QSAR study of isatin(1H-indole-2,3-dione) analogues as in vitro anti-cancer agents using the statistical analysis methods and the artificial neural network

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

Abstract— Isatin (1H-indole-2,3-dione) and its derivatives arepotent anticancer agents, these compounds inhibit cancer cells proliferation and tumorgrowth.A study of quantitative structure-activity relationship (QSAR) is applied to a set of 47 molecules derived from isatin, in order to predict the anticancer biological activity of the test compounds and find a correlation between the differentphysic-chemical parameters (descriptors) of these compounds and its biological activity, using multiple linear regression (MLR) and the artificial neural network (ANN).The topological and the electronic descriptors were computed, respectively, withACD/ChemSketchand (ChemDraw Ultra 8.0, ChemBioDraw Ultra 14.0) programs. A good correlation was found between the experimental activity and that obtained by MLR such as ($R = 0.94$ and $R^2 = 0.88$), this result could be improved with ANN such as($R = 0.97$ and$R^2 = 0.94$) with an architecture ANN (5-3-1). To test the performance of the neural network and the validity of our choice of descriptors selected by MLR and trained by ANN, we used cross-validation method (CV) such as ($R = 0.95$ and $R^2 = 0.90$) with the procedure leave-one-out (LOO).

Index Terms— Anti-cancer, Isatin derivatives, QSAR, MLR, ANN, CV

1-INTRODUCTION

At present, cancer is the main cause of diseases that cause the death of the human population in some areas of the world, and is expected to continue to be the leading cause of death in the coming years [1]. Chemotherapy, or the use of chemical agents to destroy cancer cells, is a mainstay in the treatment of malignant tumors. One of the main advantages of chemotherapy is its ability to treat widespread or metastatic cancer, whereas surgery and radiation therapies are limited to treatment of cancers for specific areas. Chemotherapy has generated much interest researchers and many ongoing efforts focused on the design and development of various anticancer drugs.

The isatin molecule (1H-indol-2,3-dione) is a polyvalent moiety that shows various biological activities[2-6],as anticancer activity, cytotoxic and antineoplasticactivites[7,8]. The N-alkylated indoles have also been reported as having anti-cancer activity.For example, the indolyl amide D-24851 has been found to be block cell cycle progression in a variety of malignant cell line including those derived from the prostate, brain, breast, pancreas and colon [9].

Quantitative structure-activity relationship (QSAR) tries to investigate the relationship between molecular descriptors that describe the unique physicochemical properties of the set of compounds of interest with their respective biological activity or chemical property [10,11].

In this work we attempt to establish a quantitative structure-activity relationship between anticancer activity of a series of 47 bioactive molecules derived from isatin and structural descriptors. Thus we can predict the anticancer activity of this group of organic compounds.

1Laboratory of Applied Chemistry, Faculty of Science and Technology, University Sidi Mohammed Ben Abdellah, P.O. Box 2202, Fez, Morocco
2ESTM, University Moulay Ismail, Meknes, Morocco

Corresponding Author: E-mail:* **khalil_fouad@yahoo.fr

Therefore we propose a quantitative model, and we try to interpret the activity of these compounds based on the different multivariate statistical analysis methods include:* The Multiple Linear Regression (MLR), which allows the selection of relevant descriptors used as input parameters for the artificial neural network (ANN). * The artificial neural network (ANN) which is a nonlinear method, which allows the prediction of the activities. * Cross-validation (CV) to validate models used with the process leave-one-out (LOO).

2-MATERIAL AND METHODS

2.1. Experimental data

The Biological data used in this study were anti-cancer activityagainst U937(inhibition of human monocyte, histiocytic lymphoma cells. (IC50)), a set of forty-seven derivatives of isatin.We have studied and analyzedthe series of isatin moleculeconsists of 47 selected derivatives that have been synthesized and evaluated for theiranticancer

activity in vitro against U937 (in terms of -log (IC_{50})) [12-14]. This in order to determine a quantitative structure-activity relationship between the anticancer activity and the structure of these molecules that are described by represented in Figure1.

