Volume 4, Issue 9, September 2014 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Anticancer Activity of Novel Molecules Based on Imidazo [4, 5-B] Pyridine. 3D-QSAR Study

Majdouline Larif* Faculty of Science, University Ibn Tofail - Kenitra

Samir Chtita Faculty of Science, University Moulay Ismail, Meknes

Azeddine Adad Faculty of Science, University Moulay Ismail, Meknes

Rachid Hmamouchi Faculty of Science, University Moulay Ismail, Meknes

Mohammed Bouachrine ESTM, University Moulay Ismail, Meknes, Morocco

Tahar Lakhlifi Faculty of Science, University Moulay Ismail, Meknes

Abstract-Our objective is to study the relationship between the activities and structure, a 3D-QSAR study is applied to a set of 20 molecules for biological activity prediction derivatives. This study was conducted using the principal component analysis PCA method, the multiple linear regression method MLR and the artificial neural network ANN. The predicted values of activities are in good agreement with the experimental results. ACP is observed in the distribution of different molecules into two different groups. The relevant descriptors obtained from the ANN showed a correlation coefficient of 0.9606 models which is a good result. As a result of quantitative structure–activity relationships, we found that the model proposed in this study is constituted of major descriptors used to describe these molecules. The obtained results suggested that the proposed combination of several calculated parameters could be useful to predict the biological activity of derivatives of imidazo[4,5-b]pyridine.

Keywords: Biological activity; 3D-QSAR; PCA; MLR; ANN; DFT study.

I. INTRODUCTION

The Aurora kinases are a family of highly conserved serine/threonine protein kinases that play a key role in regulating many pivotal processes of mitosis and completion of cell division [1-2]. The two major Aurora kinases, Aurora A and Aurora B, play distinct roles in mitosis, though they are very closely related in kinase domain sequence (71% identical) and have the identical residues lining the binding pocket for the ATP adenine ring [3]. Aurora A is involved in centrosome maturation and separation, bipolar spindle assembly, and mitotic entry

Cancer is a major public health problem and leading cause of death in many parts of the world. Deaths from cancer worldwide are projected to continue rising, with an estimated 13.1 million deaths in 2030 [4].

Quantitative structure-activity relationship (QSAR), as an important area of chemometrics, has been the subject of a series of investigations [5].

The main aim of QSAR studies is to establish an empirical rule or function relating the structural descriptors of compounds under investigation to bioactivities. This rule or function is then utilized to predict the same bioactivities of the compounds not involved in the training set from their structural descriptors. Whether the bioactivities can be predicted with satisfactory accuracy depends to a great extent on the performance of the applied multivariate data analysis method, provided the property being predicted is related to the descriptors. Many multivariate data analysis methods such as principal components analysis (PCA) and artificial neural network (ANN) have been used in QSAR studies. ANN offers satisfactory accuracy in most cases but tends to over fit the training data. There are a large number of molecular descriptors that can be used in QSAR studies. Once validated, the findings can be used to predict activities of untested compounds. Recently, computer-assisted drug design based on QSAR has been successfully employed to develop new drugs for the treatment of cancer and other diseases [6].

After a QSAR model is built and validated, it can predict the biological activity of novel molecules room their structural properties. A QSAR model can also screen potentially active molecules from a database, as described in the section on applications of the technique. Because the QSAR model can incorporate a wide range of different variables, be it physical, chemical or biological, it can also be utilized in industries apart from drug design [7], such as toxicology [8], food chemistry [9] and other fields.

II. MATERIAL AND METHODS

Material

All of the 20 compounds and associated data involved in this study were obtained from literature by [10]. The inhibitory activity data were reported as IC50 against Aurora A. The IC50 values were converted into pIC50 according to the formula in equation 1[11]. The structures of the molecules are shown and associated inhibitory activities are shown in table 1, where pI50 values for 20 inhibitors ranged from 4.538 to 7.398.

The data set implemented in this work consists of 20 imidazo[4,5-b]pyridine derivatives shown in figure 1. Half maximal inhibitory concentration (IC_{50}) is the quantity used for measuring in vitro potency of these compounds in inhibiting Aurora A kinase function.

