Available online at **www.joac.info**

ISSN: 2278-1862

Journal of Applicable Chemistry

2019, 8 (3): 1201-1212

(International Peer Reviewed Journal)

Anti-Proliferative Activity of Heterocyclic Compounds Based on Indole: QSAR study

Abdelali Idrissi Taourati[1,3], Hanane Zaki[2,3], Mounir Ghamali[3], Samir Chtita[3], Assia Belhassan[3], Farahat Guenoun[1], Mohamed Belyass[2], Tahar Lakhlifi[3] and Mohammed Bouachrine[3]*

1. Chemistry Biology laboratory Applied to the Environment (CBAE), Faculty of Science, University Moulay Ismail, Meknes, **MOROCCO**
2. Biology, Environment and Health Laboratory, Faculty of Science and Technology Errachidia, University Moulay Ismail, **MOROCCO**
3. Natural Substances and Molecular Chemistry Laboratory (NSMC), Faculty of Science, University Moulay Ismail, Meknes, **MOROCCO**

MEM (LASMAR laboratory), Faculty of Science, University Moulay Ismail, Meknes, **MOROCCO**.

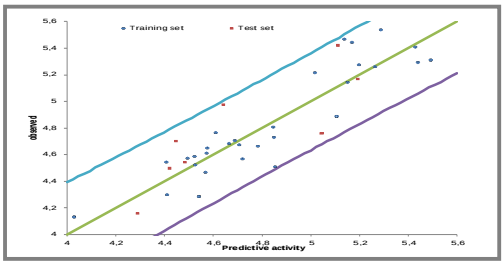
Email: bouachrine@gmail.com

Accepted on 5th April, 2019

ABSTRACT

In work, a series of thirty-eight Indole derivatives known for their antiproliferative activity were studied by QSAR methods (Quantitative-Structure-Activity-Relationship). Multiple Linear Regression (MLR) and Multiple Non-Linear Regression (MNLR) were used in model development. Refined models have been evaluated using the Leverage approach and the Inflation Factor (VIF). Some QSAR models have been subjected to numerous validation tests after their development. Given this, they gave notably encouraging outcomes upon validation (R=0.917, R^2_{cv}=0.672 for the MLR model and R=0.928, R^2_{cv}=0.721 for the MNLR model). The results show that the most important descriptors affecting the antiproliferative activity which are crucial within the layout of recent set of the indole derivatives are the low-unoccupied-molecular-orbital-energy (E_{LUMO}), the total-energy (E_T), the dipole-moment(μ), the partition-coefficient (log P) and refractive-index(n) descriptors. The influences of these descriptions on activity have been used to test and design new molecules with better activity compared to the predominant ones.

Graphical Abstract



Observed and calculated activity obtained by MLR method.

Keywords: Molecular docking, Anti-cancer, Anti-depressant, Anti-bacterial, Binding affinities.

INTRODUCTION

It is believed that cancer is the most serious disease which has long been affecting individuals in the whole world. According to the WHO (World Health Organization), no less than 15 million deaths are due to cancer every year around the globe that is almost sixty deaths every 120 min. Such a dreadful disease still influences millions of patients around the globe regardless of the highly sophisticated microscopic diagnostic methods or advanced imaging, or developed and chemo-therapeutic management [**1, 2**]. Apart from surgical treatment and irradiation techniques, chemotherapy still remains a vital option for cancer therapy.

The molecule of indole is a vital component of a number of either natural or the industrial molecules with important biological activity. It is considered among the most active heterocyclic molecules [**3-6**]. In fact, there has been large quantity of indole and derivatives as pharmacologically active molecules to develop the drug. These molecules also occur generally in many natural products such as those from vegetation [**7**].

These molecules have extensive categories of biological activities involving anticancer, antioxidant, antimicrobial, anticonvulsant, antidepressant, antileishmanial, anti-inflammatory activities and also possess anti-proliferative activity capabilities on cancer cell lines [**8-14**]. The speed and efficiency of drug discovery has shown the huge investments of leading pharmaceutical companies because of new computational techniques used that reduce the expense of each test biological and chemical synthesis of a compound. Being able to anticipate the biological activity of molecules concerning their constructive attributes, computational models are potent implements to highly devise active molecules. In this context, work on quantitative-structure-activity-relationships QSAR has been successfully carried out for the modelling of biological activities of natural and synthetic molecules [**15**]. The aim of this study is to develop QSAR models predictive of antiproliferative activity of indole and derivatives using statistical-tools; namely principal-components-analysis (PCA), multiple-linear-regression (MLR) and multiple-nonlinear-regression (MNLR). To test the stability and the performance of developed models we opt for the validation methods.

MATERIALS AND METHODS

Experimental data: In this work, we have chosen thirty-eight compounds based on Indole structure. These molecules have antiproliferative activity reported by Go *et al*. [**16**]. The activities IC_{50} (µM) (The half maximal inhibitory concentration) were converted into the corresponding pIC50 values (i.e. pIC_{50} is the negative logarithm of IC_{50} (pIC_{50}= $-\log_{10}(IC_{50})$). The derivatives of indole (Figure 1) represent the basic structure and table 1 shows the substituted forms of the studied compounds and experimental activities corresponding pIC_{50}. Additionally, the leave-one-out method was used to ensure the internal stability of the obtained models.

