Journal of materials and Environmental Sciences
ISSN : 2028-2508

JMES, 2017 Volume 8, Issue 5, Page 1532-1545

http://www.jmaterenvironsci.com/

QSAR Study of Flavonoid Derivatives as in Vitro Inhibitors Agents of Aldose Reductase (ALR2) Enzyme for Diabetic Complications

Y. Boukarai[1*], F. Khalil[1], M. Bouachrine[2]

1. LAC, Laboratory of Applied Chemistry, Faculty of Science and Technology, University Sidi Mohammed Ben Abdellah, Fez, Morocco
2. ESTM, University Moulay Ismail, Meknes, Morocco

Received 08 Nov 2016,
Revised 17 Jan 2017,
Accepted 20 Jan 2017

Keywords

- ✓ Diabetic complications;
- ✓ Aldose reductase (ALR2);
- ✓ Flavonoid derivatives;
- ✓ Phenyl-benzopyrane;
- ✓ QSAR, PCA, MLR, MNLR, ANN, CV

Y.Boukarai
boukarai.youness@gmail.com
+212665909199

Abstract

Diabetes mellitus is one of the most common chronic metabolic diseases, characterized by chronic hyperglycaemia and the development of diabetes-specific microvascular and macrovascular pathology. Prolonged hyperglycemia is a primary causal factor of several diabetic complications. The flavone (phenyl-benzopyrane) and its derivatives are potent inhibitors agents, these compounds inhibit Aldose Reductase (ALR2) enzyme. A study of quantitative structure-activity relationship (QSAR) is applied to a set of 29 molecules derived from phenyl-benzopyrane, in order to predict the ALR2 inhibitory biological activity of the test compounds and find a correlation between the different physic-chemical parameters (descriptors) of these compounds and its biological activity, using principal components analysis (PCA), multiple linear regression (MLR), multiple non-linear regression (MNLR) and the artificial neural network (ANN). We accordingly propose a quantitative model (non-linear and linear QSAR models), and we interpret the activity of the compounds relying on the multivariate statistical analysis. The topological descriptors were computed, respectively, with ACD/ChemSketch and ChemBioOffice 14.0 programs. A good correlation was found between the experimental activity and those obtained by MLR and MNLR respectively such as ($R = 0,80$ and $R^2 = 0,64$) and ($R = 0,83$ and $R^2 = 0,69$), this result could be improved with ANN such as ($R = 0,88$ and $R^2 = 0,77$) with an architecture ANN (5-2-1). To test the performance of the neural network and the validity of our choice of descriptors selected by MLR and trained by MNLR and ANN, we used cross-validation method (CV) such as ($R = 0,833$ and $R^2 = 0,693$) with the procedure leave-one-out (LOO). This study show that the MLR and MNLR have served to predict activities, but when compared with the results given by an 5-2-1 ANN model we realized that the predictions fulfilled by this latter was more effective and much better than other models. The statistical results indicate that this model is statistically significant and shows very good stability towards data variation in leave-one-out (LOO) cross validation.

Abbreviations: *QSAR, Quantitative Structure-Activity Relationship; PCA, Principal Component Analysis; MLR, Multiple Linear Regression; MNLR, Multiple Non-Linear Regression; ANN, Artificial Neural Networks; CV, Cross Validation; LOO-CV, Leave One Out Cross-Validation; R, Correlation Coefficient; R^2, Coefficient of Determination; R^2_{aj}, Adjusted Coefficient of Determination; q^2, Coefficient of Prediction; SD, Standard Deviation; MW, Molecular Weight; MR, Molar Refractivity; MV, Molar Volume; Pc, Parachor; n, Refractive Index; γ, Surface Tension; D, Density; αe, Polarizability; LogP, Lipophilic; HBA, Hydrogen Bond Acceptor; HBD, Hydrogen Bond Donor; SSE, Sum of residual (Error) Squares; SSF, Sum of regression (Factor) Squares; MSE (V_E: Error Variance), Mean Squared Error; MSF (V_F: Factor Variance), Mean Squared Factor; F, Fishers F-statistic; F value, Significance level; p-value, Critical Probability*

1. Introduction

Worldwide, diabetes is achieving higher dimensions due to change in people life style, which lead to reduced physical activity and increased obesity, the main roots of diabetic conditions. As per world health organization (WHO) diabetes webpage, 347million people have diabetes, which is projected to be the 7th leading cause of deaths by 2030 **[1]**. Diabetes mellitus is one of the most common chronic metabolic diseases, characterized by chronic hyperglycaemia and the development of diabetes-specific microvascular and macrovascular pathology. Prolonged hyperglycemia is a primary causal factor of several diabetic complications. Large prospective clinical studies show a strong relationship between glycaemia and diabetic microvascular complications in both type 1 and type 2 diabetes **[2,3]**.

Many studies have revealed a correlation between glucose metabolism via the polyol pathway and long-term diabetic complications. Aldose reductase, ALR2, is the first and rate-limiting enzyme in this pathway which normally reduces glucose to sorbitol using Nicotinamide adenine dinucleotide phosphate (NADPH) as a

cofactor, at the same time another enzyme sorbitol dehydrogenase oxidizes sorbitol to fructose. However, in diabetic condition, the glucose level in this pathway is increased and sorbitol is produced faster than being oxidized to fructose **[4]**.The accumulated sorbitol cannot cross the cell membrane easily and therefore causes swelling and cell dysfunction in a number of tissues. In addition, fructose can become phosphorylated to fructose-3-phosphate, which is broken down to 3-deoxyglucosone, ultimately forming advanced glycation end products that are capable of cellular damage **[5-7]**. These abnormal metabolic results have been reported to be responsible for diabetic complications such as cataracts **[8]**, retinopathy **[9]**, neuropathy **[10]**, and nephropathy **[11]**. The inhibition of ALR2 is a possible prevention or treatment of these effects **[12]**.

The flavonoids are of low molecular weight plant products which are abundant, ubiquitously found in a wide variety of edible plants, fruits, nuts, seeds, and plant-derived beverages, such as juice and tea **[13]**. They are also called vitamin P16 and have been described as health-promoting, disease-preventing dietary supplements **[14]**. They are relatively simple to synthesize and extremely safe and associated with low toxicity, making them excellent candidates for several interesting biological activity profiles in enzymatic systems, and as chemo-preventive agents **[15]**. They may exert an anti-hyperglycaemic effect by promoting peripheral utilization of glucose or enhancing the sensitivity of insulin in diabetic animals **[16]**. In addition, it was reported that the therapeutic benefits of flavonoids are usually linked to two properties: (i) inhibition of certain enzymes such as xanthine oxidase, ALR2 **[17]**, acetyl-cholinesterase **[18]**, Janus kinase **[19]**, Spleen Tyrosine Kinase **[20]** and (ii) antioxidant activity **[21]**, consequently their study is greatly interested in many research fields.

