Journal of Computational Methods in Molecular Design, 2014, 4 (4):140-149

Scholars Research Library
(http://scholarsresearchlibrary.com/archive.html)

ISSN : 2231- 3176
CODEN (USA): JCMMDA

Combining DFT and QSAR result for predicting the biological activity of 1-(2-ethoxyethyl)-1H-pyrazolo[4,3-d]pyrimidines as phosphodiesterase V inhibitors

B. Elidrissi[a], A. Ousaa[a], M. Ghamali[a], S. Chtita[a], M. A. Ajana[a], M. Bouachrine[b] and T. Lakhlifi[a*]

[a] Molecular Chemistry and Natural Substances Laboratory, Faculty of Science, University Moulay Ismail, Meknes, Morocco
[b] ESTM, University Moulay Ismail, Meknes, Morocco

ABSTRACT

Phosphodiesterase V acts as an attractive target for cardiovascular, Quantitative Structure–Activity Relationship (QSAR) technique is helpful for the optimization of structure requirements of twenty-six 1-(2-ethoxyethyl)-1H-pyrazolo[4,3-d]pyrimidines as phosphodiesterase V inhibitors. This work was conducted using the principal component analysis (PCA) method, the multiple linear regression method (MLR), the multiple non-linear regressions (MNLR) and the artificial neural network (ANN). The predicted results of various study compounds afford reliable prediction of IC_{50} with respect to experimental data. Density functional theory (DFT) calculations have been carried out in order to get insights into the structure, chemical reactivity and property information for the series of study compounds. This study shows that the PCA, MLR and ANN have served also to predict activities, but when compared with the results given by the RNLM, we realized that the predictions fulfilled by this latter were more effective.

Keywords: DFT study, QSAR, *1-(2-ethoxyethyl)-1H-pyrazolo[4,3-d]pyrimidine, cardiovascular*.

INTRODUCTION

Phosphodiesterase enzymes catalyse the conversion of cyclic GMP and cyclic AMP into the corresponding nucleotide monophosphates with the emergence of molecules like sildenafil and tadalafil. The phosphodiesterase V (PDE5) inhibition is now emerged as the most favorite tool for the treatment of erectile dysfunction in men. Large amount of research was carried out for the identification of new, selective PDE5 inhibitors and to investigate their usefulness in cardiovascular disorders to act as vascular smooth muscle relaxants. In human PDE5 is frequently observed in lungs, platelets, and vascular smooth muscles. PDE5 is specific for the cGMP, in its both catalytic sites, and in the two cGMP-binding allosteric sites [1-5]. In PDE5 there are two cGMP-binding domains: (i) an N-terminal domain harbouring a site for phosphorylation and (ii) a C-terminal catalytic domain of the enzyme [6-12]. The catalytic domain of PDE5 is constituted by four major parts which includes three helical subdomains, an N-terminal cycling-fold region, a linker region and a C-terminal helical bundle.

Since its introduction more than forty-five years ago [13], structure-activity relationships have been developed for various areas of applications, e.g. estimating of the different substance characteristics as well as their activities levels. Quantitative structure-activity relationships (QSAR) are widely used to predict bioactivities from chemical structure and corresponding physicochemical properties. The numerous QSAR studies have been carried out in order to explain or predict inhibitory activities of on different living systems [14-16].

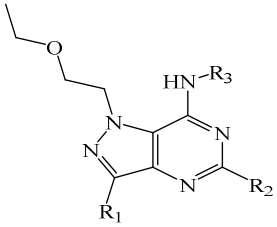
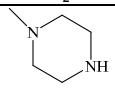
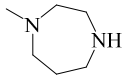
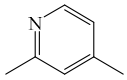
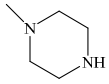
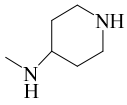
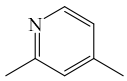
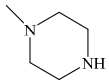
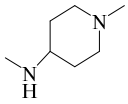
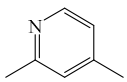
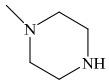
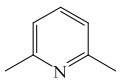
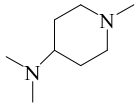
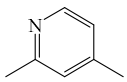
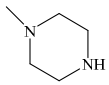
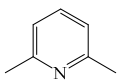
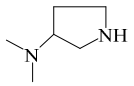
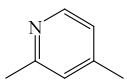
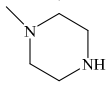
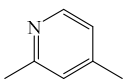
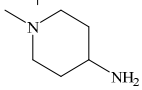
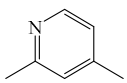
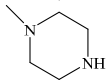
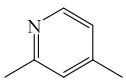
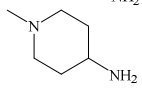
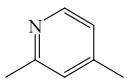
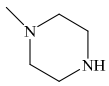
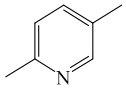
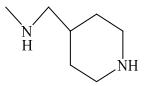
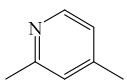
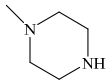
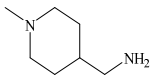
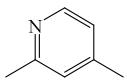
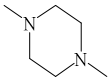
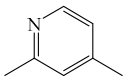
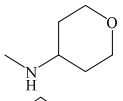


