Volume 4, Issue 2, February 2014 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

QSAR Modeling of the toxicity of pI_{50} Pyrazines Derived by Electronic Parameters Obtained by DFT

Rachid Hmamouchi
Faculty of Science, University Moulay Ismail, Meknes, Morocco

Majdouline Larif
Faculty of Science, University Ibn Tofail – Kenitra, Morocco

Azeddine Adad
Faculty of Science, University Moulay Ismail, Meknes, Morocco

Mohammed Bouachrine
ESTM, University Moulay Ismail, Meknes, Morocco

Tahar Lakhlifi*
Faculty of Science, University Moulay Ismail, Meknes, Morocco

Abstract- ***A study of structure-activity relationships was performed on a series of pyrazine derivatives. Electronic variables are obtained by calculating the DFT B3LYP level using 6,6-31G-31G (d) as basic functions. These parameters have been treated with a square principal component analysis (PCA), the method of multiple linear regression (MLR), progressive and partial regression (PLS) and artificial neural network (ANN). The predicted activities are in good agreement with the experimental results. The artificial neural network (ANN) technology, given the descriptors obtained from the MLR showed a correlation coefficient of 0.9 and ANN architecture (5-12-1), which is a good result. The results suggest that a combination of calculated parameters may be useful for predicting the biological activity of pyrazine derivatives. A QSAR analysis showed that total energy*** E_T, λ_{max} ***maximum absorption*** E_{HOMO} ***HOMO energy of the molecule plays an important role in the toxicity*** pI_{50} ***pyrazine derivatives studied.***

Keywords- structure-activity; 3D-QSAR model; RLM; ANN; PCA; PLS; DFT B3LYP study

I. INTRODUCTION

Pyrazines, nitrogen heterocycles are widely distributed in the animal and plant kingdoms, and very present in the flavor of food [1]. 2-methyl-6-phenylethynylpyridine (MPEP) is an investigational drug that has been one of the first compounds found to act as a selective antagonist of subtype of metabotropic glutamate receptor mGluR5.

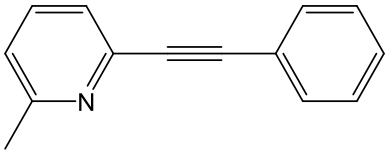


Figure 1: the 6-methyl-2-phenylethynylpyridine (MPEP) One of the best non-competitive antagonists

After initially being patented as liquid crystals for LCD screens, it was developed by the pharmaceutical company Novartis in the late 1990s [2]. It has also been shown to produce antidepressant and anxiolytic effects in animals [3, 4] and reduce the effects of morphine withdrawal.

The main significance of MPEP was as a lead compound to develop more potent and selective mGluR5 antagonists as MTEP [5], but research using MPEP continues, and recently it has been shown to reduce self -administration of nicotine, cocaine, ketamine and heroin in animals [6]. For this analysis, we have relied on a table that contains 27 molecules collected randomly from work conducted by Fabrizio Micheli, et al. Publish direct Science 2008 (phenylethynyl-pyrrolo [1,2-a] pyrazine: A new potent and selective tool in the mGluR5 antagonists arena), we insert using a software Gaussian 03 W [33] of electronic parameters such as activation energy (E_a), the total energy (E_T), the energy of the highest occupied molecular orbital (E_{HOMO}), the energy of the lowest unoccupied molecular orbital (E_{LUMO}), the dipole moment μ, the maximum absorption λ_{max} and the oscillation factor $f_{(SO)}$, which are the most relevant molecular electronic properties [7, 8]. The objective of this study is to develop predictive QSAR models for toxicity of our molecules studied. Relationship Quantitative Structure/Activity (QSAR) initiated by Hansch and Fujita (1964) have found many applications in chemistry, especially in the prediction of the chromatographic retention [9]. It is, in this case, to analyze quantitatively the relationship between chemical structure and biological activity. To do this we apply the non-linear regression, least squares (PLS) and neural networks, which produce and interpret mathematical relationships that lead to quantitative relations.

In order to model the toxicity of several organic compounds containing pyrazine, we have used several statistical tools, mainly: (XLSTAT 2009) for principal component analysis (PCA) and multiple linear regressions (MLR) and (MATLAB R2009a) for artificial neural networks (ANN).

II. METHODOLOGIES

II-1. Chemicals

Previous studies (Fabrizio Micheli et al. 2008) have introduced a number of pyrazine derivatives with the ability to enhance the power of acceptable therapeutic potential application of mGluR5 antagonists. Further work on the electronic and steric aspects, 27 molecules was studied in the (QSAR/Modilisation frame).

The following figure shows the chemicals structures of studied compound.

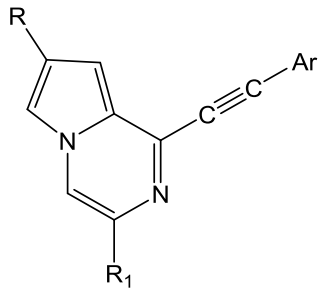


Figure 2: Chemical structure of the studied pyrazines.

The experimental toxicity of the compounds studied was obtained from previous work (2008) (Fabrizio Micheli et al. 2008) (Table 1). The range of toxicity data ranges from 6.2 to 8.00.

The following table shows the corresponding experimental activities pI_{50} each chemical structure of the compounds studied.

