

**International Journal of Innovative Research in Science,  
Engineering and Technology**

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 2, Issue 11, November 2013**

# **Studies of two different cancer cell lines activities (MDAMB-231 and SK-N-SH) of imidazo[1,2-a]pyrazine derivatives by combining DFT and QSAR results**

Samir Chtita<sup>1</sup>, Majdouline Larif<sup>2</sup>, Mounir Ghamali<sup>1</sup>, Azeddine Adad<sup>1</sup>, Hmamouchi Rachid<sup>1</sup>

Mohammed. Bouachrine<sup>3</sup> and Tahar Lakhli<sup>4\*</sup>

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

<sup>2</sup> Doctor, Separation Process Laboratory, Faculty of Science, University Ibn Tofail, Kenitra, Morocco

<sup>3</sup> Professor, ESTM, University Moulay Ismail, Meknes, Morocco.

<sup>4</sup> Professor, Molecular Chemistry and Natural Substances Laboratory, Faculty of Science, University Moulay Ismail, Meknes, Morocco

**Abstract:** To establish a quantitative structure-activity relationship for cytotoxic effects of two against different cancer cell lines, a series of thirteen imidazo[1,2-a]pyrazine derivatives molecules was submitted to a principal components analysis (PCA), to a multiple regression analysis (MRA), to a regression partial least squares (PLS), to a non-linear regression (RNLM) and to a neural network (NN). 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. The topological descriptors (Formula Weight, Molar Volume, Molecular Weight, Molar Refractivity, Parachor, Density, Refractive Index, Surface Tension and Polarizability) and the electronic descriptors (total energy (E), highest occupied molecular orbital energy ( $E_{\text{HOMO}}$ ), lowest unoccupied molecular orbital energy, ( $E_{\text{LUMO}}$ ) difference between the LUMO and the HOMO energy (Gap), total dipole moment of the molecules ( $\mu$ ), absolute hardness ( $\eta$ ), absolute electron negativity ( $\chi$ ) and reactivity index ( $\omega$ )) were computed with ACD/ChemSketch and Gaussian 03W program, respectively. This study shows that the MRA, PLS, and MNLR 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.

**Keywords:** QSAR, DFT, imidazo[1,2-a]pyrazine, cell lines.

## I. INTRODUCTION

The use of chemicals in commerce, medicine and other aspects of daily life are generally acknowledged to be a quite positive benefit; however there is continuing concern about their negative impact on human health and the environment [1]. More than 100000 chemical substances are produced and used on a commercial scale, and about 2000 new ones are introduced onto the market each year. Many of these substances have little or no adverse effects, but some may be harmful to human health and the natural environment [2]. This dichotomy in social concern has caused both regulatory agencies and chemical industries to take an interest in the potential environmental impact of a particular

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

chemical prior to its release into an ecosystem. The limited availability of experimental data necessary for the risk assessment of chemicals, and the general lack of knowledge of the properties and activities of existing substances, has led the European Commission to adopt a “White Paper on a strategy for a future Community Policy for Chemicals” [3].

Quantitative Structure-Property/Activity Relationship (QSPR/QSAR) methods are among the most practical tools in computational physical chemistry. These methods are based on the axiom that the variance in the physicochemical properties and activities of chemical compounds is determined by the variance in their molecular structures. Thus, if experimental data are available for only some chemicals in a group, one can predict the missing from molecular descriptors calculated for the whole group and suitable mathematical model [4]. The global prediction of toxicity using QSAR has been the goal of many workers who utilized a variety of approaches. This goal is alluring, but has yet to be achieved satisfactorily. There are a number of reasons for the absence of success [5]. The deficiency of available toxicity data has clearly held back progress. This lack of success has been compounded in many studies by a poor appreciation of the insufficient heterogeneity, or chemical diversity, in the dataset. Further, while some molecular properties (such as hydrophobicity) are well described, others, including electrophilic reactivity, ionization, and hydrogen bonding, are poorly parameterized. Last, mechanisms of toxic action are not fully understood or misinterpreted, or their relevance in the modelling of toxicity is ignored [6].

Discovery of new drugs for treatment of cancer has been gaining a great deal of interest mainly due to a universal resistance to conventional single drug chemotherapeutic agents. Multidrug resistance [7] characterized by resistance not only to drugs that are similar structurally and functionally but also cross-resistance to unrelated drugs like doxorubicin, vincristine, vinblastine, colchicines and actinomycin has been documented. Thus, search for novel anticancer agents with diverse chemical structure is need of the hour. Herein, we report the synthesis and evaluation of a series of imidazo [1, 2-a] pyrazines as potent anticancer agents.

Imidazo[1,2-a]pyrazines have been gaining attention in drug discovery realm especially as structural analogues of purines [8-10] (Fig. 1). Derivatives of imidazo[1,2-a]pyrazine exhibit various pharmacological activities such as antibacterial [11], anti-inflammatory [12-14], uterine relaxing activity [15], anti-bronchospastic [16], anti-ulcer [17], cardiac stimulating [18], antidepressant [19], hypoglycemic activity [20], antiproliferative activity [21], controlling allergic reactions [22], smooth muscle relaxant properties [23] and phosphodiesterase inhibitory activity [24]. They have also been shown to inhibit the receptor tyrosine kinase EphB<sub>4</sub> recently [25].

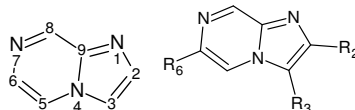


Fig. 1: Schematic diagram of imidazo[1,2-a]pyrazine skeleton.

In this work we attempt to establish a quantitative structure-activity relationship for cytotoxic effects of two against different cancer cell lines, by studying a series of 13 substituted imidazo[1,2-a]pyrazines (Figure 2) [26] have been synthesized with substitutions at 2, 3, 6-ring positions being varied

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

generating mono-, di-, tri-substituted imidazo[1,2-a]pyrazines possessing functional groups like halo, hydroxymethyl, amine, alkyl, aryl, heteroaryl etc...[27].