Figure 1: The structural template of 1-(2-ethoxyethyl)-1H-pyrazolo[4,3-d]pyrimidines

Table 1: Observed pIC_{50} of the studied compounds [19]

Mol. N°	R_1	R_2	R_3	pIC_{50}	Mol. N°	R_1	R_2	R_3	pIC_{50}
1	Et			1,155	**14**	Et			0,119
2	Me			0,398	**15**	Et			0,699
3	Et			1,155	**16**	Et			0,143
4	Me			0,187	**17**	Et			1,060
5	Et			0,180	**18**	Et			0,377
6	Me			-0,146	**19**	Me			-0,362
7	Et			0,420	**20**	Et			0,108
8	Et			-0,518	**21**	Me			0,456
9	Et			-0,740	**22**	Me			1,155
10	Me			-0,415	**23**	Me			-0,079
11	Et			0,051	**24**	Et			1,155
12	Et			-0,398	**25**	Et			1,097
13	Et			-0,079	**26**	Et			1,398

In this study, we have modelled Inhibition Concentration IC_{50} of several organic compounds based on 1-(2-ethoxyethyl)-1H-pyrazolo[4,3-d]pyrimidine (Figure 1) using several statistical tools, principal components analysis (PCA), multiple linear regression (MLR), multiple non-linear regression (MNLR) and artificial neural network (ANN) calculations, for the designing of new phosphodiesterase 5 inhibitors with vascular smooth muscle relaxant activity. The objectives of this work are to develop predictive QSAR models for the IC_{50} of our studied molecules. On the other hand, several quantum chemical methods and quantum-chemistry calculations have been performed in order to study the molecular structure and electronic 1-(2-ethoxyethyl)-1H-pyrazolo[4,3-d]pyrimidine derivatives properties [17-18]. The geometry as well as the nature of their molecular orbital, HOMO (highest occupied

molecular orbital) and LUMO (lowest unoccupied molecular orbital) is involved in the properties of biological activity of organic compounds. The more relevant molecular properties were calculated. These properties are the highest occupied molecular orbital energy E_{HOMO}, the lowest unoccupied molecular orbital energy E_{LUMO}, energy gap ΔE, dipole moment μ, total energy E_T, activation energy E_a, absorption maximum λ_{max} and factor of oscillation $f_{(SO)}$.

MATERIALS AND METHODS

Material

A dataset of 26 compounds was taken from the published phosphodiesterase V inhibitors [4-19]. The activity under investigation is the concentration of the test compound that inhibited the activity of 1-(2-ethoxyethyl)-1H-pyrazolo[4,3-d]pyrimidine derivatives by 50% (IC_{50}).

The structures and their inhibitory activities are listed in table 1. The inhibitory activity IC_{50} values were converted into the corresponding pIC_{50} (log $1/IC_{50}$) and used as dependent variables in the 3D-QSAR analyses.

Calculation of molecular descriptors

Calculation of descriptors using Gaussian 03W

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 [20,21] using the B3LYP functional [22] 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.

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 twenty-six compounds was carried out by B3LYP method employing 6–31G (d) basis set.

Calculation of descriptors using ACD/ChemSketch

ChemSketch program (Demo version 10.0) [23] was employed to calculate the others molecular descriptors, Molar Volume (MV (cm^3)), Molecular Weight (MW), Molar Refractivity (MR (cm^3)), Parachor (Pc (cm^3)), Density (D (g/cm^3)), Refractive Index (n), Surface Tension (γ(dyne/cm) and Polarizability (α (cm^3)) [24].

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

$$MV = \frac{M}{d}$$

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

$$d = \frac{M}{MV}$$

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

$$MR = \frac{n^2 - 1}{n^2 + 2}\frac{M}{d}.$$

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

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

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

- ***Surface Tension* (γ):** Calculated from calculated MV and calculated Pc:

$$\gamma = \left(\frac{P}{Vm} \right)^4$$

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

$$Pc = \left(\frac{M}{d} \right) \gamma^{\frac{1}{4}}$$

- ***Polarizability*(α):**Calculates from the Molar Refractivity as follows:

$$\alpha = 0{,}3964308 \text{ x } MR$$

Statistical analysis

Principal Components Analysis (PCA)

The molecules of 1-(2-ethoxyethyl)-1H-pyrazolo[4,3-d]pyrimidine derivatives (1 to 26) were studied by statistical methods based on the principal component analysis (PCA) [24] using the software XLSTAT 2009.
This is an essentially a descriptive statistical method which aims to present, in graphic form, the maximum informations contained in the data table 1.