TABLE 1: OBSERVED TOXICITY OF STUDIED PYRAZINE DERIVATIVES

N°	R_1	R	Ar	pI_{50}Obs
1	CH_3	H	4-Cl-phenyl	6,2
2	CH_3	H	3-OH-phenyl	6,5
3	CH_3	H	5-pyrimidinyl	5,7
4	H	CF_3	Phenyl	6,8
5	CH_3	CF_3	Phenyl	7,4
6	CH_3	CONHMe	Phenyl	6,9
7	CH_3	CH_3	4-F-phenyl	6,4
8	CH_3	CH_3	3-CF_3-phenyl	6,8
9	CH_3	CH_3	3-pyridylphenyl	5,7
10	CH_3	CH_3	3-thienylphenyl	7,2
11	CH_3	CH_3	3-OMe-phenyl	7,1
12	CH_3	CH_3	3-Cl-phenyl	7,0
13	CH_3	CH_3	2-F-phenyl	6,9
14	CH_3	CH_3	4-F-phenyl	7,0
15	CH_3	CF_3	4-pyridylphenyl	7,1
16	CH_3	CF_3	2-pyridylphenyl	7,3
17	CH_3	CF_3	3-furylphenyl	7,3
18	CH_3	CF_3	2-thienylphenyl	7,2
19	CH_3	CF_3	3-thienylphenyl	8,0
20	CH_3	CO_2Et	3-CF_3-phenyl	6,7
21	CH_3	CO_2Et	2-F-phenyl	6,5
22	CH_3	$CONH_2$	2-F-phenyl	6,3
23	CH_3	$CONH_2$	3-CF_3-phenyl	7,1
24	CH_3	$CONH_2$	3-thienylphenyl	7,3
25	CH_3	CONHMe	3-F-phenyl	7,0
26	CH_3	CONHMe	3-CF_3-phenyl	6,7
27	CH_3	CONHMe	3-thienylphenyl	7,3

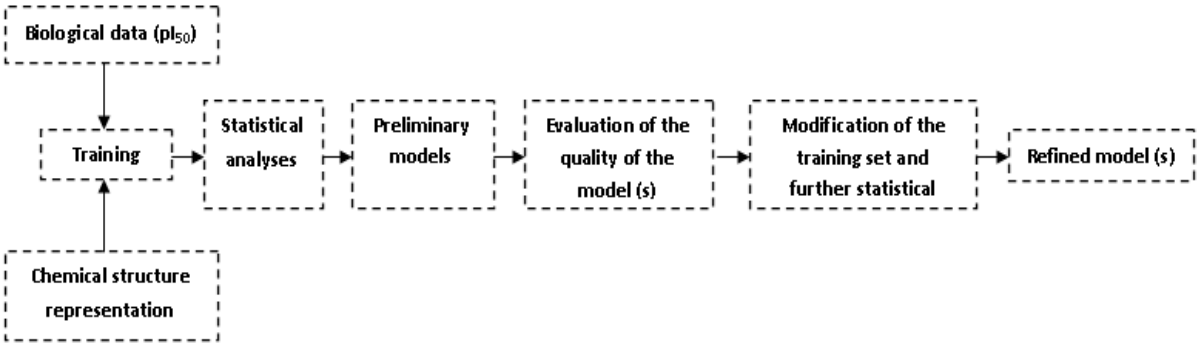
II-2. DFT Calculations

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 [10,11] using the B3LYP functional [12,13] and a 6-31G (d) basis set. The B3LYP, a version of DFT method, uses Beck'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.

II-3. Statistic analysis

II-3-1. Quantitative relationship: structure-activity relationship (QSAR) of compounds derived pyrazine

The objective of (quantitative) structure-activity relationship (QSAR) analysis is to derive empirical models that relate the biological activity of compounds to their chemical structure. In this QSAR analysis, quantitative descriptors are used to describe the chemical structure and the analysis results in a mathematical model describing the relationship between the chemical structure and biological activity. In the figure next the main steps in QSAR analysis.



II-3-1-1. Principal components analysis

The molecules of pyrazine and derivatives (1 to 27) were studied by statistical methods based on the principal component analysis (PCA) [14, 15] using the software XLSTAT 2009 and Mathlab software v 2009a.

This is 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.

II-3-1-2. Multiple Linear Regressions (RLM)

The multiple linear regression statistic technique is used to study the relation between one dependent variable and several independent variables. It is a mathematic technique that minimizes differences between actual and predicted values. The multiple linear regression model (MLR) [16, 17, 18, 19, 20, 24] was generated using the software SYSTAT, version 12 [21], to predict antifungal activities pI_{50}. It has served also to select the descriptors used as the input parameters for a back propagation network (ANN).

II-3-1-3. Partial Least Square Analysis

The PLS have two objectives: to approximate the matrix X of molecular structure descriptors to the matrix Y of dependent variables and to maximize the correlation between them. The leave-one-out (LOO) method Cramer was used to perform the cross-validated analysis. The cross-validated coefficient, q2, is calculated using the following equation:

$$q^2 = 1 - \frac{\sum (Yi - Yipred)^2}{\sum (Yi - Ymean)^2}$$

Where Yi is the i^{th} experimental toxicity value, Yipred is the i^{th} predicted toxicity Ymean is the mean of the experimental toxicity.

The optimal number of components (N) is employed to do non-validation PLS analysis to get the final model parameters such as correlation coefficient R^2 (Nguyen et al., 2008), standard deviation (S) and Fischer test value (F).

II-3-1-4. Artificial Neural Networks (ANNs)

In this study, different neural network models were tested and optimized to obtain the best model configuration for the prediction of the presence/absence of activity pI_{50}. The modelling method was based on the principles of the back propagation algorithm. The ANN model constructs a model based on examples of data with known outputs [22]. A back propagation network typically comprises three types of neuron layers: an input layer, one or more hidden layers and an

output layer each including one or several neurons. As shown in figure 3, nodes from one layer are connected to all nodes in the following layer, but no lateral connections within any layer or feed-back connections are possible. Several input neurons are used, each representing an Electronic variable. The output layer comprises one neuron, indicating the presence or absence of a toxicity pI_{50}.