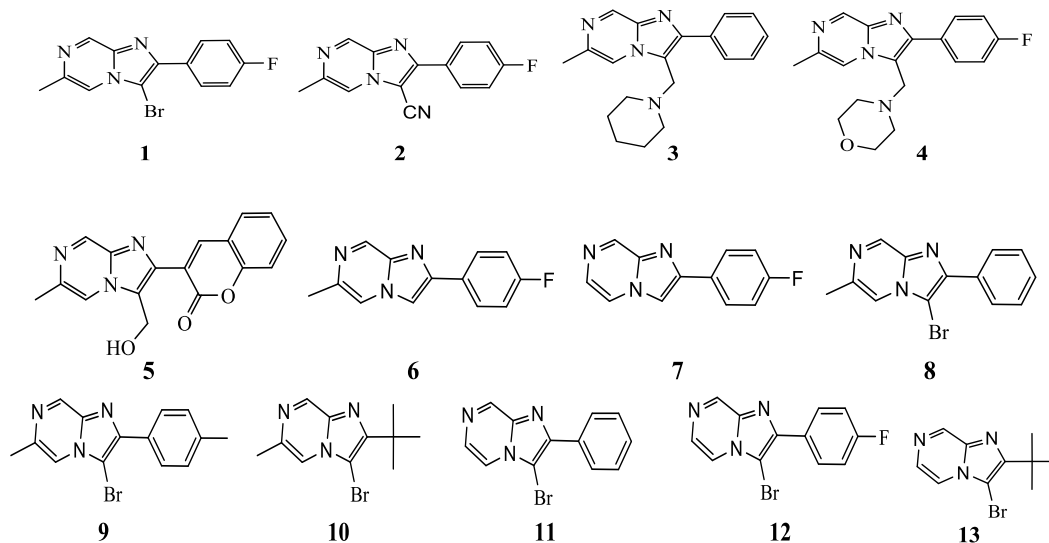


Fig.2: Structures of the study compounds.

We accordingly propose a quantitative model, and we try to interpret the activity of the compounds relying on the multivariate statistical analyses. The principal components analysis (PCA) has served to classify the compounds according to their activities and to give an estimation of the values of the pertinent descriptors that govern this classification. The multiple linear regression (MLR) has served to select the descriptors used as the input parameters for the partial least square regression (PLS), the multiples nonlinear regression (MNL), and artificial neural network (ANN). These methods (MRA, PLS, and MNL) 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.

## II. MATERIAL AND METHODS

### A. Experimental data

The experimental  $IC_{50}$  ( $\mu M$ ) cytotoxic effects of two against different cancer cell lines activities (MDAMB-231 and SK-N-SH) of imidazo[1,2-a]pyrazine derivatives are collected from recent publications [27]. The observations are converted into minus logarithm scale  $\log IC_{50}$  and are included in table 1.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

TABLE 1:  
LABELS AND INHIBITION VALUES (IC<sub>50</sub>) OF STUDY IMIDAZO[1,2-A]PYRAZINE DERIVATIVES ON HUMAN TUMOR CELL LINES.

IC <sub>50</sub> values against cancer cell lines (μM)			
	IUPAC Name	Log (MDAMB-231)	Log (SK-N-SH)
1	3-bromo-2-(4-fluorophenyl)-6-methylimidazo[1,2-a]pyrazine	2.283	1.140
2	2-(4-fluorophenyl)-6-methylimidazo[1,2-a]pyrazine-3-carbonitrile	2.064	2.016
3	6-methyl-2-phenyl-3-(piperidin-1-ylmethyl)imidazo[1,2-a]pyrazine	2.032	2.028
4	2-(4-fluorophenyl)-6-methyl-3-(morpholin-4-ylmethyl)imidazo[1,2-a]pyrazine	2.337	2.210
5	3-[3-(hydroxymethyl)-6-methylimidazo[1,2-a]pyrazin-2-yl]-2H-chromen-2-one	2.095	2.164
6	2-(4-fluorophenyl)-6-methylimidazo[1,2-a]pyrazine	2.017	2.202
7	2-(4-fluorophenyl)imidazo[1,2-a]pyrazine	2.133	2.072
8	3-bromo-6-methyl-2-phenylimidazo[1,2-a]pyrazine	1.182	0.000
9	3-bromo-6-methyl-2-(4-methylphenyl)imidazo[1,2-a]pyrazine	2.228	0.978
10	3-bromo-2-tert-butyl-6-methylimidazo[1,2-a]pyrazine	2.234	2.091
11	3-bromo-2-phenylimidazo[1,2-a]pyrazine	2.154	1.960
12	3-bromo-2-(4-fluorophenyl)imidazo[1,2-a]pyrazine	2.006	1.906
13	3-bromo-2-tert-butylimidazo[1,2-a]pyrazine	2.046	2.057

### B. Computational methods

An attempt has been made to correlate the activity of these compounds with various physicochemical parameters.

DFT (density functional theory) methods were used in this study. These methods have become very popular in recent years because they can reach similar precision to other methods in less time and less cost from the computational point of view. In agreement with the DFT results, energy of the fundamental state of a polyelectronic system can be expressed through the total electronic density, and in fact, the use of electronic density instead of wave function for calculating the energy constitutes the fundamental base of DFT [28] using the B3LYP functional [29] 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 [30].

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 thirteen compounds was carried out by B3LYP method employing 6-31G (d) basis set. ChemSketch program (Demo version 10.0) [31] was employed to calculate the others molecular descriptors.

### C. Calculation of molecular descriptors

#### *Calculation of descriptors using Gaussian 03W*

From the results of the DFT calculations, the quantum chemistry descriptors were obtained for the model building as follows: the total energy (ET (u.a)), the highest occupied molecular orbital energy (E<sub>HOMO</sub> (eV)), the lowest unoccupied molecular orbital energy (E<sub>LUMO</sub> (eV)), the energy difference

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

between the LUMO and the HOMO energy (Gap (eV)), the total dipole moment of the molecule ( $\mu$ (Debye)), absolute hardness ( $\eta$ ), absolute electron negativity ( $\chi$ ) and reactivity index ( $\omega$ ) [33].

$\eta$ ,  $\chi$  and  $\omega$  were determined by the following equations:

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

### Calculation of descriptors using ACD/ChemSketch

Advanced chemistry development's ACD/ChemSketch program [31] was used to calculate Formula Weight (PM), Molar Volume (MV (cm<sup>3</sup>)), Molecular Weight (MW), Molar Refractivity (MR (cm<sup>3</sup>)), Parachor (Pc (cm<sup>3</sup>)), Density (D (g/cm<sup>3</sup>)), Refractive Index (n), Surface Tension ( $\gamma$  (dyne/cm) and Polarizability ( $\alpha_e$  (cm<sup>3</sup>)) [32].

- **Molecular Weight (MW):** Used as the descriptor in systems such as transport studies where diffusion is the mode of operation. It is an important variable in QSAR studies pertaining to cross resistance of various drugs in multi-drug resistant cell lines.
- **Molar Volume (MV):** The molar volume calculates from additive increments. The additive atomic increments were obtained using a database of density and calculated MW:

$$MV = \frac{MW}{D}$$

- **Density (D):** The density is calculates from MW and the calculated molar volume:

$$D = \frac{MW}{MV}$$

- **Molar Refractivity (MR):** The Lorentz-Lorenz equation relates Molecular weight, density, and refractive index:

$$MR = \frac{n^2 - 1}{n^2 + 2} \frac{MW}{D}$$

- **Parachor (Pc):** The parachor is calculates from additive increments. The additive atomic increments were obtained using a database of density, surface tension, and calculated MW:

$$Pc = \left( \frac{MW}{D} \right) \gamma^{\frac{1}{4}}$$

- **Refractive Index (n):** By the Lorentz-Lorenz equation:

$$n = \sqrt{\frac{2MR + MW}{MV - MR}}$$

The refractive index calculates from the molar volume and molar refractivity, both of which are calculated as above.