Quantitative structure-activity relationship (QSAR) tries to investigate the relationship between molecular descriptors that describe the unique physicochemical properties of the set of compounds of interest with their respective biological activity or chemical property **[22,23]**.

In this work we attempt to establish a quantitative structure-activity relationship between ALR2 inhibitory activity of a series of 29 bioactive molecules derived from flavonoid (phenyl-benzopyrane) and structural descriptors. Thus we can predict the ALR2 inhibitory activity of this group of organic compounds. Therefore we propose a quantitative model, and we try to interpret the activity of these compounds based on the different multivariate statistical analysis methods include:

* The Principal Components Analysis (PCA) has served to classify the compounds according to their activities and the variability of the descriptors. It allows to reduce the number of variables not significant and make the information less redundant. * The Multiple Linear Regression (MLR) has served to select the descriptors used as the input parameters for the Multiples Non-Linear Regression (MNLR) and Artificial Neural Network (ANN). * The artificial neural network (ANN) which is a nonlinear method, which allows the prediction of the activities. * Cross-validation (CV) to validate models used with the process leave-one-out (LOO).

2. Experimental details

2.1. *Experimental data*

The Biological data used in this study were inhibitory activity against ALR2 (inhibition of aldose reductase enzyme. (IC_{50})), a set of twenty-nine derivatives of flavone (phenyl-benzopyrane). We have studied and analyzed the series of phenyl-benzopyrane molecule consists of 29 selected derivatives that have been synthesized and evaluated for their inhibitory activity in vitro against ALR2 (in terms of -log (IC_{50})) **[17]**. This in order to determine a quantitative structure-activity relationship between ALR2 inhibitory activity and the structure of these molecules that are described by their substituents 3, 5, 6, 7, 8, 2', 3', 4',and 5'.

The chemical structure of phenyl-benzopyrane is represented in **Figure1**.

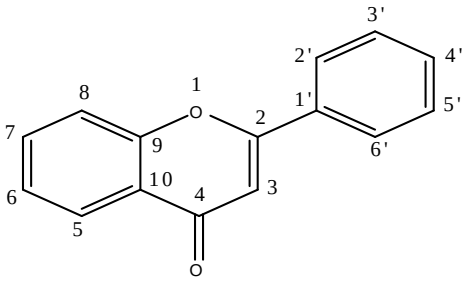

Figure1: The general structure of phenyl-benzopyrane

The chemical structures of 29 compounds of phenyl-benzopyrane used in this study and their experimental ALR2 inhibitory biological activity observed IC_{50} (Cytotoxic concentration required to inhibit the aldose reductase enzyme ALR2 than 50%) are collected from recent publications **[17]**. The observations are converted into logarithmic scale -log (IC_{50}) in molar units (M) and are included in **Table1**.

Table1: Chemical structures and activities observed of phenyl-benzopyrane derivatives against ALR2

N°	Position of substituents									pIC_{50} [a] Obs
	3	**5**	**6**	**7**	**8**	**2'**	**3'**	**4'**	**5'**	
1	OCH_3	OH	OCH_3	OH			OH	OH		7.55
2		OH	OCH_3	OH	CH_2Ph		OH	OH		7.47
3		OH	OCH_3	OCH_3	OCH_3		OH	OH		7.41
4	OCH_3	OH	OH	OH			OH	OH		7.24
5		OCH_3		OH	OCH_3		OH	OH		7.13
6		OH	OH	OH	OCH_3		OH	OH		6.92
7	OCH_3	OCH_3		OCH_3	OCH_3		OH	OH		6.77
8		OH	OH	OH			OH	OH		6.69
9		OH	OCH_3	OCH_3			OH	OH		6.77
10		OH		OCH_3	OH		OH	OH		6.64
11	OCH_3	OH		OH	OCH_3		OH	OH		6.62
12		OH		OH	OCH_3		OH	OH		6.55
13		OH	OH	OCH_3			OH	OH		6.52
14	OCH_3	OH	OCH_3	OCH_3			OH	OH		6.46
15	OCH_3	OH	OH	OCH_3			OH	OH		6.09
16		OH	OH	OCH_3	OCH_3			OH		6.07
17		OH	OH	OH	OCH_3			OH		5.92
18		OH	OH	OH	OCH_3		OCH_3	OH		5.92
19		OH	OCH_3	OCH_3				OH		5.85
20		OH	OCH_3	OH	OCH_3		OCH_3	OH		5.35
21		OCH_3	OH	OCH_3	OCH_3		OCH_3	OH		5.20
22		OH	OCH_3	OCH_3			OCH_3	OH		5.17
23		OH	OH	OH	OCH_3					5.09
24		OCH_3	OH	OCH_3	OCH_3					3.54
25		OH		OCH_3		OCH_3		OCH_3	OH	3.50
26		OCH_3		OH						3.00
27		OH		OCH_3		OCH_3		OH	OCH_3	3.00
28				OH		OH				5.78
29				OH		OH		OH		6.46

[a] **pIC_{50} = -log (IC_{50})**.

2.2. *Calculation of molecular descriptors*

Advanced chemistry development's ACD/ChemSketch program was used to calculate Molecular Weight (MW), Molar Refractivity (MR (cm^3)), Molar Volume (MV (cm^3)), Parachor (Pc (cm^3)), Density (D (g/cm^3)), Refractive Index (n), Surface Tension (γ (dyne/cm)) and Polarizability (α_e (cm^3)) **[24,25]**.

Steric, thermodynamic descriptors are calculated using ACD/ChemSketch and ChemBioOffice 14.0 **[26]** after optimization of the energy for each compound using the MM2 method (force field method with Gradient Setting Root Mean Square (RMS) 0.1 kcal mol^{-1}) **[27]**.

In this work 11 descriptors were chosen to describe the structure of the molecules constituting the series to study: the molecular weight (MW), the molar refractivity (MR (cm^3)), the molar volume (MV (cm^3)), the parachor (Pc (cm^3)), the refractive index (n), the surface tension (γ (dyne/cm)), the density (D (g/cm^3)), the

polarizability (α_e (cm^3)), the lipophilic (LogP), the hydrogen bond acceptor (HBA) and the hydrogen bond donor (HBD).

