

Prediction of biological activity of imidazo[1,2-a]pyrazine derivatives by combining DFT and QSAR results

Samir Chtita¹, Mounir Ghamali¹, Majdouline Larif², Azeddine Adad¹, Hmamouchi Rachid¹
Mohammed Bouachrine³ and Tahar Lakhli^{4*}

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

² Doctor, Laboratory of Separation Process Laboratory, Faculty of Science, University Ibn Tofail, Kenitra, Morocco

³ Professor, ESTM, University Moulay Ismail, Meknes, Morocco.

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

*Corresponding Author.

Abstract: Within the framework of the electron-topological approach the structure-activity relationship, for cytotoxic effects of two against different cancer cell lines (HepG-2: human hepatocellular liver carcinoma cell line and HCF-7: human breast adenocarcinoma cell line), was investigated in a series of thirteen imidazo[1,2-a]pyrazine derivatives by combining DFT and QSAR results using principal components analysis (PCA), multiple regression analysis (MRA), regression partial least squares (PLS), non-linear regression (RNLM) and neural network (NN). 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_{HOMO}), lowest unoccupied molecular orbital energy, (E_{LUMO}) difference between the LUMO and the HOMO energy (Gap), total dipole moment of the molecules (μ), absolute hardness (η), absolute electron negativity (χ) and reactivity index (ω)) were computed with ACD/ChemSketch and Gaussian 03W program, respectively. We accordingly propose a quantitative model, and we interpret the activity of the compounds relying on the multivariate statistical analysis.

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

I. INTRODUCTION

Breast cancer is the most commonly diagnosed non-skin cancer and ranks second on the list of cancer-related death in women [1]. Even with the recent progress in therapeutic regimens against breast cancer, 226870 American women are still being diagnosed with breast cancer and 39510 women will die from this disease as a result of metastasis annually [2], accounting for 23% of the total cancer cases and 14% of cancer related deaths. Approximately 25–50% of breast cancer cases eventually will develop metastasis. Once metastasis is diagnosed, the 5 year survival is generally less than 25% even with aggressive treatment [3, 4]. Thus, developing a new therapeutic strategy either to treat metastasis or to prevent breast cancer metastasis from occurring should be a priority.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 12, December 2013

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 [5] 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 (Fig.1). Many imidazo[1,2-a]pyrazine derivatives have been invented to pharmaceutical compositions such compounds, to processes to prepare such compounds and compositions, to the use of such compounds or pharmaceutical compositions for the prevention or treatment of neurological, psychiatric and metabolic disorders and diseases [6]. They have also been shown to inhibit the receptor tyrosine kinase EphB₄ [7].

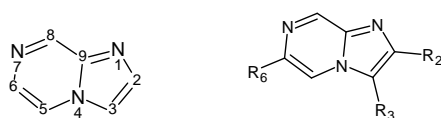


Fig. 1: Schematic diagram of imidazo[1,2-a]pyrazine skeleton.

Our group has been recently investigated the QSAR regression to predict two against different cancer cell lines activities (MDAMB-231 and SK-N-SH) and attempted to establish a quantitative structure-activity relationship for cytotoxic effects of this cancer cell lines [8] by studying a series of 13 substituted imidazo[1,2-a]pyrazines [9] have been synthesized with substitutions at 2, 3, 6-ring positions being varied generating mono-, di-, tri-substituted imidazo[1,2-a]pyrazines possessing functional groups like halo, hydroxymethyl, amine, alkyl, aryl, etc...[27]. In the current study, we attempt to establish a quantitative structure-activity relationship for cytotoxic effects of two other against cancer cell lines (HepG-2 and MCF-7), by studying the same series of thirteen imidazo[1,2-a]pyrazine derivatives table 1 [9].

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 [8].

II. MATERIAL AND METHODS

A. Experimental data

The experimental IC₅₀ (μM) cytotoxic effects of two against different cancer cell lines activities (HepG-2 and HCF-7) of imidazo[1,2-a]pyrazine derivatives (Fig. 1) are collected from recent publications [10]. The observations are converted into minus logarithm scale logIC₅₀ and are included in table 1.

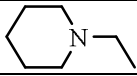
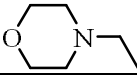
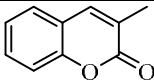
TABLE 1
LABELS AND INHIBITION VALUES (IC₅₀) OF STUDY IMIDAZO[1,2-A]PYRAZINES DERIVATIVES ON HUMAN TUMOR CELL LINES.

