Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2016, 8(3):1000-1013

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

QSAR study of 5,6-bicyclic heterocycles analogues as anti-Alzheimer's agents using the statistical analysis methods

Youness Boukarai[1*], Fouad Khalil[1] and Mohamed Bouachrine[2]

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

ABSTRACT

Alzheimer's disease (AD) is a chronic neurodegenerative disease. Current therapies of AD are only symptomatic, therefore the need for the development of new therapies to treat Alzheimer's disease effectively. The 5,6-bicyclic heterocycles and its derivatives are potent anti-Alzheimer agents, these compounds inhibit β-amyloid ($A\beta_{42}$). A study of quantitative structure-activity relationship (QSAR) is applied to a set of 34 molecules derived from 5,6-bicyclic heterocycles, in order to predict the anti-Alzheimer biological activity of the test compounds and find a correlation between the different physic-chemical parameters (descriptors) of these compounds and its biological activity, usingprincipal 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 ChemBioDraw Ultra 14.0 programs. A good correlation was found between the experimental activity and those obtained by MLR and MNLR respectively such as (R = 0,843 and R^2 = 0,712) and (R = 0,870 and R^2 = 0,758), this result could be improved with ANN such as (R = 0,924 and R^2 = 0,853) with an architecture ANN (7-2-1). 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 = 0,874 and R^2 = 0,763) 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 7-2-1 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.

Keywords: Alzheimer's disease, 5,6-bicyclic heterocycles derivatives, QSAR, PCA, MLR, MNLR ANN, CV.

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;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; MSE, Mean Squared Error; F, Fishers F-statistic; F value, Significance level; p-value, Critical Probability*

INTRODUCTION

AD is a neurodegenerative disorder associated with difficulties in memory, judgment, abstraction, and language **[1]**. More than 35 million people suffer from AD worldwide and the AD population may increase to more than 115 million by the year 2050 according to a report from Alzheimers Disease International**[2]**. Alzheimer's disease (AD)

symptoms are dementia, apraxia, aphasia, depression, short attention span, visuospatial navigation deficits, anxiety and delusions **[3]**. The majority of AD cases are sporadic, with disease onset after 65 years of age**[4]**. Key molecules involved in AD, include the presenilins, amyloid precursor protein (APP), tau, and β-amyloid **[5]**.

Amyloid protein of 40-42 amino acids derived from a double section of a normal protein in the body, APP 650-770 amino acids. The metabolism of APP is under the simultaneous action of three secretases (alpha, beta, and gamma) accordingtwo ways:

- A non-amyloidogenic pathway, under the action of alpha secretase bisecting the sequence of A-beta and a deadly soluble fragment of APP and then under that of the gamma secretase which cleaves the C-terminal fragment.
- The amyloidogenic pathway is under the action of beta-secretase which cleaves APP at the beginning of the sequence of the A-beta, and under that of the gamma secretase, which releases the free fragment, is the amyloid peptide A-beta 42 and A-beta 40.

To inhibit the amyloidogenic pathway, the objective would be to stimulate alpha secretase or inhibit beta or gamma secretase**[6-13]**.

Alzheimer's disease (AD) is the biggest unmet medical need in neurology due to the lack of disease-modifying anti-Alzheimer's drugs (DMAADs)**[14]**. Over the last decade, γ-secretase emerged as a promising target for the treatment of AD**[15]**. It has been postulated that modulation of γ-cleavage to favor the production of shorter fragments, while not affecting totalAβ levels, might be a safe approach to a disease-modifying therapy**[16]**.

γ-Secretase activity can be controlled by the inhibition of the active site of PS1 or by interference with complex assembly or substrate recognition, the latter resulting in allosteric modulation or inhibition. The allosteric mechanisms are particularly attractive targets for drug development **[17]**, as they may produce shorter, soluble, and non-toxic peptides (e.g., Aβ36–Aβ40) instead of the highly insoluble and neurotoxic Aβ42**[18]**.

γ-Secretase modulators (GSMs) modulate the cleavage of the APP C-terminal fragment such as C-99 to decrease Aβ42 and increase the shorter Aβ fragments (e.g., Aβ37/38) **[19]**.

γ-Secretase modulation is more desirable than inhibition from a therapeutic perspective and may reduce the risk of mechanism-based toxicities **[20]**. Such compounds, γ-secretase modulators (GSMs), would be good candidates for AD therapeutics **[21]**.

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]**.

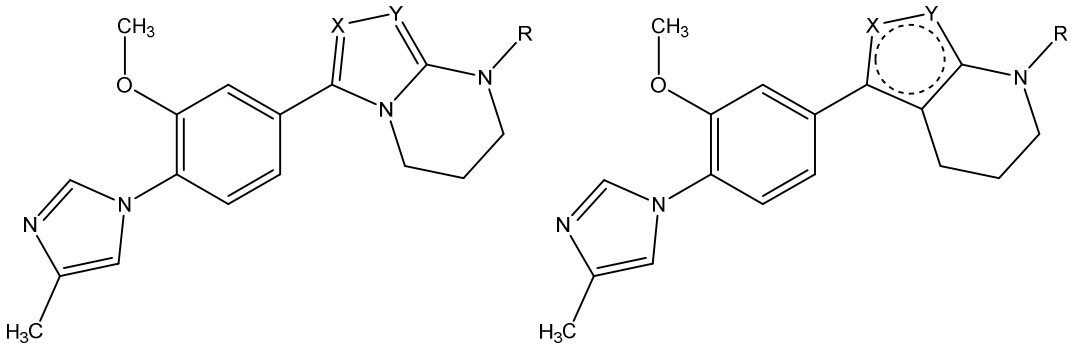
In this work we attempt to establish a quantitative structure-activity relationship between anti-Alzheimer activity of a series of 34 bioactive molecules derived from 5,6-bicyclic heterocycles and structural descriptors. Thus we can predict the anti-Alzheimer activity of this group of organic compounds. 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 to give an estimation of the values of the pertinent descriptors that govern this classification. * The Multiple Linear Regression (MLR) has served to select the descriptors used as the input parameters for the 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).