With the exception of the input neurons, which only connect one input value with its associated weight values, the net input for each neuron is the sum of all input values x_n, each multiplied by its weight w_{jn}, and a bias term z_j which may be considered as the weight from a supplementary input equalling one: $a_j = \Sigma\, w_{ji}x_i + z_j$

The output value, y_j, can be calculated by feeding the net input into the transfer function of the neuron: $y_j = f(aj)$

Many transfer functions can be used. In this study, two types of sigmoid functions have been compared: the tangential and logarithmic sigmoid transfer function [23].

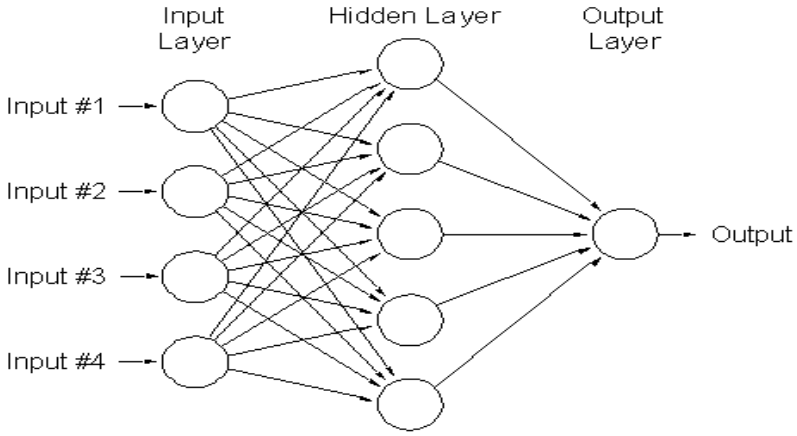


Figure 3: Illustration of a three-layered neural network with one input layer, one hidden layer and one output layer.

III. RESULTS AND DISCUSSION

A QSPR study was carried for a series of 27 derivates of pyrazine, in order to determine a quantitative relationship between structure chemical and biological activity.

The table 2 shows the values of the calculated parameters obtained by DFT/B3LYP 6-31G optimization of the studied derivates of pyrazine.

TABLE 2: VALUES OF THE THIRTEEN CHEMICAL DESCRIPTORS

Molec.	pI_{50}	E_T	E_{HOMO}	E_{LUMO}	ΔE	μ	Ea	λ_{max}	$f_{(SO)}$
1	6,2	-1185,964	-7,135	-1,857	5,278	3,023	3,126	396,59	0,161
2	6,5	-801,583	-6,996	-1,665	5,332	1,711	3,191	388,55	0,147
3	5,7	-758,442	-5,598	-2,118	3,48	4,264	3,023	410,15	0,122
4	6,8	-1024,085	-7,155	-2,052	5,103	1,46	3,333	372,01	0,199
5	7,4	-1063,406	-6,421	-1,989	4,432	2,062	3,009	375,6	0,197
6	6,9	-934,381	-5,519	-1,859	3,661	3,351	3,226	384,38	0,174
7	6,4	-864,92	-7,012	-1,674	5,338	2,569	3,173	390,79	0,156
8	6,8	-1102,723	-5,412	-1,89	3,522	4,55	3,068	404,13	0,147
9	5,7	-781,722	-7,093	-1,846	5,247	4,125	3,089	401,32	0,142
10	7,2	-1086,441	-5,248	-1,573	3,675	2,102	3,228	384,12	0,139
11	7,1	-1177,928	-7,282	-1,934	5,348	2,492	3,314	374,13	0,209
12	7,0	-1523,001	-7,453	-2,175	5,278	1,474	3,237	383,03	0,185
13	6,9	-1162,637	-5,736	-2,06	3,677	1,06	3,265	379,69	0,194
14	7,0	-1162,639	-5,735	-2,028	3,707	2,405	3,304	375,21	0,204
15	7,1	-1079,44	-6,816	-2,339	4,477	3,795	3,167	391,44	0,152
16	7,3	-1079,438	-6,613	-2,156	4,456	0,362	3,215	385,68	0,173
17	7,3	-1061,178	-7,331	-1,875	5,456	1,716	3,407	363,89	0,153
18	7,2	-1384,161	-7,018	-2,078	4,939	1,543	3,225	384,43	0,239
19	8,0	-1384,16	-6,942	-1,94	5,002	1,78	3,351	369,98	0,184
20	6,7	-1330,604	-5,732	-2,166	3,566	2,086	3,121	397,3	0,153
21	6,5	-1092,799	-7,106	-2,006	5,1	2,12	3,164	391,81	0,163
22	6,3	-994,304	-5,631	-2,021	3,61	3,231	3,174	390,59	0,157

23	7,1	-1232,108	-6,849	-2,184	4,665	1,919	3,128	396,37	0,147
24	7,3	-1215,826	-6,729	-1,892	4,837	2,771	3,255	380,96	0,144
25	7,0	-1033,608	-6,931	-2,037	4,894	2,496	3,146	394,06	0,137
26	6,7	-1271,413	-6,381	-2,141	4,241	2,669	3,107	399,01	0,136
27	7,3	-1255,131	-6,163	-1,85	4,313	2,923	3,245	382,04	0,132

III-1. Statistical analysis (implementation of the PCA)

The graphical representation of molecules and study the characteristics reports have shown that a large number of chemical and electronic parameters were significant, most of the time one (or more) link between these parameters. It therefore seems worthwhile to try to process the data statistically, using a method of multivariate analysis such as Principal Component Analysis (PCA).

III-2. Study of the eigenvalues

Here is the bar chart represents the total inertia.