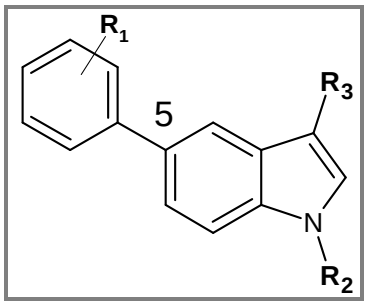
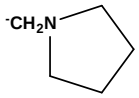
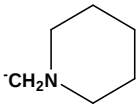
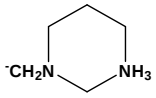
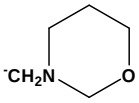
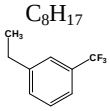
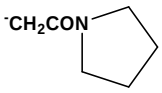
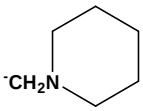
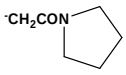
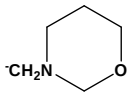
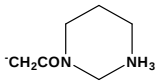


Figure 1. Chemical structure of the studied compounds based on Indole

Table 1. Observed activity of studied compounds based on Indole

S.No.	R1	R2	R3	IC_{50}	S.No.	R1	R2	R3	IC_{50}
1	m-CH_3	C_8H_{17}	CH_2CONH_2	21.8	20	m-CH_3	C_8H_{17}	$CH_2N(n\text{-}C_3H_7)_2$	13
2	0-CH_3	C_8H_{17}	CH_2CONH_2	21.2	21	m-CH_3	C_8H_{17}	$CH_2NCH_3(i\text{-}C_3H_7)$	5.3
3	m-OCH_3	C_8H_{17}	CH_2CONH_2	22.4	22	m-CH_3	C_8H_{17}		5.1
4	m-OC_2H_5	C_8H_{17}	CH_2CONH_2	18.7	23	m-CH_3	C_8H_{17}		4.9
5	H	C_8H_{17}	CH_2CONH_2	26	24	m-CH_3	C_8H_{17}		6.1
6	No5-phenyl	C_8H_{17}	CH_2CONH_2	50	25	m-CH_3	C_8H_{17}		24.5
7*	5-F	C_8H_{17}	CH_2CONH_2	70	26	m-CH_3	C_8H_{17}	$CH_2CH_2CONH_2$	20.7
8	m-CH_3		CH_2CONH_2	17.2	27	m-CH_3	C_8H_{17}	CH_2COOCH_3	30
9	m-CH_3	isoprenyl	CH_2CONH_2	28.5	28*	m-CH_3	C_8H_{17}	$CH_2NSO_2CH_3$	17.5
10*	m-CH_3	Geranyl	CH_2CONH_2	10.6	29	o- CH_3	n- C_8H_{17}	$CH_2N(C_2H_5)_2$	5.5
11	m-CH_3	C_8H_{17}	$CH_2CON(CH_3)_2$	31	30	P- CH_3	C_8H_{17}	$CH_2N(C_2H_5)_2$	3.4
12	m-CH_3	C_8H_{17}	$CH2CON(C_2H_5)_2$	27	31*	m- CH_3	isoprenyl	$CH_2N(C_2H_5)_2$	3.8
13	m-CH_3	C_8H_{17}		15.6	32	m-CH_3	isoprenyl		3.9
14*	m-CH_3	C_8H_{17}		28.9	33	m-CH_3	isoprenyl		34
15	m-CH_3	C_8H_{17}		19.7	34	No5-phenyl	isoprenyl	$-CH_2N(C_2H_5)_2$	74
16	m-CH_3	C_8H_{17}	$-CH_2NH_2$	2.9	35	5-F	n-C_8h_{17}	$-CH_2N(C_2H_5)_2$	7.2
17*	m-CH_3	C_8H_{17}	$CH_2N(CH_3)_2$	6.8	36*	5-F	isoprenyl	$-CH_2CON(C_2H_5)_2$	32
18	m-CH_3	C_8H_{17}	$CH_2N(C_2H_5)_2$	3.6	37	m-OCH_3	n-C_8h_{17}	$-CH_2CON(C_2H_5)_2$	52
19*	m-CH_3	C_8H_{17}	$CH_2N((i\text{-}C_3H_7)_2$	20.1	38	m-OC_2H_5	n-C_8h_{17}	$-CH_2CON(C_2H_5)_2$	26.8

Computational methods

Molecular descriptors: At current, several molecular descriptors can be used in QSAR studies. Once calculated, they are used to determine the predicted activity of studied compounds.

On the one hand, we used Gaussian 09W software to calculate electronic descriptors [17]. The geometries of the thirty-eight molecules of Indole derivatives have optimized with a 6-31G (d) based DFT (Density Functional Theory) technique. Then, many structural descriptors related to the results of quantum computation were selected: total-energy (ET), dipole-moment (μ), lowest-unoccupied-molecular-orbital-energy (ELUMO), highest-occupied molecular-orbital-energy(EHOMO), difference in absolute-value (ΔE), absolute-reactivity-index (ω), electronegativity (χ) and absolute-hardness (η) [18].

η, χ and ω were determined by the following equations:

$$\eta = \frac{E_{HOMO} - E_{LUMO}}{2} \; ; \; \chi = \frac{E_{HOMO} + E_{LUMO}}{2} \; ; \quad \omega = \frac{\chi^2}{2\eta}$$

Moreover, ACD/Chem Sketch program is employed to calculate the topological descriptors which are: Molar Refractivity (MR), Molar Volume (MV), Density (d), Parachor (Pc). Refractive Index (n), Polarisability (P) and Surface Tension (S) [19]. In addition, a Marvin Sketch program is used to compute the logarithm of the octanol/water partition coefficient of the test compound in its protonated state (log P).

Statistical analysis: The objective of 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 created model describing the relationship between this structure and the studied activity.