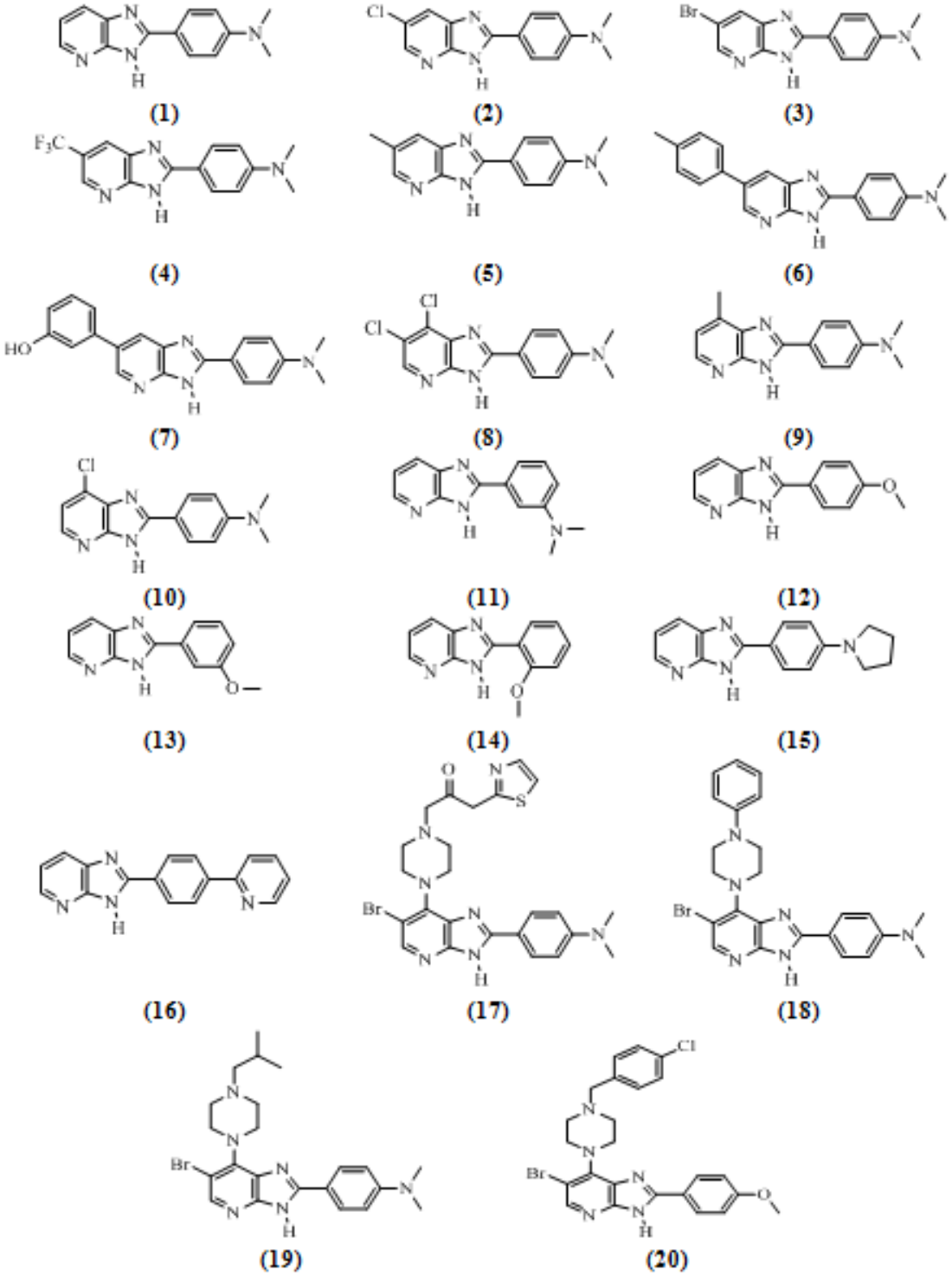

Fig. 1: Structures of imidazo[4,5-b]pyridine derivatives

Computational methods (DFT calculations)

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 [12-15], using the B3LYP functional [15,16] and a 6-31G* 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.

Calculation of molecular descriptors using gaussian 03W

From the results of the DFT calculations, the quantum chemical descriptors were obtained for the model building as follows: the total energy ($\mathbf{E_T}$ (u.a.)), the highest occupied molecular orbital energy ($\mathbf{E_{HOMO}}$ (eV)), the lowest unoccupied molecular orbital energy ($\mathbf{E_{LUMO}}$ (eV)), the energy difference between the LUMO and the HOMO energy (**Gap** (eV)), absorption maximum $\boldsymbol{\lambda_{max}}$, the total dipole moment of the molecule ($\boldsymbol{\mu}$ (Debye)), absolute hardness ($\boldsymbol{\eta}$), absolute electron negativity ($\boldsymbol{\chi}$) and reactivity index ($\boldsymbol{\omega}$) [17]. ($\boldsymbol{\eta}$) and ($\boldsymbol{\chi}$) and ($\boldsymbol{\omega}$) were determined by the following equations:

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

Principal components analysis

Twenty molecules were studied by statistical methods based on the principal component analysis (PCA) [18,19] using the software XLSTAT 2009.

Essentially a descriptive statistical method which aims to present, in graphic form, the maximum of information contained in the data (Table 1).

PCA is a statistical technique useful for summarizing all the information encoded in the structures of compounds. It is also very helpful for understanding the distribution of the compounds.

Multiple linear and nonlinear regressions (MLR and MNLR)

The multiple linear and nonlinear regression statistics techniques are used to study the relation between one dependent variable and several independent variables. The multiple linear and non linear regression models (MLR and MNLR) are generated using the software XLSTAT, version 2009, to predict pIC_{50}.

The optimal number of components (N) is employed to do validation MLR and MNLR analysis to get the final model parameters such as correlation coefficient R^2, standard deviation (S) and Fischer test value (F) [20].

Artificial neural networks (ANN)

The ANN analysis was performed with the use of Matlab software version 7.0 using a program written in C language Neural toolbox on a data set of structures of imidazo[4,5-b]pyridine derivatives [21,22]. A number of individual models of ANN were designed built up and trained. Generally the network was built for three layers; one input layer, one hidden layer and one output layer were considered [23]. The input layer was consisted of eight artificial neurons of linear activation function (Figure 2). The number of artificial neural in the hidden layer was adjusted experimentally. The hidden layer consisted of 20 artificial neural. One neuron formed the output layer of sigmoid function activation. The architecture of the applied ANN models is presented in (Figure 3).

