ΔE), Total Energy (TE) and Dipole moment (μ) were calculated from the DFT optimized structures for each molecule. The topological descriptors,were computed with chemoffice program.

3- Statistical methods

Principal Components Analysis (PCA)

Compounds based on indole-2-carboxylate c (1 to 22) were studied by statistical methods based on the principal component analysis (PCA) [15] using the software XLSTAT version 2014. This is an important descriptive statistical method which aims present, in graphic form, large information contained in a data **table 2.** 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 (MLR)

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 2014, to predict IC50. It has served also to select the descriptors used as the input parameters for a back propagation network (ANN).

Artificial Neural Networks (ANNs)

The ANNs analysis was performed with the use of Matlab software version 2009a Neural Fitting tool (nftool) toolbox on a data set of indole-2-carboxylate compounds [16].

RESULTS AND DISCUSSION

1- Data set for analysis

A QSAR study was performed of 22 indole-2-carboxylate derivativesas reported previously [X Y], to determine a quantitative relationship between the structure and cytotoxicity. The values of the 15 descriptors are shown in **Table 2.**

Table 2: The calculated quantum chemical parameters and descriptors of the studied molecules

	TorE	*EE*	*log P*	*MW*	*Kow*	*RE*	*EHomo*	*ELumo*	*ΔE*	μ	*TotE*	η	σ	χ	*ω*
M1	0.125	-21297.5	0.23	262.267	1.63	17745.9	-5.486	-1.305	-4.181	3.66	-914.260	-2.090	-0.478	-3.395	-12.055
M2	0.142	-30264.1	1.96	338.36	3.40	25889.7	-4.984	-0.915	-4.069	2.64	-953.346	-2.034	-0.491	-2.949	-8.852
M3	0.127	-26530.8	1.45	304.34	3.09	22516.4	-5.427	-1.268	-4.158	3.99	-1032.2	-2.079	-0.480	-3.347	-11.653
M4	3.51E-01	-14339	1.32	205.21	2.52	11616.1	-5.359	-1.062	-4.296	1.67	-706.243	-2.148	-0.465	-3.210	-11.073
M5	0.130	-26837.4	2.93	311.33	4.25	22812.2	-7.498	3.000	-10.498	1.17	-1006.42	-5.249	-0.190	-2.248	-13.273
M6	0.125	-22658.5	3.05	281.31	4.28	19109.2	-5.427	-0.878	-4.548	1.36	-897.988	-2.274	-0.439	-3.152	-11.303
M7	0.122	-17261.3	0.35	232.24	1.63	14185.5	-5.755	-1.499	-4.255	4.04	-799.734	-2.127	-0.469	-3.627	-14.002
M8	0.215	-16180	1.73	218.25	2.74	13398.3	-5.012	-1.041	-3.970	2.42	-725.690	-1.985	-0.503	-3.026	-9.093
M9	0.188	-14830.3	0.25	206.20	0.77	12037.9	-4.884	-0.790	-4.094	3.22	-761.578	-2.047	-0.488	-2.837	-8.242
M10	0.213	-24696.9	2.25	296.32	3.17	20926.6	-4.909	-0.816	-4.093	3.35	-953.346	-2.046	-0.488	-2.862	-8.388
M11	0.207	-26782.7	2.41	314.31	3.31	22540.9	-5.030	-0.947	-4.083	2.03	-1052.57	-2.041	-0.489	-2.989	-9.121
M12	0.240	-28487.7	2.12	326.35	3.09	24241.5	-4.901	-0.821	-4.080	3.94	-1067.84	-2.040	-0.490	-2.861	-8.350
M13	0.190	-27056.5	2.49	302.37	4.05	23201.7	-4.855	-0.770	-4.084	3.18	-996.262	-2.042	-0.489	-2.813	-8.081
M14	0.191	-19175	1.21	246.26	2.17	15944.5	-4.887	-0.795	-4.092	3.19	-838.998	-2.046	-0.488	-2.841	-8.260