IC ₅₀ values against cancer cell lines (μM)					
	R ₂	R ₃	R ₆	Log (HepG-2)	Log (MCF-7)
1	p-F-C ₆ H ₄ -	Br-	CH ₃ -	2.789	1.111

International Journal of Innovative Research in Science,
Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 12, December 2013

2	p-F-C ₆ H ₄ -	N≡C-	CH ₃ -	1.998	1.890
3	C ₆ H ₅ -		CH ₃ -	2.902	1.993
4	p-F-C ₆ H ₄ -		CH ₃ -	2.940	2.367
5		HOCH ₂ -	CH ₃ -	2.931	2.367
6	p-F-C ₆ H ₄ -	H-	CH ₃ -	2.028	1.984
7	p-F-C ₆ H ₄ -	H-	H-	2.867	2.149
8	C ₆ H ₅ -	Br-	CH ₃ -	1.786	0.792
9	p-CH ₃ -C ₆ H ₄ -	Br-	CH ₃ -	1.930	1.396
10	t-Bu	Br-	CH ₃ -	2.090	2.033
11	C ₆ H ₅ -	Br-	H-	2.036	2.744
12	p-F-C ₆ H ₄ -	Br-	H-	1.943	1.909
13	t-Bu	Br-	H-	2.279	2.451

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 [11] using the B3LYP functional [12] 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 [13].

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) [14] was employed to calculate the others molecular descriptors [8].

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_{HOMO} (eV)), the lowest unoccupied molecular orbital energy (E_{LUMO} (eV)), the energy difference between the LUMO and the HOMO energy (Gap (eV)), the total dipole moment of the molecule (μ(Debye)), absolute hardness (η), absolute electron negativity (χ) and reactivity index (ω) [15]. η, χ and ω were determined by the following equations:

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 12, December 2013

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

Calculation of descriptors using ACD/ChemSketch

Advanced chemistry development's ACD/ChemSketch program [16] was used to calculate Formula Weight (PM), Molar Volume (MV (cm³)), Molecular Weight (MW), Molar Refractivity (MR (cm³)), Parachor (Pc (cm³)), Density (D (g/cm³)), Refractive Index (n), Surface Tension (γ (dyne/cm) and Polarizability (αe (cm³)) [17].

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) [18] using the software XLSTAT. 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 [19]. 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, to predict cytotoxic effects IC₅₀ activities. Equations were justified by the correlation coefficient (R), mean squared error (MSE), fishers F-statistic (F), and significance level (F value) [20].

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. A number of individual models of ANN were designed built up and trained. Generally the network was built for tree layers; one input layer, one hidden layer and one output layer were considered [21]. The input layer was consisted of 13 artificial neurons of linear activation function. The number of artificial neural in the hidden layer was adjusted experimentally. The hidden layer consisted of 20 artificial neural. Two neurons formed the output layer of sigmoid function activation. The architecture of the applied ANN models is presented in Figure 2.

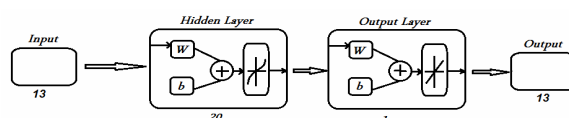


Fig. 2: The ANNs architecture.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 12, December 2013

The data subjected to ANN analysis was randomly divided into three sets: a learning set, a validation set and a testing set. Prior to that, the whole data set was scaled within the 0 to 1 range.

The set of HepG-2, and MCF-7 IC₅₀ activities were subjected to the ANN analysis. First, for the learning set of compounds, 9 inhibitors were used. ANN models were designed, built and trained. The learning set of data is used in ANNs to recognize the relationship between the input and output data. Then for the revision of ANN model designed and selected, the validation set of 2 compounds was used. Testing set with 2 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 [22]. The operation of the output layer is linear, which is given as below:

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

Where y_k is the k^{th} output layer unit for the input vector X , w_{kj} is the weight connection between the k^{th} output unit and the j^{th} hidden layer unit and b_k is the bias allows a transfer function “non-zero” given by the following equation:

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

Where y is the measured value and \bar{y} is the value predicted by the model.

The accuracy of the model was mainly evaluated by Root Mean Square Error (RMSE). Formula as follows:

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

Where n = number of compounds, p_{exp} = experimental value, p_{pred} = predicted value and summation is over all patterns in the analysed data set [23, 24]. The scripts were run on a personal PC.