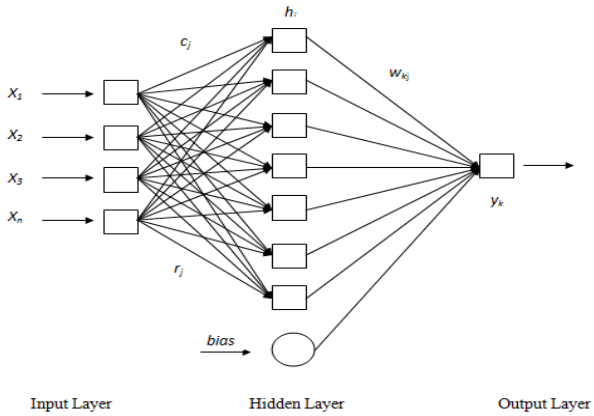

Fig. 2: Neuron Layout of ANN

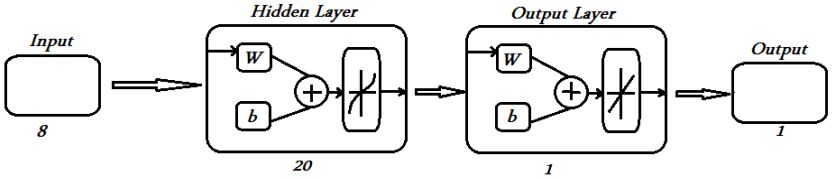

Fig. 3: The ANN architecture.

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

The set of structures of imidazo[4,5-b]pyridine derivatives [24] was subjected to the ANN analysis. First, for the learning set of compounds, i.e., Selected training set of imidazo [4,5-b]pyridine derivatives. The learning set of data is used in ANN to recognize the relationship between the input and output data. Then for the revision of the ANN model designed and selected, the validation set of 20 compounds was used. Testing set with eight 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 [25].

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 that 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 the root mean square error (RMSE). Formula is given as follows:

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

Where n = number of compounds, p_{exp} = experimental value, p_{pred} = predicted value and summation is of overall patterns in the analyzed data set [26,27]. The scripts were run on a personal PC.

III. RESULTS AND DISCUSSION

A QSAR study was carried for a series of 20 of imidazo[4,5-b]pyridine derivatives, in order to determine a quantitative relationship between structure and toxicity.

Table 1 shows the values of the calculated parameters obtained by DFT/B3LYP 6-31G* optimization of the studied Selected training set of imidazo[4,5-b]pyridine derivatives.

Table 1: Values of the twelve chemical descriptors.

Mol.	pI_{50}	Et	E_{HOMO}	E_{LUMO}	Gap	μ	χ	η	ω	Ea	λ_{max}	$f_{(SO)}$
1	5,367	-20720	-5,084	-1,009	4,075	5,04	3,046	2,037	2,277	3,8237	324,25	0,9747
2	5,244	-33235	-5,239	-1,272	3,966	7,6115	3,256	1,983	2,672	3,7197	333,32	0,9689
3	6,31	-90732	-5,237	-1,273	3,964	7,5126	3,255	1,982	2,672	3,7032	334,81	1,002
4	6,131	-29898	-5,308	-1,346	3,963	8,4867	3,327	1,981	2,793	3,7422	331,31	0,9399
5	5,161	-21791	-5,025	-0,957	4,068	4,5604	2,991	2,034	2,199	3,8127	325,19	1,0282
6	4,699	-30131	-5,024	-1,021	4,003	4,28	3,022	2,002	2,282	3,771	328,79	1,1543
7	5,046	-29060	-5,114	-1,116	3,997	6,631	3,115	1,999	2,428	3,691	335,91	1,1647
8	6,602	-45750	-5,32	-1,427	3,893	8,2146	3,373	1,946	2,923	3,632	341,37	0,9368
9	5,155	-21791	-5,03	-0,942	4,088	4,8939	2,986	2,044	2,182	3,8504	322,01	1,0083
10	5,62	-33235	-5,202	-1,217	3,985	6,2624	3,21	1,993	2,585	3,7207	333,22	0,9434
11	5,001	-206958	-5,275	-1,222	4,053	3,0197	3,248	2,026	2,604	4,1651	297,68	0,8018
12	5,18	-20191	-5,6	-1,222	4,378	2,7956	3,411	2,189	2,658	4,1578	298,2	0,8591
13	4,796	-20191	-5,851	-1,376	4,475	3,0207	3,613	2,237	2,918	4,1502	298,74	0,6459
14	4,538	-20191	-5,674	-1,274	4,4	3,7962	3,474	2,2	2,743	4,1276	300,38	0,6805
15	5,337	-22829	-4,987	-0,956	4,031	5,666	2,972	2,015	2,191	3,7784	328,14	1,0393
16	5,357	-23801	-5,799	-1,778	4,021	2,7693	3,788	2,01	3,569	4,1663	297,59	0,9855
17	7,26	-118684	-5,114	-1,125	3,989	6,5602	3,119	1,994	2,439	3,7063	334,53	0,9171
18	6,588	-104286	-4,899	-1,117	3,782	7,0894	3,008	1,891	2,392	3,7244	332,9	0,8643
19	6,678	-102277	-5,054	-1,035	4,019	6,6662	3,044	2,009	2,306	3,6773	337,16	0,859
20	7,398	-117342	-5,347	-1,332	4,016	7,067	3,339	2,008	2,777	4,0574	305,58	0,738

Principal component analysis (training set selection)

The selection of the training set is one of the most important steps in the QSAR modeling, since the establishment and optimization of a QSAR model are based on this training set. Predictability and applicability of a QSAR model also depend on the training set selection. In this part, PCA was applied to select a training set from among 20 compounds.

The set of descriptors encoding the 20 compounds and electronic and energetic parameters are submitted to PCA analysis (XSLAT 2009).