2.3. *Statistical analysis*

To explain the structure-activity relationship, these 11 descriptors are calculated for 29 molecules (**Table2**) through software ChemSketch and ChemBioOffice 14.0.

Table2: The values of the 11 chemical descriptors

Compounds	**MW**	**MR**	**MV**	**Pc**	**n**	**γ**	**D**	**α_e**	**LogP**	**HBA**	**HBD**
1	346.288	83.46	208.7	651.9	1.731	95.1	1.650	33.08	0.586	8	4
2	406.384	107.72	276.3	807.3	1.707	72.8	1.470	42.70	3.481	7	4
3	360.314	89.88	246.5	697.2	1.649	63.9	1.461	35.63	1.525	8	3
4	332.261	78.63	183.8	608.5	1.800	120	1.800	31.17	0.323	8	5
5	330.288	83.20	222.5	638.6	1.670	67.7	1.483	32.98	1.651	7	3
6	332.261	80.29	195.4	610.4	1.758	95.2	1.700	31.83	0.998	8	5
7	374.341	93.13	258.5	738.7	1.639	66.6	1.440	36.92	1.112	8	2
8	302.235	73.61	171.4	551.8	1.804	107.4	1.763	29.18	1.125	7	5
9	330.288	83.20	222.5	638.6	1.670	67.7	1.483	32.98	1.651	7	3
10	316.262	78.41	196.9	595.2	1.727	83.3	1.605	31.08	1.388	7	4
11	346.288	83.46	208.7	651.9	1.731	95.1	1.650	33.08	0.586	8	4
12	316.262	78.41	196.9	595.2	1.727	83.3	1.605	31.08	1.388	7	4
13	316.262	78.41	196.9	595.2	1.727	83.3	1.605	31.08	1.388	7	4
14	360.314	88.30	233.6	695.3	1.680	78.4	1.540	35.00	0.849	8	3
15	346.288	83.46	208.7	651.9	1.731	95.1	1.650	33.08	0.586	8	4
16	330.288	83.20	222.5	638.6	1.670	67.7	1.483	32.98	1.651	7	3
17	316.262	78.41	196.9	595.2	1.727	83.3	1.605	31.08	1.388	7	4
18	346.288	85.09	220.9	653.8	1.696	76.6	1.566	33.73	1.262	8	4
19	314.289	81.32	224.1	623.4	1.645	59.8	1.402	32.24	2.041	6	2
20	360.314	89.88	246.5	697.2	1.649	63.9	1.461	35.63	1.525	8	3
21	374.341	94.68	272.1	740.6	1.612	54.8	1.375	37.53	1.788	8	2
22	344.315	88.00	248.1	682.0	1.627	57.0	1.387	34.88	1.914	7	2
23	300.262	76.53	198.5	580.0	1.697	72.8	1.512	30.33	1.778	6	3
24	328.316	86.12	249.7	666.8	1.606	50.8	1.314	34.14	2.304	6	1
25	344.315	88.00	248.1	682.0	1.627	57.0	1.387	34.88	1.914	7	2
26	268.264	72.76	201.7	549.5	1.641	55.1	1.329	28.84	2.557	4	1
27	344.315	88	248.1	682	1.627	57	1.387	34.88	1.914	7	2
28	254.237	67.97	176.1	506.1	1.698	68.2	1.443	26.94	2.293	4	2
29	270.236	69.85	174.5	521.4	1.732	79.5	1.548	27.69	1.904	5	3

The statistical study we conducted consists of:

-The principal component analysis (PCA), the multiple linear regressions (MLR), and the non-linear regression (MNLR) available in the XLSTAT 15 software **[28]**.

-The Artificial Neural Network (ANN) and the leave-one-out cross validation (CV-LOO) are done on Matlab 7 using a program written in C language.

The structures of the molecules based on flavonoid (phenyl-benzopyrane) derivatives were studied by the principal component analysis. The PCA is a multivariate statistical technique useful for description the information encoded in the structures of the compounds. It is also very helpful for understanding the distribution of the compounds. This is an essentially descriptive statistical method which aims to present, in graphic form, the maximum of information contained in the data **Table2** and **Table3**.

Table3: The correlation matrix (Pearson (n)) between different studied descriptors

Variables	**MW**	**MR**	**MV**	**Pc**	**n**	**γ**	**D**	**α_e**	**LogP**	**HBA**	**HBD**	**pIC_{50}**
MW	**1**											
MR	0.957	**1**										
MV	0.827	0.931	**1**									
Pc	**0.968**	**0.992**	0.935	**1**								
n	-0.354	-0.530	-0.799	-0.551	**1**							
γ	-0.137	-0.364	-0.654	-0.356	0.949	**1**						
D	-0.116	-0.357	-0.650	-0.351	0.943	0.983	**1**					
α_e	0.957	**1**	0.931	**0.992**	-0.530	-0.364	-0.357	**1**				
LogP	-0.203	0.083	0.250	-0.005	-0.467	-0.682	-0.706	0.083	**1**			
HBA	0.852	0.667	0.477	0.707	0.000	0.246	0.308	0.667	-0.636	**1**		
HBD	0.040	-0.175	-0.501	-0.200	0.876	0.889	0.941	-0.175	-0.571	0.407	**1**	
pIC_{50}	0.190	0.063	-0.155	0.056	0.452	0.494	0.518	0.063	-0.376	0.348	0.551	**1**

The multiple linear regression method 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. It has served also to select the significant descriptors used as the input parameters in the multiple non-linear regression (MNLR) and artificial neural network (ANN).

The (MLR) and the (MNLR) were generated to predict cytotoxic effects IC_{50} activities of phenyl-benzopyrane derivatives. Equations were justified by the coefficient of determination (R^2), the mean squared error (MSE), the test- F (Fisher) and the significance level (p-value) **[29-31]**.

ANN is artificial systems simulating the function of the human brain. Three components constitute a neural network: the processing elements or nodes, the topology of the connections between the nodes, and the learning rule by which new information is encoded in the network. While there are a number of different ANN models, the most frequently used type of ANN in QSAR is the three-layered feed-forward network **[32]**. In this type of networks, the neurons are arranged in layers (an input layer, one hidden layer and an output layer). Each neuron in any layer is fully connected with the neurons of a succeeding layer and no connections are between neurons belonging to the same layer.
Cross-validation is a popular technique used to explore the reliability of statistical models. Based on this technique, a number of modified data sets are created by deleting in each case one or a small group of molecules, these procedures are named respectively "leave-one-out" and "leave-some-out" **[33-35]**. For each data set, an input-output model is developed. In this study we used, the leave-one-out (LOO) procedure.

