International Journal of Pharmaceutical Science Invention
ISSN (Online): 2319 – 6718, ISSN (Print): 2319 – 670X
www.ijpsi.org Volume 6 Issue 1 ǁ January 2017 ǁ PP. 06-13

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

A.Idrissi Taourati[(a)*], M. Ghamali[(b)], S. Chtita[(b)], H. Zaki[(c)] , M. Benlyass[(c)], F. Guenoun[(a)], T. Lakhlifi[(b)], M. Bouachrine[(d)]

[(a]) Biology chemistry laboratory applied to the environment. Faculty of Science. University Moulay Ismail. Meknes. Morocco
[(b)] Molecular Chemistry and Natural Substances Laboratory. Faculty of Science. University Moulay Ismail. Meknes. Morocco
[(c)] Environment & Health. FSTE. University Moulay Ismail. Meknes. Morocco
[(d)] MEM, ESTM. University Moulay Ismail. 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*

I. 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.

II. Material And Methods

2.1. 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.

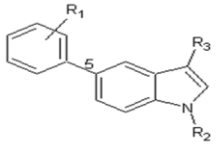


Figure 1: Chemical structure of the studied compounds

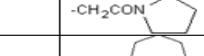
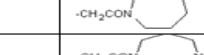
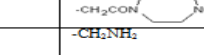
2.2. Molecular descriptors

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).

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_8H_{17}	-CH_2CONH_2	-0.176	20	m-CH_3	C_8H_{17}	-$CH_2NCH_3(i\text{-}C_3H_7)$	0.046
2	m-OCH_3	C_8H_{17}	-CH_2CONH_2	-0.279	21	m-CH_3	C_8H_{17}		0.301
3*	m-C_2H_5	C_8H_{17}	-CH_2CONH_2	-0.114	22	m-CH_3	C_8H_{17}		0.155
4	H	C_8H_{17}	-CH_2CONH_2	-0.255	23	m-CH_3	C_8H_{17}		0.046
5	No5-phenyl	C_8H_{17}	-CH_2CONH_2	-0.813	24	m-CH_3	C_8H_{17}		-0.431
6	5-F	C_8H_{17}	-CH_2CONH_2	-0.845	25	m-CH_3	C_8H_{17}	$CONH_2$	-1.477
7	m-CH_3	H	-CH_2CONH_2	-1.519	26	m-CH_3	C_8H_{17}	-$CH2CH_2CONH_2$	-0.079
8	m-CH_3		-CH_2CONH_2	-0.398	27	m-CH_3	C_8H_{17}	-CH2COOCH3	-0.255
9	m-CH_3	isoprenyl	-CH_2CONH_2	-0.886	28	o-CH_3	n- C_8H_{17}	-$CH_2N(C_2H_5)_2$	0.222
10	m-CH_3	C_8H_{17}	-CH2CON(CH3)2	-0.176	29	P-CH_3	C_8H_{17}	-$CH_2N(C_2H_5)_2$	0.222
11*	m-CH_3	C_8H_{17}	-CH2CON(C2H5)2	-0.146	30	m-CH_3	isoprenyl	-$CH_2N(C_2H_5)_2$	-0.380
12	m-CH_3	C_8H_{17}		-0.079	31*	m-CH_3	isoprenyl		-0.322
13	m-CH_3	C_8H_{17}		-0.255	32	m-CH_3	isoprenyl		-0.301
14*	m-CH_3	C_8H_{17}		-0.230	33	No5-phenyl	isoprenyl	-$CH_2N(C_2H_5)_2$	-1.826
15	m-CH_3	C_8H_{17}	-CH_2NH_2	0.155	34	5-F	isoprenyl	-CH_2CONH_2	-1.839
16	m-CH_3	C_8H_{17}	-$CH_2N(CH_3)_2$	-0.114	35*	5-F	n-C_8h_{17}	-$CH_2N(C_2H_5)_2$	-0.613
17	m-CH_3	C_8H_{17}	-$CH_2N(C_2H_5)_2$	0.155	36*	5-F	isoprenyl	-$CH_2CON(C_2H_5)_2$	-1.544
18*	m-CH_3	C_8H_{17}	-$CH_2N((i\text{-}C_3H_7)2$	-0.230	37	m-OCH_3	n-C_8h_{17}	-$CH_2CON(C_2H_5)_2$	0.097
19	m-CH_3	C_8H_{17}	-$CH_2N(n\text{-}C_3H_7)_2$	0.097					
* test set									

2.3. 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].