10 principal components were obtained.The first two principal axes are sufficient to describe the information provided by the data matrix. Indeed, the percentages of variance are 45,32% and 34,33% for the axes F1 and F2 respectively. The total information is estimated to a percentage of 79.65%. The principal component analysis (PCA) [22-29] was conducted to identify the link between the different variables. Bold values are different from 0 at a significance level of p = 0.05.

Correlations between the twenty descriptors are shown in table 2 as a correlation matrix, in figures 4 these descriptors are represented in a correlation circle.

The Pearson correlation coefficients are summarized in the following (Table 2). The obtained matrix provides information on the negative or positive correlation between variables.

Table 2: Correlation matrix (Pearson (n)) between different obtained descriptors

Variables	pI_{50}	Et	E_{HOMO}	E_{LUMO}	Gap	μ	χ	η	ω	Ea	λ_{max}	$f_{(SO)}$
pI_{50}	**1**											
Et	**-0,605**	**1**										
E_{HOMO}	0,090	-0,199	**1**									
E_{LUMO}	-0,161	0,095	**0,877**	**1**								
Gap	**-0,603**	**0,678**	-0,231	0,199	**1**							
μ	**0,650**	**-0,499**	0,216	-0,103	**-0,814**	**1**						
χ	-0,023	0,141	**-0,974**	**-0,931**	0,090	-0,143	**1**					
η	**-0,603**	**0,678**	-0,231	0,199	**1,000**	**-0,814**	0,090	**1**				
ω	0,143	-0,067	**-0,892**	**-0,998**	-0,184	0,099	**0,942**	-0,184	**1**			
Ea	**-0,484**	0,442	-0,409	-0,148	**0,728**	**-0,795**	0,320	**0,728**	0,141	**1**		
λ_{max}	**0,484**	-0,442	0,409	0,148	**-0,728**	**0,795**	-0,320	**-0,728**	-0,141	**-1,000**	**1**	
$f_{(SO)}$	-0,165	0,155	**0,572**	**0,525**	-0,200	0,077	**-0,559**	-0,200	**-0,506**	-0,283	0,283	**1**

Bold values are different from 0 at a level significant for p < 0.05
At a very significant for p < 0,01
At a highly significant to p < 0,001

Correlation circle

Principal component analysis (PCA) was also performed to detect the connection between the different variables. The principal component analysis revealed from the correlation circle (Figure 4) shows that the F1 axis (45.32%) presents the energy of the variance while the axis F2 (34.33%) of the variance is located by the other parameters of energy [17].

χ and **E_{LUMO}** are perfectly correlated (r = 0,974), both variables are redundant.

λ_{max} and □**a** are strongly negatively correlated (r = -1).

E_{LUMO} and **ω** are strongly negatively correlated (r = 0,998).

The following variables then removed are: **Gap,**□ **λ_{max}** and **ω**□

On the other hand, the correlation circle (Figure 4) indicates the correlation between electronic descriptors:

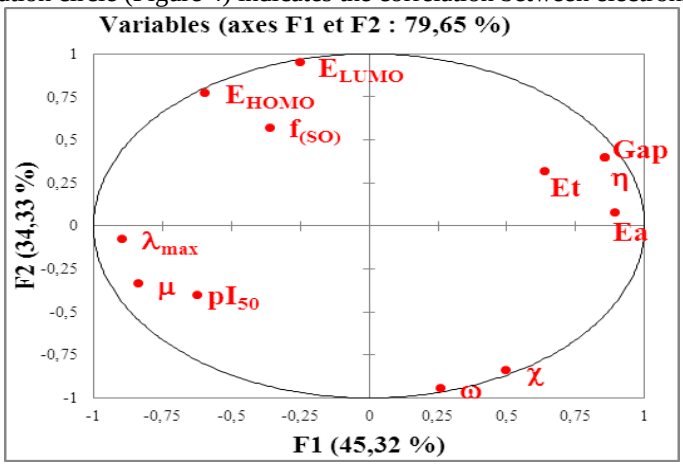

Fig. 4: Correlation circle

The Cartesian diagram (Figure 5) analyses of projections according to the plane **F1–F2** (79.65%) of the total variance of the studied molecules. We notice that the pI_{50} toxicity is strongly correlated with E_{HOMO}.

In the table 1, the total energy (Et) value of the compound **11** (Figure 5), with most probability, is wrong because it's significantly smaller if compared to the same value of compounds with similar activity.

* **G1** (one group containing molecules having the methoxy group in ortho, meta, and para (Figure 1);
* **G2** (group 2 containing chlorinated and brominated halogen molecules (Figure 1);

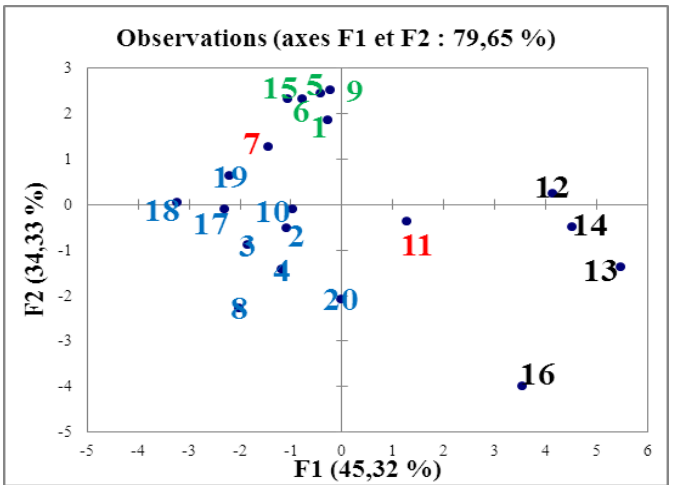

Fig. 5: Cartesian diagram showing the separation between two regions and dispersal of different molecules by groups.

