QSAR Studies of the Inhibitory Activity of a Series of Substituted Indole and Derivatives against Isoprenylcysteine Carboxyl Methyltransferase (ICMT)

A Idrissi Taourati[1*], M Ghamali[2], S Chtita[2], H Zaki[3], F Guenoun[1], T. Lakhlifi[2] and M Bouachrine[4]

[1]*Biology & Chemistry Laboratory Applied to the Environment, Faculty of Science, Moulay Ismail University, Meknes, Morocco*
[2]*Molecular Chemistry and Natural Substances Laboratory, Faculty of Science, Moulay Ismail University, Meknes, Morocco*
[3]*Environment & Health, FSTE, Moulay Ismail University, Meknes, Morocco*
[4]*MEM, ESTM, Moulay Ismail University, Meknes, Morocco*

_____

ABSTRACT

HF method, with the basis set 6-31G (d) was employed to calculate quantum some chemical descriptors of 37 substituted Indole. The best descriptors were selected to establish the quantitative structure activity relationship (QSAR) of the inhibitory activity against isoprenylcysteine carboxyl methyltransferase (Icmt), by principal components analysis (PCA), to a multiple regression analysis (MLR), to a nonlinear regression (RNLM) and to an artificial neural network (ANN). We accordingly propose a quantitative model and we interpret the activity of the compounds relying on the multivariate statistical analysis. This study shows that the MLR and have served to predict activity, but when compared with the results given by the ANN model. We concluded that the predictions achieved by this latter is more effective and much better than other models. The statistical results indicate that the model is statistically significant and shows very good stability towards data variation in the validation method. The contribution of each descriptor to the structure-activity relationship is evaluated.

Keywords: Inhibitory activity; QSAR model; HF; Indoles; ANN

_____

INTRODUCTION

The indole nucleus is an important element of many natural and synthetic molecules with significant biological activity. Indole derivatives have been a topic of substantial research interest and continue to be of the most active areas of heterocyclic chemistry.particularly due to their natural occurrence and pharmacological activities [1-4]. A large number of indole derivatives are at the fore as pharmacologically active lead compounds for drug development. Indole derivatives also occur widely in many natural products such as those from plants [5].

Icmt enzymes are of particular importance in the post-translational modification of proteins that are involved in the regulation of cell growth. Thus, effective Icmt inhibitors may be of significant therapeutic importance in oncogenesis. To determine the structural requirements responsible for high affinity of previously reported amino derivatives of indole as Icmt inhibitors, a successful pharmacophore generation. Several ICMT inhibitors have been developed and the most promising inhibitor to emerge is called Cysmethynil. It is a small molecule indole based

_____

inhibitor; this has shown significant antitumor activity in cancer cells and in a prostate cancer xenograft mouse model.ysmethynil had an in vitro IC50 of 2.4 µM in the initial screen for enzyme inhibitory activity [6-14].

The increment in the speed and the efficiency of drug discovery has seen huge investments by major pharmaceutical companies, with the primary aim of reducing cost per synthesized compound or assay. Computational models that are able to predict the biological activity of compounds by their structural properties are powerful tools to design highly active molecules. In this sense, quantitative structure–activity relationship (QSAR) studies have been successfully applied for modelling biological activities of natural and synthetic chemicals [15].

The objective of this study is to develop predictive QSAR models of the inhibitory activity of the Amino Derivatives of Indole against isoprenylcysteine carboxyl methyltransferase (Icmt) enzyme using several statistical tools: principal components analysis (PCA), multiple linear regression (MLR), nonlinear regression (RNLM) and artificial neural network (ANN) calculations. To test the performance and the stability of this model we opt for the validation method.

MATERIAL AND METHODS

Data sources

In the present study, we choose 37 Amino Derivatives of Indole as potent inhibitors of isoprenylcysteine carboxyl methyltransferase (Icmt) reported by Go and al. [16]. The activities are collected as log (1/IC50) or pIC50 values.Icmt inhibitory activity was determined and expressed in terms of IC50 in (µM) units (concentration required to reduce by 50%).The Amino Derivatives of Indole figure 1 represents the basic structure of the Amino Derivatives of Indole and the table 1 shows the substituted of the compounds studied and experimental activities corresponding pIC50. For proper validation of our data set of a QSAR model.
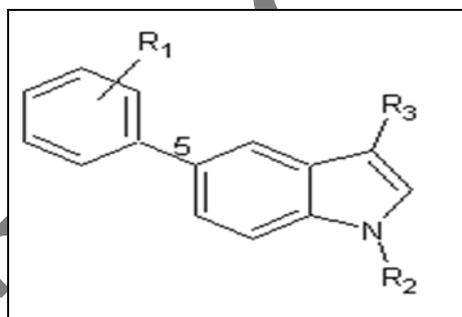