EXPERIMENTAL SECTION

The Biological data used in this study were anti-Alzheimer activity against $A\beta_{42}$ (inhibition of β-amyloid. (IC_{50})), a set of thirty-four derivatives of5,6-bicyclic heterocycles. We have studied and analyzed the series of 5,6-bicyclic heterocycles molecule consists of 34 selected derivatives that have been synthesized and evaluated for their anti-Alzheimer activity in vitro against $A\beta_{42}$(in terms of -log (IC_{50})) **[24]**. This in order to determine a quantitative structure-activity relationship between the anti-Alzheimer activity and the structure of these molecules that are described by their substituents R, X and Y.

The chemical structure of 5,6-bicyclic heterocycles is represented in **Figure1**.

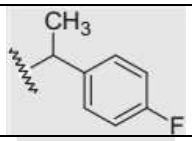
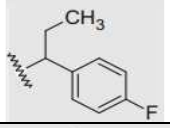
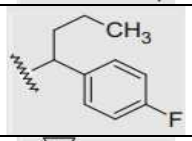
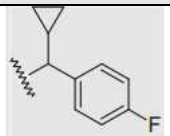
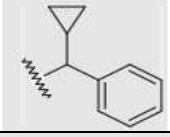
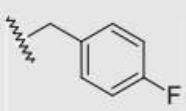
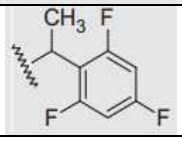


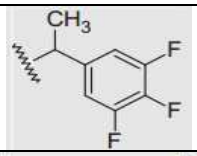
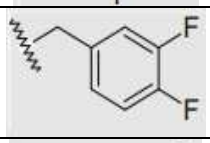
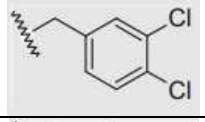
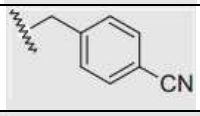
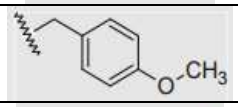
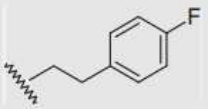
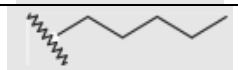
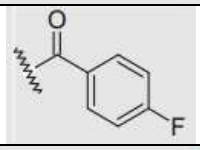
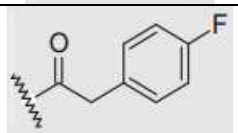
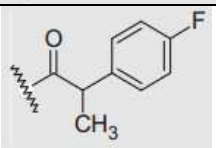
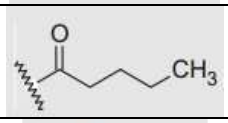
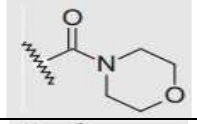
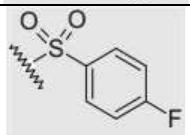
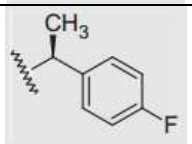
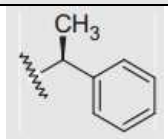
Molecules 1-21 **Molecules22-34**

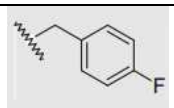
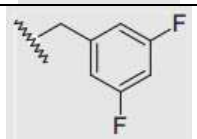
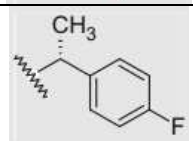
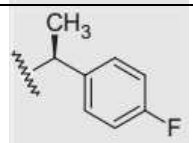
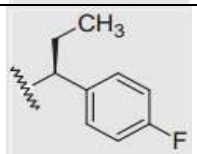
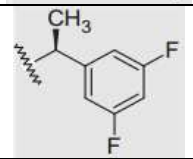
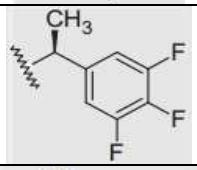
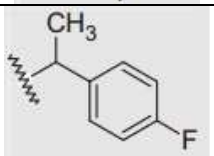
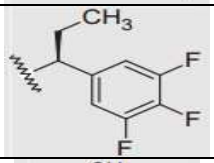
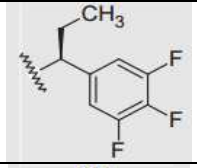
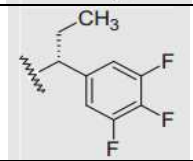
Figure1: The general structure of 5,6-bicyclic heterocycles

The chemical structures of 34 compounds of 5,6-bicyclic heterocycle sused in this study and their experimental anti-Alzheimer biological activity observed IC_{50}(Cytotoxic concentration required to inhibit theβ-amyloid$A\beta_{42}$ than 50%) are collected from recent publications**[24]**.The observations are converted into logarithmic scale-log (IC_{50})in molar units (M) and are included in **Table1**.