III. Results And Discussion

3.1. Data set for analysis

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	6.390	116.470	352.400	887.100	1.575	40.100	1.060	46.170	-1149.152	2.707	-0.276	0.113	0.389	-0.194	-0.082	-0.017
2	-0.2788	5.720	117.850	358.800	906.300	1.570	40.600	1.090	46.720	-1223.995	3.607	-0.277	0.112	0.389	-0.194	-0.082	-0.017
3*	-0.1139	6.830	121.070	368.400	925.700	1.571	39.800	1.050	47.990	-1188.187	3.080	-0.276	0.113	0.389	-0.194	-0.081	-0.017
4	-0.2553	5.880	112.040	337.200	856.000	1.578	41.500	1.070	44.410	-1110.115	2.848	-0.276	0.112	0.389	-0.194	-0.082	-0.017
5	-0.8129	4.230	86.940	269.300	679.100	1.558	40.400	1.060	34.460	-880.564	3.092	-0.275	0.132	0.407	-0.204	-0.071	-0.012
6	-0.8451	4.370	86.810	272.200	679.300	1.550	38.700	1.110	34.410	-979.415	2.169	-0.286	0.105	0.391	-0.196	-0.091	-0.021
7	-1.5185	2.350	81.050	217.000	592.400	1.669	55.400	1.217	32.130	-836.875	2.367	-0.279	0.124	0.402	-0.201	-0.078	-0.015
8	-0.3979	5.770	118.850	354.700	893.800	1.584	40.200	1.230	47.110	-1480.112	4.194	-0.281	0.107	0.389	-0.194	-0.087	-0.019
9	-0.8865	4.650	102.450	303.300	763.700	1.590	40.100	1.090	40.610	-1030.859	2.491	-0.277	0.113	0.390	-0.195	-0.082	-0.017
10	-0.1761	6.650	127.350	396.800	976.200	1.554	36.600	1.010	50.480	-1227.199	3.069	-0.274	0.116	0.390	-0.195	-0.079	-0.016
11*	-0.1461	7.360	136.570	428.900	1053.500	1.549	36.300	1.000	54.140	-1305.271	2.399	-0.273	0.117	0.390	-0.195	-0.078	-0.016
12	-0.0792	7.050	134.390	402.100	1013.700	1.583	40.300	1.070	53.270	-1304.112	3.198	-0.275	0.115	0.390	-0.195	-0.080	-0.016
13	-0.2553	7.500	139.000	418.100	1052.300	1.579	40.100	1.060	55.100	-1343.142	3.247	-0.277	0.112	0.389	-0.195	-0.083	-0.018
14*	-0.2304	6.490	142.770	428.000	1078.900	1.581	40.300	1.070	56.600	-1398.150	3.382	-0.278	0.111	0.389	-0.195	-0.083	-0.018
15	0.1549	6.020	111.000	339.100	842.800	1.568	38.100	1.020	44.000	-1036.383	1.818	-0.265	0.123	0.388	-0.194	-0.071	-0.013
16	-0.1139	6.840	121.890	383.500	932.000	1.548	34.800	0.980	48.320	-1114.439	2.020	-0.266	0.122	0.389	-0.194	-0.072	-0.013
17	0.1549	7.550	131.110	415.600	1009.200	1.543	34.700	0.970	51.970	-1192.036	2.371	-0.269	0.120	0.389	-0.195	-0.074	-0.014
18*	-0.2304	8.380	139.960	454.900	1071.400	1.540	33.300	0.970	55.480	-1270.572	2.099	-0.267	0.121	0.388	-0.194	-0.073	-0.014
19	0.0969	8.600	140.330	447.700	1086.400	1.539	34.600	0.960	55.630	-1270.265	2.063	-0.266	0.122	0.388	-0.194	-0.072	-0.013
20	0.0458	7.610	130.930	414.700	1001.700	1.544	34.000	0.970	51.900	-1192.507	2.022	-0.266	0.122	0.388	-0.194	-0.072	-0.013
21	0.3010	7.240	128.930	382.800	969.400	1.577	38.600	1.030	51.110	-1191.348	2.075	-0.267	0.121	0.387	-0.194	-0.073	-0.014
22	0.1549	7.690	133.540	404.800	1008.000	1.573	38.400	1.020	52.940	-1230.378	2.089	-0.267	0.117	0.384	-0.192	-0.075	-0.015
23	0.0458	6.680	137.310	414.700	1034.600	1.576	38.700	1.040	54.430	-1285.385	2.288	-0.268	0.116	0.384	-0.192	-0.076	-0.015
24	-0.4314	6.820	130.320	395.200	988.600	1.573	39.100	1.050	51.660	-1266.186	0.601	-0.273	0.113	0.386	-0.193	-0.080	-0.017
25	-1.4771	5.760	111.860	336.300	848.500	1.579	40.400	1.070	44.340	-1110.116	6.183	-0.286	0.095	0.381	-0.191	-0.095	-0.024
26	-0.0792	6.280	121.070	368.400	925.700	1.571	39.800	1.050	47.990	-1188.182	3.981	-0.253	0.123	0.376	-0.188	-0.065	-0.011
27	-0.2553	6.100	122.030	375.500	930.800	1.563	37.700	1.030	48.370	-1188.179	1.879	-0.274	0.119	0.393	-0.196	-0.078	-0.015
28	0.2218	8.210	131.110	415.600	1009.200	1.543	34.700	0.970	51.970	-1192.032	2.720	-0.270	0.120	0.390	-0.195	-0.075	-0.014
29	0.2218	8.210	131.110	415.600	1009.200	1.543	34.700	0.970	51.970	-1192.036	2.215	-0.265	0.130	0.395	-0.197	-0.068	-0.012
30	-0.3802	6.470	117.100	366.500	885.800	1.551	34.100	0.980	46.420	-1073.803	2.710	-0.270	0.119	0.388	-0.194	-0.075	-0.015
31*	-0.3222	6.160	114.920	339.800	846.100	1.591	38.400	1.050	45.550	-1073.061	2.717	-0.268	0.119	0.387	-0.193	-0.075	-0.014
32	-0.3010	5.540	116.310	346.200	865.200	1.586	38.900	1.080	46.110	-1147.900	1.197	-0.274	0.111	0.385	-0.192	-0.082	-0.017
33	-1.8261	4.310	87.570	283.400	677.900	1.530	32.700	0.950	34.710	-805.311	2.442	-0.269	0.123	0.392	-0.196	-0.073	-0.014
34	-1.8388	2.630	72.790	223.000	556.000	1.566	38.500	1.160	28.850	-861.123	2.284	-0.287	0.105	0.392	-0.196	-0.091	-0.021
35*	-0.6128	6.200	101.450	335.400	801.400	1.516	32.500	0.990	40.220	-1022.369	4.919	-0.280	0.111	0.392	-0.196	-0.085	-0.018
36*	-1.5441	4.450	87.440	286.300	678.000	1.522	31.400	1.000	34.660	-904.136	5.069	-0.281	0.109	0.391	-0.195	-0.086	-0.019
37	0.0969	6.880	137.960	435.300	1072.600	1.546	36.800	1.030	54.690	-1380.115	3.268	-0.274	0.116	0.390	-0.195	-0.079	-0.016