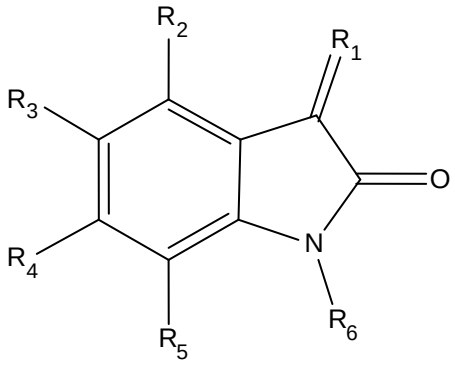

Figure1: The general structure of isatin (1H-indole-2,3-dione)

The chemical structures of 47 compounds of isatin used in this study and their experimental anti-cancer biological activity observed IC50(Cytotoxic concentration required to inhibit the growth of U937 than 50%)are collected from recent publications[12-14].The observations are converted into logarithmic scale -log (IC_{50})and are included in **Table1**.

2.2.Calculation of molecular descriptors

Advanced chemistry development's ACD/ChemSketch program was used to calculateMolar Volume (MV (cm^3)), Molecular Weight (MW), Molar Refractivity (MR (cm^3)), Parachor(Pc (cm^3)), Density (D (g/cm^3)), Refractive Index (n), Surface Tension (γ (dyne/cm)) andPolarizability (αe (cm^3))[15,16].

Steric, thermodynamic and electronic descriptors are calculated using ChemDraw Ultra 8.0 and ChemBioDraw Ultra 14.0[17,18]afteroptimization of the energy for each compound using the MM2 method (force field method with Gradient setting RootMean Square (RMS) 0.1 kcal mol-1) [19].

In this work 10 descriptors were chosen to describe the structure of the molecules constituting the series to study: the molecular weight (MW), the molar refractivity (MR (cm3)), the lipophilic (LogP), the highest occupied molecular orbital energy (E_{HOMO}(eV)), the lowest unoccupied molecular orbital energy (E_{LUMO}(eV)), the absolute ardness (η (eV)), the absolute electronegativity (χ (eV)),the repulsion energy (NRE (eV)), the hydrogen bond acceptor (HBA) and thehydrogen bond donor (HBD).
η and χ were determined by the following equations[20]:

$$\eta = (E_{LUMO} - E_{HOMO})/2 \quad \text{and}$$

$$\chi = -(E_{LUMO} + E_{HOMO})/2$$

their substituents R1, R2, R3, R4, R5 and R6.

The chemical structure of isatin (1H-indol-2,3-dione) is

2.3. Statistical analysis

To explain the structure-activity relationship, these 10 descriptors are calculated for 47 molecules (**Table2**) through software ChemSketch, ChemDraw Ultra 8.0 and ChemBioDraw Ultra 14.0.

The study we conducted consists of:

-The multiple linear regression (MLR) available in the SYSTAT 13 software [21].

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

The multiple linear regression statistic technique is used to study the relation between one dependent variable and several independent variables. It is a mathematic technique that minimizes differences between actual and predicted values. It has served also to select the descriptors used as the input parameters in the artificial neural network (ANN).

The (MLR) was generated to predict cytotoxiceffects IC_{50} activitiesof isatin derivatives.Equations were justified by the correlation coefficient (R), the Mean Squared Error (MSE), the Fishers F-statistic (F), and the significance level (F value) [22-24].

ANN is artificial systems simulating the function of the human brain. Three components constitute a neural network: the processing elements or nodes, the topology of the connections between the nodes, and the learning rule by which new information is encoded in the network. While there are a number of different ANN models, the most frequently used type of ANN in QSAR is the three-layered feed-forward network [25]. In this type of networks, the neurons are arranged in layers (an input layer, one hidden layer and an output layer). Each neuron in any layer is fully connected with the neurons of a succeeding layer and no connections are between neurons belonging to the same layer.

Cross-validation is a popular technique used to explore the reliability of statistical models. Based on this technique, a number of modified data sets are created by deleting in each case one or a small group of molecules, these procedures are named respectively “leave-one-out” and “leave-some-out” [26-28]. For each data set, an input-output model is developed. In this study we used, the leave-one-out (LOO) procedure.