- **Surface Tension ( $\gamma$ ):** Calculated from calculated MV and calculated Pc:

$$\gamma = \left( \frac{Pc}{MW} \right)^4$$

- **Polarizability ( $\alpha_e$ ):** Calculates from the Molar Refractivity as follows:

$$\alpha_e = 0,3964308 \times MR$$

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

## D. Statistical analysis

To explain the structure - activity relationship, these 16 descriptors are calculated for 13 molecules using the Gaussian 03W, Gauss View and ChemSketch software.

The study we conducted consists of:

- The principal component analysis (PCA) available in a software called XLSTAT.
- The multiple linear regressions (MLR) available in the XLSTAT software.
- The regression partial least squares (PLS) available in the XLSTAT software.
- The non-linear regression (RNLM) available in XLSTAT software.
- The Neural Network (RN) available in the software MATLAB Version 9.

The structures of the molecules based on imidazo[1,2-a]pyrazines, (1–13) were studied by statistical methods based on the principal component analysis (PCA) [34] using the software XLSTAT version Demo 2009 [35]. 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 [36]. 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 partial least squares (PLS), and the Multiples nonlinear regression (MNL) and artificial neural network (ANN).

The (MLR), the (PLS), and the (MNL) were generated using the software XLSTAT version Demo 2009 [35], to predict cytotoxic effects  $IC_{50}$  activities. Equations were justified by the correlation coefficient (R), mean squared error (MSE), fishers F-statistic (F), and significance level (F value) [35]. 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 [37]. 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 [38] and a detail of this algorithm is given elsewhere [39].

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

## III. RESULTS AND DISCUSSION

### A. Data set for analysis

The QSAR analysis was performed using the IC<sub>50</sub> of the 13 compounds against the MDAMB-231, and SK-N-SH cells (experimental values) as reported in [27], 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 (including the critical probability: p-value <0.05 for all descriptors and the model complete).

TABLE 2:

THE VALUES OF THE SIXTEEN CHEMICAL DESCRIPTORS

	PM	MR	MV	Pc	n	$\gamma$	D	$\alpha_e$	E	E <sub>HOMO</sub>	E <sub>LUMO</sub>	Gap	$\mu$	$\chi$	$\eta$	$\omega$
1	306.133	72.11	191.10	501.7	1.678	47.4	1.600	28.580	-90788.911	-6.006	-1.614	4.392	2.975	3.810	2.196	3.305
2	252.246	71.08	191.40	503.9	1.664	47.9	1.310	28.170	-23339.235	-6.368	-1.961	4.407	2.428	4.165	2.204	3.935
3	306.405	93.74	254.50	669.0	1.658	47.7	1.200	37.160	-26020.018	-5.415	-1.307	4.108	2.974	3.361	2.054	2.750
4	326.368	90.39	247.80	649.7	1.650	47.2	1.310	35.830	-29696.752	-5.777	-1.433	4.344	2.198	3.605	2.172	2.992
5	307.303	83.66	212.60	588.1	1.716	58.5	1.440	33.160	-28449.823	-5.906	-2.035	3.871	2.544	3.971	1.936	4.073
6	227.237	64.55	178.60	458.1	1.642	43.3	1.270	25.590	-20829.274	-5.932	-1.450	4.482	3.700	3.691	2.241	3.040
7	213.210	60.13	163.30	427.0	1.657	46.7	1.300	23.830	-19759.348	-6.015	-1.543	4.473	3.502	3.779	2.236	3.193
8	288.143	72.24	188.20	501.5	1.693	50.3	1.530	28.630	-88088.762	-5.999	-1.546	4.453	2.276	3.773	2.227	3.196
9	302.169	76.66	203.50	532.6	1.677	46.9	1.480	30.390	-89158.616	-5.855	-1.507	4.348	2.152	3.681	2.174	3.116
10	268.153	65.55	187.30	471.4	1.617	40.1	1.430	25.980	-86080.688	-6.052	-1.345	4.707	2.292	3.699	2.354	2.906
11	274.116	67.82	173.00	470.4	1.712	54.6	1.580	26.880	-87082.427	-6.439	-2.007	4.432	2.705	4.223	2.216	4.024
12	292.106	67.69	175.90	470.6	1.695	51.1	1.660	26.830	-89718.978	-6.084	-1.710	4.374	2.631	3.897	2.187	3.472
13	254.126	61.12	172.00	440.3	1.628	42.9	1.470	24.230	-85010.759	-6.149	-1.436	4.713	2.375	3.793	2.357	3.052

### B. Principal component analysis

The totality of the sixteen descriptors (variables) coding the thirteen molecules was submitted to a principal components analysis (PCA). Twelve principal components were obtained. The first three axes F1, F2 and F3 contributing respectively 43.4 %, 32.6 % and 15.3 % to the total variance, the total information is estimated to a percentage of 91.3%, were sufficient to describe the information represented by the data set. Table 3 shows the descriptor's contributions to F1, F2 and F3. The descriptors MR, MV, Pc,  $\alpha_e$ , E<sub>HOMO</sub>, Gap and  $\eta$  have the most significant contributions to F1, the descriptors n,  $\gamma$ , E<sub>LUMO</sub>,  $\chi$  and  $\omega$  have the most significant contributions to F2, and the descriptors PM, D,  $\mu$  and log(-E) have the most significant contributions to F3.

**International Journal of Innovative Research in Science,  
Engineering and Technology**

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 2, Issue 11, November 2013**

TABLE 3: DESCRIPTOR'S CONTRIBUTIONS TO THE FIRST TREE PRINCIPAL COMPONENTS.