III. RESULTS AND DISCUSSION

A. Data set for analysis

The QSAR analysis was performed using the IC₅₀ of the 13 compounds against the HepG-2, and MCF-7 cells (experimental values) as reported in [10], 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) [8].

TABLE 2:

THE VALUES OF THE SIXTEEN CHEMICAL DESCRIPTORS

	PM	MR	MV	Pc	n	γ	D	α_e	E	E _{HOMO}	E _{LUMO}	Gap	μ	χ	η	ω
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

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 12, December 2013

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 (PCA)

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.

The experimental matrix is composed of (16x16) by a principal component analysis (PCA) we will interpret the parameters that are highly correlated.

- **Gap** and η 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_{LUMO}** and ω are strongly negatively correlated ($r = -0.998$).

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

In the projection of the compounds in the plane of the three first axes F1, F2 and F3 (Fig. 3), 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 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 [8].

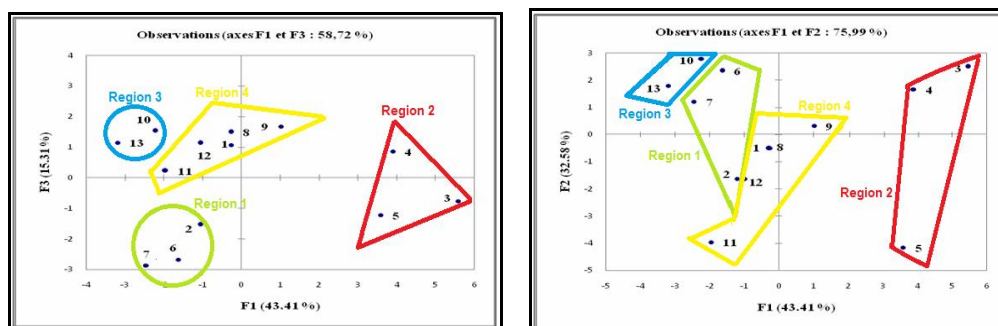


Fig. 3: Cartesian diagram according to F1F2 and F1F3: Separation between four regions

C. Multiple linear regression (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^2 , and the F-values to select the best regression performance.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 12, December 2013

Where R is the correlation coefficient; R² 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, γ and μ for **HepG-2**; γ and n for **MCF-7**.

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

- $\text{Log (HepG-2)} = 36.765 + 9.408 \cdot 10^{-03} \text{ PM} - 27.838 \text{ n} + 0.161 \gamma + 0.628 \mu$. (equation 1)

- $\text{Log (MCF-7)} = 53.909 - 38.066 \text{ n} + 0.240 \gamma$. (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 HepG-2 and MCF-7, it means that we would be taking, respectively, a lower than 0.28% and 0.29% 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. (Table 3 and 4)

TABLE 3:
ANALYSIS OF VARIANCE (GLOBAL MODEL)

	HepG-2					
	Source	DDL	Sum of squares	Mean squares	F	Pr > F
	Model	4	1.775	0.444	4.825	0.028
	Error	8	0.736	0.092		
	total corrected	12	2.510			
	MCF-7					
	Source	DDL	Sum of squares	Mean squares	F	Pr > F
	Model	2	1.848	0.924	5.178	0.029
	Error	10	1.785	0.178		
	total corrected	12	3.633			

TABLE 4
CORRELATION COEFFICIENT (R), COEFFICIENT OF DETERMINATION (R²), MEAN SQUARED ERROR (MSE), FISHERS F-STATISTIC (F) AND SIGNIFICANCES LEVEL (F VALUE)

	Log(HepG-2)	Log(MCF-7)
R²	0.707	0.509
R	0.841	0.713
MSE	0.092	0.178
F	4.825	5.178
F value	0.028	0.029

The values of predicted activities (Log (HepG-2) and Log (MCF-7)) 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 4.

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 (MNLN) and artificial neural network (ANN).