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.

3.2. 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.

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	0.810	**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	-0.380	-0.636	-0.410	-0.544	-0.454	0.749	0.792	1									
P	0.827	**0.911**	**1.000**	**0.979**	**0.996**	-0.125	-0.232	-0.410	1								
E_T	-0.714	-0.733	**-0.895**	-0.850	-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	-0.160	-0.272	-0.571	0.484	-0.206	-0.381	1					
E_{LUMO}	0.376	0.257	0.162	0.197	0.171	-0.100	-0.101	-0.445	0.162	0.104	-0.449	0.745	1				
ΔE	-0.305	-0.408	-0.444	-0.420	-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	0.420	0.426	-0.081	-0.235	-0.165	0.444	-0.433	0.104	0.337	-0.377	**-1.000**	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	-0.470	**0.899**	**0.959**	0.105	-0.105	**0.995**	1

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 ($R \geq 0.9$), were excluded.

3.3. 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 $\mathbf{E_{HOMO}}$ and the energy $\mathbf{E_{LUMO}}$.

The resulting equation is:

$$\mathbf{pIC50} = -3.774 + 0.394 \times \mathbf{LogP} + 3.733 \times \mathbf{d} - 0.091 \times \mathbf{DM} + 15.001 \times \mathbf{E_{HOMO}} + 11.934 \times \mathbf{E_{LUMO}}$$

(Equation 1)

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

$$\mathbf{MSE = 0.072 \quad F = 24.731 P < 0.0001}$$

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.