Descriptors	F1		F2		F3	
	Correlations	Contributions %	Correlations	Contributions %	Correlations	Contributions %
PM	0.740	7.887	-0.216	0.894	0.610	<b>15.178</b>
MR	0.983	<b>13.914</b>	-0.005	0.000	0.098	0.392
MV	0.945	<b>12.849</b>	0.214	0.876	0.098	0.389
Pc	0.976	<b>13.728</b>	0.062	0.074	0.075	0.230
n	0.245	0.864	-0.888	<b>15.135</b>	0.035	0.049
$\gamma$	0.375	2.027	-0.882	<b>14.917</b>	-0.093	0.351
D	-0.362	1.888	-0.552	5.847	0.649	<b>17.211</b>
$\alpha\epsilon$	0.983	<b>13.915</b>	-0.004	0.000	0.098	0.392
Log(-E)	-0.339	1.655	-0.172	0.565	0.882	<b>31.750</b>
E <sub>HOMO</sub>	0.768	<b>8.494</b>	0.530	5.388	-0.031	0.040
E <sub>LUMO</sub>	0.077	0.085	0.954	<b>17.440</b>	0.211	1.809
Gap	-0.809	<b>9.422</b>	0.460	4.062	0.274	3.057
$\mu$	-0.174	0.434	0.188	0.675	-0.767	<b>23.988</b>
$\chi$	-0.475	3.249	-0.820	<b>12.906</b>	-0.098	0.388
$\eta$	-0.809	<b>9.417</b>	0.460	4.059	0.274	3.072
$\omega$	-0.109	0.172	-0.946	<b>17.162</b>	-0.204	1.703

The principal component analysis (PCA) was conducted to identify the link between the different variables. Correlations between the sixteen descriptors are shown in table 4 as a correlation matrix, in figure 3 these descriptors are represented in a correlation circles.



# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

TABLE 4:  
THE CORRELATION MATRIX (PEARSON (N)) BETWEEN DIFFERENT OBTAINED DESCRIPTORS.

	PM	MR	MV	Pc	n	$\gamma$	D	$\alpha\epsilon$	E	E <sub>HOMO</sub>	E <sub>LUMO</sub>	Gap	$\mu$	$\chi$	$\eta$	$\omega$
PM	1															
MR	0.790	1														
MV	0.717	0.971	1													
Pc	0.757	0.995	0.986	1												
n	0.386	0.218	-0.021	0.134	1											
$\gamma$	0.396	0.343	0.118	0.279	0.943	1										
D	0.282	-0.332	-0.462	-0.395	0.515	0.335	1									
$\alpha\epsilon$	0.790	1.000	0.971	0.995	0.218	0.342	-0.332	1								
E	-0.261	0.329	0.370	0.372	-0.159	0.054	-0.860	0.329	1							
E <sub>HOMO</sub>	0.433	0.697	0.755	0.713	-0.183	-0.119	-0.480	0.698	0.331	1						
E <sub>LUMO</sub>	-0.023	0.069	0.256	0.118	-0.750	-0.786	-0.359	0.069	-0.040	0.621	1					
Gap	-0.531	-0.736	-0.591	-0.698	-0.634	-0.750	0.154	-0.735	-0.431	-0.464	0.406	1				
$\mu$	-0.579	-0.283	-0.270	-0.275	-0.086	-0.100	-0.343	-0.283	0.469	0.120	0.086	-0.042	1			
$\chi$	-0.231	-0.431	-0.566	-0.467	0.513	0.497	0.467	-0.431	-0.165	-0.904	-0.897	0.039	-0.115	1		
$\eta$	-0.530	-0.735	-0.591	-0.698	-0.634	-0.750	0.154	-0.735	-0.431	-0.464	0.406	1.000	-0.043	0.039	1	
$\omega$	-0.001	-0.099	-0.281	-0.146	0.723	0.764	0.362	-0.100	0.028	-0.648	-0.998	-0.374	-0.098	0.911	-0.374	1

- **Gap** and  $\eta$  are perfectly correlated ( $r = 1$ ), both variables are redundant.
- **MR** and  $\alpha\epsilon$  are perfectly correlated ( $r = 1$ ), both variables are redundant.
- **MV**, **Pc** and  $\alpha\epsilon$  are highly correlated ( $r$  (MV, Pc) = 0.986,  $r$  (MV,  $\alpha\epsilon$ ) = 0.971,  $r$  (Pc,  $\alpha\epsilon$ ) = 0.995).
- **E<sub>LUMO</sub>** and  $\omega$  are strongly negatively correlated ( $r = -0.998$ ).

The following variables then removed are: **Gap**, **MR**, **MV**,  $\alpha\epsilon$  and  $\omega$ .

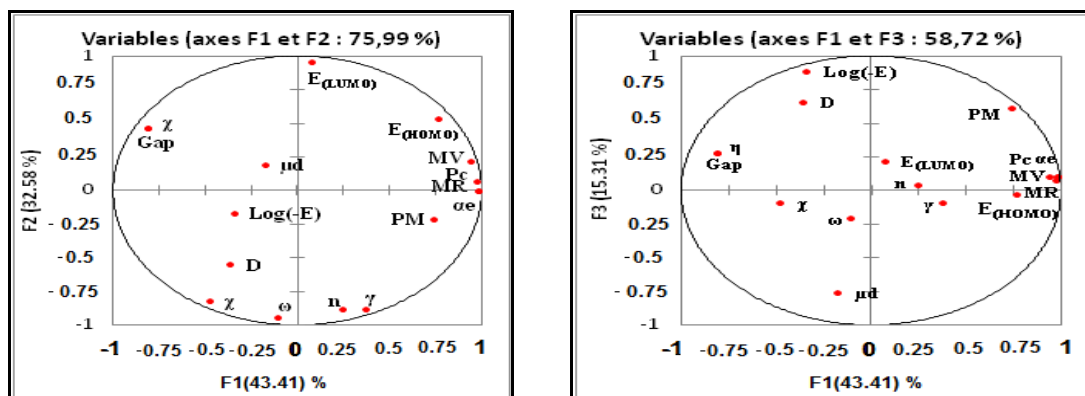


Fig. 3: Correlation circles

In the projection of the compounds in the plane of the three first axes F1, F2 and F3 (Figure 4), the compounds are distributed in four regions. Region 1 contains compounds having a values of log (-E) between 4.296 and 4.368, region 2 contains compounds having a values of log (-E) between

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

4.415 and 4.473, region 3 contains compounds having a values of log (-E) between 4.929 and 4.935 and region 4 contains compounds having a values of log (-E) between 4.940 and 4.953.