M15	0.204	-19404.4	0.81	246.26	1.75	16174.4	-4.879	-0.792	-4.087	3.14	-878.301	-2.043	-0.489	-2.836	-8.219
M16	0.190	-21813.8	1.74	262.31	2.86	18398.9	-4.865	-0.782	-4.082	3.15	-879.5	-2.041	-0.489	-2.823	-8.138
M17	0.195	-18148.3	0.85	234.25	1.93	15044	-4.861	-0.775	-4.085	3.23	-800.923	-2.042	-0.489	-2.818	-8.112
M18	0.487	-20346.2	1.17	248.28	2.24	17087.2	-4.842	-0.763	-4.079	3.31	-840.243	-2.039	-0.490	-2.802	-8.011
M19	0.199	-12967.7	0.64	190.20	1.39	10495.9	-5.084	-0.983	-4.100	1.98	-647.075	-2.050	-0.487	-3.034	-9.436
M20	0.253	-16628.9	0.52	220.22	1.40	13681.3	-4.898	-0.959	-3.939	1.95	-761.597	-1.969	-0.507	-2.928	-8.447
M21	0.478	-18274	0.85	234.25	1.93	15170.6	-4.875	-0.945	-3.929	1.95	-800.919	-1.964	-0.508	-2.910	-8.321
M22	0.105	-19840.8	1.34	248.28	2.46	16581.6	-4.867	-0.940	-3.927	1.97	-840.235	-1.963	-0.509	-2.904	-8.281

2- Principal component analysis PCA

The total of the 15 descriptors describing the 22 molecules was submitted to principalcomponents analysis (PCA) [17]. The first two principal axes are sufficient to describe the information provided by the data matrix. F1= 48.93%, F2=25.65% and the total information is estimated on 74.58%.

The principal component analysis (PCA) [18] was cared to have an idea on the relationship between the various descriptors and between various observations. The correlations between the 15descriptors are shown in **Table 3**

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

Variables	TE	EE	log P	MW	MTI	Kow	RE	E_{Homo}	E_{Lumo}	ΔE	μ	TE	η	σ	χ	ω
TE	**1**															
EE	0.272	**1**														
log P	-0.176	-0.702	**1**													
MW	-0.296	**-0.996**	0.717	**1**												
MTI	-0.304	**-0.973**	0.757	**0.983**	**1**											
Kow	-0.224	-0.762	**0.966**	0.770	0.787	**1**										
RE	-0.270	**-1.000**	0.706	**0.995**	**0.972**	0.767	**1**									
E_{Homo}	0.314	0.198	-0.299	-0.222	-0.238	-0.363	-0.195	**1**								
E_{Lumo}	-0.111	-0.281	0.454	0.284	0.302	0.422	0.283	-0.783	**1**							
ΔE	0.204	0.262	-0.414	-0.274	-0.292	-0.421	-0.262	0.920	-0.964	**1**						
μ	-0.106	-0.166	-0.344	0.147	0.081	-0.268	0.165	0.317	-0.432	0.408	**1**					
TE	0.286	**0.948**	-0.625	**-0.950**	**-0.900**	-0.675	**-0.945**	0.222	-0.271	0.265	-0.256	**1**				
η	0.204	0.262	-0.414	-0.274	-0.292	-0.421	-0.262	**0.920**	**-0.964**	**1.000**	0.408	0.265	**1**			
σ	-0.236	-0.264	0.437	0.282	0.304	0.451	0.264	**-0.942**	**0.934**	**-0.990**	-0.396	-0.273	**-0.990**	**1**		
χ	0.160	-0.234	0.400	0.211	0.224	0.280	0.240	-0.170	0.746	-0.543	-0.345	-0.191	-0.543	0.471	**1**	
ω	0.383	0.044	-0.048	-0.081	-0.089	-0.176	-0.038	0.807	-0.268	0.513	0.072	0.094	0.513	-0.582	0.440	**1**

The obtained matrix furnishes information on the high or low interrelationship between thevariables. In general good co-linearity (r>0.5) was observed between most of the variables. A perfect interrelationship was observed between RE and EE (r = -1.00) and between η and ΔE (r = 1.00). And a low interrelationship was observed between μ and MTI (r = 0.081)