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 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.

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)

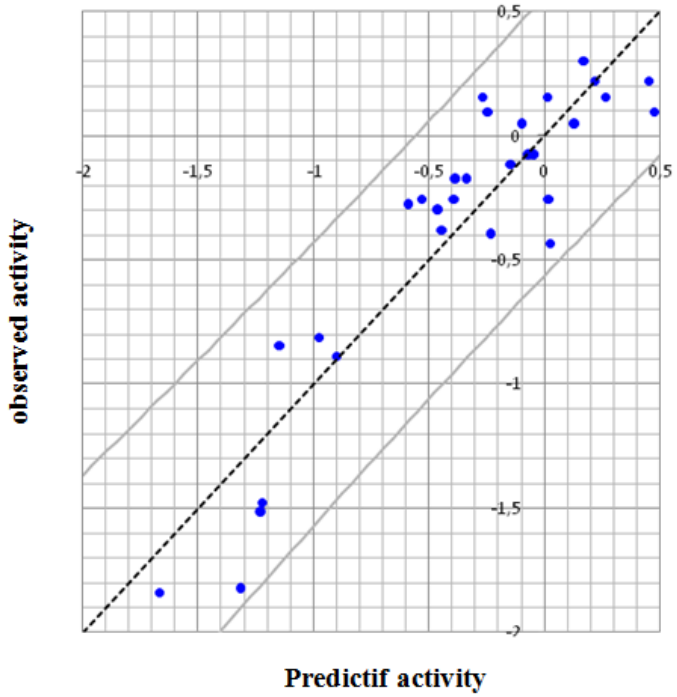


Figure 2: Graphical representation of calculated and observed activity calculated by MLR

3.4. 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:

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

(Equation 2)

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

N = 30 R=0.961 R^2 = 0.924 R^2_{cv} (R^2_{pcv} = 0.558; R^2_{ocv} =0.569) 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.

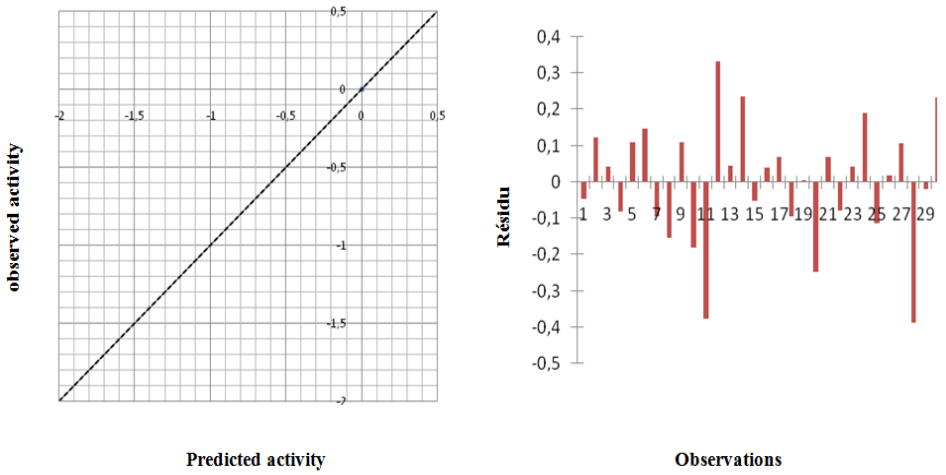


Figure 3: Graphical representation of calculated and observed activity and the residues value calculated by MNLR

3.5. 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.