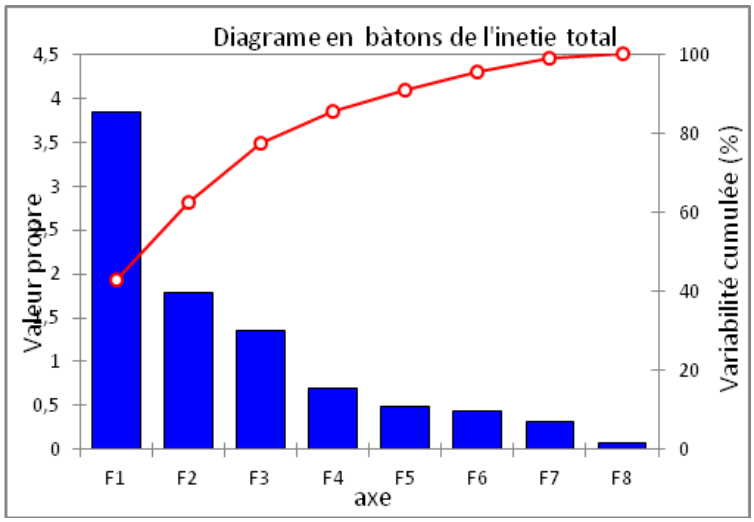


Figure 4: Total inertia diagram

We choose to study the axis 1, 2. However, it is interesting to note that the plane formed by the axes 1 and 2 has an equivalent inertia about 62.72%. We also note that Axis 3 has a relatively low inertia compared to the previous two but is interesting because it is linked to a particularly important variable to study the LUMO energy (E_{LUMO}) with a percentage of 0.7 about.

III-3. Principal Component Analysis (Training Set Selection)

A principal component analysis (PCA) was performed on a data table of nine (09) variables and 27 individuals. We pushed the analysis to two factors and 62.72% of the variance could be expressed. The number of individuals allows a critical correlation coefficient of 0.973 or $r^2 = 0.9463$.

III-4. Correlation Matrix

Examination of the table for the correlation matrix can be seen that almost all variables are correlated. There is a strong correlation between toxicity pI_{50} on the one hand and the total energy (E_T), the wavelength (λ_{max}), homo energy (E_{HOMO}) and gap (ΔE), activation energy (E_a) and the wave length (λ_{max}) of the other part.

Table 3: Correlation Matrix between variables

Variables	**pI_{50}**	**E_T**	**E_{HOMO}**	**E_{LUMO}**	**ΔE**	**μ**	**E_a**	**λ_{max}**	**$f_{(SO)}$**
pI_{50}	1								
E_T	**-0,608**	**1**							
E_{HOMO}	-0,098	0,152	**1**						
E_{LUMO}	-0,085	0,382	0,079	**1**					
ΔE	0,076	-0,054	**-0,968**	0,174	**1**				
μ	-0,514	0,403	0,324	0,101	-0,295	**1**			
E_a	0,538	-0,306	-0,271	0,148	0,305	-0,500	**1**		
λ_{max}	**-0,702**	0,299	0,270	-0,151	-0,305	0,586	**-0,827**	**1**	
$f_{(SO)}$	0,369	-0,383	-0,228	-0,184	0,179	-0,490	0,414	-0,583	**1**

- The $\mathbf{pI_{50}}$ experimental activities $\mathbf{pI_{50}}$ is negatively correlated with the $\mathbf{E_T}$ total energy (r=-0,608 and p <0,05) and negatively correlated with maximum absorption of $\boldsymbol{\lambda_{max}}$ (r=-0,702 and p<0,05) at a significant level.
- The HOMO energy $\mathbf{E_{HOMO}}$ is strongly (negative) correlated with the gap energy $\mathbf{\Delta E}$ (eV) for r=-0,968 and p <0,001 at a high level.
- The activation energy $\mathbf{E_a}$ is negatively correlated with $\boldsymbol{\lambda_{max}}$ for r=-0,827 and p<0,005 at a significant level.

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

III-5. Correlation Circle

Now represent those worn by plane 1 and 2 on a circle of correlation variables.

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 5) shows that the F_1 axis (42,78% of the variance) is mainly due to the λ_{max}, while the axis F_2 (19,93% of the variance) is located by the other parameters of ΔE (Gap energy).

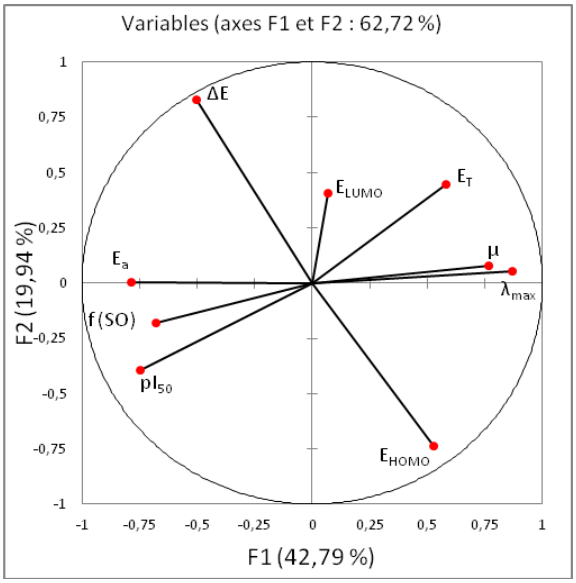


Figure 5: Correlation circle

Correlations 1 and level 2 it is interesting to note the variables supported by the axis 1 and axis variables supported by two forms a right angle meaning no linear correlation. Variables (ΔE; λ_{max}; E_{HOMO}, μ, E_a; pI_{50}) studied are close to the correlation circle and through this plan we can study the influence of the number ΔE, pI_{50}. In addition, we can see that the variables E_T, μ; λ_{max} are strongly interrelated.