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

Multiple Linear Regressions (MLR)

The multiple linear regression statistic technique is used to study the relation between one dependent variable and several independent variables. It is a mathematic technique that minimizes differences between actual and predicted values. The multiple linear regression model (MLR) [22] was generated using the software XLSTAT 2009, to predict IC_{50}. It has served also to select the descriptors used as the input parameters for a back propagation network (ANN).

Artificial neural networks (ANNs)

The ANNs analysis was performed with the use of Mathlab software v 2008a Neural Fitting tool (nftool) toolbox on a data set of 1-(2-ethoxyethyl)-1H-pyrazolo[4,3-d] pyrimidine derivatives IC_{50} activity [25].

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 [26]. The input layer consisted of fifteen 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. One neuron formed the output layer of sigmoid function activation. The architecture of the applied ANN models is presented in Figure 2.

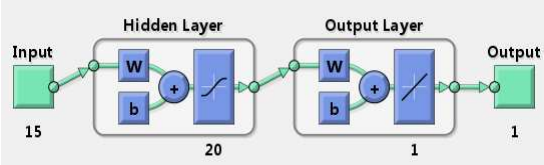


Figure 2: The ANNs architecture.

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 1-(2-ethoxyethyl)-1H-pyrazolo[4,3-d]pyrimidine derivatives of IC_{50} activity [27] were subjected to the ANN analysis. First, for the learning set of compounds, i.e., 23 1-(2-ethoxyethyl)-1H-pyrazolo[4,3-d]pyrimidine derivatives 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 the ANN model designed and selected, the validation set of five compounds was used. Testing set with five 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 [28]. The operation of the output layer is linear, which is given as below:

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

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:

$$\text{Bias} = \sum(\bar{y} - y) \qquad \textbf{(2)}$$

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:

$$\text{RMSE} = \sqrt{\frac{1}{n}.\sum_{i=1}^{n}(p_{exp} - p_{pred})^2} \qquad \textbf{(3)}$$

where n = number of compounds, p_{exp} = experimental value, p_{pred} = predicted value and summation is of overall patterns in the analyzed data set [29,30]. The scripts were run on a personal PC.

RESULTS

This study was carried for a series of 26 of 1-(2-ethoxyethyl)-1H-pyrazolo[4,3-d]pyrimidine derivatives, in order to determine a quantitative relationship between structure and IC_{50} activity. Table 2 shows the values of the calculated parameters obtained by DFT/B3LYP 6-31G* optimization and ACD/ChemSketch program of the studied compounds.

Table 2: Values of the calculated parameters obtained by DFT/B3LYP 6-31G* optimization and ACD/ChemSketch program of the studied compounds.