Table1: Chemical structure and activity observed of isatin derivatives against U937

N°	R1	R2	R3	R4	R5	R6	ExperimentalpIC50aObs
1	O	H	Br	H	Br	H2CCH=CH2	5,18
2	O	H	Br	H	Br	H2CCH2OCH3	5,46
3	O	H	Br	H	Br	H2CCH2CH(CH3)2	5,62
4	O	H	Br	H	Br	H2CC6H5	5,94
5	O	H	Br	H	Br	H2CC6H4CH3b	6,31
6	O	H	Br	H	Br	H2CC6H4OCH3b	5,74
7	O	H	Br	H	Br	H2CC6H4OCH3c	5,75
8	O	H	Br	H	Br	H2CC6H4NO2b	6,05
9	O	H	Br	H	Br	H2CC6H4NO2d	5,64
10	O	H	Br	H	Br	H2CC6H4Clb	6,01
11	O	H	Br	H	Br	H2CC6H4Brb	6,20
12	O	H	Br	H	Br	H2CC6H4Ib	5,64
13	O	H	Br	H	Br	H2CC6H4CF3b	6,10
14	O	H	H	Br	H	H2CC6H4CF3b	5,28
15	O	H	Br	H	Br	H2CC6H4COOCH3b	5,92
16	O	H	Br	H	Br	H2CC6H4C(CH3)3b	5,95
17	O	H	Br	H	Br	H2CCH=CHC6H5	5,63
18	O	H	Br	H	Br	H2CC6H4C6H5b	6,12
19	O	H	H	H	H	H	3,25
20	O	Br	H	H	H	H	3,67
21	O	H	Br	H	H	H	4,19
22	O	H	H	Br	H	H	4,13
23	O	H	H	H	Br	H	4,08
24	O	H	F	H	H	H	4,01
25	O	H	I	H	H	H	4,27
26	O	H	NO2	H	H	H	3,88
27	O	H	OCH3	H	H	H	3,38
28	O	H	Br	H	Br	H	4,98
29	O	H	Br	Br	H	H	4,94
30	O	H	I	H	I	H	5,11
31	O	H	Br	H	NO2	H	3,59
32	O	H	Br	Br	Br	H	5,17
33	N-C6H5	H	H	H	H	H	4,12
34	N-C6H5	H	Br	H	Br	H	4,86
35	O	H	H	H	H	CH3	3,62
36	O	H	Br	H	Br	H2CCH2C6H5	6,11
37	O	H	Br	H	Br	H2CCH2C6H4Brc	6,11
38	O	H	Br	H	Br	H2CCH2C6H4Brb	6,06
39	O	H	Br	H	Br	H2CCH2C6H4OCH3c	5,97
40	O	H	Br	H	Br	H2CCH2C6H4OCH3b	5,63
41	O	H	Br	H	Br	CH2C10H7e	6,72
42	O	H	Br	H	Br	CH2C10H7f	6,13
43	O	H	Br	H	Br	CH2COC6H5	5,00
44	O	H	Br	H	H	CH2COC6H4Brc	5,20
45	O	H	Br	H	Br	CH2COC6H4Brb	5,04
46	O	H	Br	H	Br	CH2COC6H4OCH3c	5,33
47	O	H	Br	H	Br	CH2COC6H4OCH3b	5,27

a pIC50 = -log (IC50).
bSubstitutions at para position.
c Substitutions at meta position.
d Substitutions at ortho position.
e 1-naphthylmethyl.
f 2-naphthylmethyl.