Figure 1: Chemical structure of the studied compounds

Molecular descriptors

At present, there are a large number of molecular descriptors that can be used in QSAR studies. Once validated, the findings can be used to predict activity of untested compounds.
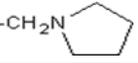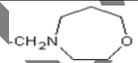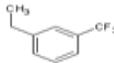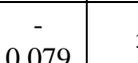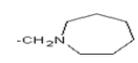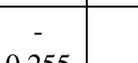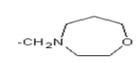
On the one hand, the computation of electronic descriptors was performed by using Gaussian 03W package [17]. The geometries of the 37 Amino Derivatives of Indole as potent inhibitors of isoprenylcysteine carboxyl methyltransferase (Icmt) are optimized with HF method a 6-31G (d) basis set. Then choose some related structural parameters from the results of quantum computation: highest occupied molecular orbital energy (E_{HOMO}), lowest unoccupied molecular orbital energy (E_{LUMO}),Their difference in absolute value (ΔE), dipole moment (DM), total energy (E_T), absolute hardness (η), absolute electronegativity (χ) and reactivity index (ω) [18] η. χ and ω were determined by the following equations:

$$\eta = \frac{EHOMO - ELUMO}{2} \quad ; \quad \chi = \frac{EHOMO + ELUMO}{2} \quad ; \quad \omega = \frac{\chi^2}{2\eta}$$

On the other hand ACD/ChemSketch program [19] is employed to calculate the topological descriptors which are: Molar Volume (MV), Molar Refractivity (MR), Parachor (Pc), Density (d), Refractive Index (n), Polarisability (P) and Surface Tension (S).

Two the other hands MarvinSketch program is employed to calculate the logarithm of the octanol/water partition coefficient of the test compound in its protonated state LogP.

_____

Table 1: Observed activity of studied amino derivatives of indole

Comp	R1	R2	R3	pIC50	Comp	R1	R2	R3	pIC$_{50}$
1	m-CH$_3$	C$_8$H$_{17}$	-CH$_2$CONH$_2$	-0.176	20	m-CH$_3$	C$_8$H$_{17}$	-CH$_2$NCH$_3$(i-C$_3$H$_7$)	0.046
2	m-OCH$_3$	C$_8$H$_{17}$	-CH$_2$CONH$_2$	-0.279	21	m-CH$_3$	C$_8$H$_{17}$	-CH$_2$N⟨pyrrolidine⟩	0.301
3*	m-C$_2$H$_5$	C$_8$H$_{17}$	-CH$_2$CONH$_2$	-0.114	22	m-CH$_3$	C$_8$H$_{17}$	-CH$_2$N⟨azepane⟩	0.155
4	H	C$_8$H$_{17}$	-CH$_2$CONH$_2$	-0.255	23	m-CH$_3$	C$_8$H$_{17}$	-CH$_2$N⟨NCH$_3$⟩	0.046
5	No5-phenyl	C$_8$H$_{17}$	-CH$_2$CONH$_2$	-0.813	24	m-CH$_3$	C$_8$H$_{17}$	-CH$_2$N⟨O⟩	-0.431
6	5-F	C$_8$H$_{17}$	-CH$_2$CONH$_2$	-0.845	25	m-CH$_3$	C$_8$H$_{17}$	CONH$_2$	-1.477
7	m-CH$_3$	H	-CH$_2$CONH$_2$	-1.519	26	m-CH$_3$	C$_8$H$_{17}$	CH2CH$_2$CONH$_2$	-0.079
8	m-CH$_3$	⟨CH$_3$/CF$_3$ benzyl⟩	-CH$_2$CONH$_2$	-0.398	27	m-CH$_3$	C$_8$H$_{17}$	-CH2COOCH3	-0.255
9	m-CH$_3$	isoprenyl	-CH$_2$CONH$_2$	-0.886	28	o- CH$_3$	n- C$_8$H$_{17}$	-CH$_2$N(C$_2$H$_5$)$_2$	0.222
10	m-CH$_3$	C$_8$H$_{17}$	-CH2CON(CH3)2	-0.176	29	P- CH$_3$	C$_8$H$_{17}$	-CH$_2$N(C$_2$H$_5$)$_2$	0.222
11*	m-CH$_3$	C$_8$H$_{17}$	CH2CON(C2H5)2	-0.146	30	m-CH$_3$	isoprenyl	-CH$_2$N(C$_2$H$_5$)$_2$	-0.380
12	m-CH$_3$	C$_8$H$_{17}$	-CH$_2$CON⟨pyrrolidine⟩	-0.079	31*	m-CH$_3$	isoprenyl	-CH$_2$N⟨azepane⟩	-0.322
13	m-CH$_3$	C$_8$H$_{17}$	-CH$_2$CON⟨azepane⟩	-0.255	32	m-CH$_3$	isoprenyl	-CH$_2$N⟨O⟩	-0.301
14*	m-CH$_3$	C$_8$H$_{17}$	-CH$_2$CON⟨NCH$_3$⟩	-0.23	33	No5-phenyl	isoprenyl	-CH$_2$N(C$_2$H$_5$)$_2$	-1.826
15	m-CH$_3$	C$_8$H$_{17}$	-CH$_2$NH$_2$	0.155	34	5-F	isoprenyl	-CH$_2$CONH$_2$	-1.839
16	m-CH$_3$	C$_8$H$_{17}$	-CH$_2$N(CH$_3$)$_2$	-0.114	35*	5-F	n-C$_8$h$_{17}$	-CH$_2$N(C$_2$H$_5$)$_2$	-0.613
17	m-CH$_3$	C$_8$H$_{17}$	-CH$_2$N(C$_2$H$_5$)$_2$	0.155	36*	5-F	isoprenyl	CH$_2$CON(C$_2$H$_5$)$_2$	-1.544
18*	m-CH$_3$	C$_8$H$_{17}$	-CH$_2$N((i-C$_3$H$_7$)2	-0.230	37	m-OCH$_3$	n-C$_8$h$_{17}$	CH$_2$CON(C$_2$H$_5$)$_2$	0.097
19	m-CH$_3$	C$_8$H$_{17}$	-CH$_2$N(n-C$_3$H$_7$)$_2$	0.097					
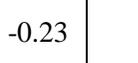
_____