The correlations between the 22 Observations are shown in **Figure 2**

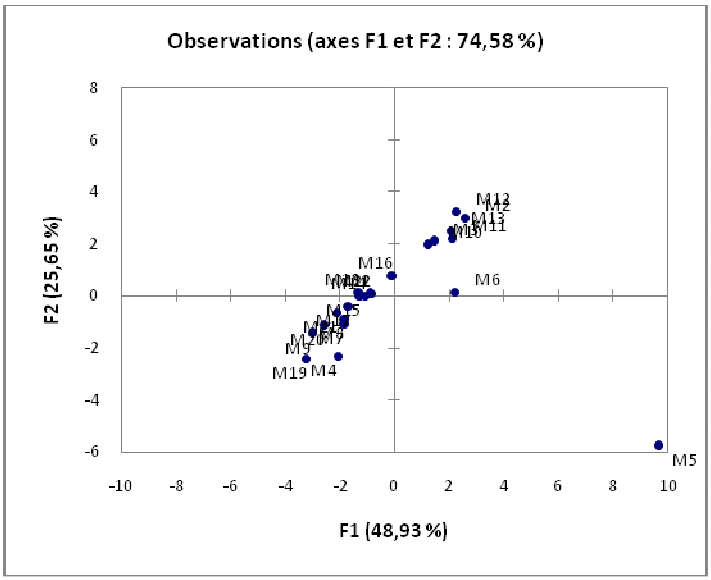


Figure 2: Cartesian diagram according to F1 and F2: Separation between two regions

A plot of observations shows that we can discern two groups of molecules: The group 1 containing the compounds with IC50 ≤ 100 and the group 2 is containing the compounds with IC50 > 100. In this representation, on the one hand the compounds M17 and M21 that should be in group 1 (low value of pIC50), present an exception. On the other hand, the compounds M6 and M5 show different behavior of two groups despite that they have value ofIC50 ≤ 100.

3- Multiple linear regressions MLR

To get an idea about the relationship between different descriptors and cytotoxicity of the molecule we have proposed three models based on the MLR. The relationship obtained in the first model ($TC50_1$) by this method is one corresponding to the linear combination of several descriptors selected: Molecular topological index **(MTI),** Molecular weight **(MW),** Energy **HOMO**, Energy **LUMO**, Dipole moment (**μ),** Softness **(σ**) and the electrophilicity index **(ω).** In the second model ($TC50_2$) the linear relationship is described by several other descriptors : Torsion Energy**(TE)** , Molecular topological index **(MTI),** Energy HOMO, Dipole moment (**μ**) and the Softness (σ) and in the last model the relationship obtained is described by :Torsion energy (***TE),*** Molecular topological index (***MTI)***, Electronic energy (***EE)***, ***log P***, Repulsion Energy (***RE),*** The electrophilicity index (**ω),** Softness **(σ),** Dipole moment (**μ**), $\mathbf{E}_{LUMO,}$ and $\mathbf{E}_{HOMO.}$

We note that there are common descriptors between the three repetition (models). MTI, $\mathbf{E}_{HOMO.,}$**(μ),**and **(σ).**

The resulting equations are:

MODEL 1

$\mathbf{TC50}_{\mathbf{1}}$ = -12158.827-0.106****MTI*** +5.187****MW*** -4837.310*$\mathbf{E}_{\mathbf{HOMO}}$ -2829.358*$\mathbf{E}_{LUMO}$+121.386***μ** +12366.099***σ** +1055.272***ω**

$$\mathbf{N = 16 \quad R^2 = 0.960 \quad R^2_{ajusté} = 0.925}$$

$$\mathbf{N_{test} = 6 \quad R_{test} = 0,178}$$

The cytotoxicity values of the studied molecules in the **model 1**increasewithincreasing MW, **μ**, **σ** and **ω**and decreasing MTI, E_{HOMO} and E_{LUMO}.