III-6. Analysis of variable space

The projection of the variables on the factorial plane F_1-F_2 shows that the F_1 axis expresses 42.78% of the variance and is mainly determined by the variables, λ_{max}, E_a, μ, pI_{50}. It therefore opposes the waters weakly mineralized waters loaded. The vertical F_2 factor expresses 19.93% of the variance. It is determined in a positive way by ΔE variables E_{HOMO}.

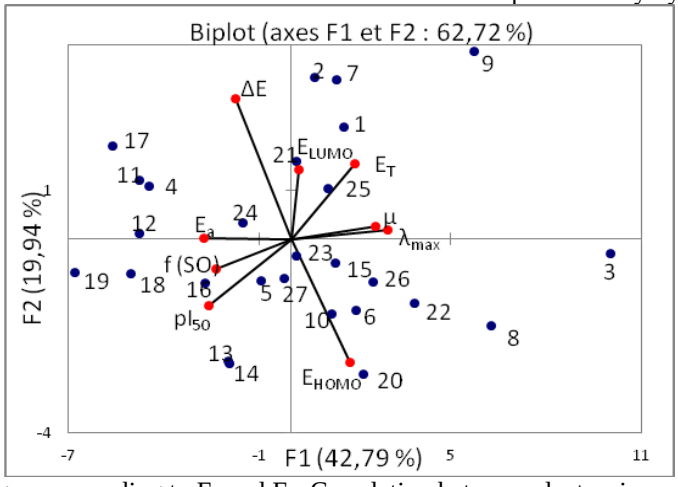


Figure 6: Cartesian diagram according to F_1 and F_2: Correlation between electronic parameters and individuals (molecules).

III-7. Space analysis of individuals

The projection of individuals on the plane F_1-F_2 revealed that a distribution of molecules in two groups: the group (-M) containing attractor motifs substituted phenyl with (λ_{max} **>400nm,** pI_{50} **< 6,** μ **> 4**) and the group (+M) containing donor motifs substituted phenyl with (λ_{max} **< 400nm,** pI_{50} **> 6,** μ **<4**).

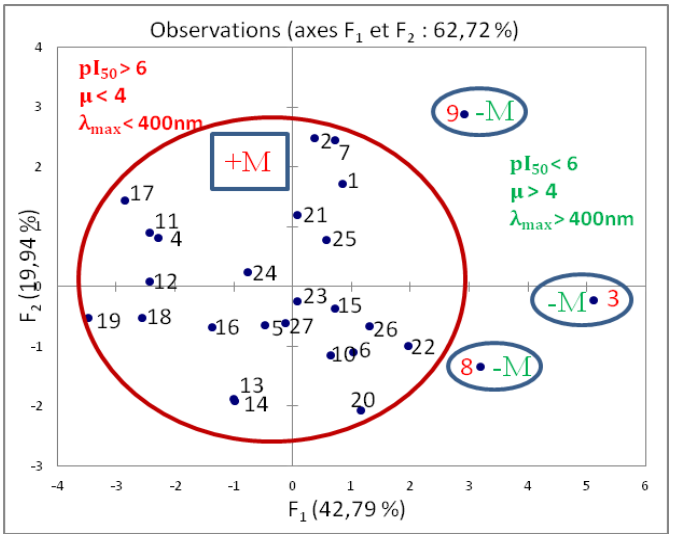


Figure 7: Cartesian diagram according to F_1 and F_2: attractor by mesomeric effect and donor by mesomeric effect both grouped in two separate regions.

III-8. Multiple linear regressions

A multiple regression analysis was performed for the data table matrices (electronic descriptors/individuals), and toxicity PI_{50} identify as the response variable. The statistical quality of the best models was evaluated on the basis of R^2 (coefficient of determination) and RMSE (coefficient indicates the accuracy of the model).

III-9. Multiple linear regression of the variable toxicity (MLR)

Modeling toxicity pI_{50} value of all training compounds (compounds 27 pyrazine derivatives) led to the best value corresponding to the linear combination of the following descriptors: total energy E_T, λ_{max} maximum absorption.

$$pI_{50} = 15{,}6842 - 1{,}19634.10^{-3} E_T - 0{,}02623.\lambda_{max} \quad (1)$$

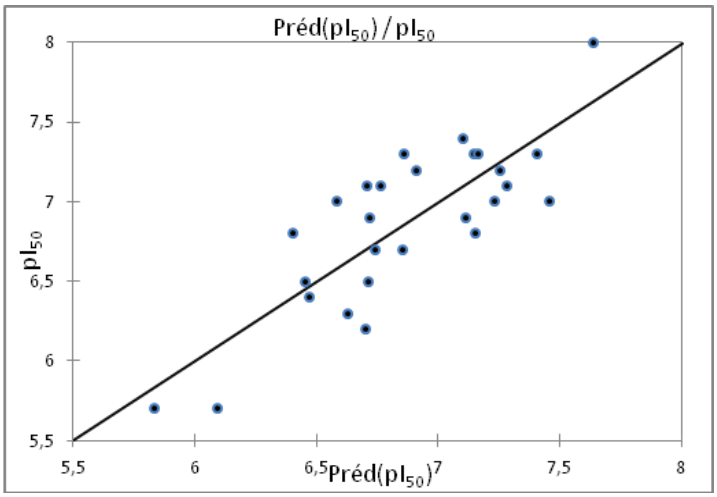


Figure 8: Graphical representation of calculated and observed toxicity by MLR.

For our 27 compounds, the correlation between experimental toxicity and calculated one based on this model is quite significant (Fig. 8) as indicated by statistical values:

N = 27 **R = 0,817** **R^2 = 0,667** **RMSE = 0,307**

The figure 8 shows a very regular distribution of toxicity values depending on the experimental values.