Statistical Analysis

The objective of quantitative structure-activity relationship (QSAR) analysis is to derive empirical models that relate the biological activity of compounds to their chemical structures. In this QSAR analysis, quantitative descriptors are used to describe the chemical structure and the analysis results in a mathematical model describing the relationship between the chemical structure and biological activity. To explain the structure-activity relationship, these 16 descriptors are calculated for 37 molecules using the Gaussian03W, ChemSketch and MarvinSketch programs.

The quantitative descriptors of the substituted Amino Derivatives of Indole are studied by statistical methods based on the principal component analysis (PCA) [20] using the software XLSTAT version 2014 [21]. PCA is a useful statistical technique for summarizing all the information encoded in the structures of the compounds. It is also very helpful to understand the distribution of the compounds [22]. This is an essentially descriptive statistical method which aims to present in graphic form, the maximum of information contained in the data table 1 and table 2.

The multiple linear regression (MLR) analysis with descendent selection and elimination of variables is employed to model the structure-activity relationship. It is a mathematic technique that minimizes difference between actual and predicted values. Also it is served to select the descriptors used as the input parameters in the multiples nonlinear regression (MNLR) and artificial neural network (ANN).

MLR and MNLR are generated using the software XLSTAT version 2014.To predict pIC50 Equations are justified by the correlation coefficient (R), the mean squared error (MSE), the fisher's criterion (F) and the significance level (P).

The ANN analysis is performed with the use of Matlab software version 2009a Neural Fitting tool (nftool) toolbox on a data set of the compounds [23]. A number of individual models of ANN are designed built up and trained. 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 [24]. 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.

According to the supervised learning adopted, the networks are taught by giving them examples of input patterns and the corresponding target outputs. Through an iterative process, the connection weights are modified until the network gives the desired results for the training set of data. A back propagation algorithm is used to minimize the error function. This algorithm has been described previously with a simple example of application [25] and a detail of this algorithm is given elsewhere [26].

The validation of predictive power of a QSAR model, two basic principles (internal validationand external validation) are available. The cross validation is one of the most popular methods for internal validation. In the paper, the internal predictive capability of the model is evaluated by Leave-p-out cross-validation (LPOCV with p=3) and Leave-one-out cross-validation (LOOCV is a particular case of leave-p-out cross-validation with p = 1) (R^2_{pcv}; R^2_{ocv}). A good R^2_{cv} often indicates a good robustness and high internal predictive power of a QSAR model. However, recent studies [27].So, the external validation remains the only way to determine both the generalizability and the true predictive power of QSAR models for new chemicals. For this reason, the statistical external validation was applied to the models as described by Globarikh and Tropsha [28].

RESULTS AND DISCUSSION

Data set for analysis

A QSAR study was carried for a series of the 37 Amino Derivatives of Indole, as potent inhibitors of isoprenylcysteine carboxyl methyltransferase (Icmt) as reported in [16].In order to determine a quantitative relationship between structure and biological activity, the values of the 16 descriptors as shown in table 2.

Principal component analysis

The totality of the 14 descriptors coding the 37 molecules is submitted to a principal components analysis (PCA) [29]. The first three principal axes are sufficient to describe the information provided by the data matrix. Indeed, the percentages of variance are 48.61%, 21.93% and 13.72% for the axes F1, F2 and F3.Respectively, the total information is estimated to a percentage of 84.26%. The principal component analysis (PCA) [30] was conducted to identify the link between the different variables. The correlations between the fourteen descriptors are shown in Table 3, as a correlation matrix.