Multiple linear regressions (MLR)

In order to propose a mathematical model and to evaluate quantitatively the substituent's physicochemical effects on the activity pI_{50} of the totality of the set of these 20 molecules, we submitted the data matrix constituted obviously from the 11 physicochemical variables corresponding to the 20 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. Where R is the correlation coefficient; R^2 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 20 molecules to quantitatively evaluate the effect of substituent.

$$\mathbf{pI_{50}} = 3{,}689\text{-}7{,}016.10^{-6}\mathbf{.Et} + 0{,}284.\boldsymbol{\mu} \quad \textbf{(Equation 1)}$$

$$\mathbf{N = 20} \qquad \mathbf{R^2 = 0{,}624} \qquad \mathbf{R=0{,}790} \qquad \mathbf{RMCE = 0{,}555}$$

Table 3: Analyses of variance

Source	DDL	Sum of squares	Mean square	F	Pr > F
Model	2	8,699	4,349	14,113	< 0,0001
Error	17	5,239	0,308	-	-
Total corrected	19	13,938	-	-	-

As a remark (Table 3), the model the values are different from 0 at a significant level p <0.05 for **Pr** <0,001 with $\mathbf{F_{(2,19)}}$ = 14,113. The figure 6 shows a very regular distribution of toxicity values depending on the experimental values [17].

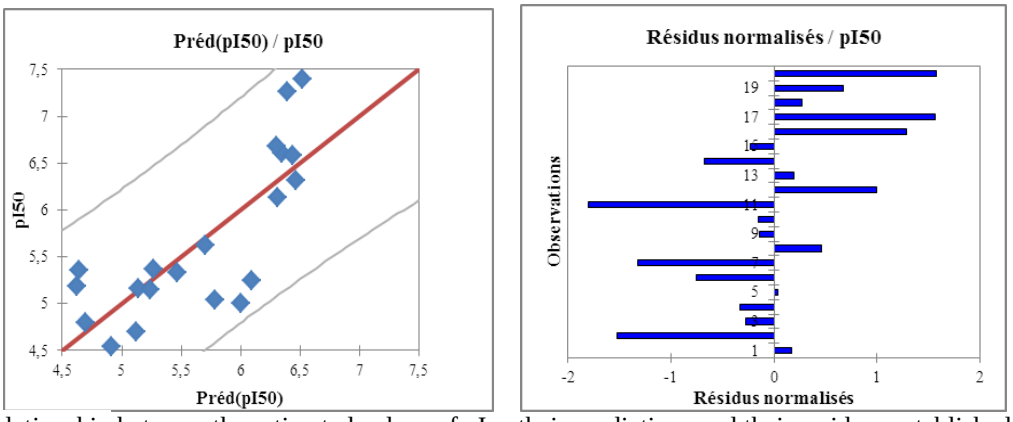

Fig. 6: Relationship between the estimated values of pI_{50}, their predictions and their residues established by MLR

Multiples non linear 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 20 molecules. The coefficients R, R^2, and the F-values are used to select the best regression performance.

$$pI_{50} = 4{,}779 - 4{,}256.10^{-5}.\,Et - 0{,}3094.\,\mu - 1{,}839.10^{-1}.\,Et^2 + 3{,}685.10^{-2}.\,\mu^2 \quad \textbf{(Equation 2)}$$

$$\mathbf{N = 20 \qquad R^2 = 0{,}831 \qquad R = 0{,}912 \qquad RMCE = 0{,}396}$$

With MLNR was obtained significantly better correlation coefficient R = 0,912 (Figure 7) shows a very uniform distribution of the toxicity observed values depending on the experimental values and the correlation between the experimental results and calculated alter them pI_{50}. The residual values tended to zero which is why we did not graph for prediction residuals [17,30].

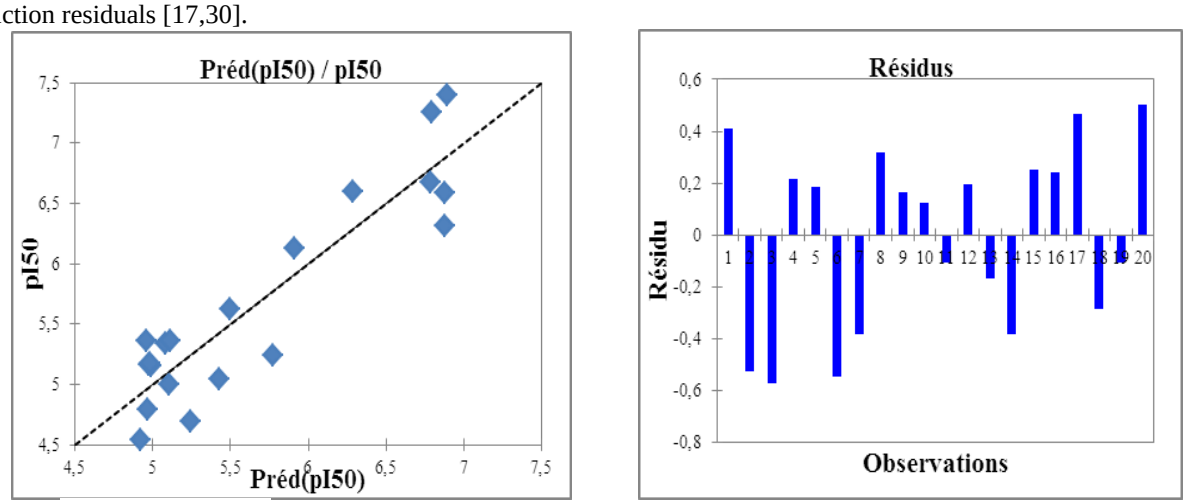

Fig. 7: Relationship between the estimated values of pI_{50} and their Predictions established by MNLR