Table2:The values of the 10 chemical descriptors

	MW	MR	LogP	EHOMO	ELUMO	η	χ	NRE	HBA	HBD
1	344,99000	67,99900	2,92600	-9,52065	-1,87192	3,824	5,696	12493,20000	2,00000	0,00000
2	363,00500	69,88000	2,07800	-9,45669	-1,94679	3,754	5,701	14572,50000	3,00000	0,00000
3	375,06000	77,25800	3,80500	-9,48114	-1,83318	3,823	5,657	16083,10000	2,00000	0,00000
4	395,05000	83,44900	3,96600	-9,38780	-1,87934	3,754	5,633	17947,40000	2,00000	0,00000
5	409,07700	88,49000	4,45300	-9,27097	-1,86221	3,704	5,566	19626,70000	2,00000	0,00000
6	425,07600	89,91200	3,84000	-9,36581	-1,94403	3,710	5,654	21423,80000	3,00000	0,00000
7	425,07600	89,91200	3,84000	-9,46207	-1,89278	3,784	5,677	21446,30000	3,00000	0,00000
8	441,05400	0,000000	2,87200	-9,71348	-2,17846	3,767	5,945	22692,10000	3,00000	1,00000
9	441,05400	0,000000	2,87200	-9,61100	-1,79782	3,906	5,704	23212,50000	3,00000	1,00000
10	429,49200	88,25400	4,52400	-9,48591	-1,97870	3,753	5,732	19486,70000	2,00000	0,00000
11	473,94600	91,07200	4,79500	-9,49694	-1,98373	3,756	5,740	19440,90000	2,00000	0,00000
12	520,94600	95,85700	5,32400	-9,51527	-1,97847	3,768	5,746	19382,30000	2,00000	0,00000
13	463,04800	89,42300	4,88700	-9,64969	-2,06174	3,793	5,855	25253,90000	5,00000	0,00000
14	384,15200	81,80000	4,05800	-9,51923	-1,99328	3,762	5,756	23251,50000	5,00000	0,00000
15	453,08600	94,97700	3,78600	-9,50834	-1,93164	3,788	5,719	24216,40000	3,00000	0,00000
16	451,15800	102,1150	5,67100	-9,25160	-1,83878	3,706	5,545	25237,90000	2,00000	0,00000
17	421,08800	93,76800	4,48200	-9,49410	-1,83538	3,829	5,664	20071,80000	2,00000	0,00000
18	471,14800	108,5850	5,64100	-8,85293	-1,87098	3,490	5,361	26081,70000	2,00000	0,00000
19	147,13300	38,69400	0,01600	-9,42574	-1,65145	3,887	5,538	6414,220000	2,00000	1,00000
20	226,02900	46,31700	1,16900	-9,58258	-1,79338	3,894	5,687	7528,970000	2,00000	1,00000
21	226,02900	46,31700	1,16900	-9,53783	-1,85593	3,840	5,696	7400,850000	2,00000	1,00000
22	226,02900	46,31700	1,16900	-9,63385	-1,85124	3,891	5,742	7393,620000	2,00000	1,00000
23	226,02900	46,31700	1,16900	-9,56071	-1,81863	3,871	5,689	7496,630000	2,00000	1,00000
24	165,12300	38,91100	0,49800	-9,61602	-1,91009	3,852	5,763	7539,100000	3,00000	1,00000
25	273,02900	51,10200	1,69700	-9,57914	-1,84884	3,865	5,713	7350,170000	2,00000	1,00000
26	193,13700	0,000000	-0,02400	-10,0261	-2,43532	3,795	6,230	9566,360000	3,00000	2,00000
27	177,15900	45,15700	0,21400	-9,39187	-1,73895	3,826	5,565	8877,580000	3,00000	1,00000
28	304,92500	53,94000	1,99800	-9,67216	-2,00843	3,831	5,840	8568,440000	2,00000	1,00000
29	304,92500	53,94000	1,99800	-9,69359	-2,02015	3,836	5,856	8517,500000	2,00000	1,00000
30	398,92500	63,51000	3,05500	-9,71980	-1,99390	3,862	5,856	8459,860000	2,00000	1,00000
31	272,03300	0,000000	0,86400	-9,98068	-2,36068	3,810	6,170	11202,20000	3,00000	2,00000
32	383,82100	61,56300	2,82700	-9,77524	-2,13602	3,819	5,955	9828,410000	2,00000	1,00000
33	222,24700	65,42600	2,46100	-8,81104	-1,01986	3,895	4,915	12850,00000	2,00000	1,00000
34	380,03900	80,67100	4,11900	-9,04455	-1,35481	3,844	5,199	15708,70000	2,00000	1,00000
35	161,16000	43,59100	0,58000	-9,15225	-1,59984	3,776	5,376	7730,140000	2,00000	0,00000
36	409,07700	88,20400	4,24600	-9,33509	-1,88291	3,726	5,609	18909,40000	2,00000	0,00000
37	487,97300	95,82700	5,07500	-9,47335	-1,94685	3,763	5,710	20401,00000	2,00000	0,00000
38	487,97300	95,82700	5,07500	-9,41377	-1,95928	3,727	5,686	20280,90000	2,00000	0,00000
39	439,10300	94,66700	4,12000	-9,39041	-1,92212	3,734	5,656	22407,10000	3,00000	0,00000
40	439,10300	94,66700	4,12000	-9,32384	-1,92889	3,697	5,626	22156,90000	3,00000	0,00000
41	445,11000	99,89900	4,96300	-8,63729	-1,84743	3,394	5,242	23829,20000	2,00000	0,00000
42	445,11000	99,89900	4,96300	-8,66361	-1,84208	3,410	5,252	23498,00000	2,00000	0,00000
43	423,06000	88,73800	3,15300	-9,49587	-1,93897	3,778	5,717	20399,50000	3,00000	0,00000
44	501,95600	96,36100	3,98100	-9,58504	-2,01542	3,784	5,800	21905,80000	3,00000	0,00000
45	501,95600	96,36100	3,98100	-9,59270	-2,00546	3,793	5,799	21808,10000	3,00000	0,00000
46	453,08600	95,20100	3,02600	-9,54714	-1,98876	3,779	5,767	23979,70000	4,00000	0,00000
47	453,08600	95,20100	3,02600	-9,53993	-1,98219	3,778	5,761	23738,10000	4,00000	0,00000