A Idrissi Taourati *et al*

J. Chem. Pharm. Res., 2016, 8(11):159-172

_____

Table 2: Values of the obtained parameters of the studied substituted indole

Comp	pIC$_{50}$	Log P	MR	MV	PC	n	S	d	P	E$_T$	DM	E$_{HOMO}$	E$_{LUMO}$	ΔE	η	χ	ω
1	-0.1761	6.39	116.47	352.4	887.1	1.575	40.1	1.06	46.17	-1149.152	2.707	-0.276	0.113	0.389	-0.194	-0.082	-0.017
2	-0.2788	5.72	117.85	358.8	906.3	1.57	40.6	1.09	46.72	-1223.995	3.607	-0.277	0.112	0.389	-0.194	-0.082	-0.017
3*	-0.1139	6.83	121.07	368.4	925.7	1.571	39.8	1.05	47.99	-1188.187	3.08	-0.276	0.113	0.389	-0.194	-0.081	-0.017
4	-0.2553	5.88	112.04	337.2	856	1.578	41.5	1.07	44.41	-1110.115	2.848	-0.276	0.112	0.389	-0.194	-0.082	-0.017
5	-0.8129	4.23	86.94	269.3	679.1	1.558	40.4	1.06	34.46	-880.564	3.092	-0.275	0.132	0.407	-0.204	-0.071	-0.012
6	-0.8451	4.37	86.81	272.2	679.3	1.55	38.7	1.11	34.41	-979.415	2.169	-0.286	0.105	0.391	-0.196	-0.091	-0.021
7	-1.5185	2.35	81.05	217	592.4	1.669	55.4	1.217	32.13	-836.875	2.367	-0.279	0.124	0.402	-0.201	-0.078	-0.015
8	-0.3979	5.77	118.85	354.7	893.8	1.584	40.2	1.23	47.11	-1480.112	4.194	-0.281	0.107	0.389	-0.194	-0.087	-0.019
9	-0.8865	4.65	102.45	303.3	763.7	1.59	40.1	1.09	40.61	-1030.859	2.491	-0.277	0.113	0.39	-0.195	-0.082	-0.017
10	-0.1761	6.65	127.35	396.8	976.2	1.554	36.6	1.01	50.48	-1227.199	3.069	-0.274	0.116	0.39	-0.195	-0.079	-0.016
11*	-0.1461	7.36	136.57	428.9	1053.5	1.549	36.3	1	54.14	-1305.271	2.399	-0.273	0.117	0.39	-0.195	-0.078	-0.016
12	-0.0792	7.05	134.39	402.1	1013.7	1.583	40.3	1.07	53.27	-1304.112	3.198	-0.275	0.115	0.39	-0.195	-0.08	-0.016
13	-0.2553	7.5	139	418.1	1052.3	1.579	40.1	1.06	55.1	-1343.142	3.247	-0.277	0.112	0.389	-0.195	-0.083	-0.018
14*	-0.2304	6.49	142.77	428	1078.9	1.581	40.3	1.07	56.6	-1398.15	3.382	-0.278	0.111	0.389	-0.195	-0.083	-0.018
15	0.1549	6.02	111	339.1	842.8	1.568	38.1	1.02	44	-1036.383	1.818	-0.265	0.123	0.388	-0.194	-0.071	-0.013
16	-0.1139	6.84	121.89	383.5	932	1.548	34.8	0.98	48.32	-1114.439	2.02	-0.266	0.122	0.389	-0.194	-0.072	-0.013
17	0.1549	7.55	131.11	415.6	1009.2	1.543	34.7	0.97	51.97	-1192.036	2.371	-0.269	0.12	0.389	-0.195	-0.074	-0.014
18*	-0.2304	8.38	139.96	454.9	1071.4	1.54	33.3	0.97	55.48	-1270.572	2.099	-0.267	0.121	0.388	-0.194	-0.073	-0.014
19	0.0969	8.6	140.33	447.7	1086.4	1.539	34.6	0.96	55.63	-1270.265	2.063	-0.266	0.122	0.388	-0.194	-0.072	-0.013
20	0.0458	7.61	130.93	414.7	1001.7	1.544	34	0.97	51.9	-1192.507	2.022	-0.266	0.122	0.388	-0.194	-0.072	-0.013
21	0.301	7.24	128.93	382.8	969.4	1.577	38.6	1.03	51.11	-1191.348	2.075	-0.267	0.121	0.387	-0.194	-0.073	-0.014
22	0.1549	7.69	133.54	404.8	1008	1.573	38.4	1.02	52.94	-1230.378	2.089	-0.267	0.117	0.384	-0.192	-0.075	-0.015
23	0.0458	6.68	137.31	414.7	1034.6	1.576	38.7	1.04	54.43	-1285.385	2.288	-0.268	0.116	0.384	-0.192	-0.076	-0.015
24	-0.4314	6.82	130.32	395.2	988.6	1.573	39.1	1.05	51.66	-1266.186	0.601	-0.273	0.113	0.386	-0.193	-0.08	-0.017
25	-1.4771	5.76	111.86	336.3	848.5	1.579	40.4	1.07	44.34	-1110.116	6.183	-0.286	0.095	0.381	-0.191	-0.095	-0.024
26	-0.0792	6.28	121.07	368.4	925.7	1.571	39.8	1.05	47.99	-1188.182	3.981	-0.253	0.123	0.376	-0.188	-0.065	-0.011
27	-0.2553	6.1	122.03	375.5	930.8	1.563	37.7	1.03	48.37	-1188.179	1.879	-0.274	0.119	0.393	-0.196	-0.078	-0.015
28	0.2218	8.21	131.11	415.6	1009.2	1.543	34.7	0.97	51.97	-1192.032	2.72	-0.27	0.12	0.39	-0.195	-0.075	-0.014
29	0.2218	8.21	131.11	415.6	1009.2	1.543	34.7	0.97	51.97	-1192.036	2.215	-0.265	0.13	0.395	-0.197	-0.068	-0.012
30	-0.3802	6.47	117.1	366.5	885.8	1.551	34.1	0.98	46.42	-1073.803	2.71	-0.27	0.119	0.388	-0.194	-0.075	-0.015
31*	-0.3222	6.16	114.92	339.8	846.1	1.591	38.4	1.05	45.55	-1073.061	2.717	-0.268	0.119	0.387	-0.193	-0.075	-0.014
32	-0.301	5.54	116.31	346.2	865.2	1.586	38.9	1.08	46.11	-1147.9	1.197	-0.274	0.111	0.385	-0.192	-0.082	-0.017
33	-1.8261	4.31	87.57	283.4	677.9	1.53	32.7	0.95	34.71	-805.311	2.442	-0.269	0.123	0.392	-0.196	-0.073	-0.014
34	-1.8388	2.63	72.79	223	556	1.566	38.5	1.16	28.85	-861.123	2.284	-0.287	0.105	0.392	-0.196	-0.091	-0.021
35*	-0.6128	6.2	101.45	335.4	801.4	1.516	32.5	0.99	40.22	-1022.369	4.919	-0.28	0.111	0.392	-0.196	-0.085	-0.018
36*	-1.5441	4.45	87.44	286.3	678	1.522	31.4	1	34.66	-904.136	5.069	-0.281	0.109	0.391	-0.195	-0.086	-0.019
37	0.0969	6.88	137.96	435.3	1072.6	1.546	36.8	1.03	54.69	-1380.115	3.268	-0.274	0.116	0.39	-0.195	-0.079	-0.016