Artificial neural networks (ANN)

In order to increase the probability of good characterization of studied compounds, neural networks (ANN) can be used to generate predictive models of quantitative structure-propriety relationships (QSPR) between a set of molecular descriptors obtained from the MLR, and observed activities. The ANN calculated activities model were developed using the properties of several studied compounds. Some authors [31,32]. Have proposed a parameter ρ, leading to determine the number of hidden neurons, which plays a major role in determining the best ANN architecture defined as follows:

ρ = (Number of data points in the training set/Sum of the number of connections in the NN) In order to avoid over fitting or under fitting, it is recommended that $1.8<\rho<2.3$ [29,33]. The output layer represents the calculated activity values log (Ki). The architecture of the ANN used in this work (6-3-1), ρ = 2.18.

The correlation between ANN calculated and experimental activities and the residues values are very significant as illustrated in Figure 8 and as indicated by R and R^2 values.

$$\mathbf{N = 20 \qquad R = 0.9606}$$

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

Validation

Before using a QSAR model to predict the activity of new compounds, we should validate it using a validation method. In this paper we validated our model with cross validation using LOO procedure. The correlation of the observed activities with the CV calculated ones and the residues values are illustrated in figure 8.

These values show that the relationship between the estimated values of pIC_{50} and their residues established by artificial neural networks are illustrated in figure 8.

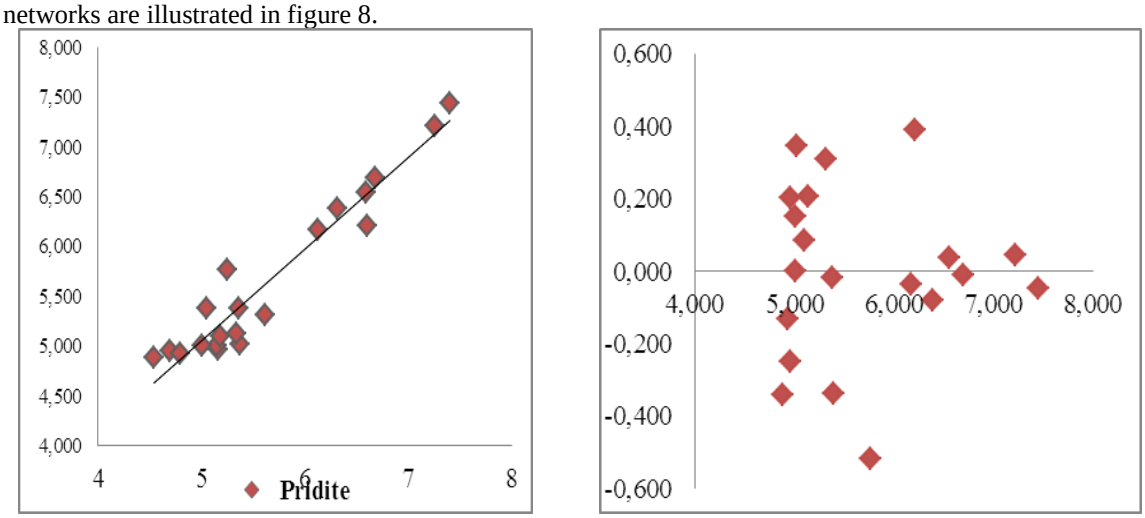

Figure 8: Relationship between the estimated values of pI_{50} and their residues established by ANN

The obtained determination coefficient (R^2) value is 0,9606 for this data. It confirms that the artificial neural network results were the best to build the quantitative structure activity relationship models. In this part, we investigated the best linear QSAR regression equations established in this study. Based on this result, a comparison of the quality of de MLR, MNLR and ANN models shows that the ANN models have substantially better.

Predictive capability because the ANN approach gives better results than MLR and MNLR, ANN was able to establish a satisfactory relationship between the molecular descriptors and the activity of the studied compounds (Table 4).

We note from table 4 that the higher the value of R^2 increases more and RMSE values decrease MCE what could be explained by the best correlation [34,35].

Table 4: Statistical results of comparative all models based on the N = 20 compounds

Statistical result	**MLR**	**MNLR**	**ANN**
R^2	0,624	0,831	0,9606
MCE	0,308	0,157	-
RMSE	0,555	0,396	-

R^2: Determination coefficient; S: Standard error of estimated; F: Fischer test value; RMSE: Root mean square error.

Table 5 shows the comparison of observed values with the calculated values of the MLR, MNLR and ANN.

Table 5: Observed values and calculated of pI_{50} according to different methods.