III-10. Multiple nonlinear regression of the variable toxicity (MNLR)

We have also used the technique of nonlinear regression model to improve the structure-toxicity in a quantitative way, taking into account several parameters. This is the most common tool for the study of multidimensional data. The resulting equation is:

$$pI_{50} = 28{,}951 - 2{,}64.10^{-3}.E_T - 5{,}583.E_{HOMO} + 13{,}985.E_{LUMO} - 0{,}338.\mu - 313{,}801.E_a + 2{,}409.\lambda_{max}\ 1{,}701.f_{(SO)} - 5{,}872.10^{-7}.E^2_T - 2{,}34.10^{-3}.E^2_{HOMO} + 2.E^2_{LUMO} - 0{,}636\Delta E^2 + 6{,}928.10^{-2}.\mu^2 + 49{,}666.E^2_a - 3{,}096.10^{-3}.\lambda^2_{max} + 2{,}294.f^{\ 2}_{(SO)} \quad (2)$$

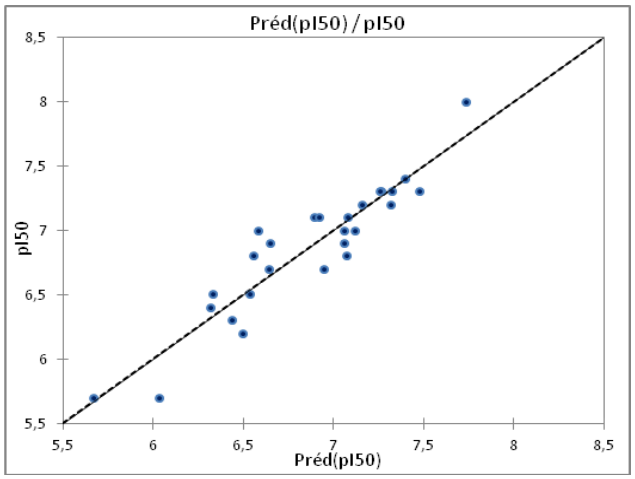


Figure 9: Graphical representation of calculated and observed toxicity by MNLR.

The obtained parameters describing the electronic aspect of the studied molecules are:

N = 27 R = 0,93 R^2 = 0,864 RMSE = 0,304

The toxicity value pI_{50} predicted by this model is somewhat similar to that observed. Figure 9 shows a very regular distribution of toxicity values based on the observed values.

III-11. Partial Least Square Regression (PLS)

To linearly correlate the molecule descriptors: the total energy, energy E_{HOMO}, energy E_{LUMO}, energy gap ΔE, the dipole moment μ, activation energy E_a, absorption maximum λ_{max} and the factor of oscillation $f_{(SO)}$ to pI_{50}, the following equations was used:

$$pI_{50} = 6{,}862 - 3{,}760.10^{-4}.E_T - 6{,}913.10^{-2}.E_{HOMO} + 0{,}136.E_{LUMO} - 6{,}005.10^{-2}.\mu + 0{,}707\ E_a - 8{,}869.10^{-3}.\lambda_{max} + 2{,}698.f_{(SO)} + 0{,}090.\Delta E \quad (3)$$

N = 27 R = 0,731 R^2 = 0,534 RMSE = 0,327

The figure 10 shows a very regular distribution of pI_{50} values depending on the experimental values.

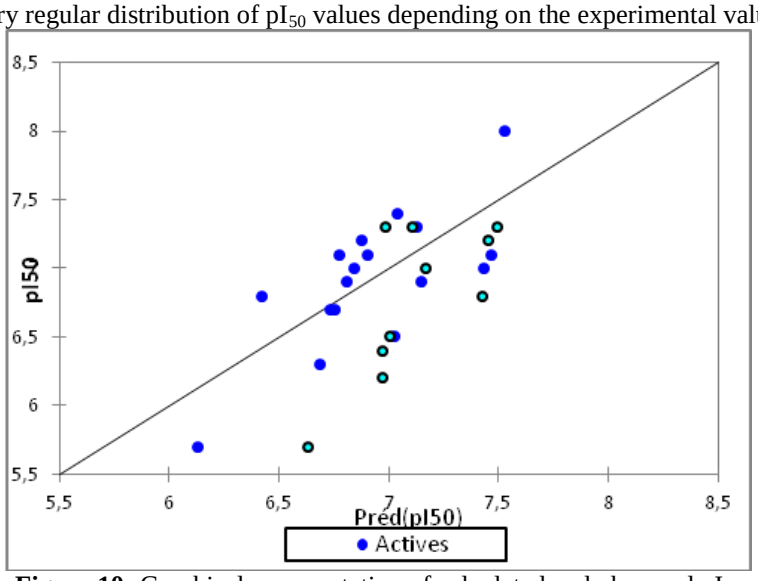


Figure 10: Graphical representation of calculated and observed pI_{50}

The obtained coefficient of correlation in equation (2) is quite interesting (**0,864**). To optimize the error standard deviation and better finish building our model, we involve in the next part artificial neural networks (ANN).

As part of this conclusion, we can say that the toxicity values obtained from nonlinear regression are highly correlated to that of the observed toxicity comparing to results obtained by MLR method.

III-12. Artificial neural networks

In order to increase the probability of good characterization of studied compounds, neural network (ANN) can be used to generate predictive models of quantitative structure–activity relationships (QSAR) between a set of molecular descriptors obtained from the MLR and observed activity. The ANN calculated toxicity model were developed using the properties of several studied compounds. The correlation between ANN calculated and experimental toxicity values is very significant as illustrated in figure 9 and as indicated by R and R^2 values.

N = 27 R = 0,973 R^2 = 0,946 RMSE = 0,032

These values show that the relationship between the estimated values of pI_{50} and their residues established by artificial neural networks are illustrated in figure 11.