We used the Principal-Component-Analysis (PCA) [20] method to study the correlation between chemical descriptors using the software XLSTAT edition 2014 [21]. PCA seems to be an important statistical method for summarizing all the data encoded in the rubrics of the compounds [22]. It is clear to understand the distribution of the compounds. This is a pertinently descriptive statistical way that targets to present in graphic form the maximum of information contained in the data (Tables 1 and 2).

The multiple linear regression (MLR) analysis with descendent collection and elimination of variables is employed to create the model of structure-activity relationship. It is a mathematic way that limits difference between actual and predicted values. Also, it is maintained to select the descriptors used as the input parameters in the multiples nonlinear regression (MNLR).

We used the XLSTAT edition 2014 software to prepare the two models; MLR and MNLR, which are used to study the antiproliferative activity (pIC_{50}) by the determination of statistical parameters, such as: the correlation coefficient (R); the fisher's criterion (F); the mean squared error (MSE) and the significance level (P).

Validation of QSAR models: To validate the predicted QSAR models, two main factors are available: internal validation and external validation. The cross validation is one of the most popular method used for internal validation. Along with this, we tend to evaluate the model's internal predictive capability by Leave-one-out procedure (LOOCV). A good R^2_{cv} shows a considerable robustness and high internal predictive power of a QSAR model. Yet, recent studies show that the external validation remains the only way to maintain both the generalizability and the true predictive power of QSAR models for new chemicals [23]. For this reason, the statistical external validation was useful to the models as described by Globarikh and Tropsha [24].

RESULTS AND DISCUSSION

Data analysis: A QSAR study was shown for a series of the thirty-eight particles of indole derivatives, which have antiproliferative activity as demonstrated in table 1. It is to show a quantitative-relationship between chemical structure and studied activity. The result of sixteen descriptors is presented in table 2.

Principal component analysis: To identify the relationship between the diverse descriptors, we have conducted the method of Principal-Component-Analysis (PCA) [25]. Table 3 presents the correlations between the sixteen descriptors. This resulting table carries information on the high or low inter-relationship between the different descriptors. In general, a good co-linearity between variables is

observed when (r> 0.5). The lowest correlation is observed between MV and η (r=-0.030) and the highest correlation is observed between (P) and (MR) (r=1). Also, to reduce the redundancy present in our data, strongly correlated descriptors (r ≥ 0.9) are excluded.

Table 2. Values of chemical descriptors of the studied substituted Indole.