Table1: Chemical structure and activity observed of 5,6-bicyclic heterocycles derivatives against $A\beta_{42}$

N°	R	X	Y	Experimental pIC_{50} [a] Obs
1	H	N	N	4,70
2		N	N	6,29
3		N	N	6,81
4		N	N	6,84
5		N	N	4,79
6		N	N	4,70
7		N	N	6,19
8		N	N	6,47

9		N	N	6,94
10		N	N	6,26
11		N	N	6,90
12		N	N	5,60
13		N	N	5,85
14		N	N	6,21
15		N	N	5,85
16		N	N	4,71
17		N	N	5,72
18		N	N	6,31
19		N	N	5,68
20		N	N	4,70
21		N	N	5,65
22		O	N	6,33
23		O	N	6,27

24		O	N	6,45
25		O	N	6,36
26		N	O	5,86
27		N	O	6,35
28		N	O	6,06
29		N	O	6,03
30		N	O	6,27
31		N	NH	6,79
32		N	NH	6,70
33		N	NH	6,91
34		N	NH	6,68

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

Calculation of molecular descriptors

Advanced chemistry development's ACD/Chem Sketch 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)) **[25,26]**.

Steric, thermodynamic descriptors are calculated using ACD/Chem Sketch and Chem Bio Draw Ultra 14.0**[27]**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}) **[28]**.

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).

Statistical analysis

To explain the structure-activity relationship, these 11 descriptors are calculated for 34 molecules (**Table2**) through software Chem Sketch and Chem Bio Draw Ultra 14.0.

Table2: The values of the 11 chemical descriptors

	MW	MR	MV	Pc	n	γ	D	α_e	LogP	HBA	HBD
1	310,353	87,230	221,2	604,3	1,718	55,6	1,40	34,58	1,496	7	1
2	432,493	122,14	329,3	865,4	1,663	47,6	1,31	48,42	4,493	8	0
3	446,519	126,75	345,3	904,0	1,655	46,9	1,29	50,24	4,980	8	0
4	460,546	131,36	361,4	942,6	1,647	46,2	1,27	52,07	5,397	8	0
5	458,530	128,99	333,7	895,4	1,699	51,8	1,37	51,13	4,880	8	0
6	440,540	129,12	330,9	895,2	1,708	53,5	1,33	51,19	4,722	7	0
7	418,466	117,72	314,1	834,3	1,672	49,7	1,33	46,66	4,175	8	0
8	468,474	121,88	335,0	865,8	1,647	44,5	1,39	48,31	4,810	10	0
9	468,474	121,88	335,0	865,8	1,647	44,5	1,39	48,31	4,810	10	0
10	436,457	117,59	317,0	834,5	1,664	48,0	1,37	46,61	4,333	9	0
11	469,366	127,05	329,8	891,8	1,696	53,4	1,42	50,36	5,133	7	0
12	425,485	124,37	324,0	879,9	1,693	54,3	1,31	49,3	4,050	8	0
13	430,502	123,66	332,9	884,4	1,665	49,8	1,29	49,02	3,891	8	0
14	432,493	122,33	330,2	872,9	1,662	48,8	1,30	48,49	4,455	8	0
15	380,486	110,99	306,8	804,2	1,643	47,1	1,24	44,00	3,943	7	0
16	432,450	118,57	311,3	840,0	1,686	52,9	1,38	47,00	3,199	8	0
17	446,476	123,18	327,4	878,6	1,676	51,8	1,36	48,83	3,143	8	0
18	460,503	127,60	342,6	909,7	1,667	49,7	1,34	50,58	3,710	8	0
19	394,470	111,84	304,0	809,9	1,656	50,3	1,29	44,34	2,632	7	0
20	423,468	114,82	293,6	815,9	1,710	59,6	1,44	45,52	0,887	8	0
21	468,503	122,86	318,2	876,8	1,698	57,6	1,47	48,70	0,000	9	0
22	432,490	121,13	336,4	869,3	1,639	44,6	1,28	48,02	5,320	7	0
23	414,499	121,26	333,5	869,1	1,647	46,1	1,24	48,07	5,162	6	0
24	418,463	116,70	321,2	838,2	1,646	46,3	1,30	46,26	5,002	7	0
25	436,453	116,57	324,0	838,4	1,638	44,7	1,34	46,21	5,160	8	0
26	432,490	121,13	336,4	869,3	1,639	44,6	1,28	48,02	5,161	7	0
27	432,490	121,13	336,4	869,3	1,639	44,6	1,28	48,02	5,161	7	0
28	446,516	125,74	352,4	907,9	1,632	44,0	1,26	49,84	5,647	7	0
29	450,480	121,00	339,2	869,5	1,631	43,1	1,32	47,96	5,319	8	0
30	468,470	120,87	342,1	869,7	1,624	41,7	1,36	47,91	5,477	9	0
31	431,505	122,79	334,6	869,9	1,655	45,6	1,28	48,68	5,088	7	1
32	481,512	127,14	356,3	908,9	1,632	42,3	1,35	50,40	5,891	9	1
33	481,512	127,14	356,3	908,9	1,632	42,3	1,35	50,40	5,891	9	1
34	481,512	127,14	356,3	908,9	1,632	42,3	1,35	50,40	5,891	9	1

The study we conducted consists of:

-The principal component analysis (PCA), the multiple linear regressions (MLR), and the non-linear regression (MNLR) available in the XLSTATsoftware**[29]**.

-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 5,6-bicyclic heterocycles derivatives were studied by statistical methods based on the principal component analysis (PCA). PCA is a statistical technique useful for summarizing all the information encoded in the structures of the compounds. It is also very helpful for understanding the distribution of the compounds.

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**.