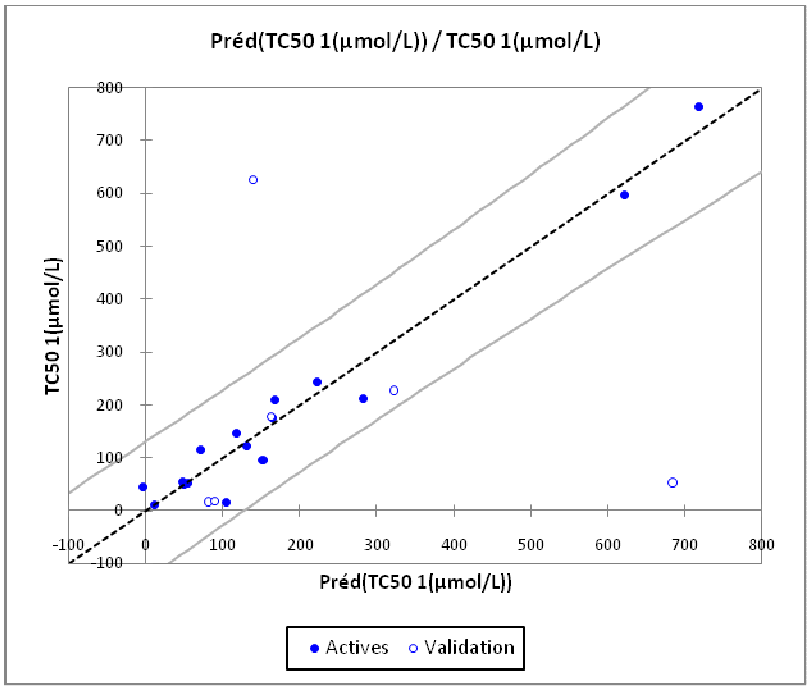


Figure 3: Correlations of observed and predicted activities $TC_{50(1)}$ calculated using MLR

MODEL 2

$TC50_2$ = -8935.541-502.587* ***TE*** -2.690E-02****MTI*** -1018.445*E_{HOMO}+118.678* **µ** -8032.220* **σ**

N = 16 R^2 = 0.956 $R^2_{ajusté}$= 0.933

N_{test}= 6 R_{test} = 0,239

The cytotoxicity values of the studied molecules in the **model 2** increase with increasing **µ**, and decreasing MTI, E_{HOMO}, TE and the **σ**.

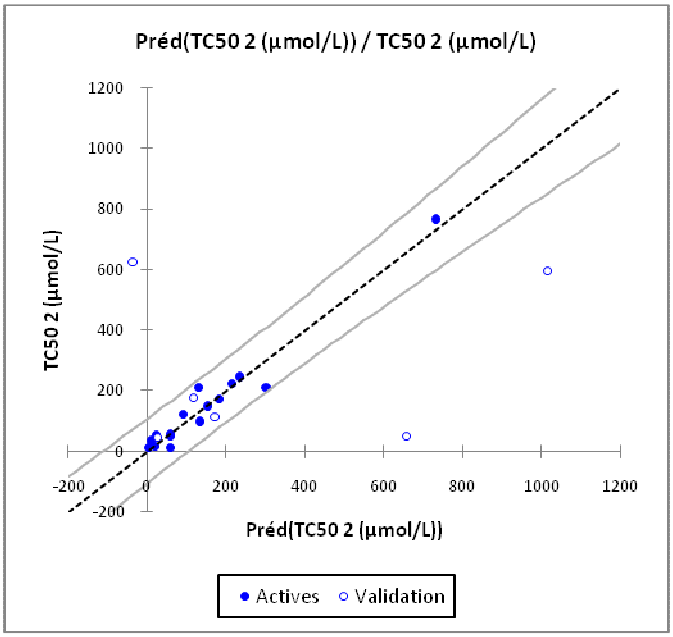


Figure 4: Correlations of observed and predicted activities $TC_{50(2)}$ calculated using MLR

MODEL 3

$TC50_3$ = 11316.571+1313.708* ***TE*** +8.310E-02*$\boldsymbol{MTI}$ -0.416*$\boldsymbol{EE}$ -216.426*$\boldsymbol{log\ P}$-0.469*$\boldsymbol{RE}$ +5613.389*$\mathbf{E_{HOMO}}$+3204.650*$\mathbf{E_{LUMO}}$-90.950* **μ** -14521.789* **σ** -1345.023* **ω**

$$\mathbf{N = 16 R^2 = 0.979 R^2_{ajusté} = 0.937}$$

$$\mathbf{N_{test} = 6 R_{test} = 0,47}$$

The cytotoxicity values of the studied molecules in the **model 3**increase with increasing ***TE***, ***MTI,*** $\mathbf{E_{HOMO}}$, $\mathbf{E_{LUMO}}$, and decreasing **μ**, **σ**, ***EE, log P*** , ***RE*** and **ω**.