N°	pIC_{50}	MW	MR (cm^3)	MV (cm^3)	Pc (cm^3)	n	γ (dyne/ cm)	D (g/ cm^3)	E_T (ua)	E_{HOMO} (ev)	E_{LUMO} (ev)	ΔE (ev)	μ (D)	E_a (ev)	λ_{max} (nm)	f(so)
1	1,155	395,5	113,33	306,3	817,1	1,661	50,6	1,29	-34750	-5,096	-1,09	4,006	2,48	2,214	560,1	0,0139
2	0,398	382,46	107,17	278,8	759,7	1,694	55	1,37	-33052,9	-5,511	-1,58	3,935	2,53	3,499	354,4	0,0024
3	1,155	396,49	111,77	294,9	798,3	1,682	53,6	1,34	-35186,6	-5,007	-1,07	3,934	3,53	2,149	576,8	0,0134
4	0,187	396,49	111,59	294	790,8	1,683	52,3	1,34	-35194,8	-5,141	-1,24	3,903	3,86	3,072	403,6	0,0244
5	0,18	410,52	116,2	310,1	829,4	1,672	51,1	1,32	-36257,6	-5,195	-1,22	3,975	1,30	3,472	357,1	0,0006
6	-0,146	396,49	111,59	294	790,8	1,683	52,3	1,34	-35194,6	-5,189	-1,31	3,881	5,29	3,561	348,2	0,002
7	0,420	410,52	116,2	310,1	829,4	1,672	51,1	1,32	-36265,3	-5,202	-1,32	3,881	5,38	3,065	404,6	0,0216
8	-0,519	410,52	116,2	310,1	829,4	1,672	51,1	1,32	-36257,5	-5,182	-1,22	3,964	2,8	2,112	587,1	0,0156
9	-0,74	410,52	116,2	310,1	829,4	1,672	51,1	1,32	-36257,5	-5,005	-0,98	4,028	2,34	3,253	381,2	0,0694
10	-0,415	410,51	116,92	315,3	836,2	1,663	49,4	1,3	-36257,7	-5,004	-1,08	3,926	4,43	3,554	348,8	0,0021
11	0,051	424,54	121,52	331,3	874,8	1,654	48,5	1,28	-37328,2	-5,21	-1,25	3,957	1,89	2,761	449,1	0,0083
12	-0,398	424,54	120,62	325,2	860,5	1,663	48,9	1,30	-37328,4	-5,153	-1,21	3,942	1,40	3,494	354,8	0,0024
13	-0,079	424,54	120,62	325,2	860,5	1,663	48,9	1,30	-37328	-4,904	-0,96	3,939	4,73	3,549	349,3	0,006
14	0,119	424,54	120,81	326,1	868	1,662	50,1	1,30	-37328,1	-4,915	-1,09	3,825	4,01	3,406	364,0	0,0577
15	0,699	424,54	119,91	320	853,6	1,672	50,6	1,32	-37328,1	-5,213	-1,25	3,964	2,41	3,384	366,4	0,0611
16	0,143	438,57	125,23	341,3	899,1	1,654	48,1	1,28	-38398,6	-5,059	-1,05	4,013	4,44	3,221	384,9	0,0236
17	1,060	452,6	130,56	362,5	944,5	1,639	46	1,24	-39468,7	-4,900	-0,99	3,91	4,79	2,633	470,8	0,0055
18	0,377	424,54	120,62	325,2	860,5	1,663	48,9	1,30	-36257,9	-5,028	-1,09	3,94	5,6	3,737	331,8	0,0463
19	-0,362	410,52	115,06	302,1	816,8	1,686	53,3	1,35	-36257,9	-5,028	-1,09	3,94	5,6	3,452	359,2	0,0032
20	0,108	424,54	119,67	318,2	855,4	1,675	52,2	1,33	-37328,4	-5,008	-1,09	3,916	3,48	3,742	331,4	0,0754
21	0,456	424,54	119,91	320	853,6	1,672	50,6	1,32	-37328,1	-5,038	-1,07	3,97	4,77	3,326	372,8	0,0545
22	1,155	424,54	119,67	318,2	855,4	1,675	52,2	1,33	-37310,3	-4,846	-1,00	3,843	4,08	3,754	330,3	0,062
23	-0,079	411,5	113,63	305,8	814,4	1,665	50,3	1,34	-36798,6	-5,112	-1,12	3,996	3,55	4,051	306,1	0,0659
24	1,155	439,55	123,21	334,7	892,5	1,657	50,5	1,31	-38948	-5,092	-1,28	3,812	5,27	3,448	359,6	0,0635
25	1,097	452,55	125,13	331,5	899,7	1,679	54,2	1,36	-40414,7	-4,859	-0,92	3,939	5,85	1,932	641,7	0,0078
26	1,398	453,54	124,07	332	898,2	1,67	53,5	1,36	-40956,1	-5,079	-1,12	3,956	5,43	3,319	373,6	0,0197

The set of fifteen descriptors encoding the 26 of 1-(2-ethoxyethyl)-1H-pyrazolo[4,3-d]pyrimidine compounds, electronic, energetic and topologic parameters are submitted to PCA analysis [30]. The first three principal axes are sufficient to describe the information provided by the data matrix. Indeed, the percentages of variance are 43.73%; 15.80% and 14.55% for the axes F1, F2 and F3, respectively. The total information is estimated to a percentage of 84.08%. The principal component analysis (PCA) [31] 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 fifteen

descriptors are shown in table 3 as a correlation matrix and in figure 3 these descriptors are represented in a correlation circle.

The Pearson correlation coefficients are summarized in the following table 3. The obtained matrix provides information on the negative or positive correlation between variables.

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

	pIC_{50}	MW	MR	MV	Pc	n	γ	D	E_T	E_{HOMO}	E_{LUMO}	ΔE	μ	E_a	λ_{max}	f(so)
pIC_{50}	**1**															
MW	0,367	**1**														
MR	0,301	**0,962**	**1**													
MV	0,270	0,917	**0,984**	**1**												
Pc	0,344	**0,962**	**0,995**	**0,987**	**1**											
n	-0,096	-0,577	-0,720	-0,831	-0,746	**1**										
γ	0,197	-0,360	-0,555	-0,674	-0,547	0,905	**1**									
D	0,059	-0,317	-0,547	-0,668	-0,551	0,898	**0,942**	**1**								
E_T	-0,374	**-0,974**	-0,897	-0,843	-0,904	0,496	0,246	0,194	**1**							
E_{HOMO}	0,192	0,546	0,563	0,532	0,556	-0,342	-0,235	-0,271	-0,540	**1**						
E_{LUMO}	0,088	0,504	0,523	0,503	0,515	-0,355	-0,273	-0,284	-0,501	0,928	**1**					
ΔE	-0,273	-0,110	-0,104	-0,077	-0,107	-0,036	-0,098	-0,035	0,101	-0,189	0,191	**1**				
μ	0,292	0,392	0,306	0,245	0,298	0,007	0,097	0,134	-0,397	0,452	0,300	-0,400	**1**			
E_a	-0,311	-0,047	-0,085	-0,084	-0,115	0,047	-0,103	0,104	0,082	-0,092	-0,161	-0,180	0,053	**1**		
λ_{max}	0,317	0,046	0,061	0,051	0,091	0,001	0,161	-0,029	-0,087	0,134	0,204	0,184	-0,031	**-0,984**	**1**	
f(so)	0,070	0,169	0,126	0,110	0,125	-0,068	-0,026	0,045	-0,151	0,241	0,190	-0,132	0,021	0,400	-0,365	**1**