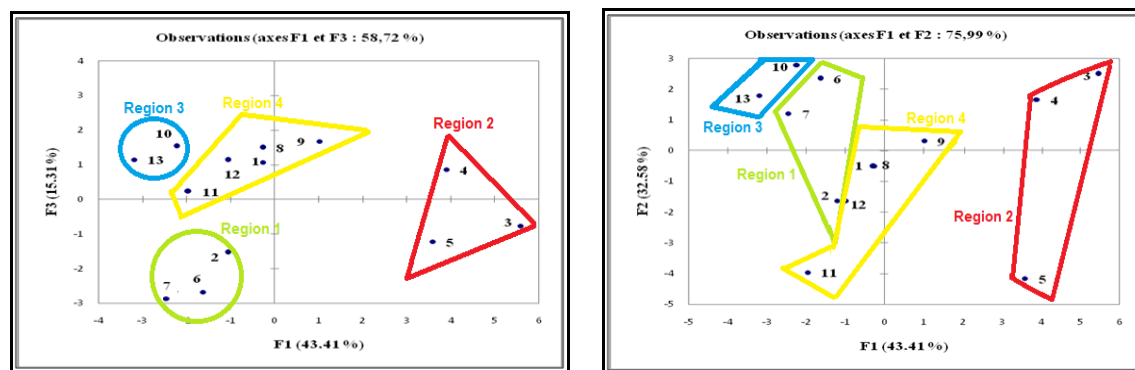


Fig. 4: Cartesian diagram according to F1F2 and F1F3: Separation between four regions

### C. Multiple Linear Regressions MLR

In order to propose a mathematical model and to evaluate quantitatively the substituent's physicochemical effects on the two activities of the totality of the set of these 13 molecules, we submitted the data matrix constituted obviously from the 14 physicochemical variables corresponding to the 13 molecules, to a progressive multiple regression analysis. This method used the coefficients R, R<sup>2</sup>, and the F-values to select the best regression performance.

Where R is the correlation coefficient; R<sup>2</sup> is the coefficient of determination; MSE is the mean squared error; F is the Fisher F-statistic.

Treatment with multiple linear regressions is more accurate because it allows you to connect the structural descriptors for each activity of 13 molecules to quantitatively evaluate the effect of substituent. The selected descriptors are:

- PM, n,  $\mu$  and  $\chi$  for MDAMB-231 ;  $\gamma$  and n for SK-N-SH

The QSAR models built using multiple linear regression (MLR) method is represented by the following equation:

- $\text{Log (MDAMB-231)} = 11.124 + 1.002 \cdot 10^{-02} \text{ PM} - 10.508 \text{ n} + 0.456 \mu + 1.173 \chi$  (equation 1).
- $\text{Log (SK-N-SH)} = 66.790 - 46.547 \text{ n} + 0.263 \gamma$  (equation 2).

The Fisher's F test is used. Given the fact that the probability corresponding to the F value is lower than 0.05 for SK-N-SH, it means that we would be taking a lower than 0.28% risk in assuming that the null hypothesis is wrong. Therefore, we can conclude with confidence that the models do bring a significant amount of information. For MDAMB-123, the F value (F value = 0.265) is up than 0.05, the model is not significant. (Tables 5 and 6)

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

TABLE 5:  
ANALYSIS OF VARIANCE (GLOBAL MODEL)

MDAMB-231	Source	DDL	Sum of squares	Mean square	F	Pr > F
	Model	4	0.434	0.109	1.600	0.265
	Error	8	0.543	0.068		
total corrected		12	0.977			
SK-N-SH	Source	DDL	Sum of squares	Mean square	F	Pr > F
	Model	2	2.733	1.366	5.657	0.023
	Error	10	2.415	0.242		
	total corrected		12	5.148		

TABLE 6:  
CORRELATION COEFFICIENT (R), COEFFICIENT OF DETERMINATION (R<sup>2</sup>), MEAN SQUARED ERROR (MSE), FISHERS F-STATISTIC (F) AND SIGNIFICANCES LEVEL (F VALUE)

	Log(MDAMB-231)	Log(SK-N-SH)
<b>R<sup>2</sup></b>	0.444	0.531
<b>R</b>	0.667	0.729
<b>MSE</b>	0.068	0.242
<b>F</b>	1.600	5.657
<b>F value</b>	0.265	0.023

The values of predicted activities (Log (MDAMB-231) and Log (SK-N-SH) calculated from equations (1 and 2), and the observed values are given in table 10. The correlations of predicted and observed are illustrated in figure 5.

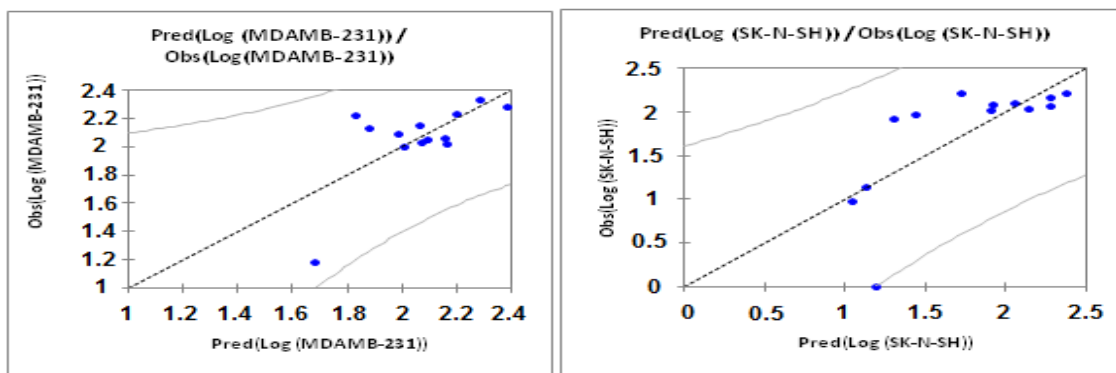


Figure 5: the correlations of observed and predicted activities calculated using MLR

The descriptors proposed in equations (1 and 2) by MLR were, therefore, used as the input parameters in the partial least squares (PLS), and the Multiples nonlinear regression (MNL) and artificial neural network (ANN).

#### D. Partial least squares PLS

Partial Least Squares regression (PLS) is an efficient and optimal for a criterion method based on covariance. It is recommended in cases where the number of variables is high, and where it is likely

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

**Vol. 2, Issue 11, November 2013**

that the explanatory variables are correlated. (<http://www.xlstat.com/en/products-solutions/pls.html>).

We submitted the data matrix constituted obviously from the descriptors proposed by MLR corresponding to the 13 molecules, to the partial least squares (PLS). This method used the coefficients R, R<sup>2</sup>, and the F-values to select the best regression performance.

The QSAR models built using partial least squares (PLS) method is represented by the following equation:

- $\text{Log (MDAMB-231)} = 6.829 + 9.261 \cdot 10^{-04} \text{ PM} - 3.0768 \text{ n} + 0.064 \mu - 1.649 \cdot 10^{-02} \chi$  (equation 3)
- $\text{Log (SK-N-SH)} = 66.790 - 46.547 \text{ n} + 0.263 \gamma$  (equation 4)

The correlation coefficient (R), the coefficient of determination (R<sup>2</sup>), the Mean Squared Error (MSE) and Standard deviation (S) for the two models.(Table 7).

TABLE 7:  
CORRELATION COEFFICIENT (R), COEFFICIENT OF DETERMINATION (R<sup>2</sup>), MEAN SQUARED ERROR (MSE) AND STANDARD DEVIATION (S).