3. Results and Discussion

3.1. Data set for analysis
The QSAR analysis was performed using the -log (IC_{50}) of the 29 selected molecules that have been synthesized and evaluated for their inhibitory activity in vitro against ALR2 (experimental values) **[17]**. The exploitation of experimental data observed by the use of mathematical and statistical tools is an effective method to find new chemical compounds with high ALR2 inhibitory activity. The values of the 11 chemical descriptors as shown in Table2. The principle is to perform in the first time, a PCA, which allows us to choose the relevant descriptors from several correlated descriptors (dependent), then perform a decreasing study of MLR based on the elimination of no significant descriptors until a valid model (including the critical probability: **p-value < 0.05** for all descriptors and the model complete).

3.2. Principal Components Analysis (PCA)

The totality of the 11 descriptors (variables) coding the 29 molecules was submitted to a principal components analysis. 12 principal components were obtained (**Figure2**). The first two components F1 and F2 contributing respectively 53.37 % and 35.37 % to the total variance, the total information is estimated to a percentage of 88.74%.

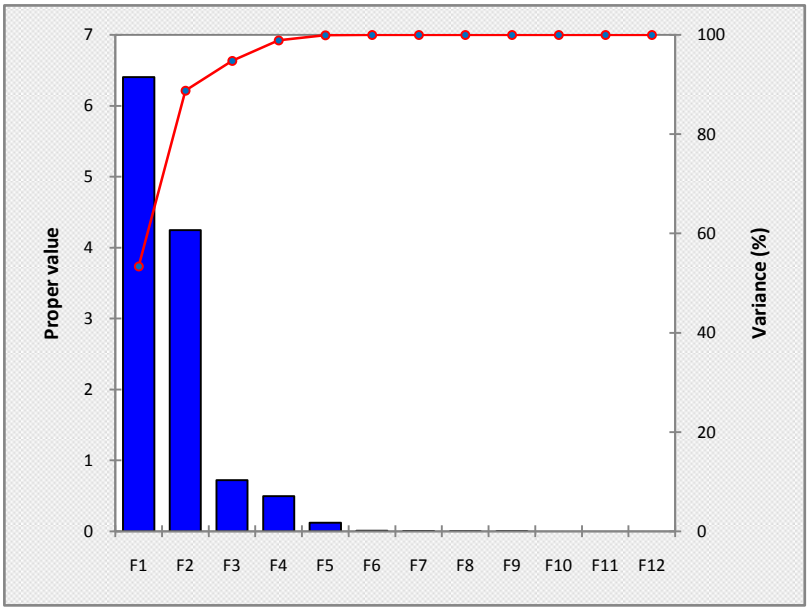

Figure2: The principal components and there variances

The Pearson correlation coefficients are summarized in the above Table3. The obtained matrix provides information on the negative or positive correlation between variables. The principal component analysis (PCA) was conducted to identify the link between the different variables. Correlations between the 11 descriptors are shown in Table3 as a correlation matrix and in **Figure3** these descriptors are represented in a correlation circle.

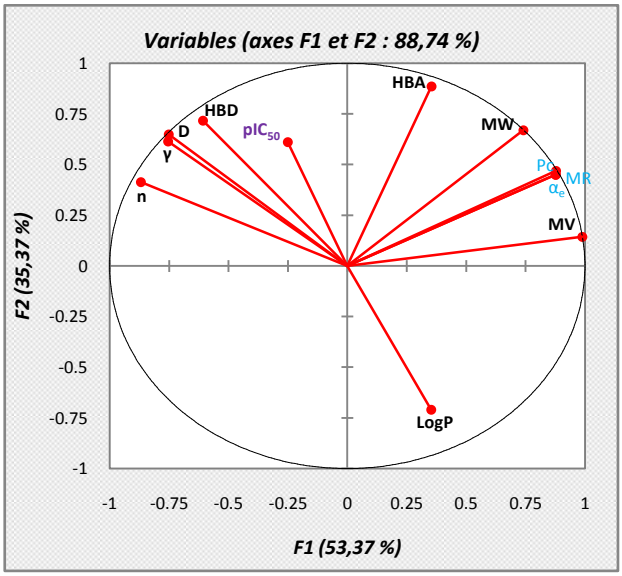

Figure3: Correlation circle

In order to reduce the number of parameters not significant and according to the matrix and the circle of correlations we observed the following correlations:

- MR and α_e are perfectly correlated (r = 1), both variables are redundant.
- Pc, MR and α_e are highly correlated (r (Pc, MR) = 0,992; r (Pc, α_e) = 0,992).
- Pc and MW are highly correlated (r (Pc, MW) = 0,968).

Therefore, the variables (ae) and (Pc) were then not retained
In addition, we have tried also to eliminate the descriptors γ or D, (because they are also correlated) but we are not able to find adequate model. We therefore retained the descriptors γ and D and eliminated only (αe) and (Pc).

3.3. Multiple Linear Regression (MLR)
In order to propose a mathematical model linking the descriptors and activity, and to evaluate quantitatively the substituent's physicochemical effects on the activity of the totality of the set of these 29 molecules, we presented the data matrix which is the corresponding physicochemical variables different substituent's from 29 molecules to a multiple linear regression analysis. This method used the coefficients R^2, R^2_{aj}, q^2, and the p-values to select the best regression performance.
Treatment with multiple linear regression is more accurate because it allows you to connect the structural descriptors for each activity of 29 molecules to quantitatively evaluate the effect of substituent. The selected descriptors are: **MW**, **MR, n, γ** and **D**.
The QSAR model built using multiple linear regression (MLR) method is represented by the following equation:

$$\mathbf{pIC_{50\ MLR}} = -201{,}588 + 0{,}414\ \mathbf{MW} - 1{,}515\ \mathbf{MR} + 191{,}552\ \mathbf{n} - 0{,}137\ \boldsymbol{\gamma} - 76{,}928\ \mathbf{D} \qquad (\mathbf{Equation\ 1})$$

3.4. Validation criteria of the MLR model (ANOVA: Analysis Of Variance)
In order to validate the correlation equation provided by the statistical method of multiple linear regression (MLR), different criteria may be used **[36]**.

3.4.1. Overall assessment of the regression
Table 4 summarizes the results of the regression analysis of variance such as Fisher's value (Fexp) and overall p-value of the model.