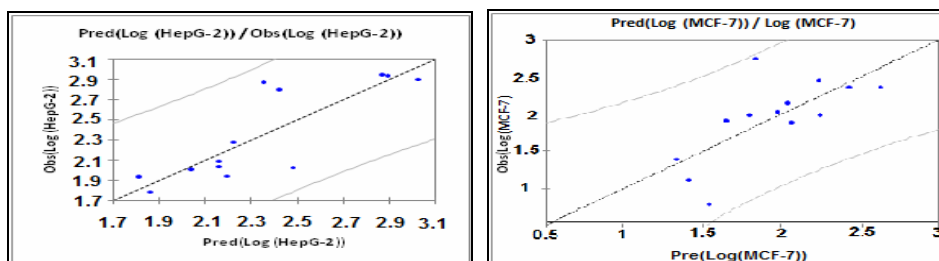


Fig. 4: the correlations of observed and predicted activities calculated using MLR

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 that the explanatory variables are correlated. [8]

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, R², 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:

- $\text{Log (HepG-2)} = -0.177 + 6.306 \cdot 10^{-03} \text{ PM} - 0.996 \text{ n} + 0.027\gamma + 0.426 \mu.$ (equation 3).
- $\text{Log (MCF-7)} = 49.688 - 35.504 \text{ n} + 0.239 \gamma.$ (equation 4)

The correlation coefficient (R), the coefficient of determination (R²), the Mean Squared Error (MSE) and Standard deviation (S) for the two models. (Table 5)

TABLE 5:
CORRELATION COEFFICIENT (R), COEFFICIENT OF DETERMINATION (R²), MEAN SQUARED ERROR (MSE) AND STANDARD DEVIATION (S).

	Log(HepG-2)	Log(MCF-7)
R²	0.291	0.491
R	0.540	0.701
MSE	0.137	0.142
S	0.402	0.410

The values of predicted activities (Log (HepG-2) and Log (MCF-7)) 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 5.

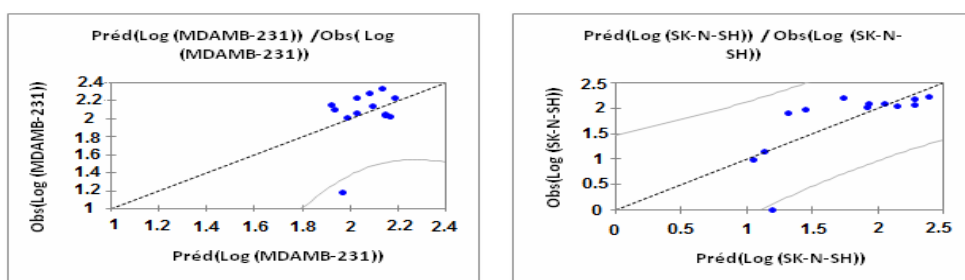


Fig. 5: Correlations of observed and predicted activities calculated using PLS

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 12, December 2013

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 (MNLr)

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 to the data matrix constituted obviously from the descriptors proposed by MLR corresponding to the 13 molecules. The coefficients R, R², 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 X₁, X₂, X₃, X₄,... : represent the variables.

The resulting equations were:

- Log (HepG-2) = -380.741 - 0.208 PM + 505.997 n - 0.758 γ + 8.469μ + 3.890 10⁻⁰⁴ PM² - 153.499 n² + 7.892 10⁻⁰³ γ² - 1.479 μ² (equation 5).
- Log (MCF-7) = 1332.832 - 1602.410 n + 1.301 γ + 468.812 n² - 1.086 10⁻⁰² γ² (equation 6).

The correlation coefficient (R), the coefficient of determination (R²), the Mean Squared Error (MSE) and Standard deviation (S) for the two models. (Table 6)

TABLE 6
CORRELATION COEFFICIENT (R), COEFFICIENT OF DETERMINATION (R²), AND MEAN SQUARED ERROR (MSE)

	Log(HepG-2)	Log(MCF-7)
R²	0.956	0.625
R	0.978	0.791
MSE	0.110	1.361

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 6.

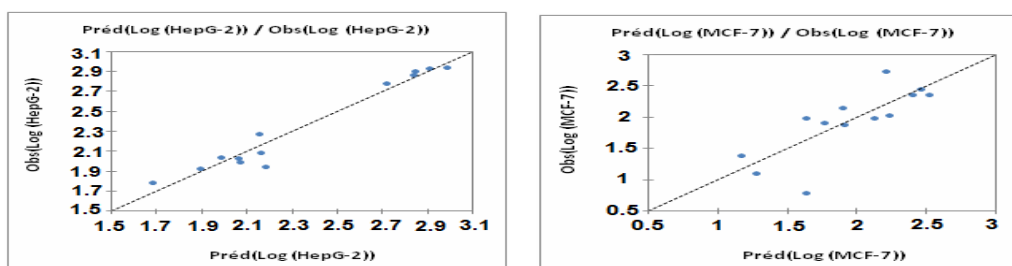


Fig.6: Correlations of observed and predicted activities calculated using MNLr

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 7), 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 12, December 2013

TABLE 7:
CORRELATION COEFFICIENT (R) AND COEFFICIENT OF DETERMINATION (R²)

	Log(HepG-2)	Log(MCF-7)
R²	0.872	0.992
R	0.934	0.996

The values of predicted activities and the observed values are given in table 8. The correlations of predicted and observed are illustrated in figure 7.