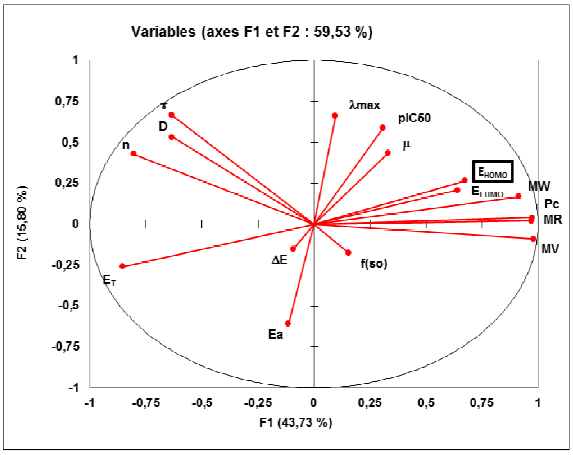


Figure 3: Correlation circle

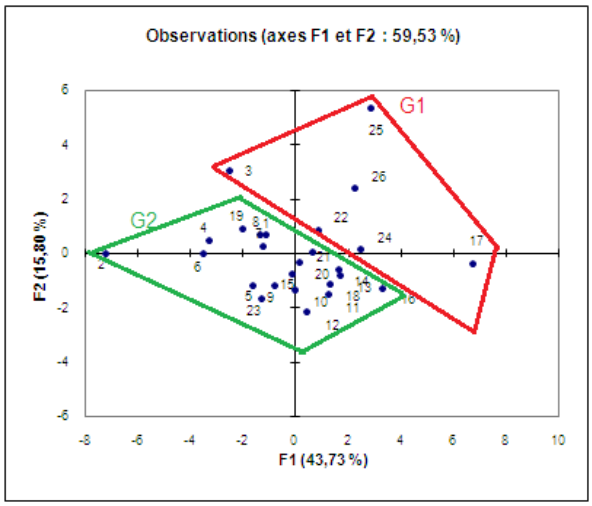


Figure 4: Cartesian diagram according to F1-F2

Correlation circle

The principal component analysis (PCA) was also performed to detect the connection between the different variables. The principal component analysis revealed the correlation circle (Figure 4) shows that the F1 axis (43.73% of the variance) appears to represent the Molar Refractivity (**MR**), and the F2 axis (15.80% of the variance) seems to represent the absorption maximum ($\boldsymbol{\lambda_{max}}$).

Multiple linear regressions (MLR)

To establish quantitative relationships between activity inhibitor pIC_{50} and selected descriptors, our array data were subjected to a multiple linear and nonlinear regression. Only variables whose coefficients are significant were retained.

Multiple linear regression of the variable pIC_{50} (MLR)

Many attempts have been made to develop a relationship with the indicator variable of activity inhibitor pIC_{50}, but the best relationship obtained by this method is only one corresponding to the linear combination of several descriptors: the molecular weight (**MW**), the parachor (**Pc**) the refractive index (**n**), the surface tension ($\boldsymbol{\gamma}$), the density (**D**) and the total energy ($\mathbf{E_T}$).

The resulting equation is:

$$\mathbf{pIC_{50}}=323{,}24+0{,}52\times\mathbf{MW}-0{,}22\times\mathbf{Pc}-123{,}87\times\mathbf{n}+1{,}76\times\boldsymbol{\gamma}-154{,}35\times\mathbf{D}+7{,}93\times10^{-04}\times\mathbf{E_T} \qquad (4)$$

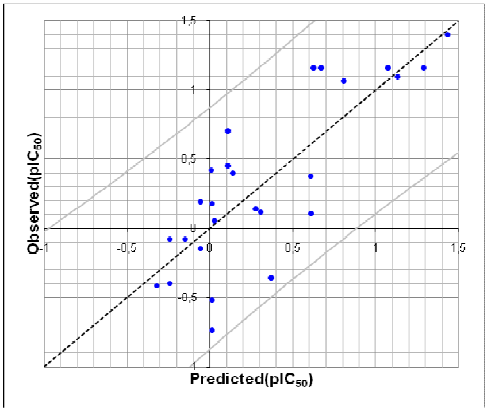


Figure 5: Graphical representation of calculated and observed pIC_{50} by MLR

For our 26 compounds, the correlation between experimental and calculated pIC_{50}one based on this model are quite significant (Figure 5) as indicated by statistical values:

$$\mathbf{N = 26 \quad R = 0.811 \quad R^2 = 0.658 \quad RMSE = 0.169}$$

The figure 5 shows a very regular distribution of pIC_{50} values depending on the experimental values.