3-RESULTS AND DISCUSSION

3.1.Data set for analysis

The QSAR analysis was performed using the -log (IC_{50}) of the 47 selected molecules that have been synthesized and evaluated for their anticancer activity in vitro against U937 (experimental values) [12-14].The exploitation of experimental data observed by the use of mathematical and statistical tools is an effective method to find new chemical compounds with high anticancer activity.The values of the 10 chemical descriptors as shown in Table2.

The principle is to perform in the first time, a study of MLR based on the elimination of descriptors aberrant until a valid model (including the critical probability: **p-value < 0.05** for all descriptors and the model complete).In this study we worked only with 7 descriptors (**MW, MR, LogP, η, NRE, HBA and HBD**) among the 10 calculated.

3.2.Multiple Linear Regression (MLR)

In order to propose a mathematical model linking the descriptors and activity, and to evaluate quantitatively the substituent's physicochemical effects on the activity of the totality of the set of these 47 molecules,we presented the data matrix which is the corresponding physicochemical variables different substituent's from 47 molecules to a multiple linear regression analysis.This method used the coefficients R, R2, MSE and the F-values to select the best regression performance. Where R is the correlation coefficient; R2 is the coefficient of determination; MSE is the mean squared error; F is the Fisher F-statistic.

Treatment with multiple linear regression is more accurate because it allows you to connect the structural descriptors for each activity of 47 molecules to quantitatively evaluate the effect of substituent.The selected descriptors are:
MW, η, MR, HBD, and LogP.

The QSAR model built using multiple linear regression (MLR) method is represented by the following equation:

pIC_{50}MLR= 10,035 + 0,003MW-1,535η-0,013 MR -0,497 HBD+ 0,370 LogP

(Equation 1)

N= 47 , R = 0,940 , R^2=0,884 , F=42,588 , MSE= 0,113

Higher correlation coefficient and lower mean squared error (MSE) indicate that the model is more reliable. And the Fisher's F test is used. Given the fact that the probability corresponding to the F value is much smaller than **0.05**, it mean that we would be taking a lower than 0.01 % risk in assuming that the null hypothesis is wrong. Therefore, we can conclude with confidence that the model do bring a significant amount of information.
The elaborated QSAR model reveals that the anticancer activity could be explained by a number of electronic and topologic factors.The negative correlation of the Absolute Ardness (η), the Molar Refractivity (MR) and the Hydrogen Bond Donor (HBD) with the ability to displace the isatin activity reveals that a decrease in the value of pIC_{50},While the positive correlation of the descriptors (Molecular Weight (MW) and Lipophilic (LogP)) with the ability to displace the isatin activity reveals that an increase in the value of pIC_{50}.