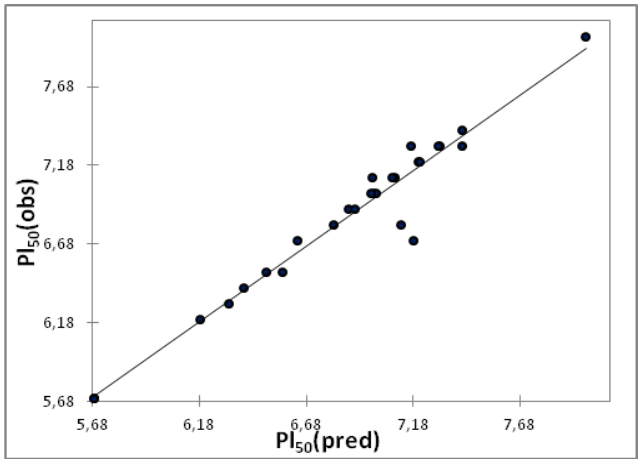


Figure 11: Correlation between the calculated and experimental inhibition pI_{50}

The statistic of the tree steps of the calculation by the ANNs: Training, validation and test are illustrated in table 4.

Table 4: Values obtained by ANNs.

	Samples	**RMSE**	**R**	**R^2**
Training	19	$1{,}05.10^{-4}$	0.999	0,998
Validation	4	$8{,}32.10^{-3}$	0.998	0,954
Test	4	$8.85.10^{-2}$	0.704	0,49

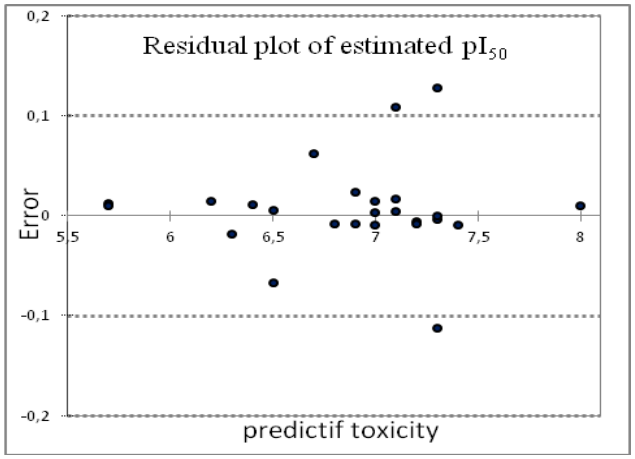


Figure 12: Relationship between the estimated values of pI_{50} and their residues established by artificial neural networks

The obtained squared correlation coefficient (R^2) value is 0,946 for this data set of pyrazines. 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 CPA, MLR and ANN models shows that the ANN models have substantially better predictive capability because the ANN approach gives better results than MLR. ANN was able to establish a satisfactory relationship between the molecular descriptors and the activity of the studied compounds.

IV. CONCLUSION

Theoretical calculations were carried out on a series of pyrazines driven to evaluate the stability structures and reactivity properties. The position of the substituent phenyl donors attractors by mesomeric play a vital role in deciding the stability

of the structure / reactivity driven pyrazines. Global descriptors such as total energy E_T, λ_{max} maximum absorption, E_{HOMO} HOMO energy, provide necessary information on the overall responsiveness driven pyrazines.

Comparison of key statistical terms as R or R^2 different models obtained using different statistical descriptors and different tools have been shown in the table 5. The study of the quality of the MLR and ANN models showed that the ANN results with substantially better predictive capacity than the other methods. With the ANN approach we have established a relationship between several descriptors (E_{HOMO}, E_{LUMO}...) and the toxicity of satisfactory ways. Finally, we can conclude that the descriptors studied (E_{HOMO}, E_{LUMO}...) which are rich enough in electronics and chemical information to encode the structural features can be used with other topological descriptors for the development of predictive QSAR models.

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

N°	R_1	R	Ar	pI_{50}				
				Obs.	**RLM**	**RNLM**	**PLS**	**ANN**
1	CH_3	H	4-Cl- Phenyl	6,2	6,701	6,498	6,971	6,186
2	CH_3	H	3-OH- Phenyl	6,5	6,452	6,335	7,005	6,567
3	CH_3	H	5-pyrimidnyl	5,7	5,834	5,67	6,133	5,687
4	H	CF_3	Phenyl	6,8	7,152	7,075	7,428	7,126
5	CH_3	CF_3	Phenyl	7,4	7,105	7,4	7,038	7,409
6	CH_3	CONHMe	Phenyl	6,9	6,72	6,65	6,811	6,876
7	CH_3	CH_3	4-F	6,4	6,469	6,322	6,97	6,389
8	CH_3	CH_3	3-CF3	6,8	6,404	6,561	6,42	6,808
9	CH_3	CH_3	3-pyridyl	5,7	6,093	6,035	6,63	5,69
10	CH_3	CH_3	3-thienyl	7,2	6,909	7,321	6,874	7,205
11	CH_3	CF_3	3-OMe- Phenyl	7,1	7,281	6,898	7,468	7,096
12	CH_3	CF_3	3-Cl- Phenyl	7	7,46	7,063	7,433	6,996
13	CH_3	CF_3	2-F- Phenyl	6,9	7,116	7,061	7,148	6,91
14	CH_3	CF_3	4-F- Phenyl	7	7,234	7,121	7,168	7,009
15	CH_3	CF_3	4-pyridyl	7,1	6,709	7,079	6,774	7,083
16	CH_3	CF_3	2-pyridyl	7,3	6,86	7,265	7,13	7,172
17	CH_3	CF_3	3-furyl	7,3	7,409	7,328	7,497	7,303
18	CH_3	CF_3	2-thienyl	7,2	7,257	7,161	7,452	7,208
19	CH_3	CF_3	3-thienyl	8	7,636	7,737	7,526	7,989
20	CH_3	CO_2Et	3-CF3- Phenyl	6,7	6,855	6,644	6,755	6,638
21	CH_3	CO_2Et	2-F- Phenyl	6,5	6,715	6,54	7,025	6,494
22	CH_3	$CONH_2$	2-F- Phenyl	6,3	6,629	6,438	6,685	6,318
23	CH_3	$CONH_2$	3-CF3- Phenyl l	7,1	6,762	6,924	6,901	6,991
24	CH_3	$CONH_2$	3-thienyl	7,3	7,147	7,262	7,108	7,3
25	CH_3	CONHMe	3-F- Phenyl	7	6,585	6,584	6,843	6,985
26	CH_3	CONHMe	3-CF3- Phenyl l	6,7	6,74	6,949	6,737	7,181
27	CH_3	CONHMe	3-thienyl	7,3	7,165	7,477	6,984	7,412