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

Variables	**MW**	**MR**	**MV**	**Pc**	**n**	**γ**	**D**	**α_e**	**LogP**	**HBA**	**HBD**	**pIC50**
MW	**1**											
MR	0,866	**1**										
MV	0,851	0,916	**1**									
Pc	0,865	0,986	0,955	**1**								
n	-0,344	-0,250	-0,615	-0,373	**1**							
γ	-0,368	-0,276	-0,620	-0,361	**0,952**	**1**						
D	0,228	-0,126	-0,316	-0,202	0,536	0,499	**1**					
α_e	0,866	**1**	**0,916**	**0,986**	-0,250	-0,276	-0,127	**1**				
LogP	0,443	0,512	0,709	0,533	-0,700	-0,833	-0,520	0,512	**1**			
HBA	0,583	0,239	0,258	0,231	-0,140	-0,176	0,561	0,239	0,006	**1**		
HBD	-0,017	-0,161	-0,061	-0,179	-0,107	-0,225	0,107	-0,161	0,139	0,130	**1**	
pIC50	0,488	0,401	0,617	0,469	-0,702	-0,707	-0,267	0,401	0,623	0,252	0,180	**1**

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. It has served also to select the 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 5,6-bicyclic heterocycles derivatives. Equations were justified by the correlation coefficient (R), the Mean Squared Error (MSE), the Fishers F-statistic (F), and the significance level(F value) **[30,31]**.

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.

RESULTS AND DISCUSSION

Data set for analysis

The QSAR analysis was performed using the -log (IC_{50}) of the 34 selected molecules that have been synthesized and evaluated for their anti-Alzheimer activity in vitro against $A\beta_{42}$(experimental values) **[24]**. 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 anti-Alzheimer activity.The values of the 11 chemical descriptors as shown in Table2.

The principle is to perform in the first time, a main component analysis (PCA), which allows us to eliminate descriptors that are highly correlated(dependent), then perform a decreasing study of MLR based on the elimination of descriptors aberrant until a valid model (including the critical probability:**p-value < 0.05** for all descriptors and the model complete).

Principal Components Analysis (PCA)

The totality of the 11 descriptors (variables) coding the 34 molecules was submitted to a principal components analysis (PCA). 12 principal components were obtained (**Figure2**).

The first three axes F1, F2 and F3 contributing respectively 52.2 %, 20.9 % and 13.7 % to the total variance, the total information is estimated to a percentage of 86.8%.

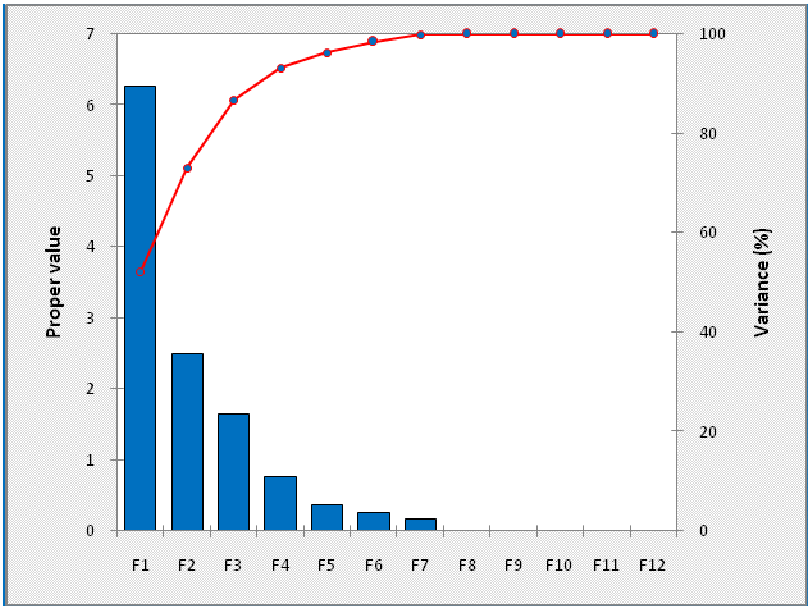


Figure2: The principal components and there variances

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.