_____

The obtained matrix provides information on the high or low interrelationship between the variables. In general, good co-linearity (r > 0.5) was observed between most of the variables. A high interrelationship was observed between (MR) and (P) (r = 1) and a low interrelationship was observed between E_T and DM (r = −0.016). Additionally, to decrease the redundancy existing in our data matrix, the descriptors that are highly correlated (R ≥ 0.9), were excluded.

Table 3: Correlation matrix between different obtained descriptors

	pIC50	Log P	MR	MV	Pc	n	S	d	P	E_T	DM	E_{HOMO}	E_{LUMO}	ΔE	η	χ	ω
pIC50	1																
Log P	0.834	1															
MR	0.827	0.911	1														
MV	0.81	0.949	0.979	1													
Pc	0.829	0.927	0.996	0.991	1												
n	-0.128	-0.385	-0.125	-0.311	-0.198	1											
S	-0.194	-0.473	-0.232	-0.398	-0.282	0.941	1										
d	-0.38	-0.636	-0.41	-0.544	-0.454	0.749	0.792	1									
P	0.827	0.911	1	0.979	0.996	-0.125	-0.232	-0.41	1								
E_T	-0.714	-0.733	-0.895	-0.85	-0.889	0.011	0.081	0.052	-0.895	1							
DM	-0.323	-0.157	-0.179	-0.158	-0.159	-0.114	-0.023	0.145	-0.179	0.016	1						
E_{HOMO}	0.602	0.555	0.484	0.502	0.481	-0.16	-0.272	-0.571	0.484	-0.206	-0.381	1					
E_{LUMO}	0.376	0.257	0.162	0.197	0.171	-0.1	-0.101	-0.445	0.162	0.104	-0.449	0.745	1				
ΔE	-0.305	-0.408	-0.444	-0.42	-0.426	0.081	0.235	0.165	-0.444	0.433	-0.104	-0.337	0.377	1			
η	0.305	0.408	0.444	0.42	0.426	-0.081	-0.235	-0.165	0.444	-0.433	0.104	0.337	-0.377	-1	1		
χ	0.523	0.434	0.345	0.373	0.348	-0.139	-0.199	-0.543	0.345	-0.053	-0.444	0.933	0.935	0.024	-0.024	1	
ω	0.512	0.407	0.323	0.351	0.326	-0.127	-0.178	-0.527	0.323	-0.035	-0.47	0.899	0.959	0.105	-0.105	0.995	1

Multiple linear regressions MLR

Many attempts have been made to develop a relationship with the indicator variable of biological activity pIC50, but the best relationship obtained by this method is only one corresponding to the linear combination of several descriptors selected: octanol/water partition coefficient of the test compound in its protonated state (LogP), Density (d), dipole moment (DM), the energy E_{HOMO} and the energy E_{LUMO}.