ACKNOWLEDGMENT

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

REFERENCES

[1] Parliment, T.H., Epstein, M.F., Organoleptic properties of some alkyl-substituted alkoxy- and alkylthiopyrazines. J. Agric. Food Chem., 21: 714-716, 1973.

[2] Micheli, F. "Methylphenylethynylpyridine (MPEP) Novartis". Current opinion in investigational drugs (London, England: 2000) 1 (3): 355–9, 2000.

[3] Pilc, A; Kłodzińska, A; Brański, P; Nowak, G; Pałucha, A; Szewczyk, B; Tatarczyńska, E; Chojnacka-Wójcik, E; Wierońska, JM. "Multiple MPEP administrations evoke anxiolytic- and antidepressant-like effects in rats". Neuropharmacology, 43 (2): 181–7, 2002.

[4] Kłodzińska, A; Tatarczyńska, E; Chojnacka-Wójcik, E; Pilc, A. "Anxiolytic-like effects of group I metabotropic glutamate antagonist 2-methyl-6-(phenylethynyl)-pyridine (MPEP) in rats" Polish journal of pharmacology 52 (6): 463–6, 2000.
[5] Lea Pm, 4th; Faden, AI. "Metabotropic glutamate receptor subtype 5 antagonists MPEP and MTEP" CNS Drug Reviews 12 (2): 149–66, 2006.
[6] Van Der Kam, EL; De Vry, J; Tzschentke, TM. "Effect of 2-methyl-6-(phenylethynyl)pyridine on intravenous self-administration of ketamine and heroin in the rat". Behavioural Pharmacology, 18(8): 717–24, 2007.
[7] K Laarej; M Bouachrine; S Radi; S. Kertit and B Hammouti, E-Journal of Chemistry, 7(2), 419-424, 2010.
[8] H Zarrok; H Oudda; A Zarrouk; R Salghi; B Hammouti; M Bouachrine, Der Pharma Chemica, 3 (6): 576-590, 2011.
[9] Kaliszan, R. Quantitative relationships between molecular structure and chromatographic retention. CRC Crit. Rev. Anal. Chem., 16: 323-383, 1986.
[10] Adamo, C.; Barone, V., Chem. Phys. Lett., 330, 152–160, 2000.
[11] Gaussian 03, Revision B.01, M. J. Frisch, and al., Gaussian, Inc., Pittsburgh, PA, 2003.
[12] Becke, A. D., Phys., 98, 1372, 1993.
[13] Lee, C., Yang, W., Parr, R. G., Phys. Rev., B. 37, 785-789, 1988.
[14] 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., Chemosphere, 86: 718–726, 2012.
[15] Taurino, A.M., Dello, D., Monaco, S. Capone, M. Epifani, R. Rella, P. Siciliano, L. Ferrara, G. Maglione, A. Basso, D. Balzarano., Sensors and Actuators B 95, 123–131, 2003
[16] M Larif; A Adad; R Hmammouchi; AI Taghki; A Soulaymani; A Elmidaoui; M Bouachrine; T Lakhlifi, in press in Arabian Journal of Chemistry, http://dx.doi.org/10.1016/j.arabjc.2012.12.033, 2013.
[17] Azeddine Adad, Majdouline Larif, Rachid Hmamouchi, Abdelhafid Idrissi Taghki, Mohammed Bouachrine and Tahar Lakhlifi, Journal of Chemica Acta, 2, 115-118, 2013.
[18] Azeddine Adad, Rachid Hmamouchi, Mohammed Bouachrine and Tahar Lakhlifi, Journal of Chemica Acta, 2, 98-104, 2013.
[19] A Adad; R Hmamouchi; A I Taghki; A Abdellaoui; M Bouachrine and T Lakhlifi, Journal of Chemical and Pharmaceutical Research, 5(7): 28-41, 2013.
[20] R. Hmamouchi; A. I. Taghki; M. Larif; A. Adad; A. Abdellaoui; M. Bouachrine and T. Lakhlifi, Journal of Chemical and Pharmaceutical Research, 5(9): 198-202, 2013.
[21] STATITCF Software, Technical Institute of cereals and fodder, Paris, France, 1987.
[22] Demuth, H., Hugan, M., Beal M. Neural Network Toolbox. For use with MATHLAB, User Guid's, Version 9, 2011.
[23] Turkan N., Génie, gènes et neurones, Revue de l'Université de Moncton, 26 (1), 205-221, 1993.
[24] S. Chtita, M. Ghamali, M. Larif, A. Adad, R. Hmamouchi, M. Bouachrine and T. Lakhlifi, International Journal of Innovative Research in Science, Engineering and Technology, pp. 2319-8753, 2013.