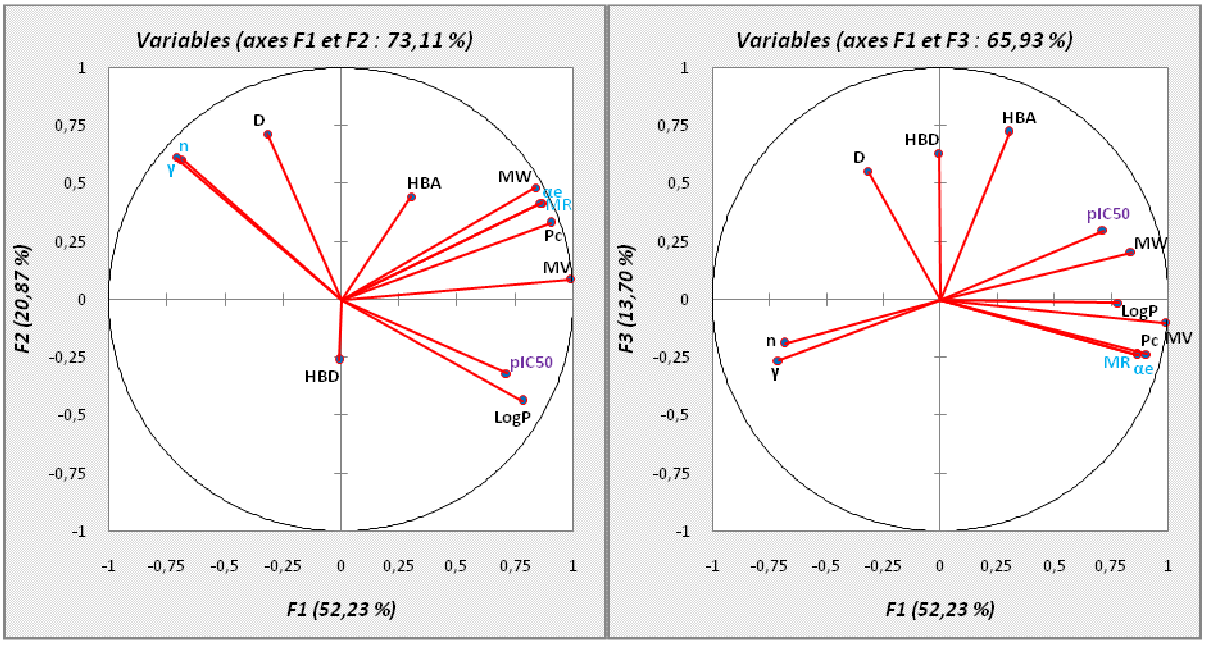


Figure3: Correlation circles

(MR, α_e) are perfectly correlated (r=1).
α_e, MV and Pc are highly correlated(r (α_e, MV) = 0,916, r (α_e, Pc) = 0,986).
(γ) and (n)are highly correlated(r (γ, n) = 0,952).
The following variables then removed are: (α_e) and (n).

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 34 molecules,we presented the data matrix which is the corresponding physicochemical variables different substituent's from 34 molecules to a multiple linear regression analysis. This method used the coefficients R, R^2, MSE and the F-values to select the best

regression performance. Where R is the correlation coefficient; R^2 is the coefficient of determination; MSE is the mean squared error; F is the Fisher F-statistic.

Treatment with multiple linear regression is more accurate because it allows you to connect the structural descriptors for each activity of 34 molecules to quantitatively evaluate the effect of substituent.The selected descriptors are:

MW, **MR**, **MV, Pc, γ, D** and**HBD**.
The QSAR model built using multiple linear regression (MLR) method is represented by the following equation:
$\mathbf{pIC_{50MLR}}$= - 34,488- 0,199**MW**- 0,208**MR** - 0,360**MV**+ 0,268**Pc**-1,151γ+ 71,430**D** + 0,679**HBD**

N = 34R = 0.843R^2 = 0.712 F = 9.182 MSE = 0.170 (**Equation 1**)

Higher correlation coefficient and lower mean squared error (MSE) indicate that the model is more reliable. And the Fisher's F test is used. 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 with confidence that the model do bring a significant amount of information.

The elaborated QSAR model reveals that the anti-Alzheimer activity could be explained by a number of topologic factors. The negative correlation of the Molecular Weight (MW), the Molar Refractivity (MR),the Molar Volume (MV)and the Surface Tension (γ) with the ability to displace the 5,6-bicyclic heterocycles activity reveals that a decrease in the value of pIC_{50}, While the positive correlation of the descriptors (Parachor (Pc), Density (D) and Hydrogen Bond Donor (HBD)) with the ability to displace the 5,6-bicyclic heterocycles 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 **Table4**. 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.

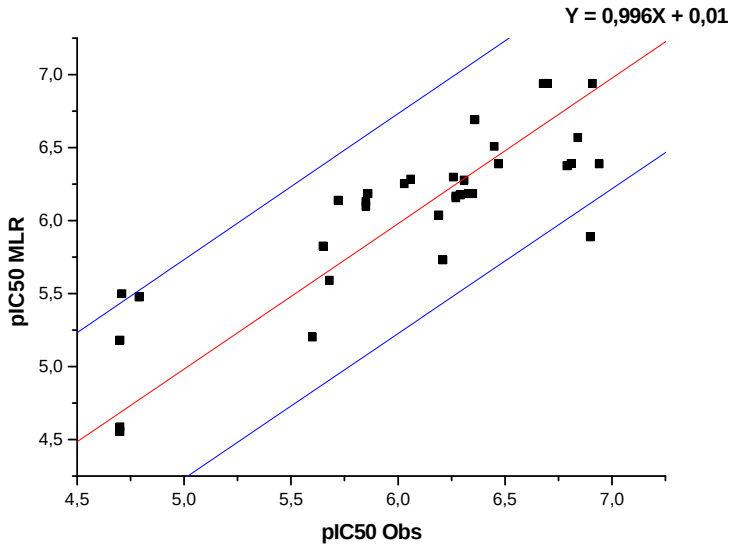


Figure4: Correlations of observed and predicted activities calculated using MLR

Table4: The observed, the predicted activities (pIC_{50}), according to different methods MLR, MNLR, ANN and CV for the 34 derivatives of 5,6-bicyclic heterocycles