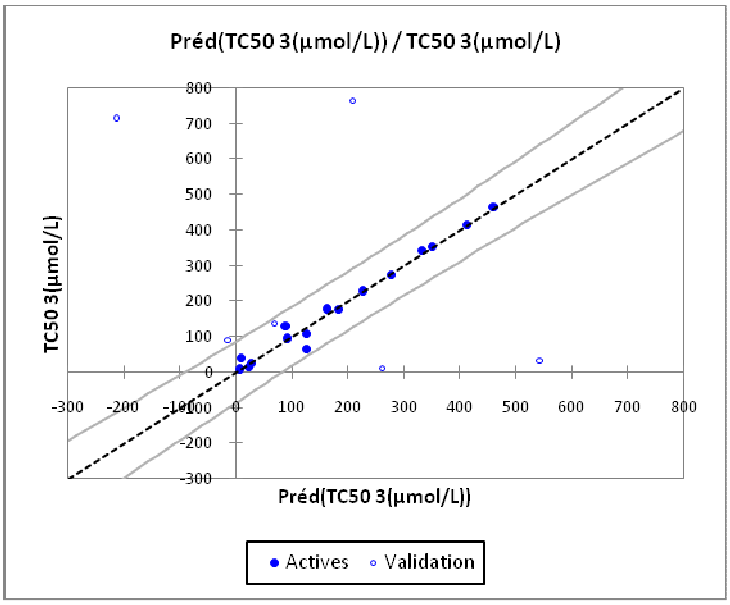


Figure 5: Correlations of observed and predicted activities $TC_{50(3)}$ calculated using MLR

The three models (1 ; 2 ; and 3) show a very significant correlation coeficient R^2 =**0.960** ; R^2 = **0.956**; R^2 = **0.979**respectively, with different combinations of molecules by choosing N = 16 observation and 6 observations to in tern testing of the model, the 6 molecules for testing were selected randomly and change in moving from one model to another, which explains the differences of these three models. Based on the relationship between the chosen descriptors and the structure of molecules .We will try to define the closest models to the reality.

Analyzing these results we notice that the model 2 includes the various descriptors that repeat themselves in the three models (Torsion energy, molecular topological index, E_{HOMO}, Softness**,** Dipole moment**)**, despite the random change of the combination of molecules in each model, those descriptors rest with large influence in each model which allows us to suppose that these descriptors have great influence on the cytotoxicity of the molecule.

On the other hand The **table 3 , figure 3, figure 4 and figure 5** shows the different results of the experimental and predicted cytotoxicity. The comparison between the three models shows that the **model 3** is the most reliable regarding the predicted activity.

Table 3: The observed, predicted activities (TC_{50}), according to method MLR for the 22 derivatives of indole-2-carboxylate

Observation	TC50 1(µmol/L)	Préd(TC50 1(µmol/L))	TC50 2 (µmol/L)	Préd(TC50 2(µmol/L))	TC50 3(µmol/L)	Préd(TC50 3(µmol/L))
M1	763.360	717.139	763.360	732.746	**763.360**	**209.387**
M2	45.590	-3.971	45.590	26.769	341.630	332.756
M3	**50.690**	**685.204**	**50.690**	**658.311**	24.380	27.833
M4	**225.460**	**321.585**	225.460	214.712	225.460	227.085
M5	49.550	49.671	49.550	60.529	92.060	92.055
M6	54.840	48.029	54.840	22.947	**136.980**	**67.823**
M7	597.710	620.620	597.710	1014.966	414.440	412.605
M8	212.020	282.395	212.020	304.923	**711.970**	**-212.465**
M9	174.950	163.742	174.950	183.638	174.950	162.011
M10	**14.510**	**80.630**	14.510	59.554	108.640	126.265
M11	16.370	103.904	16.370	20.290	273.570	277.321
M12	52.910	54.725	52.910	59.121	354.200	350.983
M13	11.750	11.726	11.750	4.606	14.130	23.606
M14	210.330	167.817	210.330	131.770	130.280	90.210
M15	**177.770**	**162.806**	177.770	118.307	177.770	183.215
M16	122.330	131.363	122.330	92.645	40.760	8.293
M17	94.960	153.366	94.960	134.882	65.850	125.966
M18	**625.850**	**138.620**	**625.850**	**-33.334**	465.600	461.882
M19	243.260	221.814	243.260	236.024	**90.790**	**-12.672**
M20	145.680	117.956	145.680	154.948	**11.230**	**261.602**
M21	**18.290**	**88.788**	31.670	10.616	**31.670**	**544.321**
M22	115.440	70.755	**115.440**	**175.318**	5.770	7.403