S.No.	PIC_5	MR	MV	PC	n	log *P*	S	d	P	μ	E_T	E_{HOM}	E_{LUM}	ΔE	η	χ	ω
1	4.66	116.4	352.4	887.1	1.575	6.389	40.10	1.060	46.17	1.704	-31497.418	-5.305	-0.493	4.812	-2.406	-2.899	-1.747
2	4.67	116.4	352.4	887.1	1.575	6.389	40.10	1.060	46.17	1.518	-31497.305	-5.345	-0.524	4.821	-2.410	-2.935	-1.787
3	4.65	117.8	358.8	906.3	1.57	5.718	40.60	1.090	46.72	2.230	-33545.241	-5.311	-0.499	4.812	-2.406	-2.905	-1.754
4	4.72	121.0	368.4	925.7	1.571	6.834	39.80	1.050	47.99	1.939	-32567.926	-5.302	-0.491	4.812	-2.406	-2.897	-1.744
5	4.58	112.0	337.2	856.0	1.578	5.876	41.50	1.070	44.41	1.765	-30426.792	-5.322	-0.516	4.805	-2.403	-2.919	-1.773
6	4.30	86.94	269.3	679.1	1.558	4.229	40.40	1.060	34.46	1.787	-24002.197	-5.186	-1.782	3.404	-1.702	-3.484	-3.567
7*	4.15	86.81	272.2	679.3	1.55	7.371	38.70	1.110	34.41	2.001	-26837.213	-5.472	-0.651	4.821	-2.411	-3.062	-1.944
8	4.76	118.8	354.7	893.8	1.584	6.178	40.20	1.230	47.11	3.433	-40543.359	-5.465	-0.647	4.818	-2.409	-3.056	-1.939
9	4.54	102.4	303.3	763.7	1.59	4.402	40.10	1.080	40.61	1.802	-28218.738	-5.570	-0.933	4.637	-2.318	-3.251	-2.280
10*	4.97	125.3	382.7	949.3	1.568	6.310	37.80	1.040	49.67	1.900	-33571.471	-5.335	-0.482	4.852	-2.426	-2.908	-1.743
11	4.50	127.3	396.8	976.2	1.554	6.480	36.60	1.010	50.48	2.917	-33638.007	-5.280	-0.411	4.869	-2.434	-2.845	-1.663
12	4.56	136.5	428.9	1053.	1.549	7.550	36.30	1.000	54.14	2.241	-35779.136	-5.263	-0.398	4.865	-2.432	-2.830	-1.647
13	4.80	134.3	402.1	1013.	1.583	7.240	40.30	1.070	53.27	3.205	-38323.404	-4.917	-0.791	4.126	-2.063	-2.854	-1.974
14*	4.53	139.0	418.1	1052.	1.579	7.687	40.10	1.060	55.10	3.588	-36816.661	-5.007	-0.739	4.268	-2.134	-2.873	-1.934
15	4.70	142.7	428.0	1078.	1.581	6.684	40.30	1.070	56.60	3.205	-38323.404	-4.917	-0.791	4.126	-2.063	-2.854	-1.974
16	5.53	111.0	339.1	842.8	1.568	6.684	38.10	1.020	44.00	2.124	-28410.575	-5.102	-0.254	4.848	-2.424	-2.678	-1.479
17*	5.16	121.8	383.5	932.0	1.548	7.500	34.80	0.980	48.32	2.454	-30551.365	-5.153	-0.297	4.856	-2.428	-2.725	-1.529
18	5.44	131.1	415.6	1009.	1.543	8.214	34.70	0.970	51.97	2.411	-32684.165	-5.161	-0.275	4.886	-2.443	-2.718	-1.512
19*	4.69	139.9	445.9	1071.	1.54	9.047	33.30	0.970	55.48	2.470	-34833.322	-5.158	-0.312	4.846	-2.423	-2.735	-1.543
20	4.88	140.3	447.7	1086.	1.539	9.259	34.60	0.960	55.63	2.025	-34825.692	-5.143	-0.292	4.851	-2.426	-2.718	-1.523
21	5.27	130.9	414.7	1001.	1.544	8.273	34.00	0.970	51.90	2.452	-32692.371	-5.137	-0.286	4.852	-2.426	-2.712	-1.516
22	5.29	128.9	388.8	969.4	1.577	7.906	38.60	1.030	51.11	2.453	-32659.528	-5.148	-0.324	4.824	-2.412	-2.736	-1.552
23	5.31	133.5	404.8	1008.	1.573	8.350	38.40	1.020	52.94	2.537	-33729.954	-4.511	-0.276	4.235	-2.117	-2.393	-1.353
24	5.21	137.3	414.7	1034.	1.576	7.347	38.70	1.040	54.43	2.409	-35236.644	-4.493	-0.442	4.050	-2.025	-2.468	-1.503
25	4.61	130.3	395.2	988.6	1.573	7.281	39.10	1.050	51.66	1.161	-34707.175	-4.760	-0.526	4.233	-2.117	-2.643	-1.650
26	4.68	121.0	368.4	925.7	1.571	6.834	39.80	1.050	47.99	3.210	-32567.806	-4.081	-1.294	2.787	-1.393	-2.688	-2.592
27	4.52	122.0	375.5	930.8	1.563	6.503	37.70	1.030	48.37	1.660	-32567.791	-5.322	-0.592	4.729	-2.365	-2.957	-1.849
28*	4.75	126.7	377.7	971.3	1.585	6.480	43.70	1.120	50.24	2.519	-44418.762	-3.980	0.270	4.251	-2.125	-1.855	-0.810
29	5.26	131.1	415.6	1009.	1.543	8.214	34.70	0.970	51.97	2.792	-32684.051	-5.201	-0.300	4.900	-2.450	-2.751	-1.544
30	5.46	131.1	415.6	1009.	1.543	8.214	34.70	0.970	51.97	2.292	-32684.166	-5.149	-0.271	4.878	-2.439	-2.710	-1.506
31*	5.42	117.1	366.5	885.8	1.551	6.473	34.10	0.980	46.42	2.717	-29440.132	-5.187	-0.322	4.865	-2.433	-2.754	-1.559
32	5.40	114.9	339.8	846.1	1.591	6.165	38.40	1.050	45.55	2.709	-29414.649	-5.181	-0.378	4.803	-2.401	-2.780	-1.609
33	4.46	116.3	346.2	865.2	1.586	5.540	38.90	1.080	46.11	1.475	-31462.296	-4.773	-0.580	4.193	-2.097	-2.677	-1.709
34	4.13	87.57	283.4	677.9	1.53	4.312	32.70	0.950	34.71	2.378	-22079.483	-5.172	-0.244	4.928	-2.464	-2.708	-1.488
35	5.14	101.4	335.4	801.4	1.516	6.196	32.50	0.990	40.22	4.355	-28025.089	-5.366	-0.469	4.897	-2.449	-2.917	-1.738
36*	4.49	87.44	286.3	678.0	1.522	4.455	31.40	1.000	34.66	4.494	-24781.076	-5.392	-0.517	4.875	-2.437	-2.955	-1.791
37	4.28	137.9	435.3	1072.	1.546	6.879	36.80	1.030	54.69	2.800	-37826.959	-5.273	-0.407	4.865	-2.433	-2.840	-1.658
38	4.57	142.5	451.3	1111.	1.544	7.236	36.70	1.020	56.52	2.654	-38897.622	-5.267	-0.399	4.867	-2.434	-2.833	-1.649

Multiple Linear Regressions (MLR): To develop a relationship with the indicator variable of antiproliferative activity, a big number of attempts with method lop down have been conducted. Therefore, using this method, we have obtained such a considerable link that is only one corresponding to the multiple-linear-regressions of five descriptors, which are as follows: the total-energy (ET), the energy (ELUMO), the dipole-moment (μ), the partition-coefficient (log P) and the refractive-index (n).

The resulting equation is:

$$\mathbf{pIC_{50}} = -11.783 + 0.00007\times\mathbf{E_T} + 0.703\times\mathbf{E_{LUMO}} + 0.299\times\boldsymbol{\mu} + 0.244\times\mathbf{log}\ \boldsymbol{P}_{+} 10.807\times\mathbf{n}$$

(Equation 1)

Table 3. Correlation between calculated descriptors of the studied substituted Indoles