	Log(MDAMB-231)	Log(SK-N-SH)
<b>R<sup>2</sup></b>	0.103	0.531
<b>R</b>	0.320	0.729
<b>MSE</b>	0.067	0.186
<b>S</b>	0.282	0.491

The values of predicted activities (Log (MDAMB-231) and Log (SK-N-SH)) calculated from equations (3 and 4), and the observed values are given in table 10. The correlations of predicted and observed are illustrated in figure 6.

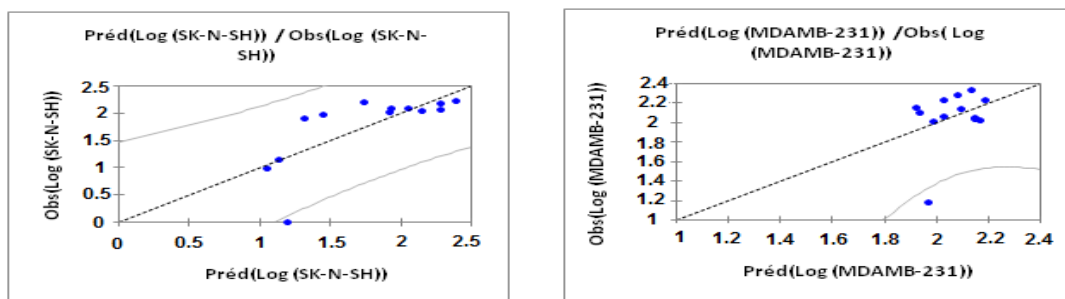


Fig. 6: Correlations of observed and predicted activities calculated using PLS

Despite the good results we have obtained by multiple linear regressions and partial least squares (PLS), it is likely that any non-linear relationship took place. Nonlinear regression performed by XLSTAT software and the neural network are suitable concepts to accomplish this task.

### E. Multiples nonlinear regression (MNLN)

We have used also the technique of nonlinear regression model to improve the structure - activity relationship to quantitatively evaluate the effect of substituent. It takes into account several parameters. This is the most common tool for the study of multidimensional data. We have applied

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

to the data matrix constituted obviously from the descriptors proposed by MLR corresponding to the 13 molecules. The coefficients R, R<sup>2</sup>, and the F-values are used to select the best regression performance.

We used a pre-programmed function of XLSTAT following:

$$Y = a + (b X_1 + c X_2 + d X_3 + e X_4 \dots) + (f X_1^2 + g X_2^2 + h X_3^2 + i X_4^2 \dots)$$

Where A, b, c, d,... : represent the parameters and X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>, X<sub>4</sub>,... : represent the variables.

The resulting equations were:

- $\text{Log (MDAMB-231)} = 167.448 - 6.079 \cdot 10^{-02} \text{PM} - 178.715 \text{n} + 0.712 \mu - 2.573 \chi + 1.280 \cdot 10^4 \text{PM}^2 + 50.609 \text{n}^2 - 7.304 \cdot 10^{-02} \mu^2 + 0.4931 \chi^2$  (equation 5).

- $\text{Log (SK-N-SH)} = 18.889 + 16.006 \text{n} + 1.179 \cdot 10^{-02} \gamma - 17.9487 \text{n}^2 + 2.387 \cdot 10^{-03} \gamma^2$  (equation 6)

The correlation coefficient (R), the coefficient of determination (R<sup>2</sup>), the Mean Squared Error (MSE) and Standard deviation (S) for the two models. (Table 8)

TABLE 8:  
CORRELATION COEFFICIENT (R), COEFFICIENT OF DETERMINATION (R<sup>2</sup>), AND MEAN SQUARED ERROR (MSE)

	Log(MDAMB-231)	Log(SK-N-SH)
<b>R<sup>2</sup></b>	0.564	0.537
<b>R</b>	0.751	0.733
<b>MSE</b>	0.426	2.384

The values of predicted activities calculated from equations (5 and 6), and the observed values are given in table 10. The correlations of predicted and observed are illustrated in figure 7.

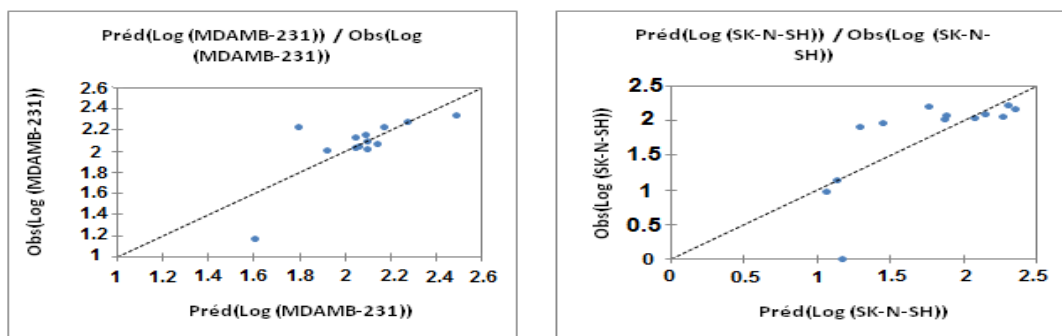


Fig. 7: Correlations of observed and predicted activities calculated using MNL

### F. Artificial neural networks (ANN)

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 correlations coefficients and Standard Error of Estimate, obtained with the Neural network (Table 9), show that the selected descriptors by MLR are pertinent and that the model proposed to predict activity is relevant.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

TABLE 9:  
CORRELATION COEFFICIENT (R) AND COEFFICIENT OF DETERMINATION (R<sup>2</sup>)

	Log(MDAMB-231)	Log(SK-N-SH)
<b>R<sup>2</sup></b>	0.943	0.996
<b>R</b>	0.971	0.998

The values of predicted activities and the observed values are given in table 10.

The obtained squared correlation coefficient (R<sup>2</sup>) value confirms that the neural network result 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 ACP, MLR, PLS, MNLr and ANN models shows that the ANN models have substantially better predictive capability because the ANN approach gives better results than MLR, PLS and MNLr. ANN was able to establish a satisfactory relationship between the molecular descriptors and the activity of the studied compounds.

The values of predicted activities calculated using ANN and the observed values are given in table 10. The correlations of predicted and observed are illustrated in figure 8.