4- Artificial Neural Networks (ANN)

In order to increase the probability of good characterization of studied compounds, artificial 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 the observed activity. The ANN calculated activities model were developed using the properties of several studied compounds. Some authors **[19, 20]** have proposed a parameter **ρ**, leading to determine the number of hidden neurons, which play a major role in determining the best ANN architecture are defined as follows:

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

In order to avoid over fitting or under fitting, it is recommended that **1.8 < ρ < 2.3**. The output layer represents the calculated activity values TC_{50}. ANN architectures used in this work are :- (**7-1-1**) for $\mathbf{TC_{50\ (1)}}$, (**5-2-1**) for $\mathbf{TC_{50\ (2)}}$ and (**10-1-1**) for $\mathbf{TC_{50\ (3)}}$.

The values of predicted activities $\mathbf{TC_{50\ (1)\ ANN}}$, $\mathbf{TC_{50\ (2)\ ANN}}$ and $\mathbf{TC_{50\ (3)\ ANN}}$, calculated using ANN and the observed values are given in **Table4**. The correlations of predicted and observed activities are illustrated in **Figures 6**, **7** and **8**.

The correlations between ANN calculated and experimental activities are very significant as illustrated in Figures 6,7 and 8, and as indicated by R and R^2 values.

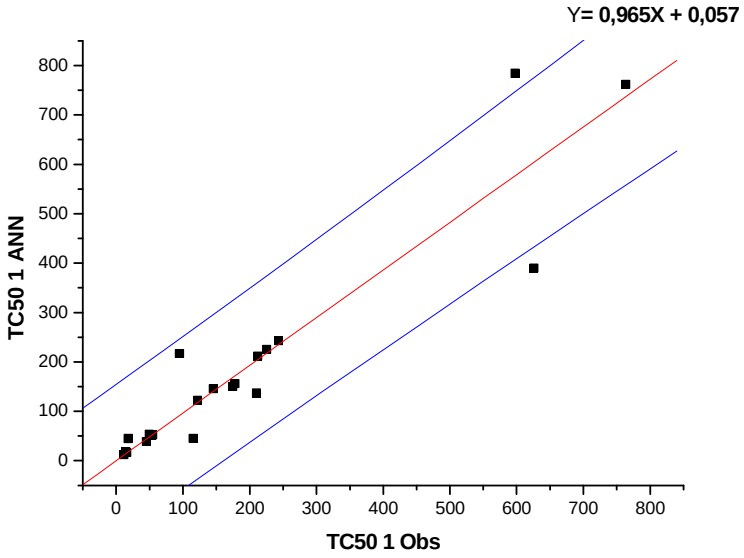


Figure 6: Correlations of observed and predicted activities $TC_{50(1)}$calculated using ANN

N = 22R = 0.951R^2 = 0.904

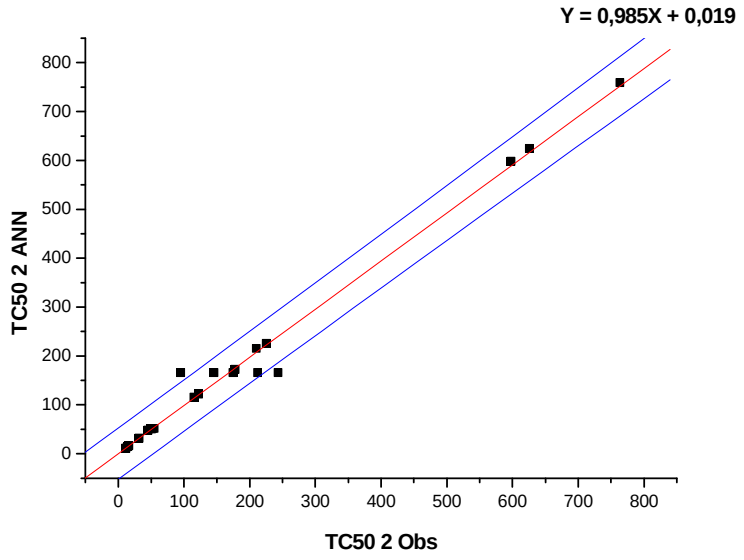


Figure 7: Correlations of observed and predicted activities $TC_{50\ (2)}$ calculated using ANN

N = 22R = 0.992R^2 = 0.984

The obtained squared correlation coefficient (R^2) value confirms that the artificial neural network result were the best to build the quantitative structure activity relationship models.