Var	PIC_{50}	MR	MV	Pc	n	log *P*	S	P	M	E_T	E_{HOMO}	E_{LUMO}	ΔE	η	χ	ω
PIC_{50}	**1**															
MR	0.134	**1**														
MV	0.164	**0.977**	**1**													
PC	0.140	**0.994**	**0.990**	**1**												
n	-0.122	0.166	-0.045	0.083	**1**											
log *P*	0.122	0.548	0.610	0.567	-0.268	**1**										
S	-0.138	0.069	-0.119	0.024	0.882	-0.335	**1**									
P	0.134	**1.000**	**0.977**	**0.994**	0.166	0.548	0.069	**1**								
μ	-0.059	0.143	0.193	0.162	-0.261	0.125	-0.262	0.143	**1**							
E_T	0.084	-0.448	-0.420	-0.456	-0.119	-0.398	-0.189	-0.448	-0.201	**1**						
E_{HOMO}	0.100	0.320	0.247	0.300	0.318	0.148	0.325	0.320	0.104	-0.249	**1**					
E_{LUMO}	0.445	0.220	0.266	0.239	-0.197	0.238	-0.182	0.220	-0.041	-0.142	0.160	**1**				
ΔE	0.217	-0.123	-0.030	-0.093	-0.405	0.035	-0.400	-0.123	-0.116	0.115	-0.742	0.542	**1**			
η	-0.217	0.123	0.030	0.093	0.405	-0.035	0.400	0.123	0.116	-0.115	0.742	-0.542	**-1.000**	**1**		
χ	0.330	0.361	0.334	0.357	0.117	0.246	0.130	0.361	0.051	-0.263	0.820	0.696	-0.225	0.225	**1**	
ω	0.425	0.283	0.315	0.294	-0.140	0.282	-0.151	0.283	-0.059	-0.172	0.261	**0.983**	0.445	-0.445	0.760	**1**

Multiple Linear Regressions (MLR): To develop a relationship with the indicator variable of antiproliferative activity, a big number of attempts with method lop down have been conducted. Therefore, using this method, we have obtained such a considerable link that is only one corresponding to the multiple-linear-regressions of five descriptors, which are as follows: the total-energy (ET), the energy (ELUMO), the dipole-moment (μ), the partition-coefficient (log P) and the refractive-index (n).

The resulting equation is:

$$\mathbf{pIC_{50}} = -11.783 + 0.00007\times\mathbf{E_T} + 0.703\times\mathbf{E_{LUMO}} + 0.299\times\boldsymbol{\mu} + 0.244\times\mathbf{log\ P} + 10.807\times\mathbf{n}$$

(Equation 1)

N = 30; R=0.917; R^2 = 0.840; R^2_{cv} = 0.672; MSE = 0.03; F = 25.237; P <0.0001

As statistical results of model (MLR): N is the number of training compounds; R is the correlation coefficient; R^2 is the determination coefficient; R^2_{cv} is the coefficient of cross-validation; (MSE) is the mean-squared-error; F is the fisher's criterion and P is the significance level.

The results model are statistically good, the cross validation analysis was performed using leave-one-out (LOO) method [26]. The value of $R^2_{cv} > 0.5$ is enough for qualifying the model. The additional relationship coefficient that link variables in the model is the variance-inflation-factor (VIF) as exhibited in Table 4. These VIF were explained as $1/(1- R^2)$, where R was the multiple correlation coefficients for one autonomous variable against all the other descriptors in the model:

When VIF = 1, no inter-correlation exists among each variable;
When VIF ranges from 1 to 5, the correlation equation is satisfactory;
If VIF > 10, the regression equation is unstable and recheck is essential.

The other statistical index that allows determining the influence of each descriptor on the activity of the molecules studied is the t-test values. The t-test-values of six descriptors displayed in (Eq. 1) are 7.140, 3.838, 5.566, 6.289 and 5.290, respectively, are also listed in table 4.

Accordingly, the three important descriptors that influence the antiproliferative activity are as follows:

- The log P generally simulates organic compounds lipophilicity that seems tenable to the description of the organic compound's passage through membranes. The partition coefficient has a

positive sign in the model that results into the activity which can be increased with the lipophilicity of the Indole derivatives.

- The high total-energy E_T is generally associated with the high chemical stability of the compound. Therefore, the inhibitory activity varies with the total-energy E_T of the substituted Indole. Consequently, there is chemical stability against proliferative.

- The average charge separation in a molecular system is characterized by the moment-dipolar μ, which can signify the electronic information of the compounds. In addition, this descriptor can partially reflect the molecular-polarity. According to the result of this model, μ has favourable effect towards the activity value as evidenced by the positive regression coefficient. Increasing the value of μ makes the molecules of Indoles to participate in certain dipole-dipole or polar types of interaction with targets in cells, and leading to greater activity.

Table 4. VIF and t-test values of descriptors for MLR model

S. No.	E_T	E_{LUMO}	μ	log *P*	n
t-test value	**7.140**	3.838	**5.566**	**6.289**	5.290
VIF	1.387	1.707	1.281	2.001	1.497

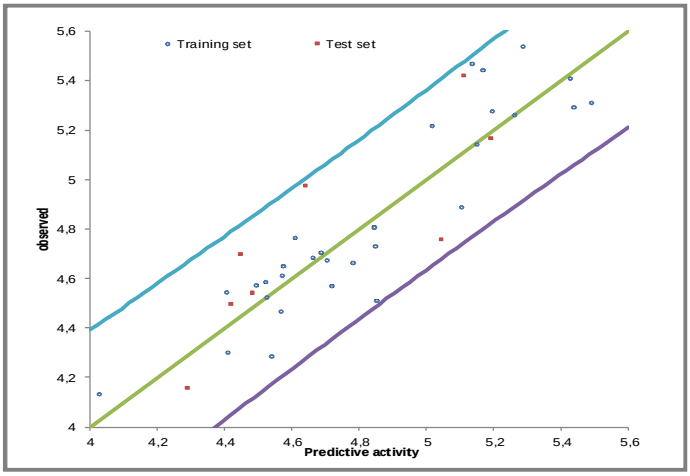


Figure 2. Observed and calculated activity obtained by MLR method.