N°	pIC_{50Obs}	pIC_{50MLR}	pIC_{50MNLR}	$pIC_{50\ ANN}$	pIC_{50CV}
1	4,700	4,587	4,716	4,688	4,689
2	6,290	6,177	6,236	6,284	6,331
3	6,810	6,390	6,384	6,284	6,700
4	6,840	6,567	6,677	6,749	6,740
5	4,790	5,476	5,279	4,856	5,686
6	4,700	5,179	4,944	4,690	5,239
7	6,190	6,038	6,040	6,108	6,243
8	6,470	6,389	6,592	6,284	6,379
9	6,940	6,389	6,592	6,284	6,285
10	6,260	6,298	6,206	6,272	5,995
11	6,900	5,888	5,951	6,283	6,590
12	5,600	5,202	5,375	5,585	4,720
13	5,850	6,099	6,105	6,284	6,093
14	6,210	5,731	5,910	6,284	5,782
15	5,850	6,130	6,107	6,284	6,170
16	4,710	5,499	5,617	4,689	4,900
17	5,720	6,137	6,081	6,284	5,650
18	6,310	6,276	6,237	6,290	6,280
19	5,680	5,591	5,337	5,710	5,210
20	4,700	4,556	4,331	4,688	4,260
21	5,650	5,825	6,044	5,647	5,770
22	6,330	6,185	6,134	6,284	6,108
23	6,270	6,155	6,346	6,284	6,144
24	6,450	6,508	6,316	6,284	6,113
25	6,360	6,690	6,466	6,284	6,310
26	5,860	6,185	6,134	6,284	6,090
27	6,350	6,185	6,134	6,284	6,150
28	6,060	6,283	6,461	6,302	5,860
29	6,030	6,252	6,223	6,284	6,140
30	6,270	6,167	6,190	6,248	6,210
31	6,790	6,375	6,560	6,759	6,570
32	6,700	6,939	6,834	6,795	6,674
33	6,910	6,939	6,834	6,795	6,567
34	6,680	6,939	6,834	6,795	6,363

Multiples Non-Linear Regression (MNLR)

We have used also the technique of nonlinear regression model to improve thestructure-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 34 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:

$\mathbf{pIC_{50MNLR}}$= 110,979– 0,202**MW** +2,923 **MR** - 1,533**MV**+0,285**Pc**-3,053 γ- 20,206 **D**+ 0,561 **HBD** +1,076E-04(**MW**)^2- 1,252E-02(**MR**)^2+7,759E-04 (**MV**)^2+1,027E-02(γ)^2+ 22,360 (**D**)^2

$$\mathbf{N = 34 R = 0.870 R^2 = 0.758 MSE = 0.195} \qquad \textbf{(Equation 2)}$$

With the optimal MNLR model, the values of predicted activities $\mathbf{pIC_{50\ MNLR}}$ calculated from equation2 and the observed values are given in Table4. 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.

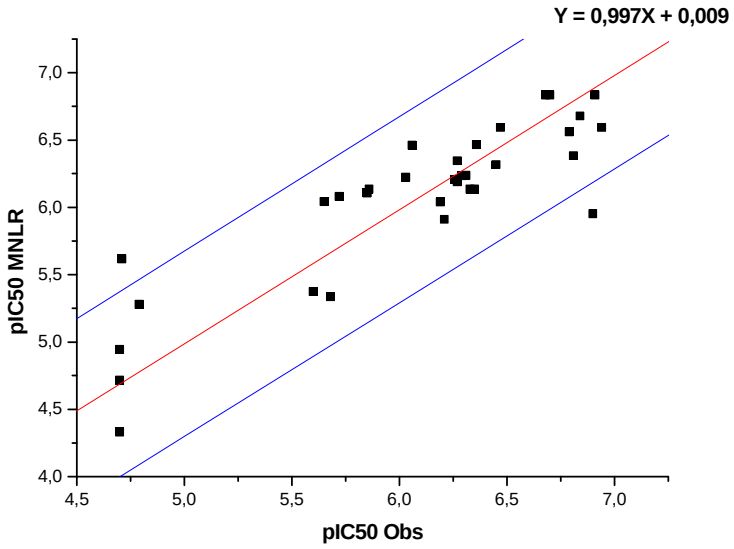


Figure5: Correlations of observed and predicted activities calculated using MNLR

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 **[36,37]** 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[38]**.The output layer represents the calculated activity values pIC_{50}. The architecture of the ANN used in this work (**7-2-1**), **ρ =1.8**.

The values of predicted activities $\mathbf{pIC_{50\ ANN}}$ calculated using ANN and the observed values are given in Table4. 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.

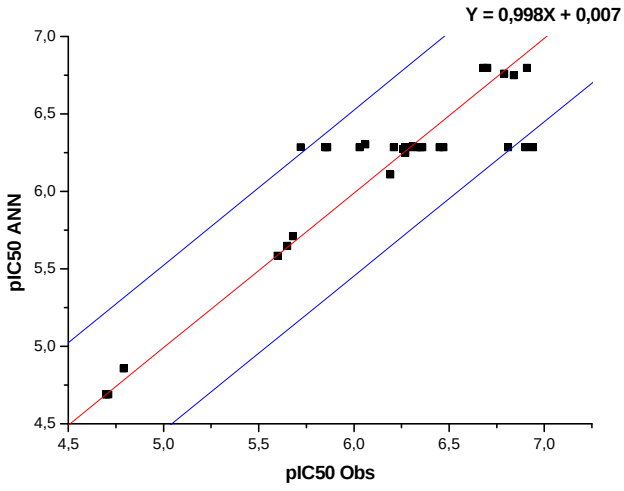


Figure6: Correlations of observed and predicted activities calculated using ANN

N = 34R = 0.924R^2 = 0.853

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.

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.

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 Table4. 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.