The resulting equation is:

$$\text{pIC50} = -3.774 + 0.394 \times \textbf{LogP} + 3.733 \times \textbf{d} - 0.091 \times \textbf{DM} + 15.001 \times \textbf{E}_{\textbf{HOMO}} + 11.934 \times \textbf{E}_{\textbf{LUMO}} \text{ (Equation 1)}$$

$$N = 30 \quad R = 0.915 \quad R^2 = 0.837 \quad R^2_{cv} \quad (R^2_{pcv} = 0.698; R^2_{ocv} = 0.678)$$

$$MSE = 0.072 \quad F = 24.731 \quad P < 0.0001$$

In the equation, N is the number of compounds, R is the correlation coefficient, MSE is the mean squared error, F is the fisher's criterion, P is the significance level.

Higher correlation coefficient and lower mean squared errors indicate that the model is more reliable. And P is much smaller than 0.05 means that the regression equation is statistically significance. The QSAR model expressed by Eq. (1) is cross validated by its appreciable R^2cv values R^2cv ($R^2pcv = 0.698$; $R^2ocv = 0.678$) obtained by the Leave-p-out (LpO) Leave-one-out (LOO) methods. The value of R^2cv is greater than 0.5 is the basic requirement for qualifying a QSAR model to be valid one [27].

The correlation coefficients between variables in the model were calculated by variance inflation factor (VIF) as shown in Table 4. The VIF was defined as $1/(1 − R^2)$, where R was the multiple correlation coefficients for one independent variable against all the other descriptors in the model. Models with a VIF greater than 5 were unstable and were eliminated. Models with a VIF values between 1 and 4 means the models can be accepted, as can be seen from Table 4. The VIF values of the five descriptors are all smaller than 5.0, indicating that there is no collinearity among the selected descriptors and the resulting model has good stability.

The elaborated QSAR model reveals that the inhibitory activity against isoprenylcysteine carboxyl methyltransferase (Icmt) may be explained by a number of electronic and physic-chemical factors. The positive correlation of the electronic descriptors (DM, E_{HOMO}, E_{LUMO}) and the topologic descriptors (LogP, d) with the activity pIC50 shows that an increase in the values of these factors indicates an increase in the value of the activity pIC50.

Table 4: the variance inflation factors (VIF) of descriptors in QSAR model

Statistic	LogP	d	DM	E_{HOMO}	E_{LUMO}
Tolerance	0.4499	0.4093	0.8371	0.3054	0.3879
VIF	2.2225	2.4431	1.1945	3.2749	2.5782

The correlation of the predicted and observed activity and the residual graph of absolute numbers are illustrated in figure 2. The descriptors proposed in Eq. (1) by MLR are, therefore, used as the input parameters in the multiple nonlinear regressions (MNLR) and artificial neural network (ANN)
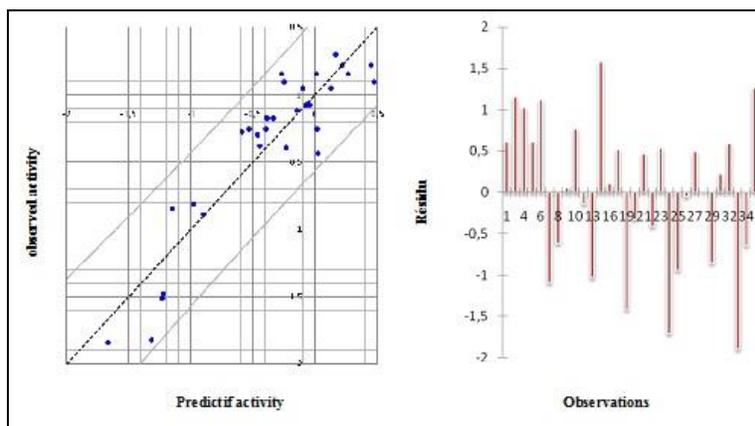
Figure 2: Graphical representation of calculated and observed activity and the residues values calculated by MLR

Multiple nonlinear regressions MNLR

We have also used the technique of nonlinear regression model to improve the structure activity in a quantitative way, taking into account several parameters. This is the most common tool for the study of multidimensional data. We have applied to the data matrix constituted obviously from the descriptors proposed by MLR corresponding to the 30 compound training set.

The resulting equation is:

$$\text{pIC50} = -19.765 + 0.847 \times \text{LogP} + 29.985 \times \text{d} + 0.361 \times \text{DM} - 71.331 \times E_{HOMO} - 184.19807 \times E_{LUMO} - 0.038 \times (\text{Log P})^2 - 11.960 \times \text{d}^2 - 0.080 \times \text{DM}^2 + 162.216 \times E^2_{HOMO} - 826.624 \times E^2_{LUMO} \text{ (Equation 2)}$$

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

$$N = 30 \quad R = 0.961 \quad R^2 = 0.924 \quad R^2_{cv} (R^2_{pcv} = 0.558; R^2_{ocv} = 0.569) \quad MSE = 0.042$$

The QSAR model expressed by Eq. (2) is cross validated by its appreciable R^2cv values R^2cv (R^2pcv = 0. 558 - R^2ocv =0. 569) obtained by the Leave-p-out (LpO) and Leave-one-out (LOO) methods. The value of R^2cv is greater than 0.5 is the basic requirement for qualifying a QSAR model to be valid one [27].