Multiple nonlinear regressions (MNLR): Nonlinear regression model is used to develop the structure-activity relationship taking into consideration numerous factors. This is the most common device for the study of multidimensional data. We used the data matrix constituted obviously from the descriptors proposed by MLR model in relation to the 38 elements.

The resulting equation is:

$$\mathbf{pIC_{50}} = 281.842 + 0.00018\times\mathbf{E_T} + 2.057\times\mathbf{E_{LUMO}} + 0.242\times\boldsymbol{\mu} + 0.408\times\mathbf{log}\ \boldsymbol{P} - 364.320\times\mathbf{n} + 1.783.10^{-9}\times\mathbf{E^2_T} + 0.909\times\mathbf{E^2_{LUMO}} - 0.0021\times\boldsymbol{\mu}^2 - 0.013\times\mathbf{logP^2} + 120.564\times\mathbf{n^2}$$

(Equation 2)

The statistical indicators obtained are:

N = 30; R=0.928; R^2 = 0.863; r^2_{cv}=0. 721; MSE = 0. 032

The internal stability of the model expressed by Eq. 2 is provided by the Leave-one-out (LOO) method. The value of r^2_{cv} is greater than 0.500, which confirms the stability of the model [26]. Figure 3 demonstrates correlation between calculated and observed activities.

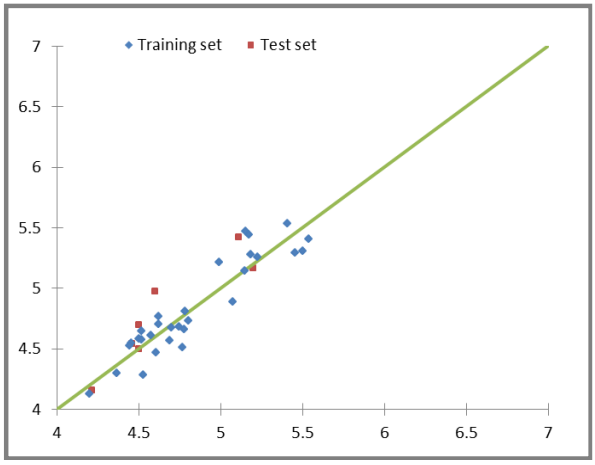


Figure 3. Observed and calculated activity by MNLR method
(The Y axis is the observed activity and the X axis is the predictive activity).

External validation: To estimate and to test the predictive power of the MLR and MNLR models, we must use a set of compounds that have not been used as the training set to establish the QSAR model. The two models established in the computation procedure using the thirty-substituted Indole are used to calculate the activity of the remaining eight compounds. Table 5 presents the main performance indicators of the two models.

Table 5. Performance comparison between models obtained by MLR and RNLM methods

Model	Training set				Test set	
	R	R^2	R^2_{cv}	MSE	R_{ext}	R^2_{ext}
MLR	0.917	0.840	0.672	0.030	0.839	0.704
MNLR	0.928	0.863	0.721	0.032	0.922	0850

The evaluation of best model regression equations established in this study. As a matter of fact of this result, a comparison of the considerable quality of the MLR and MNLR models shows that MNLR approach gives relatively, the better results than those of MLR method. As a general outlook, the two models tend to build up a considerable link between the molecular descriptors and the studied compounds' activity.

Domain of applicability: We used the domain of applicability to confirm the reliability of the QSAR model and its power to predict other molecules. The predicted compounds found in this area can be considered reliable. The Williams graph appears in figure 4 presents the domain of applicability (DA), in which standardized residuals and leverage-values (h_i) are plotted. It is based on the result of the leverage (h_i). For each compound, the QSAR model seems to calculate the following activity:

$$h_i = x_i(X^T X)^{-1} x_i^T \quad (i=1,\ldots n)$$

Here, X is the variable matrix resulted from the thirty training set variable values and xi is the row vector of the descriptor compounds i; the index T refers to the matrix/vector transposed. As a whole, the critical leverage h* is fixed at 3(k + 1)/N, where N is the number of training compounds, and k is

the number of model parameters. The prediction of the compound can be considered as unreliable if the leverage value h of a compound is higher than the critical value (h*) i.e., h > h*(26).

Figure 4 presents the Williams graph for the MLR model. The leverage values (hi) of any compound are below the critical value (h*=0.53) with the exception of compounds 6 and 5 as outliers. In addition, the standardized residues of all compounds in the training and test assemblies are less than three standard deviation units (± 3σ). Therefore, the applicability domain has confirmed the stability of the model, and the activity predicted by this model is reliable.

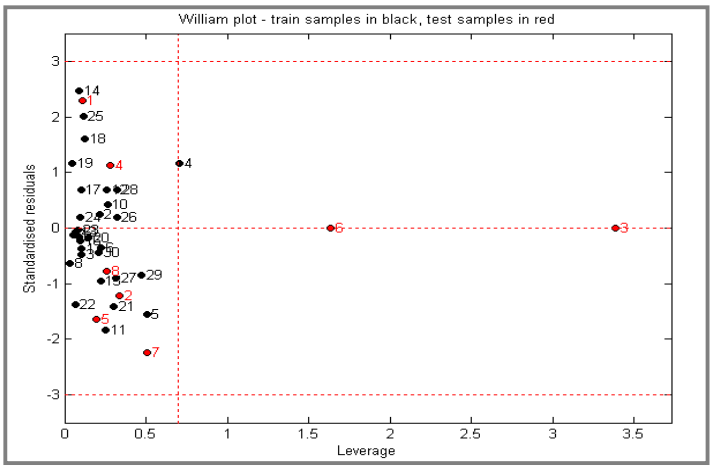


Figure 4. Williams plot for the training set and external validation for the antiproliferative activity of indole compounds, listed in Table 1 (h* = 0.53 and residual limits ±3).