Molec.	**pI_{50} (Obs.)**	**Pred(pI_{50}/) RLM**	**Pred(pI_{50})/ RNLM**	**Pred(pI_{50})/ ANN**
1	5,367	5,270	4,958	5,020
2	5,244	6,09	5,77	5,763
3	6,310	6,465	6,881	6,389
4	6,131	6,315	5,915	6,168
5	5,161	5,141	4,974	4,960
6	4,699	5,119	5,245	4,950
7	5,046	5,781	5,429	5,385
8	6,602	6,349	6,286	6,211
9	5,155	5,236	4,987	5,003
10	5,620	5,705	5,498	5,310
11	5,001	6,001	5,109	5,001
12	5,18	4,627	4,986	5,096
13	4,796	4,691	4,965	4,928
14	4,538	4,912	4,920	4,879
15	5,337	5,463	5,085	5,130
16	5,357	4,645	5,113	5,373
17	7,260	6,39	6,795	7,217
18	6,588	6,439	6,875	6,550
19	6,678	6,305	6,782	6,690
20	7,398	6,525	6,894	7,446

IV. CONCLUSION

In this work, we studied the QSAR regression to predict the toxicity of a series of 20 compounds as R or R^2 different models obtained using different statistical tools and various descriptors was shown in table 5. The study of the quality of ANN, MLR and MNLR models showed that the ANN result has substantially better predictive capability than the other methods. With the ANN approach we have established a relationship between several descriptors (E_{HOMO}, $E_{LUMO\ ...}$) and toxicity in satisfactory manners.

Finally, we can conclude that one of the studied descriptors (E_{HOMO}, $E_{LUMO,}$ . ..), which is sufficiently rich in chemical and electronic information to encode the structural features, may be used with other topological 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] Ando, R., Ikegami, H., Sakiyama, M., Ooike, S., Hayashi, M., Fujino, Y., Abe, D., Nakamura, H., Mishina, T., Kato, H., Iwase, Y., Tomozane, H., Morioka, M. 3-Cyano-6-(5-methyl-3-pyrazoloamino)pyridines: Selective Aurora A kinase inhibitors. Bioorg. Med. Chem. Lett., 20, 4709–4711, 2010.

[2] Coumar, M.S., Chu, C.Y., Lin, C.W., Shiao, H.Y., Ho, Y.L., Reddy, R., Lin, W.H., Chen, C.H., Peng, Y.H., Leou, J.S., Lien, T.W., Huang, C.T., Fang, M.Y., Wu, S.H., Wu, J.S., Chittimalla, S.K., Song, J.S., Hsu, J.T.A., Wu, S.Y., Liao, C.C., Chao, Y.S., Hsieh, H.P. Fast-Forwarding Hit to Lead: Aurora and Epidermal Growth Factor Receptor Kinase Inhibitor Lead Identification. J. Med. Chem. 53, 4980-4988, 2010.

[3] Aliagas-Martin, I., Burdick, D., Corson, L., Dotson, J., Drummond, J., Fields, C., Huang, O.W., Hunsaker, T., Kleinheinz, T., Krueger, E., Liang, J., Moffat, J., Phillips, G., Pulk, R., Rawson, T.E., Ultsch, M., Walker, L., Wiesmann, C., Zhang, B., Zhu, B.Y., Cochran, A.G. A Class of 2,4-Bisanilinopyrimidine Aurora A Inhibitors with Unusually High Selectivity against Aurora B. J. Med. Chem. 52, 3300–3307, 2009.

[4] Xiang xiang Wu, Huahui Zeng, Xin Zhu, Qiujuan Ma, Yimin Hou, Xuefen Wu,. Novel pyrrolopyridinone derivatives as anticancer inhibitors towards Cdc7: QSAR studies based on dockings by solvation score approach. European Journal of Pharmaceutical Sciences, 50 : 323–334, 2013.

[5] Esposito, E.X., Hopfinger, A. J., Madura, J. D. Methods for applying the quantitative structure-activity relationship paradigm Methods Mol. Biol., 275, pp. 131-214, 2004.

[6] Perkins, R., Fang, H., Tong, W., Welsh, W.J. Quantitative structure-activity relationship methods: perspectives on drug discovery and toxicology. Environ. Toxicol. Chem., 22, pp. 1666- 1679, 2003.

[7] Du, Q.S., Huang, R.B., Chou, K.C. Recent advances in QSAR and their applications in predicting the activities of chemical molecules, peptides and proteins for drug design. Curr. Protein Pept. Sci., 9, pp. 248-260, 2008.

[8] Bradbury, S.P. Quantitative structure-activity relationships and ecological risk assessment: an overview of predictive aquatic toxicology research. Toxicol. Lett., 79, pp. 229-237, 1995.

[9] Martinez-Mayorga, K., Medina-Franco, J.L. Chemoinformatics-applications in food chemistry. Adv. Food. Nutr. Res., 58, pp. 33-56, 2009.

[10] Behnam Mohseni Bababdani, Mehdi Mousavi Gravitational search algorithm: A new feature selection method for QSAR study of anticancer potency of imidazo[4,5-b]pyridine derivatives. Chemometrics and Intelligent Laboratory Systems, 122, 1–11, 2013.

[11] Cichero, E., Cesarini, S., Spallarossa, A., Mosti, L., Fossa, P. Computational studies of the binding mode and 3D-QSAR analyses of symmetric formimidoester disulfides: A new class of non-nucleoside HIV-1 reverse transcriptase inhibitor. J. Mol. Model., 15:357–367, 2009.

[12] Adamo, C., Barone, V. A TDDFT study of the electronic spectrum of s-tetrazine in the gas-phase and in aqueous solution. Chem. Phys. Lett., 330, pp. 152-160, 2000.

[13] Parac, M., Grimme, S. All calculations were done by GAUSSIAN 03 W software. J. Phys. Chem., A, 106, pp. 6844-6850, 2004.