The correlations of predicted and observed activities and the residual graph of absolute numbers are illustrated in Figure 3.
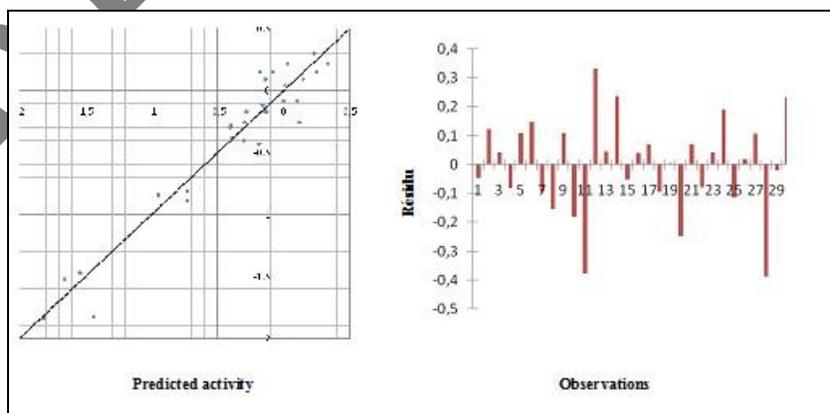
Figure 3: Graphical representation of calculated and observed activity and the residues value calculated by MNLR

Artificial neural networks ANN

Neural networks (ANN) can be used to generate predictive models of quantitative structure– activity relationship (QSAR) between a set of molecular descriptors obtained from the MLR and observed activity. The ANN calculated activity model is developed using the properties of several studied compounds. The correlation of predicted and observed activities and the residual graph of absolute numbers are illustrated in figure 4.
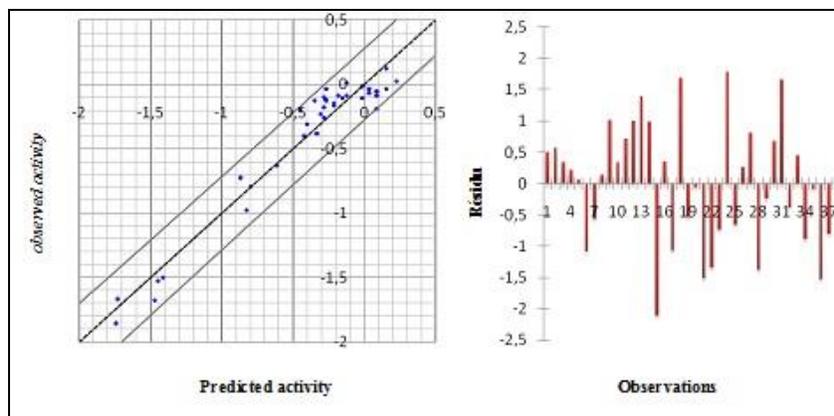

Figure 4: Graphical representation of calculated and observed activity and the residues values calculated by ANN

$$N = 37 \qquad R^2 = 0.971 \qquad R^2cv = 0.911 \qquad MSE = 0.017$$

The obtained squared correlation coefficient (R) value is 0.971 for this data set of substituted Amino of Indole. It confirms that the artificial neural networks (ANN) results are the best to build the quantitative structure activity relationship model. Further, the high R^2cv value ($R^2cv = 0.911$) shows that the obtained QSAR model can to predict the inhibitory activity against isoprenylcysteine carboxyl methyltransferase (Icmt).

External validation

To estimate the predictive power of the MLR, MNLR and ANN models, we must use a set of compounds that have not been used as the training set to establish the QSAR model. The models established in the computation procedure using the 30 substituted Amino of Indole are used to predict the activity of the remaining 7 compounds The main performance parameters of the three models are shown in Table 5. As seen from this table, the statistical parameters of the ANN model are better than the other models.

Table 5: Performance comparison between models obtained by MLR, RNLM and ANN

Model	Training set					Test set	
	R	R^2	R^2cv		MSE	R_{ext}	R^2_{ext}
			LPO	LOO			
MLR	0.915	0.837	0.698	0.678	0.072	0.915	0.837
MNLR	0.961	0.924	0.558	0.569	0.042	0.945	0.894
ANN	0.985	0.971	0.911		0.017	0.987	0.975

We assessed the best linear QSAR regression equations established in this study. Based on this result, a comparison of the quality of the MLR and MNLR models shows that the ANN model has a significantly better predictive capability because the ANN approach yields better results than those of MLR and MNLR. ANN establishes a satisfactory relationship between the molecular descriptors and the activity of the studied compounds.