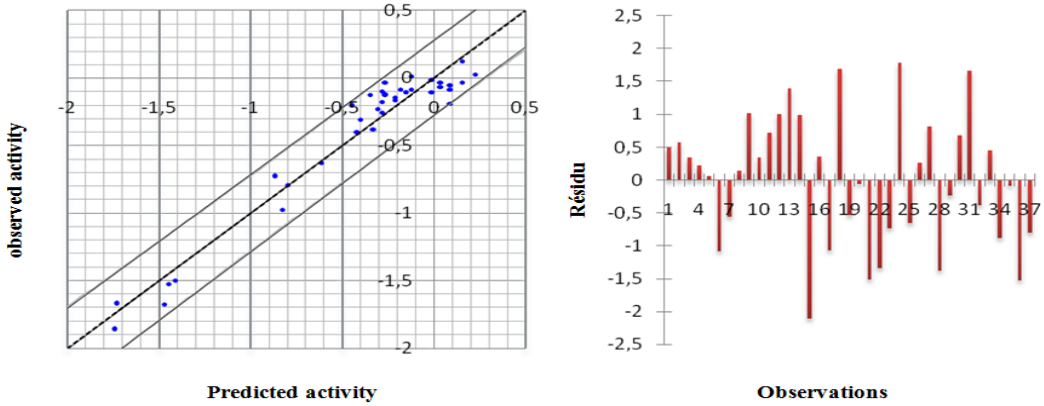


Figure 4: Graphical representation of calculated and observed activity and the residues values calculated by ANN

N = 37 R^2 = 0.971 R^2cv = 0. 911 **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).

3.6 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 remaining7 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^2_{cv}		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.

IV. 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	-0.130	-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	-0.900	-0.732	-0.730
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	-0.160	-0.108
17	0.155	0.015	-0.080	-0.052
18*	-0.230	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	-0.440	-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	-1.230	-0.626
36*	-1.544	-1.666	-2.012	-1.677
37	0.097	-0.242	-0.135	-0.069
		* test set		