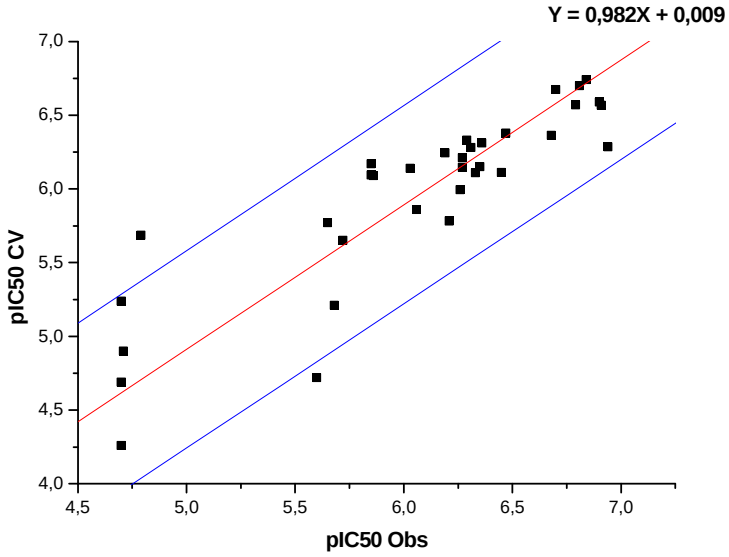


Figure7: Correlations of observed and predicted activities calculated using CV

N = 34R = 0.874R^2 = 0.763

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 5,6-bicyclic heterocycles accordingly to Table1 and could add further knowledge in the improvement of the search in the domain of anti-Alzheimer agents.

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 **R_{CV}= 0.874**. 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.

CONCLUSION

In this study, three different modelling methods, MLR, MNLR and ANN were used in the construction of a QSAR model for the anti-Alzheimer agents 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 anti-Alzheimer 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 Table4. 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 anti-Alzheimer agents (compounds based on 5,6-bicyclic heterocycles derivatives).

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 5,6-bicyclic heterocycles using cross validation, obtained a correlation coefficient (**R = 0.845**) **[39]**. In this study the correlation coefficient obtained from the MLR (**R_{MLR} = 0.843**), by using a variety of descriptors, is very important and this coefficient improved by using MNLR and ANN respectively(**R_{MNLR} = 0.870**) and (**R_{ANN} = 0.924**) so the proposed model is very significant and its performance is tested by cross-validation method CV (**R_{CV} = 0.874**).

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.

Acknowledgment
We gratefully acknowledge to the group of Applied Chemistry Laboratory.

REFERENCES

[1] S.D. Kim, J. Kim, (**2008**). *Cell Stress and Chaperones* 13, 401–412.
[2] Z.-Y. Sun, T. Asberom, T. Bara, C. Bennett, D. Burnett, I. Chu,J. Clader, M. Cohen-Williams, D. Cole, M. Czarniecki, J. Durkin,G. Gallo, W. Greenlee, H. Josien, X. Huang, L. Hyde, N. Jones,I. Kazakevich, H. Li, X. Liu, J. Lee, M. MacCoss, M.B. Mandal,T. McCracken, A. Nomeir, R. Mazzola, A. Palani, E.M. Parker, D.A. Pissarnitski, J. Qin, L. Song, G. Terracina, M. Vicarel,J. Voigt, R. Xu, L. Zhang, Q. Zhang, Z. Zhao, X. Zhu,Z. Zhu, (**2012**). J. *Med. Chem*. 55, 489–502.
[3] B.K. Harvey, C.T. Richie, B.J. Hoffer, M. Airavaara, (**2011**). *J. Neural Transm*. 118, 27–45.
[4] K. Iijima-Ando, K. Iijima, (**2010**). *Brain Struct.Funct*. 214, 245–262.
[5] Y. Balaraman, A.R. Limaye, A.I. Levey, S. Srinivasan, (**2006**). *Cell. Mol. Life Sci*. 63, 1226–1235.
[6] Y. Luo, B. Bolon, S. Kahn, et al. *Nat Neurosci* (**2001**); 4: 231-32.
[7] I. Hussain, J. Hawkins, D. Harrison, et al. *J. Neurochem* (**2007**); 100: 802-9.
[8] ID. Hills, JP.Vacca. *CurrOpin Drug DiscovDevel*(**2007**); 4: 383-91.
[9] M. Hull, M. Berger, M. Heneka. *Drugs* (**2006**); 66: 2075-93.
[10] G.Evin, MF. Sernee, Masters CL. *CNS Drugs* (**2006**); 20: 351-72.
[11] A. Petit, F. Bihel, C. Alves Da Costa, et al. *Nat Cell Biol* (**2001**); 3: 507-11.
[12] E. Siemers, M. Skinner, RA. Dean, et al. *Clinneuropharmacol* (**2005**); 28:126-32.
[13] R. Etcheberrigaray, M. Tan, I. Dewatchter, et al. *ProcNatlAcadSci USA* (**2004**); 101: 1141-46.
[14] A. Kreft, B. Harrison, S. Aschmies, K. Atchison, D. Casebier, D.C. Cole, G. Diamantidis, J. Ellingboe, D. Hauze, Y. Hu, D. Huryn, M. Jin, D. Kubrak, P. Lu, J. Lundquist, C. Mann, R. Martone, W. Moore, A. Oganesian, A. Porte, D.R. Riddell, J. Sonnenberg- Reines, J.R. Stock, S.-C. Sun, E. Wagner, K. Woller, Z. Xu, H. Zhou, J.S .Jacobsen, (**2008**). *Bioorg. Med. Chem. Lett*. 18, 4232–4236.
[15] Z. Xin, H. Peng, A. Zhang, T. Talreja, G. Kumaravel, L. Xu, E. Rohde, M.-Y. Jung, M.N. Shackett, D. Kocisko, S. Chollate, A.W. Dunah, P.A. Snodgrass-Belt, H.M. Arnold, A.G. Taveras, K.J. Rhodes, R.H. Scannevin, (**2011**). *Bioorg. Med. Chem. Lett*. 21, 7277–7280.