Table4: Variance analysis

Source	SS	df	MS	Fexp	p-value
Regression	**29.372**	**5**	**5.874**	**7.851**	**0.000**
Residual	**17.210**	**23**	**0.748**	**-**	**-**
Total	**46.582**	**28**	**6.622**	**-**	**-**

The results seem excellent and the model parameters effect is significant because we achieved lower overall p-value at α = 0.05 level (p-value =0.000).

-To a threshold of (**0.05**) comparing $\boldsymbol{F_{exp}}$ obtained by the theoretical calculation and that obtained from Fisher's table $\boldsymbol{F_{(p,N-p-1)}}$ for one degree of freedom (**p, N-p-1**) with **p = 5** and **N = 29**, such as **(N-p-1) = 23**.
-We Accept H_1 if $\boldsymbol{F_{exp} > F_{(5,23)}}$.
-We then find $\boldsymbol{F_{(5,23)}}$ **= 2.64** and $\boldsymbol{F_{exp}}$ **= 7.851**, so we accept H_1 and H_0 is rejected.

-A good correlation between the target activity and initial activity if **R** is closer to **1**.
-A non-linear correlation between the target activity and initial activity if **R** is closer to **0**.
-In our case we have **R = 0.80**, so a good correlation was shown between the observed activity and that obtained by **MLR**.

-In our case we have $\mathbf{R^2}$ **= 0.64**, this figure means that **64%** of the variable Y (activity) is attributable to the variation in the variable X (descriptors), which indicates that this model is statistically explanatory.

-In our case we have $\mathbf{R^2_{aj} = 0.56}$, so the overall quality of the MLR is best. This indicates that this model is statistically significant.

-In our case we have $\mathbf{q^2 = 0.98 > 0.6}$, so the predictive power of this model is very significant, which shows that the model proposed in this paper is able to predict activity with a great performance, and that the selected descriptors are pertinent. This means that the prediction of the new compounds is feasible.

-we can enjoy the performance of the predictive power of this model to explore and propose new molecules could be active.

-N: (**N = 29**) number of data points considered.
-p: (**p = 5**) number of restrictions on the degrees of freedom (equal to the number of parameters).
-In our case we have **SD = 0.86**, so the correlation between the observed activity and that obtained by MLR is best.

N = 29	**R = 0.80**	**R^2 = 0.64**	**F = 7.851**	**MSE = 0.748**

Given the fact that the probability corresponding to the F value is much smaller than **0.05**, it mean that we would be taking a lower than 0.01 % risk in assuming that the null hypothesis is wrong. Therefore, we can conclude that the model do bring a significant amount of information.

The elaborated QSAR model reveals that the ALR2 inhibitory activity could be explained by a number of topologic factors. The negative correlation of the Molar Refractivity (MR), the Surface Tension (γ) and the Density (D) with the ability to displace the flavonoid (phenyl-benzopyrane) activity reveals that a decrease in the value of pIC_{50}, While the positive correlation of the descriptors (Molecular Weight (MW) and the Refractive Index (n)) with the ability to displace the phenyl-benzopyrane activity reveals that an increase in the value of pIC_{50}.

With the optimal MLR model, the values of predicted activities $\mathbf{pIC_{50\ MLR}}$ calculated from equation1 and the observed values are given in **Table5**. The correlations of predicted and observed activities are illustrated in **Figure4**. The descriptors proposed in equation1 by MLR were, therefore, used as the input parameters in the multiples non-linear regression (MNLR) and artificial neural network (ANN).
The correlation between MLR calculated and experimental activities are very significant as illustrated in Figure4 and as indicated by R and R^2 values.

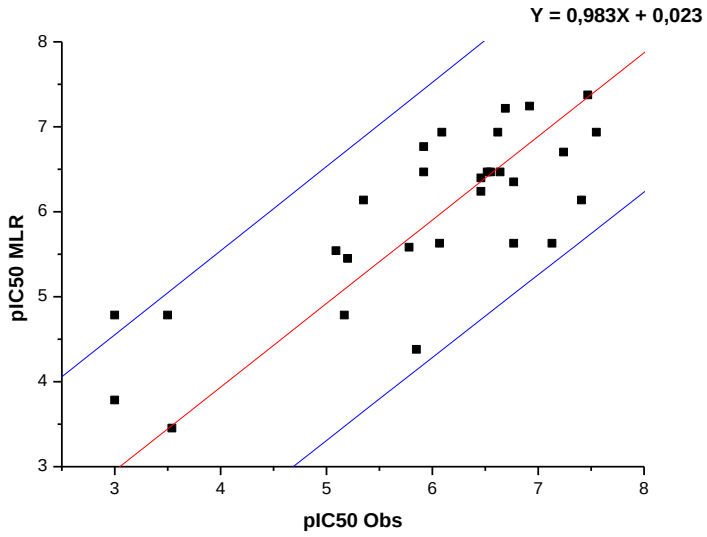

Figure4: Correlations of observed and predicted activities calculated using MLR

Table5: The observed, the predicted activities (**pIC_{50}**), according to different methods MLR, MNLR, ANN and CV-MLR for the 29 derivatives of flavonoid (phenyl-benzopyrane)

N°	**$pIC_{50\ Obs}$**	**$pIC_{50\ MLR}$**	**$pIC_{50\ MNLR}$**	**$pIC_{50\ ANN}$**	**$pIC_{50\ CV\text{-}MLR}$**
1	7.55	6.936	6.963	6.557	6.62
2	7.47	7.374	7.473	6.557	7.55
3	7.41	6.136	6.260	6.533	6.33
4	7.24	6.701	7.052	6.557	6.14
5	7.13	5.630	5.945	6.557	7.28
6	6.92	7.242	6.478	6.557	6.94
7	6.77	6.350	6.061	6.554	7.66
8	6.69	7.216	6.770	6.557	6.78
9	6.77	5.630	5.945	6.557	7.11
10	6.64	6.469	6.419	6.557	7.23
11	6.62	6.936	6.963	6.557	6.31
12	6.55	6.469	6.419	6.557	6.74
13	6.52	6.469	6.419	6.557	6.63
14	6.46	6.399	6.589	6.557	6.38
15	6.09	6.936	6.963	6.557	7.20
16	6.07	5.630	5.945	6.557	6.01
17	5.92	6.469	6.419	6.557	5.73
18	5.92	6.766	6.731	6.557	6.47
19	5.85	4.380	4.689	5.831	4.49
20	5.35	6.136	6.260	6.533	4.81
21	5.20	5.451	4.982	5.191	6.15
22	5.17	4.785	4.844	3.891	5.94
23	5.09	5.541	5.887	6.557	4.22
24	3.54	3.452	3.061	3.554	4.36
25	3.50	4.785	4.844	3.891	4.35
26	3.00	3.783	3.308	2.988	3.17
27	3.00	4.785	4.844	3.891	3.70
28	5.78	5.581	5.604	5.781	6.68
29	6.46	6.241	6.542	6.557	5.24