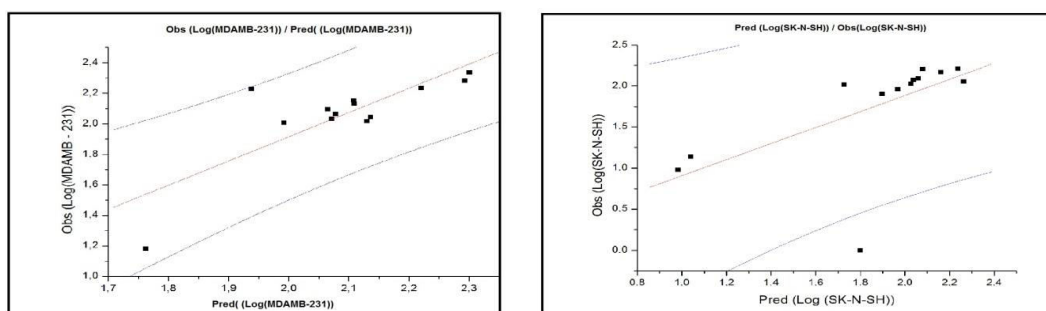


Fig. 8: Correlations of observed and predicted activities calculated using ANN

TABLE 10:  
OBSERVED, PREDICTED ACTIVITIES AND RESIDUE ACCORDING TO DIFFERENT METHODS.

IC <sub>50</sub> values against cancer cell lines (µM)									
	Log(MDAMB-231)								
	Observed	MLR		PLS		MNLr		ANN	
		Pred.	Resid.	Pred.	Resid.	Pred.	Resid.	Pred.	Resid.
<b>1</b>	2.283	2.384	-0.101	2.077	0.206	2.272	0.011	2.293	-0.010
<b>2</b>	2.064	2.158	-0.094	2.030	0.035	2.140	-0.076	2.078	-0.014
<b>3</b>	2.032	2.070	-0.038	2.146	-0.115	2.041	-0.009	2.072	-0.040
<b>4</b>	2.337	2.286	0.051	2.136	0.202	2.486	-0.149	2.301	0.036
<b>5</b>	2.095	1.988	0.107	1.931	0.164	2.098	-0.003	2.065	0.030
<b>6</b>	2.017	2.163	-0.146	2.163	-0.146	2.097	-0.080	2.130	-0.113
<b>7</b>	2.133	1.878	0.255	2.090	0.043	2.044	0.089	2.109	0.024
<b>8</b>	1.182	1.683	-0.501	1.970	-0.789	1.602	-0.420	1.762	-0.580
<b>9</b>	2.228	1.828	0.400	2.026	0.201	1.791	0.437	1.938	0.290
<b>10</b>	2.234	2.202	0.032	2.188	0.046	2.170	0.064	2.221	0.013
<b>11</b>	2.154	2.067	0.087	1.919	0.235	2.091	0.063	2.108	0.046
<b>12</b>	2.006	2.010	-0.004	1.989	0.017	1.917	0.089	1.992	0.014
<b>13</b>	2.046	2.094	-0.048	2.145	-0.098	2.062	-0.016	2.136	-0.090
	Log (SK-N-SH)								

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

1	1.140	1.131	0.009	1.131	0.009	1.134	0.006	1.04	0.100
2	2.016	1.914	0.102	1.914	0.102	1.869	0.147	1.727	0.289
3	2.028	2.141	-0.113	2.141	-0.113	2.083	-0.055	2.027	0.001
4	2.210	2.382	-0.172	2.382	-0.172	2.311	-0.101	2.236	-0.026
5	2.164	2.277	-0.113	2.277	-0.113	2.364	-0.200	2.160	0.004
6	2.202	1.73	0.472	1.730	0.472	1.767	0.435	2.080	0.122
7	2.072	1.925	0.147	1.925	0.147	1.889	0.183	2.037	0.035
8	0.000	1.194	-1.194	1.194	-1.194	1.177	-1.177	1.799	-1.799
9	0.978	1.046	-0.068	1.046	-0.069	1.06	-0.082	0.983	-0.005
10	2.091	2.054	0.037	2.054	0.037	2.154	-0.063	2.06	0.031
11	1.960	1.439	0.521	1.439	0.520	1.447	0.513	1.967	-0.007
12	1.906	1.311	0.595	1.311	0.595	1.29	0.616	1.898	0.008
13	2.057	2.277	-0.220	2.277	-0.220	2.277	-0.220	2.263	-0.206

## IV. CONCLUSION

In this work we have investigated the QSAR regression to predict toxicity of several compounds based on imidazo[1,2-a]pyrazine derivatives.

Comparison of key statistical terms like R or R<sup>2</sup> of different models obtained by using different statistical tools and different descriptors has been shown in table 10.

The studies of the quality of the MLR, PLS, RNLM and ANN models have shown that:

- The PLS method gave low coefficients of determination (R<sup>2</sup>), thus it was had no efficiency in predicting the values of activities.
- The nonlinear regression and the neural network ANN results have substantially better predictive capability than the other methods.
- With ANN approach, we have established a relationship between several descriptors and inhibition values (IC<sub>50</sub>) of Imidazo[1,2-a]pyrazines on human tumor cell lines (Log (MDAMB-231) and Log (SK-N-SH)) in satisfactory manners.

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

## ACKNOWLEDGMENT

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

## REFERENCES

- [1] J.D Gough., L.H. Hall, Modeling the toxicity of amide herbicides using the electrotopological state, Environ. Toxicol. Chem., Vol. 18, pp. 1069–1075, 1999.
- [2] A. Sabljic, W.T. Piver, Quantitative modelling of environmental fate and impact of commercial, 1992.
- [3] Commission Of The European Communities, Strategy for a future Chemicals Policy, 2001.
- [4] W. Karcher, J. Devillers, SAR and QSAR in environmental chemistry and toxicology: scientific tool or wishful thinking In: Karcher, W., Devillers, J. (Eds.), Practical Applications of Quantitative Structure–Activity Relationships (QSAR) in Environmental Chemistry and Toxicology. Kluwer Academic, Dordrecht, the Netherlands, pp. 1–12, 1990.
- [5] S. Arulmozhiraja, M. Morita, Structure–activity relationships for the toxicity of polychlorinated dibenzofurans: approach through density functional theory-based descriptors, Chem. Res. Toxicol., Vol. 17, pp. 348, 2004.
- [6] M.T.D Cronin, N. Manga, J.R Seward, G.D. Sinks, T.W. Schultz, Parametrization of electrophilicity for the prediction of the toxicity of aromatic compounds, Chem. Res. Toxicol, Vol. 14, pp. 1498, 2001.
- [7] M.M. Gottesman, Cancer Res., Vol. 53, pp. 747, 1993.
- [8] D.L. Temple Jr., J.P. Yevich, J.D. Catt and All, J. Med. Chem., Vol. 23, pp. 1188–1198, 1980.
- [9] S.W. Schneller, J.K. Luo, J. Org. Chem, Vol. 45, pp. 4045–4048, 1980.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