A comparison of the quality of 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.

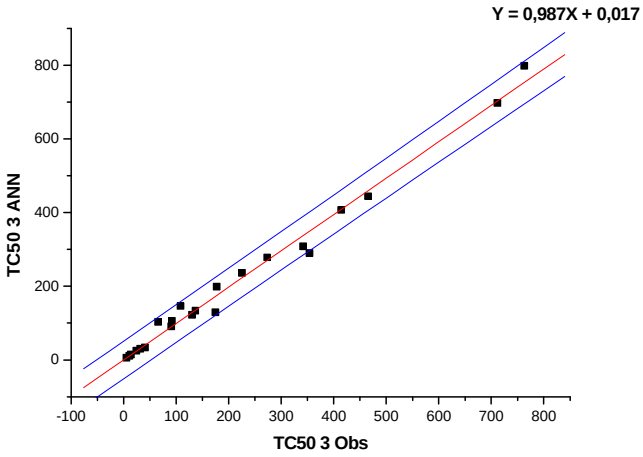


Figure 8: Correlations of observed and predicted activities $TC_{50(3)}$ calculated using ANN

N = 22R = 0.993R^2 = 0.986

In this study, three different modelling methods, MLR and ANN were used in the construction of a QSAR model for 22 derivatives of indole-2-carboxylate and the resulting models were compared. It was shown that the artificial neural network ANN results have substantially better predictive capability than the MLR , yielding regression model with improved predictive power, we have established a relationship between several descriptors and the activity **TC_{50}** in satisfactory manners. Furthermore, we can conclude that studied descriptors, which are sufficiently rich in chemical, electronic and topological information to encode the structural feature and have a great influence on the activity may be used with other descriptors for the development of predictive QSAR models.

Thus, thanks to QSAR studies, especially with the ANN that has allowed us to improve the correlation between the observed biological activity and the predicted activity, we can enjoy the performance of the predictive power of this model to explore and propose new molecules that could be active.

Table 4: The observed, the predicted activities (TC_{50}), according to method ANN for the 22 derivatives of indole-2-carboxylate

N°	$TC_{50\ (1)\ Obs}$	$TC_{50(1)\ ANN}$	$TC_{50\ (2)Obs}$	$TC_{50\ (2)\ ANN}$	$TC_{50\ (3)Obs}$	$TC_{50\ (3)\ ANN}$
M1	763,36	762,07	763,36	758,75	763,36	798,54
M2	45,590	38,480	45,590	47,660	341,63	308,31
M3	50,690	51,890	50,690	50,580	24,380	25,120
M4	225,46	225,42	225,46	225,37	225,46	236,37
M5	49,550	53,900	49,550	51,550	92,060	105,87
M6	54,840	51,890	54,840	51,550	136,98	133,07
M7	597,71	784,15	597,71	598,13	414,44	406,91
M8	212,02	211,83	212,02	165,72	711,97	697,75
M9	174,95	150,03	174,95	165,72	174,95	129,15
M10	14,510	17,770	14,510	14,730	108,64	146,04
M11	16,370	17,090	16,370	16,360	273,57	278,54
M12	52,910	51,400	52,910	51,550	354,20	290,06
M13	11,750	11,980	11,750	11,550	14,130	14,420
M14	210,33	136,83	210,33	215,22	130,28	122,34
M15	177,77	155,91	177,77	172,54	177,77	198,88
M16	122,33	122,17	122,33	122,17	40,760	33,360
M17	94,960	216,62	94,960	165,76	65,850	103,68
M18	625,85	389,76	625,85	624,45	465,60	444,73
M19	243,26	242,71	243,26	165,72	90,790	90,360
M20	145,68	145,54	145,68	165,72	11,230	11,310
M21	18,290	44,660	31,670	31,630	31,670	30,640
M22	115,44	45,210	115,44	115,18	5,7700	5,7600

CONCLUSION

QSAR analysis on the 22 molecules derived from indole-2-carboxylate using Principal component analysis (PCA), Multiple linear regressions (MLR) and Artificial Neural Networks **(**ANN) allowed as to select the most influencing molecular descriptors which are Torsion energy, molecular topological index, E_{HOMO}, Softness**,** and Dipole moment

and predict the cytotoxicity of the compounds. The studied descriptors which are rich in chemical, topological and electronic informationthat encoded the molecular structure may be used to predict new molecules with moderate cytotoxicity.