3.5. Multiple 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. We have applied to the data matrix constituted obviously from the descriptors proposed by MLR corresponding to the 29 molecules. The coefficients R, R^2, and the F-values are used to select the best regression performance. We used a pre-programmed function of XLSTAT following:

$$\mathbf{Y = a + (bX_1 + cX_2 + dX_3 + eX_4 \ldots) + (fX_1^2 + gX_2^2 + hX_3^2 + iX_4^2 \ldots)}$$

Where a, b, c, d… represent the parameters and X_1, X_2, X_3, X_4…: represent the variables. The resulting equations:

$pIC_{50\ MNLR}$ = 3,970 - 0,223 **MW** + 2,171 **MR** - 335,208 **n** - 0,400 **γ** + 194,231 **D** +6,139E-04 (**MW**)^2 - 1,602E-02 (**MR**)^2 + 146,150E-04 (**n**)^2 + 2,101E-03 (**γ**)^2 - 81,460 (**D**)^2 (**Equation 2**)

N = 29 R = 0.83 R^2 = 0.69 MSE = 0.829

With the optimal MNLR model, the values of predicted activities **$pIC_{50\ MNLR}$** calculated from equation2 and the observed values are given in Table5. The correlations of predicted and observed activities are illustrated in **Figure5**.
The correlation between MNLR calculated and experimental activities are very significant as illustrated in Figure5 and as indicated by R and R^2 values.

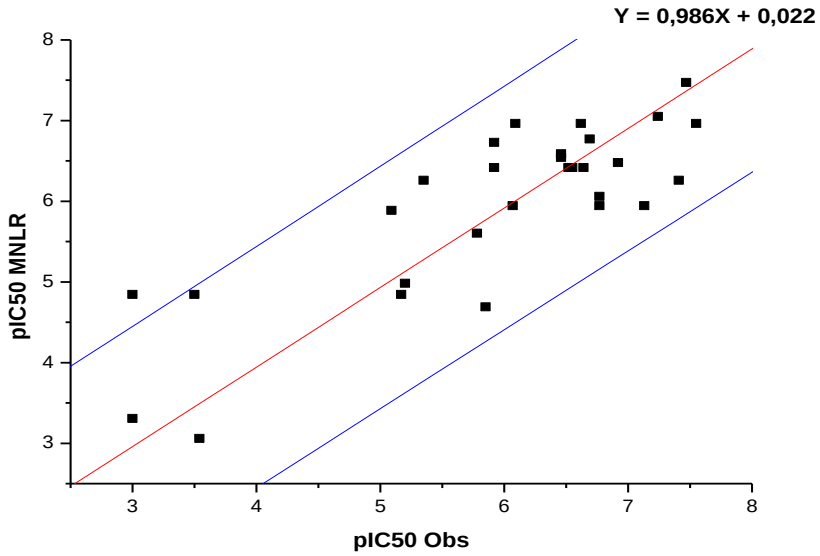

Figure 5: Correlations of observed and predicted activities calculated using MNLR

3.6. Artificial Neural Networks (ANN)
In order to increase the probability of good characterization of studied compounds, artificial 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 activities model were developed using the properties of several studied compounds. Some authors **[37,38]** 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 ANN)

In order to avoid over fitting or under fitting, it is recommended that **1.8 < ρ < 2.3 [39]**. The output layer represents the calculated activity values pIC_{50}. The architecture of the ANN used in this work (**5-2-1**), **ρ =1.93**.
The values of predicted activities **$pIC_{50\ ANN}$** calculated using ANN and the observed values are given in Table5. The correlations of predicted and observed activities are illustrated in **Figure6**.
The correlation between ANN calculated and experimental activities are very significant as illustrated in Figure6 and as indicated by R and R^2 values.
The obtained squared correlation coefficient (R^2) value confirms that the artificial neural network result were the best to build the quantitative structure activity relationship models.
It is important to be able to use ANN to predict the activity of new compounds. To evaluate the predictive ability of the ANN models, '**Leave-one-out**' is an approach particularly well adapted to the estimation of that ability.

3.7. Cross Validation (CV)
To test the performance of the neural network and the validity of our choice of descriptors selected by MLR and trained by MNLR and ANN, we used cross-validation method (CV) with the procedure leave-one-out (LOO). In this procedure, one compound is removed from the data set, the network is trained with the remaining compounds and used to predict the discarded compound. The process is repeated in turn for each compound in the data set.

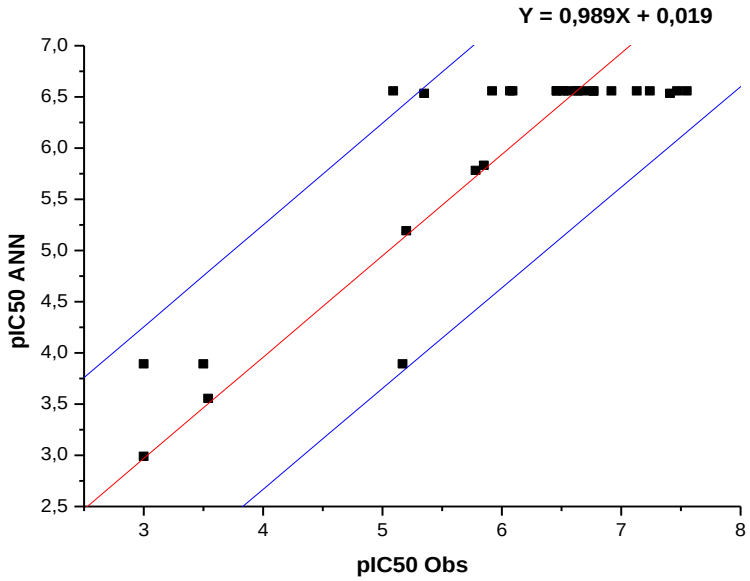

Figure6: Correlations of observed and predicted activities calculated using ANN

N = 29 R = 0.88 R^2 = 0.77

In this paper the 'leave-one-out' procedure was used to evaluate the predictive ability of the ANN.
The values of predicted activities **$pIC_{50\ CV}$** calculated using CV and the observed values are given in Table5. The correlations of predicted and observed activities are illustrated in **Figure7**.
The correlation between CV calculated and experimental activities are very significant as illustrated in Figure7 and as indicated by R and R^2 values.