[16] C. Fischer, S.L. Zultanski, H. Zhou, J.L. Methot, W.C. Brown, D.M. Mampreian, A.J. Schell, S. Shah, H. Nuthall, B.L. Hughes, N. Smotrov, C.M. Kenific, J.C. Cruz, D. Walker, M. Bouthillette, G.N. Nikov, D.F. Savage, V.V. Jeliazkova-Mecheva, D. Diaz, A.A. Szewczak, N. Bays, R.E. Middleton, B. Munoz, M.S. Shearman, (**2011**). *Bioorg. Med. Chem. Lett*. 21, 4083–4087.
[17] R. Narlawar, B.I.P. Revuelta, K. Baumann, R. Schubenel, C. Haass, H. Steiner, B. Schmidt, (**2007**). *Bioorg. Med. Chem. Lett.* 17, 176–182.
[18] J.P. Caldwell, C.E. Bennett, T.M. McCracken, R.D. Mazzola, T. Bara, A. Buevich, D.A. Burnett, I. Chu, M. Cohen-Williams, H. Josein, L. Hyde, J. Lee, B. McKittrick, L. Song, G. Terracina, J. Voigt, L. Zhang, Z. Zhu, (**2010**). *Bioorg. Med. Chem. Lett*. 20, 5380–5384.
[19] C. Fischer, S.L. Zultanski, H. Zhou, J.L. Methot, W.C. Brown, D.M. Mampreian, A.J. Schell, S. Shah, H. Nuthall, B.L. Hughes, N. Smotrov, C.M. Kenific, J.C. Cruz, D. Walker, M. Bouthillette, G.N. Nikov, D.F. Savage, V.V. Jeliazkova-Mecheva, D. Diaz, A.A. Szewczak, N. Bays, R.E. Middleton, B. Munoz, M.S. Shearman, (**2011**). *Bioorg. Med. Chem. Lett*. 21, 4083–4087.
[20] Z. Xin, H. Peng, A. Zhang, T. Talreja, G. Kumaravel, L. Xu, E. Rohde, M.-Y. Jung, M.N. Shackett, D. Kocisko, S. Chollate, A.W. Dunah, P.A. Snodgrass-Belt, H.M. Arnold, A.G. Taveras, K.J. Rhodes, R.H. Scannevin, (**2011**). *Bioorg. Med. Chem. Lett*. 21, 7277–7280.
[21] M. Kurosumi, Y. Nishio, S. Osawa, H. Kobayashi, T. Iwatsubo, T. Tomita, H. Miyachi, (**2010**). *Bioorg. Med. Chem. Lett*. 20, 5282–5285.
[22] C. Nantasenamat, C. Isarankura-Na-Ayudhya, T. Naenna& V. Prachayasittikul, *J. Excli*. 8 (**2009**) 74-88.
[23] C. Nantasenamat, C. Isarankura-Na-Ayudhya& V. Prachayasittikul, *J. Expert Opin. Drug Discov*. 5(7) (**2010**) 633-654.
[24] J. Qin, , P. Dhondi, X. Huang, M. Mandal, Z. Zhao, D. Pissarnitski, W. Zhou, R. Aslanian, Z. Zhu, W. Greenlee, J. Clader, L. Zhang, M. Cohen-Williams, N. Jones, L. Hyde, A. Palani, (**2011**). *Bioorg. Med. Chem*. Lett. 21, 664–669.
[25] Advanced Chemistry Development Inc., Toronto, Canada. (**2009**).
[26] ACD/ChemSketch Version 4.5 for Microsoft Windows User's Guide.
[27] ACD/Labs Extension for ChemBioDraw Version 14.0 for Microsoft Windows User's Guide.
[28] A, N. L. Conformational Analysis 130. MM2. A Hydrocarbon Force Field Utilizing V1 and V2 Torsional Terms, *J. Am. Chem. Soc*. Vol. 99, pp.8127-8134, (**1977**).
[29] XLSTAT **2015** Add-in software (XLSTAT Company).
[30] Y. Boukarai, F. Khalil, M. Bouachrine, 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 Research* (**2015**).
[31] S. Chtita, R. Hmamouchi, M. Larif, M. Ghamali, M. Bouachrine, T. Lakhlifi, *J. of Taibah Univ. for Sci*. (**2015**).
[32] V.J. Zupan& J. Gasteiger, Neural Networks for Chemists - An Introduction, VCH Verlagsgesellschaft, Weinheim/VCH Publishers, New York. 106(12) (**1993**) 1367-1368.
[33] B. Efron, *J. Am.Stat. Assoc*. 78 (**1983**) 316-331.
[34] M.A. Efroymson, Multiple regression analysis, In Mathematical Methods for Digital Computers, Ralston, A., Wilf, H.S., Eds,WileyNewYork, (**1960**).
[35] D.W. Osten, *J. Chemom*. 2(**1998**) 39-48.
[36] S-S. So & W.G. Richards, *J. Med. Chem*. 35 (1992) 3201-3207.
[37] T.A. Andrea & H. Kalayeh, *J. Med. Chem*. 34 (**1991**) 2824–2836.
[38] M. Elhallaoui, Modélisatricemoléculaire et étude QSAR d'antagonistes non compétitifs du récepteur NMDA par les méthodesstatistiques et le réseau de neurones, (**2002**) 106.
[39] K.S. Bhadoriya, M.C. Sharma, S. Sharma, S.V. Jain, M.H. Avchar, *Arabian Journal of Chemistry* (**2014**).