CONCLUSION

In this work we investigated the QSAR regression to predict the inhibitory activity of the Amino Derivatives of Indole against isoprenylcysteine carboxyl methyltransferase (Icmt). The studies of the quality of the three models built in the paper have good stabilities and great predictive powers. Moreover, compared to the MLR, the MNLR models, the ANN model is better, which is an effective tool to predict the inhibitory activity of the Amino Derivatives of Indole. Further, With ANN approach, we have established a relationship between several descriptors

and inhibition values pIC50 of several organic compounds based on substituted Amino of Indole in satisfactory manners. The accuracy and predictability of the proposed models were illustrated by the comparison of key statistical indicators like R or R^2 of different models obtained by using different statistical tools and different descriptors has been shown in table 5 and to validate these results we generate which are illustrated in Table 6.

Finally, we can conclude that the 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.

Table 6: Observed values and calculated values of pIC_{50} according to different methods

N°	pIC_{50}		pIC_{50} (calc.)	
	(obs.)	MLR	NMLR	ANN
1	-0.176	-0.338	-0.13	-0.147
2	-0.279	-0.589	-0.402	-0.231
3*	-0.114	-0.233	-0.068	-0.109
4	-0.255	-0.531	-0.297	-0.256
5	-0.813	-0.977	-0.733	-0.791
6	-0.845	-1.145	-0.954	-0.974
7	-1.519	-1.226	-1.665	-1.524
8	-0.398	-0.231	-0.304	-0.399
9	-0.886	-0.9	-0.732	-0.73
10	-0.176	-0.383	-0.286	-0.167
11*	-0.146	-0.066	-0.062	-0.088
12	-0.079	-0.042	0.101	0.011
13	-0.255	0.022	0.122	-0.098
14*	-0.23	-0.372	-0.179	-0.129
15	0.155	-0.269	-0.179	-0.191
16	-0.114	-0.142	-0.16	-0.108
17	0.155	0.015	-0.08	-0.052
18*	-0.23	0.407	0.155	-0.036
19	0.097	0.478	0.149	-0.032
20	0.046	0.133	0.006	-0.016
21	0.301	0.175	0.231	0.024
22	0.155	0.267	0.248	-0.088
23	0.046	-0.099	0.042	-0.106
24	-0.431	0.027	-0.183	-0.208
25	-1.477	-1.225	-1.547	-1.499
26	-0.079	-0.066	-0.002	-0.089
27	-0.255	-0.388	-0.299	-0.177
28	0.222	0.219	0.031	-0.032
29	0.222	0.451	0.335	0.122
30	-0.38	-0.44	-0.398	-0.309
31*	-0.322	-0.277	-0.084	-0.124
32	-0.301	-0.458	-0.409	-0.379
33	-1.826	-1.317	-1.439	-1.671
34	-1.839	-1.667	-1.821	-1.862
35*	-0.613	-0.962	-1.23	-0.626
36*	-1.544	-1.666	-2.012	-1.677
37	0.097	-0.242	-0.135	-0.069
		* test set		

ACKNOWLEDGMENT

REFERENCES

[1] Q Hayat; S Hayat; M Irfan; A Ahmad. *Environmental and experimental botany*, **2010**, 68(1), 14-25.

[2] Chuanfu; An; Z Mou. *J integrative plant biology*, **2011**, 53(6), 412-428.

[3] AC Vlot; DMA Dempsey; DF Klessig. *Annual review of phytopathology*, **2009**, 47, 177-206.

[4] S Manivannan; R Karthikeyan; M Muthubharathi. *Int J chem tech Res*, **2015**,7, 2936-2941.

[5] M Mayavathi; V Chitra; Sathish; DR singh. *Int J ChemTech Res*,**2015**,7:2956-2962.

[6] A Manaf; J Athra; G Sager. *J university of Anbar for pure science,* **2013**,7(2),1.

[7] BE Ezema; SA Agada; PC Uzoewulu. *Int J ChemTech Res*, **2016**,1,169-174.

[8] MEL Amane; M Bouhdada. *Int J ChemTech Res*, **2014**, 2, 1430-1439.

[9] M Abirami; V Nadaraj. *Int J PharmTech Res*, **2015**, 4, 558-561.

[10] ASP Azzouz ; RT Ali. *National J Chem*, **2010**, 37, 158-168,

[11] R Vijayalakshmi; YNS Ramya; AD Mani; MD Dhanaraju. *Int J PharmTech Res*, **2016**, 6, 301-306.

[12] FM Benachenhou; RK Slimane. *Arab . J. Chem .* **2012**, 5, 245-250.

[13] A Adabiardakani; M Hakimi; H Kargar,.*AP J* ,**2012**, 2 (11), 472-476.

[14] PB Mohite; VH Bhaskar. *Int J pharm Tech Res*, 2011, 3(3), 1557-1566 .

[15] FA El-Samahy; M Elsedik; T Aysha; A Magda. *Int J PharmTech Res*, **2016**, 6, 436-445.

[16] LT Daood. *Raf.gour.sci*,**2008**,19(3), 24-37.