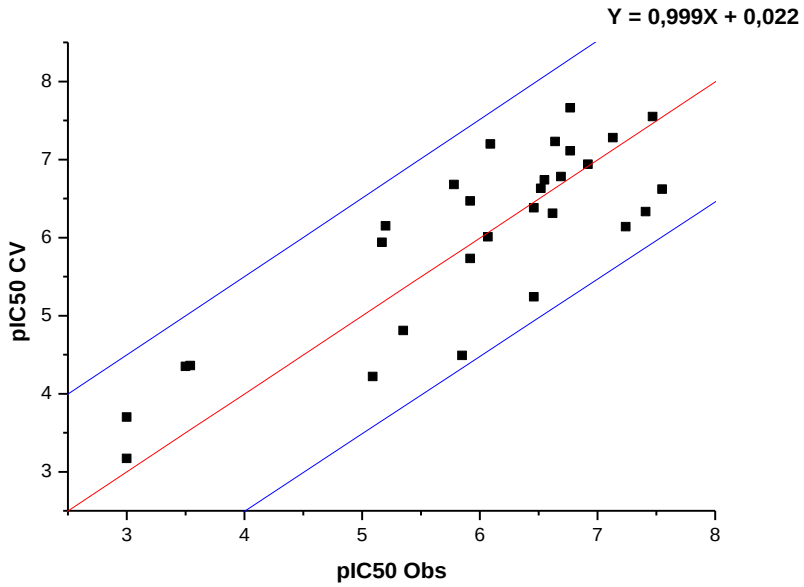

Figure7: Correlations of observed and predicted activities calculated using CV

N = 29 R = 0.833 R^2 = 0.693

The good results obtained with the cross validation, shows that the model proposed in this paper is able to predict activity with a great performance, and that the selected descriptors are pertinent.
The results obtained by MLR and MNLR are very sufficient to conclude the performance of the model. Even if it is possible that this good prediction is found by chance we can claim that it is a positive result. So, this model could be applied to all derivatives of flavonoid (phenyl-benzopyrane) accordingly to Table1 and could add further knowledge in the improvement of the search in the domain of inhibitors of aldose reductase (ALR2) enzyme for diabetic complications.
A comparison of the quality of 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. A good correlation was obtained with cross validation $\mathbf{R_{CV} = 0.833}$. So the predictive power of this model is very significant. The results obtained in this study, showed that models MLR, MNLR and ANN are validated, which means that the prediction of the new compounds is feasible.

Conclusions

In this study, three different modelling methods, MLR, MNLR and ANN were used in the construction of a QSAR model for the inhibitors of aldose reductase (ALR2) enzyme and the resulting models were compared. It was shown the artificial neural network ANN results have substantially better predictive capability than the MLR and MNLR, yields a regression model with improved predictive power, we have established a relationship between several descriptors and the ALR2 inhibitory activity in satisfactory manners. The good results obtained with the cross validation CV, shows that the model proposed in this paper is able to predict activity with a great performance, and that the selected descriptors are pertinent.
The accuracy and predictability of the proposed models were illustrated by the 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 Table5. It was also shown that the proposed methods are a useful aid for reduction of the time and cost of synthesis and activity determination of inhibitors of (ALR2) (compounds based on phenyl-benzopyrane).
Furthermore, we can conclude that studied descriptors, which are sufficiently rich in chemical and topological information to encode the structural feature and have a great influence on the activity may be used with other descriptors for the development of predictive QSAR models.
Previous studies QSAR already performed on the same set of flavonoid (phenyl-benzopyrane) using partial least squares (PLS), obtained a correlation coefficient (**R = 0.87**) **[40]**. In this study the correlation coefficient obtained from the MLR ($\mathbf{R_{MLR} = 0.80}$), by using a variety of descriptors, is very important and this coefficient improved by using MNLR and ANN respectively ($\mathbf{R_{MNLR} = 0.83}$) and ($\mathbf{R_{ANN} = 0.88}$) so the proposed model is very significant and its performance is tested by cross-validation method CV ($\mathbf{R_{CV} = 0.833}$).
Thus, grace to QSAR studies, especially with the ANN that has allowed us to improve the correlation between the observed biological activity and that predicted, we can enjoy the performance of the predictive power of this model to explore and propose new molecules could be active.

Acknowledgments-We gratefully acknowledge to the group of Applied Chemistry Laboratory.

References

1. Data Accessed From Official Diabetes Webpage of WHO *http://www.who.int/mediacentre/factsheets/fs312/en/on* October 19, (2014).
2. The Diabetes Control and Complications Trial Research Group, The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus, *N. Engl.J. Med.* 329(1993) 977-986.
3. UK Prospective Diabetes Study (UKPDS) Group, Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33), *Lancet* 352 (1998)837-853.
4. Kador P.F., Robison W.G., Kinoshita J.H., The pharmacology of aldose reductase inhibitors, *Annu. Rev. Pharmacool. Toxicol.* 25 (1985) 691-714.
5. Szwergold B.S., Kappler F., Brown T.R., Identification of fructose 3-phosphatein the lens of diabetic rats, *Science* 247 (1990) 451-454.