Multiple nonlinear regression of the variable pIC_{50} (MNLR)

We have used also the technique of nonlinear regression model to improve the predicted IC_{50} in a quantitative way. It takes into account several parameters. This is the most common tool for the study of multidimensional data. We have applied it to table 2 containing 26 molecules associated with fifteen variables.

The resulting equation is:

$$\mathbf{pIC_{50}}=3501{,}25-6{,}57\times\mathbf{MW}+14{,}23\times\mathbf{MR}+0{,}31\times\mathbf{MV}-0{,}97\times\mathbf{Pc}-5638{,}91\times\mathbf{n}-0{,}35\times\boldsymbol{\gamma}+2889{,}11\times\mathbf{D}-3{,}77\times10^{-02}\times\mathbf{E_T}-$$
$$188{,}88\times\mathbf{E_{HOMO}}+689{,}65\times\mathbf{E_{LUMO}}-1{,}58\times\boldsymbol{\mu}-3{,}82\times\mathbf{E_a}-2{,}31\times10^{-02}\times\lambda_{max}+20{,}10\times\mathbf{f(so)}+8{,}31\times10^{-03}\times\mathbf{MW}^2-$$
$$4{,}43\times10^{-02}\times\mathbf{MR}^2+1{,}94\times10^{-03}\times\mathbf{MV}^2-2{,}10\times10^{-04}\times\mathbf{Pc}^2+1462{,}29\times\mathbf{n}^2+6{,}43\times10^{-02}\times\boldsymbol{\gamma}^2-1141{,}42\times\mathbf{D}^2-$$
$$5{,}06\times10^{-07}\times\mathbf{E_T}^2+57{,}75\times\mathbf{E_{HOMO}}^2-37{,}32\times\mathbf{E_{LUMO}}^2-98{,}10\times\boldsymbol{\Delta E}^2+0{,}26\times\boldsymbol{\mu}^2 \qquad (5)$$

The obtained parameters describing the topologic and the electronic aspects of the studied molecules are:

$$\mathbf{N = 26 \quad R = 0.999 \quad R^2 = 0.999}$$

The pIC_{50} value predicted by this model is somewhat similar to that observed. The figure 6 shows a very regular distribution of IC_{50} values based on the observed values.

With MLNR was obtained significantly better correlation coefficient R = 0, 999. Figure 6 shows a very uniform distribution of the IC_{50} observed values depending on the experimental values and the correlation between the experimental results and calculated alter them pIC_{50}. The residual values tended to zero which is why we did not graph for prediction residuals.

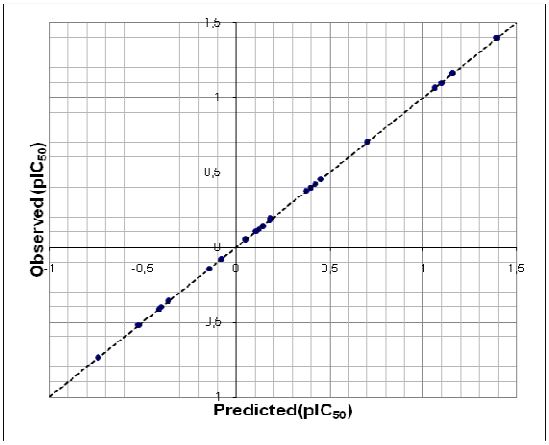


Figure 6: Graphical representation of calculated and observed pIC_{50}.

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–activity relationships (QSAR) between a set of molecular descriptors obtained from the MLR and observed activity. The ANN calculated pIC_{50} model was developed using the properties of several studied compounds. The correlation between ANN calculated and experimental pIC_{50} values are very significant as illustrated in figure 7 and as indicated by R and R^2 values.

N = 26 R = 0.928 R^2 = 0.861 RMSE = 0.658

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 7.