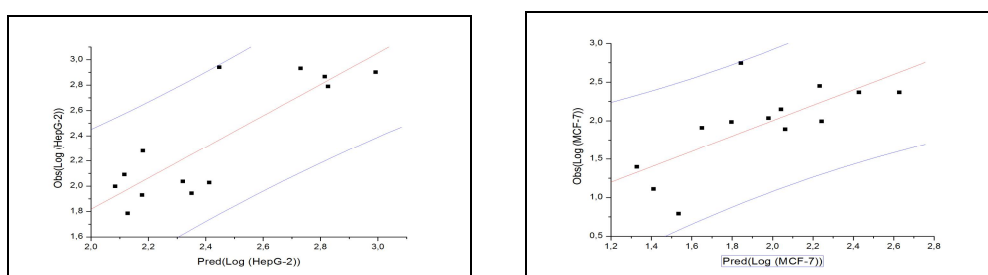


Fig.7: Correlations of observed and predicted activities calculated using ANN

The obtained squared correlation coefficient (R²) 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.

TABLE 8:
OBSERVED, PREDICTED ACTIVITIES AND RESIDUE ACCORDING TO DIFFERENT METHODS.

IC ₅₀ values against cancer cell lines (µM)									
	Log (HepG-2)								
	Observed	MLR		PLS		MNLR		ANN	
		Pred.	Resid.	Pred.	Resid.	Pred.	Resid.	Pred.	Resid.
1	2.789	2.423	0.366	2.624	0.164	2.720	0.069	2.826	-0.037
2	1.998	2.043	-0.045	2.079	-0.081	2.069	-0.071	2.084	-0.086
3	2.902	3.030	-0.128	2.654	0.248	2.847	0.055	2.992	-0.09
4	2.940	2.873	0.067	2.444	0.496	2.988	-0.048	2.447	0.493
5	2.931	2.891	0.040	2.710	0.221	2.915	0.016	2.731	0.200
6	2.028	2.479	-0.451	2.361	-0.333	2.064	-0.036	2.412	-0.384
7	2.867	2.352	0.516	2.265	0.603	2.845	0.023	2.815	0.052
8	1.786	1.864	-0.078	2.277	-0.491	1.681	0.105	2.128	-0.342
9	1.930	1.817	0.113	2.237	-0.307	1.894	0.036	2.178	-0.248
10	2.090	2.161	-0.071	1.958	0.131	2.160	-0.071	2.116	-0.026
11	2.036	2.164	-0.128	2.468	-0.431	1.989	0.047	2.320	-0.284
12	1.943	2.197	-0.254	2.472	-0.530	2.186	-0.244	2.350	-0.407
13	2.279	2.225	0.053	1.970	0.309	2.159	0.120	2.181	0.098
	Log (MCF-7)								
1	1.111	1.410	-0.300	1.436	-0.325	1.270	-0.160	1.410	-0.299
2	1.890	2.063	-0.173	2.053	-0.163	1.902	-0.012	2.063	-0.173
3	1.993	2.244	-0.251	2.218	-0.225	2.120	-0.127	2.244	-0.251
4	2.367	2.428	-0.061	2.382	-0.015	2.397	-0.030	2.428	-0.061
5	2.367	2.628	-0.261	2.739	-0.372	2.512	-0.145	2.628	-0.261

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 12, December 2013

6	1.984	1.797	0.187	1.735	0.249	1.631	0.353	1.797	0.187
7	2.149	2.042	0.107	2.014	0.134	1.893	0.256	2.042	0.107
8	0.792	1.535	-0.743	1.596	-0.804	1.634	-0.842	1.535	-0.743
9	1.396	1.328	0.068	1.352	0.044	1.161	0.235	1.328	0.068
10	2.033	1.980	0.052	1.858	0.175	2.231	-0.198	1.980	0.053
11	2.744	1.844	0.900	1.949	0.795	2.212	0.533	1.844	0.900
12	1.909	1.651	0.258	1.716	0.193	1.765	0.144	1.651	0.258
13	2.451	2.234	0.217	2.136	0.315	2.456	-0.006	2.234	0.217