Proposed compounds: As suggested, four Indoles molecules from the results of the MLR and MNLR models. Considerable activity values are obtained: 8.61 to 21.49 (µM) for MNLR model and 44.90 to 104.57 (µM) for MLR model, the best results given by the MNLR model. The values of the parameters shown by calculations presented in table 6 for the Indole compounds proposed from the results of two models Equations (1) and (2).

Table 6. Proposed compounds

S. No.	R1	R2	R3	E_T	E_{LUMO}	µ	log *P*	n	RLM		RNLM	
									PIC_{50}	IC_{50}	PIC_{50}	IC_{50}
1	m-CH_3	C_8H_{17}	$CCH_3COOHNH_2$	-26040.350	-1.653	2.911	4.210	1.559	3.98	104.57	4.67	21.49
2	m-CH_3	C_8H_{17}	$CC_2H_3COOHNH_2$	-27105.104	-1.741	2.877	4.650	1.572	4.08	82.83	4.96	11.01
3	m-CH_3	C_8H_{17}	$CC_3H_7COOHNH_2$	-28169.690	-1.698	2.772	5.170	1.551	3.91	123.93	4.69	20.48
4	m-CH_3	C_8H_{17}	$CC_4H_9COOHNH_2$	-28234.242	-1.698	3.756	5.610	1.555	4.35	**44.90**	5.07	**8.61**

APPLICATION

Quantitative structure-activity relationship (QSAR) study is important in silico methods to minimize the time and cost needed in drug designing process in comparison to traditional drug discovery methods. The aim of this work is to approach of validated QSAR modeling which can also be applied to the discovery of new Indole derivatives as antiproliferative activity in the field of pharmaceutical research to design and synthesize new molecules likely to become anticancer drugs.

CONCLUSION

Along with this research, the multiple-linear and multiple-nonlinear regressions have been introduced to build up a quantitative structure-propriety relation (QSAR) model of indole derivatives for the

antiproliferative-activity. A certain relationship between several descriptors and inhibition values pIC_{50} of several organic compounds based on substituted of Indole in satisfactory manners has been well established. The robustness of the two models constructed used in this study has good internal stabilities along big predictive powers, as has been assessed by the internal and external validations. The accuracy and predictability of the proposed models which are illustrated by the comparison of key statistical indicators like R or R^2 of different models obtained using different statistical tools and different descriptors are exhibited in table 5, to validate these results, we generate those, which are illustrated in table 7.

As conclusion, the most important finding from this research is that we have been able to design and propose new compounds with considerable values of the existing compounds (Table 6) by adding suitable substituents, shown by calculating their propriety and by using the regression equations. Hence, the proposed models will reduce the time and cost of synthesis as well as the determination of the antiproliferative activity of Indole derivatives.

Finally, as we tend to carry on this research in the future, we propose other molecules that have more significant activities.

Table 7. Observed and calculated values of pIC_{50} according to different methods

Comp No.	pIC_{50}	pIC_{50} (calc.)		Comp No.	pIC_{50}	pIC_{50} (calc.)	
	(obs.)	MLR	NMLR		(obs.)	MLR	NMLR
1	4.662	4.783	4.775	20	4.886	5.103	5.075
2	4.674	4.705	4.696	21	5.276	5.196	5.184
3	4.650	4.576	4.514	22	5.292	5.439	5.454
4	4.728	4.846	4.802	23	5.310	5.489	5.499
5	4.585	4.521	4.500	24	5.215	5.016	4.991
6	4.301	4.411	4.366	25	4.611	4.572	4.571
7*	4.155	4.292	4.214	26	4.684	4.661	4.747
8	4.764	4.611	4.619	27	4.523	4.524	4.444
9	4.545	4.406	4.451	28*	4.757	5.043	6.676
10*	4.975	4.641	4.597	29	5.260	5.263	5.222
11	4.509	4.851	4.765	30	5.469	5.134	5.152
12	4.569	4.717	4.685	31*	5.420	5.110	5.110
13	4.807	4.844	4.780	32	5.409	5.427	5.535
14*	4.539	4.483	4.459	33	4.469	4.567	4.604
15	4.706	4.687	4.622	34	4.131	4.026	4.195
16	5.538	5.288	5.408	35	5.143	5.149	5.144
17*	5.167	5.191	5.201	36*	4.495	4.422	4.502
18	5.444	5.167	5.173	37	4.284	4.540	4.525
19*	4.697	4.447	4.502	38	4.572	4.493	4.513

* test set

ACKNOWLEDGEMENTS

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

REFERENCES

[1]. M. Nakhjiri, M. Safavi, E. Alipour, S. Emami, A. F. Atash, M. Jafari-Zavareh, S. K. Ardestani, M. Khoshneviszadeh, A. Foroumadi, A. Shafiee, Asymmetrical 2,6-bis(benzylidene) cyclohexanones: Synthesis, cytotoxic activity and QSAR study, *J. Eur. Med. Chem.*, **2012**, 50, 113-123.

[2]. Y. L. Chen, S. Z. Lin, J. Y. Chang, Y. L. Cheng, N. M. Tsai, S. P. Chen, W. L. Chang, H. J. Harn, In vitro and in vivo studies of a novel potential anticancer agent of isochaihulactone on human lung cancer A549 cells. Biochem Pharmacol, *Bioch. Phar.*, 2006, 72, 308-319.