Acknowledgements

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

References

[1]. R.E. Dolle. K.H. Nelson. Comprehensive survey of combinatorial library synthesis: _ 1998. J. Comb. Chem. 1(1999) 235–282.
[2]. R. G. Franzen. Recent advances in the preparation of heterocycles on solid support. J. Comb. Chem. 2(2000)195–214.
[3]. Dolle. R.E. Comprehensive survey of combinatorial library synthesis:_ 2000. J. Comb. Chem. 3(2001) 477–517.
[4]. S. Hanessian. G. McNaughton-Smith. H.G. Lombart.W.D.Lubell. Design and synthesis of conformationally constrained amino acids as versatile scaffolds and peptide mimetics. Tetrahedron .53(1997) 12789–12854.
[5]. Roberts.M.F.; Wink. M. Alkaloids: Biochemistry.Ecology.and Medicinal Applications; Plenum Press. New York. NY. USA.1998.
[6]. J.-L. Leow. R. Baron. P.J. Casey.M.-L. Go. Quantitative Structure-Activity Relationship (QSAR) of indoloacetamides as inhibitors of human isoprenylcysteine carboxyl methyltransferase.Bioorg. Med. Chem. Lett. 17 (2007) 1025–1032.
[7]. M.S. Buchanan. A.R. Carroll. G.A. Fechner. A. Boyle. M. Simpson. R. Addepalli. V.M. Avery. P.I. Forster. G.P. Guymer. T. Cheung. H. Chen. R.J. Quinn. Small-molecule inhibitors of the cancer target.isoprenylcysteine carboxyl methyltransferase.fromHoveaparvicalyx.Phytochem. 69 (2008) 1886–1889.
[8]. M. Wang. M.S. Hossain. W. Tan. B. Coolman. J. Zhou. S. Liu. P.J. Casey. Inhibition of isoprenylcysteinecarboxylmethyltransferase induces autophagic-dependent apoptosis and impairs tumor growth.Oncog. 29 (2010) 4959–4970.
[9]. J.A. Bergman. K. Hahne.C.A. Hrycyna. R.A. Gibbs. Lipid and sulfur substituted prenylcysteineanalogs as human Icmt inhibitors.Bioorg. Med. Chem. Lett. 21 (2011) 5616–5619
[10]. J.D. Majmudar. H.B. Hodges-Loaiza. K. Hahne. J.L. Donelson. J. Song. L. Shrestha. M.L. Harrison.C.A. Hrycyna. R.A. Gibbs. Amide-Modified Prenylcysteine based Icmt Inhibitors: Structure Activity Relationships.Kinetic Analysis and Cellular Characterization.Bioorg. Med. Chem. 20 (2012) 283–295.
[11]. M.S. Buchanan. A.R. Carroll. G.A. Fechner. A. Boyle. M.M. Simpson. R. Addepalli. V.M. Avery.N.A. Hooper John. N. Su. H. Chen. R. J. Quinn.the First Natural Product Inhibitor of Isoprenylcysteine Carboxyl Methyltransferase.a New Cancer Target.Bioorg. Med. Chem. Lett. 17 (2007) 6860–6863.
[12]. R.A. Baron. P.J. Casey.Analysis of the kinetic mechanism of recombinant human isoprenylcysteinecarboxylmethyltransferase (Icmt).Biochem. (2004) 5-19
[13]. M.S. Buchanan. A.R. Carroll. G.A. Fechner. A. Boyle. M. Simpson. R. Addepalli. V.M. Avery.N.A. Hooper John. T. Cheung. H. Chen. R.J. Quinna.Aplysamine 6.anAlkaloidal Inhibitor of Isoprenylcysteine Carboxyl Methyltransferase from the Sponge Pseudoceratina sp. J. Nat. Prod. 71 (2008) 1066–1067.
[14]. B. Thorisson. Inactivation of isoprenylcysteine carboxyl methyltransferase reduces proliferation of BRAF and NRAS mutated human melanoma cells. Master of Science Thesis.Chalmers University of Technology.Göteborg. Sweden. (2011).
[15]. H. Kubinyi. QSAR: hansch analysis and related approaches. In: Methods and Principles in Medicinal Chemistry. R. Mannhold. P. Krokgsgaard-Larsen. H. Timmerman (Eds.).Weinheim.VCH. (1993).
[16]. M.-L. Go. J.L. Leow. S.K. Gorla. A.P. Schuller. M. Wang. P.J. Casey. Amino Derivatives of IndoleAs Potent Inhibitors of Isoprenylcysteine Carboxyl Methyltransferase. J. Med. Chem. 53 (2010) 6838-6850.
[17]. M.J. Frisch & al. Gaussian 03.Revision B.01.Gaussian.Inc.. Pittsburgh. PA. (2003).
[18]. U. Sakar. R. Parthasarathi. V. Subramanian & P.K. Chattaraji.Toxicity analysis of polychlorinated dibenzofurans through global. J. Mol. Des. IECMD. (2004) 1-24.
[19]. Advanced Chemistry Development Inc.. Toronto. Canada. (2009). (www.acdlabs.com/resources/freeware/chemsketch/).
[20]. M. Larif. A. Adad. R. Hmammouchi.A.I. Taghki. A. Soulaymani.A. Elmidaoui. M. Bouachrine& T. Lakhlifi.Biological activities of triazine derivatives combining DFT and QSAR results.Arab. J. Chem. (2015).http://dx.doi.org/10.1016/j.arabjc.2012.12.033 (in press).
[21]. XLSTAT 2014 software (XLSTAT Company). http://www.xlstat.com.
[22]. S. Chtita. M. Larif. M. Ghamali. M. Bouachrine & T. Lakhlifi. DFT-based QSAR Studies of MK801 derivatives for non competitive antagonists of NMDA using electronic and topological descriptors.Journal of taibah university for chemistry 9(2) (2014) 143-154.
[23]. A. Adad. M. Larif.R. Hmamouchi. M. Bouachrine& T. Lakhlifi.Two different antibacterial activities against Staphylococcus aureus and Bacillus subtilis of 1.3-disubstituted-1H-naphtho[1.2-e][1.3]oxazine derivatives. Studies by combining DFT and QSAR results. J. Comp. Meth. Mol. Des. 4(3) (2014) 72-83.
[24]. Hmamouchi. R. Larif. M. Adad. A. Bouachrine. M. Lakhlifi.T. (2014). QSAR Modeling of the toxicity of pI50 Pyrazines Derived by Electronic Parameters Obtained by DFT. International Journal of Advanced Research in Computer Science and Software Engineering.Volume 4.Issue 2. 241- 251
[25]. D. Cherqaoui. D. Villemin.Use of a neural network to determine the boiling point of alkanes. J. Chem. Soc. Faraday. Trans. 90 (1994) 97-102.
[26]. J.A. Freeman. D.M. Skapura. Addition Wesley Publishing Company.Reading (1991).
[27]. A. Golbraikh. A. Tropsha. Beware of q2. J. Mol. Graphics Model. 20 (2002) 269-276.
[28]. A. Globarikh. A. Tropsha. Beware of q2! . J. Mol. Graphics. Model 20 (2002) 269–276.
[29]. STATITCF Software.Technical Institute of cereals and fodder.Paris. France. (1987).
[30]. A. Ousaa. B. Elidrissi. M. Ghamali. S. Chtita.M. Bouachrine& T. Lakhlifi. Acute toxicity of halogenated phenols: Combining DFT and QSAR studies. J. Comp. Meth. Mol. Des. 4(3) (2014) 10-18