- [10] R.A. Glennon, M.E. Rogers, J.D. Smith, M.K.E.J. Said, L. Egle, J. Med. Chem., Vol. 24, pp. 658–661, 1981.
- [11] Y. Rival, G. Grassy, G. Michel, Chem. Pharm. Bull. Vol. 40, pp. 1170–1176, 1992.
- [12] E. Abignente, F. Arena, P.De. Caprariis, L. Parente, II Farmaco, Ed. Sci., Vol. 30, pp. 815–822, 1975.
- [13] E. Abignente, F. Arena, P. De. Caprariis, II Farmaco, Ed. Sci., Vol. 36, pp. 61–80, 1981.
- [14] M.G. Rimoli, L. Avallone, P. De. Caprariis and All, Eur. J. Med. Chem., Vol. 32, pp. 195–203, 1997.
- [15] O. Vitse, P.A. Bonnet, J. Bompard, H. Viols, G. Subra, J.P. Chapat, G. Grassy, J. Heterocycl. Chem. Vol. 34, pp. 701–707, 1997.
- [16] P.J. Zimmermann, C. Brehm, W. Buhr, A.M. Palmer, J. Volz, W.A. Simon, Bioorg. Med. Chem., Vol. 16, pp. 536–541, 2008.
- [17] P.A. Bonnet, A. Michel, F. Laurent and All, J. Med. Chem., Vol. 35, pp. 3353–3358, 1992.
- [18] C. Sablayrolles, G.H. Cros, J.C. Milhavet, E. Rechencq, J.P. Chapat, M. Boucard, J.J. Serrano, J. Med. Chem., 27, pp. 206–212, 1984.
- [19] W.C. Lumma, W.C. Randall, E.L. Cresson, J.R. Huff, R.D. Hartman, T.F. Lyon, J. Med. Chem., Vol. 26, pp. 357–363, 1983.
- [20] L.C. Meurer, R.L. Tolman, E.W. Chapin, R. Saperstein, P.P. Vicario, M.M. Zrada, M. MacCoss, J. Med. Chem. Vol. 35, pp. 3845–3857, 1992.
- [21] K. Zurbonsen, A. Michel, P.A. Bonnet, M.N. Mathieu, C. Chevillard, Gen. Pharmac. Vol. 32, pp. 135–141, 1999.
- [22] A. Brown, A. Heenderson, C. Lane, M. Lansdell, G. Maw, S. Monaghan, Bioorg. Med. Chem. Lett. Vol. 16, pp. 4697–4700, 2006.
- [23] A. Michel, F. Lauret, J.P. Chapat, M. Boucard, P.A. Bonnet, Arzneimittel Forschung, Vol. 45, pp. 1288, 1991.
- [24] (a) Laurent, T. Pockock, K.R.F. Elliot, K. Portet, V. Benezech, O. Vitse, Fundam. Clin. Pharmacol. Vol. 10, pp. 65, 1996; (b) S.J. Cook, K. Archer, A. Martin and All, Br. J. Pharmacol. Vol. 114, pp. 143–151, 1995.
- [25] S.A. Mitchell, M.D. Danca, P.A. Blomgren, J.W. Darrow, K.S. Currie, J.E. Kropf, S.H. Lee, S.L. Gallion, M. J-Xiong, D.A. Pippin, R.W. DeSimone, D.R. Brittelli, A. Bourret, M. Hill-Dezewi, P.M. Maciejewski, L.L. Elkin, Bioorg. Med. Chem. Lett, Vol. 19, pp. 6991–6995, 2009.
- [26] A. Zahlou, S.Chtita, M. Ghamali, L. Bejjit, T. Lakhlifi and M. Bouachrine, Electronic and photovoltaic properties of new materials based on Imidazo[1,2-a]pyrazine Computational investigations, accepted in journal Functional Materials, 2013.
- [27] Shailaja M., Manjula A., Venkateshwarlu S., Vittal Rao B. Anthony A., Structure activity relationship studies of imidazo[1,2-a]pyrazine derivatives against cancer cell lines, Vol. 45, Issue 11, pp. 4697-5540, November 2010.
- [28] Adamo and Baron, 2000; Parac and Grimme, 2003; Gaussian 03, 2003.
- [29] Becke, 1993; Lee et al., 1988.
- [30] A. Adad, R. Hmammouchi, A. I. Taghki, A. Abdellaoui, M. Bouachrine and T.Lakhlifi : Atmospheric half-lives of persistent organic pollutants (POPs) study combining DFT and QSPR results, J. of Chemical and Pharmaceutical Research, Vol. 5, Issue 7, pp. 28-41, 2013.
- [31] Advanced Chemistry Development Inc., Toronto, Canada (2009). [www.acdlabs.com/resources/freeware/chemsketch/](http://www.acdlabs.com/resources/freeware/chemsketch/).
- [32] ACD/ChemSketch Version 4.5 for Microsoft Windows User's Guide.
- [33] Sakar U, Parthasarathi R, Subramanian V and Chattaraji PK, Toxicity analysis of polychlorinated dibenzofurans through global and local electrophilicities, Internet Electronic Journal of Molecular Design IECMD, pp. 1-24, 2004.
- [34] M. Larif, A. Adad, R. Hmammouchi, A.I. Taghki, A. Soulaymani, A. Elmidaoui, M. Bouachrine, T. Lakhlifi, Biological activities of triazine derivatives. Combining DFT and QSAR results, article in press in Arabian Journal of Chemistry, 2013.
- [35] XLSTAT 2009 version Demo 2009 Add-in software (XLSTAT Company). <http://www.xlstat.com>.
- [36] R. Hmammouchi, A. I. Taghki, M. Larif, A. Adad, A. Abdellaoui, M. Bouachrine and T. Lakhlifi: Combining DFT and QSAR result for predicting the biological activity of the phenylsuccinimide derivatives, J. of Chem. and Pharm. Research, Vol. 5, Issue 9, pp. 198-209, 2013.
- [37] J. Zupan, J. Gasteiger, Neural Networks for Chemists an Introduction, VCH Publishers Weinheim, 1993.
- [38] D. Cherqaoui, D. Villemin, Use of neural network to determine the boiling point of alkanes, J. Chem. Soc. Faraday. Trans., Vol 90, pp. 97-102, 1994.
- [39] Freeman, J. A., Skapura, D. M., Neural Networks Algorithms, Applications, and Programming Techniques. Addison Wesley Publishing Company, Reading, 1991.

## BIOGRAPHY



**Samir Chtita** was born on 12 December 1982 in Taounate. He is a professor of computer science in secondary schools since 2006. He holds a Bachelor degree in organic chemistry in 2005 and a Master degree in Chemistry and Structural Computing in 2011 from University college of Sidi Mohammed Ben Abdallah, Fez. He will prepare her PhD in Molecular Chemistry and Natural Substances Laboratory in Faculty of Science, University Moulay Ismail, Meknes.