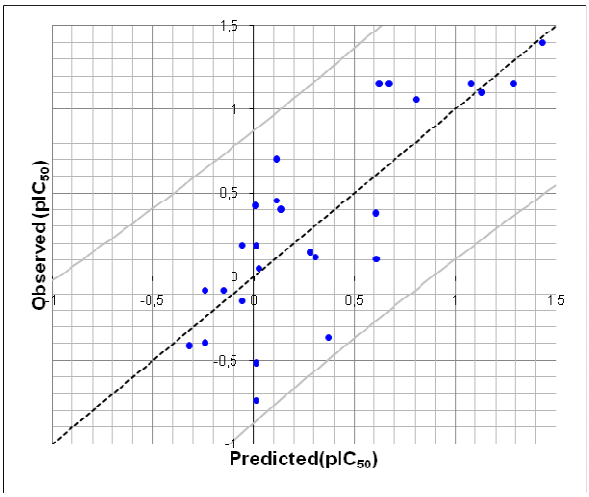


Figure 7: Graphical representation of calculated and observed IC_{50} (pIC_{50})

DISCUSSION

Principal component analysis

* The Molecular Weight **MW** is well correlated with the Molar Refractivity **MR** (r= 0.962 and p <0.05) and with the Parachor **Pc** (r=0.962 and p<0.05), and is negatively correlated with the Total Energy $\mathbf{E_T}$(r=-0,974 and p <0.05) at a significant level.

* The Molar refractivity **MR** is strongly correlated with the Parachor **Pc** (r =0.995 and p < 0.05) and the molar volume **MV** (r= 0.984 and p <0.05) at a high level.

* The Energy of activation $\mathbf{E_a}$ is negatively correlated with the maximum of absorption $\boldsymbol{\lambda}_{\mathbf{max}}$ (r= -0.984 and p <0.05) at a significant level.
* The Density **D** is positively correlated with the surface tension $\boldsymbol{\gamma}$ (r =0.942 and p< 0.001) at a significant level.
Analysis of projections according to the planes F1–F2 and F1-F3 (59.53% and 58.28% of the total variance respectively) of the studied molecules (Figure 4) shows that the molecules are dispersed, according to the structure of the R groups of pyrazolo[4,3-d] pyrimidines, in two classes of compounds belonging to two groups: **group 1**, the substitution of electron withdrawing groups in pyrazol ring and sterically less bulkier groups could give more potential inhibition (positive contribution, pIC_{50}>1.06) and **group 2**, the compound without the substitution of electron withdrawing groups on the aromatic yield less potent phosphodiesterase V inhibitors and the substitution at the R_1 should be small, the increase in chain length could decrease the activity(low activity pIC_{50}<1,06).

Statistical Analysis
The obtained Artificial Neural Networks ANN and Multiple Nonlinear Regression MNR correlation coefficient R values are respectively 0.928 and 0.999 for this data set of 1-(2-ethoxyethyl)-1H-pyrazolo [4,3-d] pyrimidine derivatives. They confirm that the Artificial Neural Networks ANN or Multiple Nonlinear Regression (MNLR) 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 the CPA, MLR, MNLR and ANN models shows that the ANN and MNLR models have substantially better predictive capability because the ANN and MNLR approach give better results than MLR. ANN and MNLR were able to establish a satisfactory relationship between the molecular descriptors and the activity of the studied compounds.

We have investigated the QSAR regression to predict the inhibitory activities (Inhibition Concentration IC_{50}) of several organic compounds based on 1-(2-ethoxyethyl)-1H-pyrazolo[4,3-d]pyrimidines. Comparison of key statistical terms like R or R^2 of different models obtained by using different statistical tools and different descriptors has been shown in table 5.

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

N°	Obs (pIC_{50})	Pred (pIC_{50})		
		MLR	MNLR	ANN
1	1,155	1,290	1,155	1,155
2	0,398	0,138	0,398	0,381
3	1,155	0,673	1,155	1,173
4	0,187	-0,058	0,187	0,025
5	0,180	0,013	0,180	-0,195
6	-0,146	-0,058	-0,146	0,016
7	0,420	0,006	0,420	-0,200
8	-0,519	0,013	-0,519	-0,195
9	-0,740	0,013	-0,740	-0,195
10	-0,415	-0,322	-0,415	-0,479
11	0,051	0,027	0,051	0,055
12	-0,398	-0,243	-0,398	-0,086
13	-0,079	-0,242	-0,079	-0,086
14	0,119	0,306	0,119	0,264
15	0,699	0,112	0,699	0,180
16	0,143	0,280	0,143	0,497
17	1,060	0,805	1,060	0,949
18	0,377	0,606	0,377	0,458
19	-0,362	0,372	-0,362	0,099
20	0,108	0,613	0,108	0,489
21	0,456	0,112	0,456	0,180
22	1,155	0,628	1,155	0,497
23	-0,079	-0,151	-0,079	-0,070
24	1,155	1,078	1,155	1,019
25	1,097	1,132	1,097	1,227
26	1,398	1,433	1,398	1,418

CONCLUSION

In this work, the study of the quality of the CPA, MLR, MNLR and ANN models shows that the ANN and MNLR result have substantially better predictive capability than the other methods. With ANN approach, we have established a relationship between several descriptors and the inhibition values (pIC_{50}) of 1-(2-ethoxyethyl)-1H-pyrazolo[4,3-d]pyrimidine as phosphodiesterase V inhibitors in satisfactory manners.