6. Gonzalez A.M., Sochor M., Hothersall J.S., McLean P., Effect of aldose reductase inhibitor (sorbinil) on integration of polyol pathway, pentose phosphate path-way, and glycolytic route in diabetic rat lens, *Diabetes* 35 (1986) 1200-1205.
7. WellsKnecht K.J., Brinkmann E., Wells-Knecht M.C., Litchfield J.E., Ahmed M.U., Reddy S., Zyzak D.V., Thorpe S.R., Baynes J.W., New biomarkers of Maillardreaction damage to proteins, *Nephrol. Dial. Transpl*. 11 (1996) 41-47.
8. Robinson Jr.W.G., Kador P.F., Kinoshita J.H., Retinal capillaries: basement mem-brane thickening by galactosemia prevented with aldose reductase inhibitor,*Science* 221 (1983) 1177-1179.
9. Engerman R.L., Pathogenesis of diabetic retinopathy, *Diabetes* 38 (1989)203-1206.
10. Young R.J., Ewing D.J., Clarke B.F., A controlled trial of sorbinil, an aldose reductase inhibitor, in chronic painful diabetic neuropathy, *Diabetes* 32 (1983)938-942.
11. Dunlop M, Aldose reductase and the role of the polyol pathway in diabetic nephropathy, *Kidney Int*. 58 (2000) S3-S12.
12. Varma S.D., Kinoshita J.H., Inhibition of lens aldose reductase by flavonoids - their possible role in the prevention of diabetic cataracts, *Biochem. Pharmacol*.25 (1976) 2505-2513.
13. Seyoum A., Asres K., El-Fiky F.K., Structure–radical scavenging activity relation-ships of flavonoids, *Phytochemistry* 67 (2006) 2058-2570.
14. Havsteen B.H., The biochemistry and medical significance of the flavonoids,*Pharmacol. Ther*. 96 (2002) 67-202.
15. Moon Y.J., Wang X., Morris M.E., Dietary flavonoids: effects on xenobiotic and carcinogen metabolism, *Toxicol. In Vitro* 20 (2006) 187-210.
16. Xie W., Wang W., Su H., Xing D., Pan Y., Du L., Effect of ethanolic extracts of Ananascomosus L. leaves on insulin sensitivity in rats and HepG2, Biochem.Physiol. C: *Toxicol. Pharmacol*. 143 (2006) 429-435.
17. Mercader A.G., Duchowicz P.R., Fernandez F.M., Castro E.A., Bennardi D.O., Autino J.C., Romanelli G.P., QSAR prediction of inhibition of aldose reductase for flavonoids, *Bioorg. Med. Chem.* 16 (2008) 7470-7476.
18. Sheng R., Lin X., Zhang J., Chol K.S., Huang W., Yang B., He Q., Hu Y., Design,synthesis and evaluation of flavonoid derivatives as potent AChEinhibitors, *Bioorg. Med. Chem.* 17 (2009) 6692-6698.
19. Kim M.K., Chong Y., Towards selective inhibitors of Janus kinase 3: identificationof a novel structural variation between Janus kinases 2 and 3, *Bull. Korean Chem. Soc*. 33 (2012) 4207-4210.
20. Shichijo M., Yamamoto N., Tsujishita H., Kimata M., Nagai H., Kokubo T., Inhibi-tion of syk activity and degranulation of human mast cells by flavonoids, *Biol.Pharm. Bull.* 12 (2003) 1685-1690.
21. Cotelle N., Role of flavonoids in oxidative stress, Curr. Top.*Med. Chem.* 1 (2001) 569-590.
22. Nantasenamat C., Isarankura-Na-Ayudhya C., Naenna T. &Prachayasittikul V., A practical overview of quantitative structure-activity relationship, *J. Excli.* 8 (2009) 74-88.
23. Nantasenamat C., Isarankura-Na-Ayudhya C. &Prachayasittikul V., Advances in computational methods to predict the biological activity of compounds, *J. Expert Opin. Drug Discov*. 5(7) (2010) 633-654.
24. Advanced Chemistry Development *Inc., Toronto, Canada*. (2009). *http://www.acdlabs.com/resources/freeware/chemsketch.*
25. ACD/ChemSketch Version 4.5 for Microsoft Windows User's Guide.
26. ACD/Labs Extension for ChemBioOffice Version 14.0 for *Microsoft Windows User's Guide*.
27. A N. L. Conformational Analysis 130. MM2. A Hydrocarbon Force Field Utilizing V1 and V2 Torsional Terms, *J. Am. Chem. Soc. Vol*. 99, (1977), pp.8127-8134
28. XLSTAT 2015 Add-in software (XLSTAT Company). *www.xlstat.com*.
29. Boukarai Y., Khalil F., Bouachrine M., QSAR study of isatin (1H-indole-2-3-dione) analogues as in vitro anti-cancer agents using the statistical analysis methods, *International Journal of Scientific & Engineering ResearchVolume 6, Issue 11,* (2015) 159-166.
30. Chtita S., Hmamouchi R., Larif M., Ghamali M., Bouachrine M., Lakhlifi T., QSPR studies of 9-anilioacridine derivatives for their DNA drug binding properties based on density functional theory using statistical methods: Model, validation and influencing factors, *J. of Taibah Univ. for Sci.* (2015).
31. Boukarai Y., Khalil F., Bouachrine M., QSAR study of 5,6-bicyclic heterocycles analogues as anti-Alzheimer's agents using the statistical analysis methods, *Journal of Chemical and Pharmaceutical Research*, 8(3) (2016), 1000-1013.
32. Zupan V.J. & Gasteiger J., Neural Networks for Chemists - An Introduction, VCH Verlagsgesellschaft, *Weinheim/VCH Publishers, New York*. 106(12) (1993) 1367-1368.

33. Efron B., Estimating the error rates of a predictive rule: improvement on cross-validation, *J.Am.Stat. Assoc.* 78 (1983) 316-331.
34. Efroymson M.A., Multiple regression analysis, In Mathematical Methods for Digital Computers, Ralston, A., *Wilf, H.S., Eds,WileyNewYork*, (1960).
35. OstenD.W., Selection of optimal regression models via cross-validation, *J. Chemom.* 2 (1998) 39-48.
36. Boukarai Y., Khalil F., Bouachrine M., DFT-based QSAR Studies of cytotoxicity of phenothiazine derivatives as in vitro anti-cancer agents using the statistical analysis methods, *Der PharmaChemica*, 2016, 8(18):171-185.
37. So S-S. &RichardsW.G., Application of neural networks: quantitative structure-activity relationships of the derivatives of 2, 4 diamino (substituted-benzyl) pyrimidines as DHFR inhibitors, *J. Med. Chem.* 35 (1992) 3201-3207.
38. Andrea T.A. &Kalayeh H., Applications of neural networks in quantitative structure-activity relationships of dihydrofolatereductase inhibitors, *J. Med. Chem*. 34 (1991) 2824-2836.
39. Elhallaoui M., Modélisatrice moléculaire et étude QSAR d'antagonistes non compétitifs du récepteur NMDA par les méthodes statistiques et le réseau de neurones, *Thesis de Doctorat, Fes, Morocco* (2002) 106.
40. Vyas B., Singh M., Kaur M., Bahia M.S., Jaggi A.S., Silakari O., Singh B., Aldose reductase inhibitors for diabetic complications: Receptor induced atom-based 3D-QSAR analysis, synthesis and biological evaluation, *Journal of Molecular Graphics and Modelling* 59 (2015) 59-71.

(2017) ; http://www.jmaterenvironsci.com