[3]. R. E. Dolle, K. H. Nelson, Comprehensive survey of combinatorial library synthesis, *J. Comb. Chem.*, 1999, 1, 235-282.
[4]. Atul H. Bapodra, Application of Functionalized Indole Derivatives as Kinase Inhibitor and Potential Anticancer Agents, *J. Applicable Chem.*, **2018**, 7(2), 299-308.
[5]. Vivekanand Biradar, Mahendra raj Karekal, Mruthyunjayaswamy Bennikallu Hire mathada, *J. Applicable Chem.*, **2014**, 3(1), 99-109.
[6]. 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*, **1997**, 53, 12789-12854.
[7]. M. Wink, M. F. Roberts, Alkaloids: Biochemistry, Ecology, and Medicinal Applications, Plenum Press, New York, **1998**.
[8]. S. Narsimha, N. S. Kumar, B. K. Swamy, N. V. Reddy, S. A. Hussain, M. S. Rao, Indole-2-carboxylic acid derived mono and bis 1,4 disubstituted 1,2,3-triazoles: characterization and evaluation of anticancer, antibacterial, and DNA-cleavage activities, *Bioorg. Med. Chem. Lett.*, **2016**, 26, 16391644.
[9]. S. Daly, K. Hayden, I. Malik, N. Porch, H. Tang, S. Rogelj, L. Frolova, K. Lepthien, A. Kornienko, I. V. Magedov, Unprecedented C-2 Arylation of Indole with Diazonium Salts: Syntheses of 2,3-Disubstituted Indoles and their Antimicrobial Activity, *Bioorg. Med. Chem. Lett.*, **2011**, 21, 4720-4723.
[10]. B. S. Holla, M. Mahalinga, M. S. Karthikeyan, B. Poojary, P. M. Akberali, N. S. Kumari, Synthesis, characterization and antimicrobial activity of some substituted 1,2,3-triazoles, *J. Eur. Med. Chem.*, **2005**, 40, 1173-1178.
[11]. M. F. Mady, G. E. Awad, K. B. Jørgensen, Ultrasound-assisted synthesis of novel 1,2,3-triazoles coupled diaryl sulfone moieties by the CuAAC reaction, and biological evaluation of them as antioxidant and antimicrobial agents, *J. Eur. Med. Chem.*, **2014**, 84, 433-443.
[12]. P. S. Rao, C. Kurumurthy, B. Veeraswamy, G. S. Kumar, Y. Poornachandra, C. G. Kumar, S. B. Vasamsetti, S. Kotamraju, B. Narsaiah, Synthesis of novel 1,2,3-triazole substituted-*N*-alkyl/aryl nitrone derivatives, their anti-inflammatory and anticancer activity, *J. Eur. Med. Chem.*, **2014**, 80, 184-191.
[13]. H. N. Dogan, A. Duran, S. Rollas, G. Sener, M. K. Uysal, D. Gülen, Synthesis of new 2,5-disubstituted-1,3,4-thiadiazoles and preliminary evaluation of anticonvulsant and antimicrobial activities, *Bioorg. Med. Chem.*, **2002**, 9, 2893-2898.
[14]. F. Poorrajab, S. K. Ardestani, S. Emami, M. Behrouzi-Fardmoghadam, A. Shafiee, A. Foroumadi, Nitroimidazolyl-1,3,4-thiadiazole-based anti-leishmanial agents: synthesis and in vitro biological evaluation, *J. Eur. Med. Chem.*, **2009**, 44, 1758-1762.
[15]. H. Timmerman, R. Mannhold, P. Krogsgaard-Larsen, Chemometric methods in molecular design, John Wiley and Sons, **2008**.
[16]. M. L. Go, J. L. Leow, S. K. Gorla, A. P. Schüller, M. Wang, P. J. Casey, Amino derivatives of indole as potent inhibitors of isoprenylcysteine carboxyl methyltransferase, *J. Eur. Med. Chem.*, **2010**, 53, 6838-6850.
[17]. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G.Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery, J.J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, and D.J. Fox, Gaussian 09: Gaussian, *Inc.*, *Wallingford CT*, **2009**.

[18]. U. Sarkar, J. Padmanabhan, R. Parthasarathi, V. Subramanian, P.K. Chattaraj, Toxicity analysis of polychlorinated dibenzofurans through global and local electrophilicities, *J. Mol. Str.*, **2006**, 758, 119-125.
[19]. Advanced Chemistry Development, Inc Toronto Canada, **2009**.
[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, *J. Arab. Chem.*, **2017**, 10, 946-955.
[21]. XLSTAT Company, **2014**.
[22]. A. Idrissi Taourati, M. Ghamali, S. Chtita, H. Zaki, M. Benlyass, F. Guenoun, T. Lakhlifi, M. Bouachrine, QSAR Studies of the Inhibitory Activity of a Series of Substituted Indole and Derivatives Againt Isoprenylcysteine Carboxyl Methyltransferase (Icmt). *Int. J. Phar. Sc. Inv.*, **2017**, 6(1), 6-13.
[23]. J. A. Freeman. D. M. Skapura, Addition Wesley Publishing Company, Reading, **1991**.
[24]. A. Golbraikh, A. Tropsha, Beware of q2, *J. Mol. Graph. Mod.*, **2002**, 20, 269-276.
[25]. STATITCF Software. Technical Institute of cereals and fodder, Paris, **1987**.
[26]. N. Frimayanti, M. L. Yam, H. B. Lee, R. Othman, S. M. Zain, N. A. Rahman, Validation of quantitative structure-activity relationship (QSAR) model for photosensitizer activity prediction. *J. Inter .Mol. Sci.*, **2011**, 12, 8626-8644.