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

REFERENCES

[1] A J Kochanowska-Karamyan, M T Hamann, *Chem Rev*, **2010**, 110, 4489–4497.
[2] A Arora, K Seth, N Kalra, Y Shukla, *Appl Pharmacol*, **2005**, 202, 237-243.
[3] Y Higashio, T Shoji, *Appl Catal A*, **2004**, 260, 251-259.
[4] H E M Dobson, J Arroyo, G Bergström, I Groth, *Biochem Syst Ecol*, **1997**, 25, 685-706.
[4] M Artico, R Silvestri, S Massa, A G Loi, S Corrias, G Piras, P La Colla, *J Med Chem*, **1996** ,39 ,522-530.
[6] H Chai, X Liang, L Li, C Zhao, P Gong, Z Liang, W Zhu, H Jiang, C Luo, *J Mol Model*, **2011**, 17 1831–1840.
[7] S Xue, L Ma, R Gao, Y Li Z Li, *Acta Pharm Sinic B*, **2014**, 4, 313-321.
[8] M Niemyjska, D Maciejewska, I Wolska, P Truszkowski, *J Mol Struct*, **2012**, 1026, 30–35.
[9] X Ji, S Xue, Y Zhan, J Shen, L Wu, J Jin, Z Wang, Z Li, *Eur J Med Chem*, **2014**, 83, 409-418.
[10] R G Parr, W Yang, *Oxford University Press*, **1989**, 350. ISBN 0-19-509276-7.
[11]R G Parr, W Yang, *Oxford University Press*, **1989**, ISBN 0-19-504279-4.
[12]M J Frisch, G W Trucks Schlegel , H B Scuseria, G E Robb, M A Cheeseman, J R Zakrzewski, V G Montgomery, J A Stratmann, R E Burant, J R Dapprich, S Millam, J M Daniels, A D Kudin, K N Strain, M C Farkas, O Tomasi, J Barone, V Cossi, M Cammi, R Mennucci, B Promelli, C Adamo, C Clifford, S J Ocherski , A Petersson , P Y Ayala Q Cui, K Morokuma, D K Malick ,A D Rabuck , K Raghavachari , J B Foresman ,J Cioslowski , J V Ortiz ,B B Stefanov, G Liu , A Liashenko ,P Piskorz, I Komaromi ,R Gomperts, R L Martin ,D J Fox, T Keith, M A Al-Laham , C Y Peng, A Nanayakkara, C Gonzalez ,M Challacombe ,P M W Gill , B G Johnson , W Chen ,M W Wong ,J L Andres , M Head-Gordon, E S Replogle, J A Pople, GAUSSIAN98 , Revision A 7 *Gaussian Inc Pittsburgh PA*, **2003**.
[13] C Lee, W Yang, R G Parr, *Phys Rev B*, **1993**, 37, 785.
[14]A D Becke, *J Chem Phys*, **1992**, 96 ,9489 ; A D Becke, *J Chem Phys*, **1993**, 98, 1372.
[15] M Larif, A Adad , R Hmamouchi, A I Taghki, A Soulaymani , A Elmidaoui, M Bouachrine, T Lakhlifi article in press in *Arabian Journal of Chemistry* **2013**.
[16] A Adad, R Hmamouchi, A I Taghki, A Abdellaoui, M Bouachrine, T Lakhlifi, *J of Chemical and Pharmaceutical Research*, **2013**, 5, 28-41.
[17] STATITCF Software, *Technical Institute of cereals and fodder*, Paris, France, **1987**.
[18] A Ousaa, B Elidrissi, M Ghamali, S Chtita, M Bouachrine, T Lakhlifi, *J Comp Meth Mol Des*, **2014**, 4, 10-18.
[19] S-S So , W G Richards, *J Med Chem*, **1992**, 35,3201-3207.
[20] T A Andrea, H Kalayeh, *J Med Chem*, **1991** , 34 , 2824–2836.