[14] Gaussian 03, Revision B.01, M. J. Frisch, and al., Gaussian, Inc., Pittsburgh, PA, 2003.

[15] Becke, A. D. A new mixing of Hartree–Fock and local density-functional theories. J. Chem. Phys., 98, pp. 1372, 1993.

[16] Lee, C., Yang, W., Parr, R. G. Development of the Colle-Salvetti conelation energy formula into a functional of the electron density., Phys. Rev., B., 37, pp. 785-789,1988.

[17] Larif, M., Chtita, S., Adad, A., Hmamouchi, R., Bouachrine, M. and Lakhlifi T. Predicting biological activity of anticancer molecules 3-aryl-4-hydroxyquinolin-2-(1H)-one by DFT-QSAR models. International Journal of Advanced Research in Computer Science and Software Engineering; 3 (4), pp. 1-6, 2013.

[18] Hogarh, J. N., Seike, N., Kobara, Y., Habib, A., Namd, J. J., Lee, J. S. Qilu Li, Liu, X., Jun Li, Zhang, G., Masunaga, S. Passive air monitoring of PCBs and PCNs across East Asia: A comprehensive congener evaluation for source characterization. Chemosphere, 86, pp. 718–726, 2012.

[19] Taurino, A.M., Dello Monaco, D., Capone, S., Epifani, M., Rella, R., Siciliano, P., Ferrara, L., Maglione, G., Basso, A., Balzarano, D. Analysis of dry salami by means of an electronic nose and correlation with microbiological methods. Sensors and Actuators, B 95, pp. 123–131, 2003.

[20] Rücker, C., Rücker, G., Meringer, M.,y-Randomization and Its Variants in QSPR/QSAR, J. Chem. Inf. Model., 47, pp. 2345-2357, 2007.

[21] Demuth, H., Hugan, M., Beal, M. Neural Network Toolbox. For use with MATHLAB, User Guid's, Version 9, 2011.

[22] Larif, M., Adad, A., Hmamouchi, R., Taghki, A. I., Soulaymani, A., Elmidaoui, A., Bouachrine, M., Lakhlifi, T. Biological activities of triazine derivatives. Combining DFT and QSAR results, in press in Arabian Journal of Chemistry, 2013.

[23] Zupan, J., Gasteiger, J. Neural Networks for Chemistry and Drug Design: An Introduction, second ed., VCH, Weinheim, 1999.

[24] Chimizou, R., Iwamura, H., Fujita, T., Agric Food Chem., 36, pp. 1276, 1988.

[25] Turkkan, N. Génie, gènes et neurones, Revue de l'Université de Moncton, 26 (1), pp. 205-221, 1993.

[26] Lee, P. Y., Toxicity and quantitative structure–activity relationships of benzoic acids to *Pseudokirchneriella subcapitata*, Chen C. Y. J. Hazard. Mater., 165, pp. 156-161, 2009.

[27] Jing, G., Zhou, Z., Zhuo, J., Quantitative structure-activity relationship (QSAR) study of toxicity of quaternary ammonium compounds on *Chlorella pyrenoidosa* and *Scenedesmus quadricauda*, Chemosphere, 86, pp. 76-82, 2012.

[28] Jonathan, N. H., Nobuyasu, S., Yuso, K., Ahsan, H., Jae-Jak, N., Jong-Sik, L. Q. L., Xiang, L., Jun; L., Gan, Z., Shigeki, M. Passive air monitoring of PCBs and PCNs across East Asia: A comprehensive congener evaluation for source characterization, Chemosphere, 86, pp. 718–726, 2012.

[29] Elhallaoui, M., Elasri, M., Ouazzani, F., Mechaqrane, A. and Lakhlifi, T. Quantitative Structure-Activity Relationships of Noncompetitive Antagonists of the NMDA Recetor: A Study of a series of MK801 Derivative Molecules Using Statistical Methods and Neural Network. Int., J. Mol. Sci, 4, pp. 249-262, 2003.

[30] Hmamouchi, R., Idrissi, Taghki, A. Larif, M., Adad, A., Abdellaoui, A., Bouachrine, M. and Lakhlifi, T. Combining DFT and QSAR result for predicting the biological activity of the phenylsuccinimides derivatives, Journal of Chemical and Pharmaceutical Research, 5(9):198-209, 2013.

[31] Andrea, T.A., Kalayeh, H.J., Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors, J. Med. Chem., 34, 2824, 1991.

[32] So, S., Richards, G. J. Med. Chem., 35, 3207, 1992.

[33] Chtita, S., Larif, M., Ghamali, M., Adad, A., Hmamouchi, R., Bouachrine, M. and Lakhlifi, T., 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, International Journal of Innovative Research in Science, Engineering and Technology, 2 (11), pp. 6586-6601, 2013.

[34] Adad, A., Larif., M.Hmamouchi, R., Taghki, A., Bouachrine, M., Lakhlifi T., DFT-QSAR models to predict biological activities of carbamates, s-alkylcarbamathioates, ureas and chloroacetamides derivatives, 2(2) 115-118, 2013.

[35] Adad, A., Hmamouchi, R. Larif, M., Bouachrine, M. and Lakhlifi, T. Binding affinities (AhR) of polychlorinated biphenyls (PCBs), dibenzo-p-dioxins (PCDDs) and dibenzofurans (PCDFs) study combining DFT and QSAR results. International Journal of Advanced Research in Computer Science and Software Engineering, 4 (4) 202-213, 2014.