Finally, we can conclude that one 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] M Sausbier; R Schubert; V Voigt; C Hirneiss; A Pfeifer; M Korth; T Kleppisch; P Ruth; F Hofmann, *Circ. Res*, **2000**, 87, 825–830.
[2] L M McAllister-Lucas; W K Sonnenburg; A Kadlecek; D Seger; H L Trong; J L Colbran; M K Thomas; K A Walsh; S H Francis; J D Corbin; J A Beavo, *J. Biol. Chem.*, **1993**, 268, 22863–22873.
[3] T T Chang; H J Huang; K J Lee; H W Yu; H Y Chen; F U Tsai; S Mao-Feng; C C Yu-Chian, *J. Biomol. Struct. Dyn.*, **2010**, 28, 309–321.
[4] M B Tollefson; B A Acker; E J Jacobsen; R O Hughes; J K Walker; D N A Fox; M J Palmer; S K Freeman; Y Yu; B R Bond, *Bioorg. Med. Chem. Lett*, **2010,** 20, 3120–3124.
[5] G F Yang; H T Lu; Y Xiong; C G Zhan, *Bioorg. Med. Chem.*, **2006,** 14, 1462–1473.
[6] C Chen; Y Chang; D Bau; H Huang; F Tsai; C Tsai, C Y Chen, *Acta Pharmacol. Sin.*, **2009,** 30, 1186–1194.
[7] T Matsumoto; T Kobayashi; K Kamata, *J. Smooth Muscle Res.*, **2003,** 39, 67–86.
[8] T Reffelmann; R A Kloner, *Cardiovasc. Res.*, **2009,** 83, 204–212.
[9] M Guazzi; *Circ. Heart Fail*, **2008,** 1, 272–280.
[10] M J Palmer, A S Bell, D N A Fox, D G Brown, *Curr. Top. Med. Chem*, **2007**, 7, 405–419.
[11] P Srivani; E Srinivas; R Raghu; G Narahari Sastry, *J. Mol. Graph. Model*, **2007,** 26, 378–390.
[12] D P Rotella, *Drug Discov.*, **2002,** 1, 674–682.
[13] C Hansch; P P Maloney; T Fujita; R M Muir, *Nature*, **1962,** 194, 178–180,.
[14] J D Fegade, S S Rane, R Y Chaudhari, V R Patil, *Digest J. Nanomater. Biostruct*, **2009,** 4, 145–154.
[15] K K Jhaa; A Samadc; Y Kumar; M Shaharyar; R Khosad; J Jaina; S Bansald, *IJPR*, **2009,** 8, 163–167.
[16] M C Sharma; S Sharma, *Int. J. Chem. Tech. Res.*, **2010,** 2, 606–614.
[17] K Laarej; M Bouachrine; S Radi; S Kertit; B Hammouti, *E-J. Chem.*, **2010**, 7 (2), 419–424.
[18] S Chtita; M Larif; M Ghamali; A Adad; R Hmamouchi; M Bouachrine; T Lakhlifi, *IJIRSET*, **2013,** 2 (11), 6586-6601.
[19] P Choudhari; M Bhatia; *Journal of Saudi Chemical Society*, **2012**, http://dx.doi.org/10.1016/j.jscs.2012.02.008.
[20] Gaussian 03; M J Revision B.01; M J Frisch and al., *Gaussian, Inc., Pittsburgh, PA*, **2003**.
[21] C Lee; W Yang; R G Parr; *Phys. Rev.*, **1988,** 37, 785-789.
[22] S Chtita; M Larif; M Ghamali and al., *IJIRSET*, **2013**, 2(11), 6586-6601.
[23] J N Hogarh; N Seike; Y Kobara; YA Habib; J J Namd; J S Lee; X Liu; L Jun; G Zhang; S Masunaga, *Chemosphere*, **2012**, 86, 718–726.
[24] M Ghamali; S Chtita; R Hmamouchi; A Adad; M Bouachrine; T Lakhlifi, *IJARCSSE*, **2014**, 4(1), 536–546.
[25] S Huang; R Li; P J Connolly; S Emanuel and S A Middleton, *Bioorganic & Med. Chem. Lett.*, **2006,** 16, 4818–4821.
[26] J Zupan; J Gasteiger, *Weinheim*, 1999.
[27] R Chimizou; H Iwamura; T Fujita, *Agric. Food Chem.*, **1988**, 36, 1276.
[28] N Turkkan, *Revue de l'Université de Moncton*, **1993,** 26 (1), 205–221.
[29] G Jing; Z Zhou; J Zhuo, *Chemosphere*, **2012**, 86, 76-82.
[30] P Y Lee; C Y J Chen, *Hazard. Mater.*, **2009**, 165, 156-161.
[31] B Elidrissi; A Ousaa; M Ghamali; S Chtita; M A Ajana; M Bouachrine; T Lakhlifi, *JCMMD*, **2014**, 28-37.