IV. CONCLUSION

In this work, we have investigated the QSAR regression to predict two against different cancer cell lines activities Log (HepG-2) and Log (MCF-7) of several compounds based on imidazo[1,2-a]pyrazine derivatives, and we have compared the key statistical terms like R or R² 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 (R²), 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 (IC₅₀) of Imidazo[1,2-a]pyrazines on human tumor cell lines (Log (HepG-2) and Log (MCF-7)) 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] R. Siegel, D. Naishadham, A. Jemal, Cancer statistics, CA: Cancer J. Clin., Vol. 63, pp. 11–30, 2013.
- [2] R. Siegel, D. Naishadham, A. Jemal, Cancer statistics, CA: Cancer J. Clin., Vol. 62(1), pp. 10–29, 2012.
- [3] S.A. Rabbani, A.P. Mazar, Evaluating distant metastases in breast cancer, Cancer Metastasis Rev., Vol. 26, pp. 663-674, 2007.
- [4] S. Valastyan, R.A. Weinberg, Tumor metastasis: molecular insights and evolving paradigms, Cell, Vol. 147 (2), pp. 275–292, 2011.
- [5] M.M. Gottesman, Cancer Res., Vol. 53, p. 747, 1993.
- [6] Bartolomé-Nebreda and All, Imidazo[1,2-a]pyrazine derivatives and their use for the prevention or treatment of neurological, psychiatric and metabolic disorders and diseases, United States Patent Application Publication, Dec. 27, 2012.
- [7] S.A. Mitchell, M.D. Danca, P.A. Blomgren, J.W. Darrow, and All, Med. Chem. Lett., Vol. 19, pp. 6991–6995, 2009.
- [8] S. Chtita, M. Larif, M. Ghamali and All, 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, IJIRSET, Vol. 2, Issue 11, 6586-6601, 2013.
- [9] A. Zahlou, S.Chtita, M. Ghamali and All, Electronic and photovoltaic properties of new materials based on Imidazo[1,2-a]pyrazine. Computational investigations, accepted in journal Functional Materials, 2013.
- [10] Shailaja M., Manjula A., Venkateswarlu S. and All, Structure activity relationship studies of imidazo[1,2-a]pyrazine derivatives against cancer cell lines, 2010.
- [11] Adamo and Baron, 2000; Parac and Grimme, 2003; Gaussian 03, 2003.
- [12] Becke, 1993; Lee et al., 1988.
- [13] A. Adad, R. Hmamouchi, A. I. Taghki, and al., Atmospheric half-lives of persistent organic pollutants (POPs) study combining DFT and QSPR results, J. of Chemical and Pharmaceutical Research, Vol. 5(7), pp. 28-41, 2013.
- [14] Advanced Chemistry Development Inc., Toronto, Canada (2009). www.acdlabs.com/resources/freeware/chemsketch/.
- [15] 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.
- [16] Advanced Chemistry Development Inc., Toronto, Canada (2009). www.acdlabs.com/resources/freeware/chemsketch/.
- [17] ACD/ChemSketch Version 4.5 for Microsoft Windows User's Guide.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 12, December 2013

- [18] Larif, M. ; Adad, A. ; Hmamouchi, R. and al., Biological activities of triazine derivatives. Combining DFT and QSAR results, article in press in Arabian Journal of Chemistry, 2013.
- [19] R. Hmamouchi, A. I. Taghki, M. Larif and al., Combining DFT and QSAR result for predicting the biological activity of the phenylsuccinimide derivatives, J. of Chemical and Pharmaceutical Research, Vol. 5(9), pp. 198-209, 2013.
- [20] XLSTAT 2009 version Demo 2009 Add-in software (XLSTAT Company). <http://www.xlstat.com>.
- [21] Zupan, J., Gasteiger, J., second ed VCH, Weinheim, 1999.
- [22] Turkkán N., Revue de l'Université de Moncton, Vol. 26 (1), pp. 205-221, 1993.
- [23] Lee, P. Y., Chen C. Y. J., Hazard. Mater. Vol. 165, pp. 156-161, 2009.
- [24] Jing, G., Zhou, Z., Zhuo, J., Chemosphere, Vol. 86, pp. 76-82, 2012.

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 Abdelah, Fez. He will prepare her PhD in Molecular Chemistry and Natural Substances Laboratory in Faculty of Science, University Moulay Ismail, Meknes.