With the optimal MLR model, the values of predicted activities pIC_{50} MLR calculated from equation1 and the observed values are given in Table3. The correlations of predicted and observed activities are illustrated in Figure2. The descriptors proposed in equation1 by MLR were, therefore, used as the input parameters in the artificial neural network (ANN).

The correlation between MLR calculated and experimental activities are very significant as illustrated in Figure2 and as indicated by R and R^2 values.

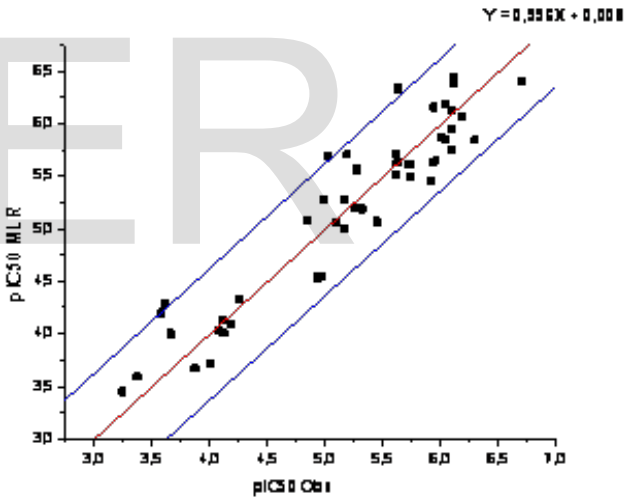

Figure2: Correlations of observed and predicted activities calculatedusing MLR

Table3: The observed, the predicted activities (pIC50), according to different methods MLR, ANN and CV for the 47 derivatives of isatin

N°	pIC50Obs	pIC50MLR	pIC50ANN	pIC50CV
1	5,18	5,27	5,20	5,26
2	5,46	5,06	5,42	5,38
3	5,62	5,51	5,62	5,83
4	5,94	5,63	5,75	5,87
5	6,31	5,84	5,93	6,07
6	5,74	5,61	5,72	5,68
7	5,75	5,49	5,60	5,66
8	6,05	5,85	5,98	6,03
9	5,64	5,62	5,54	5,86
10	6,01	5,86	5,90	6,04
11	6,20	6,06	5,94	5,88
12	5,64	6,33	6,01	5,29
13	6,10	5,94	5,93	6,06
14	5,28	5,56	5,79	5,09
15	5,92	5,45	5,52	5,93
16	5,95	6,15	6,09	5,98
17	5,63	5,62	5,78	5,28
18	6,12	6,44	6,14	6,08
19	3,25	3,44	3,06	3,85
20	3,67	4,00	3,89	3,93
21	4,19	4,08	4,25	3,92
22	4,13	4,00	3,92	4,15
23	4,08	4,03	4,10	3,94
24	4,01	3,72	3,91	4,07
25	4,27	4,33	4,45	3,83
26	3,88	3,67	3,73	3,69
27	3,38	3,59	3,50	3,57
28	4,98	4,53	4,72	4,34
29	4,94	4,53	4,70	4,29
30	5,11	5,06	5,30	5,05
31	3,59	4,20	3,90	3,08
32	5,17	4,99	5,28	5,15
33	4,12	4,13	3,86	4,08
34	4,86	5,08	5,08	4,35
35	3,62	4,28	3,88	3,97
36	6,11	5,74	5,85	6,04
37	6,11	6,12	5,97	6,03
38	6,06	6,18	6,00	5,88
39	5,97	5,64	5,77	6,07
40	5,63	5,70	5,81	5,83
41	6,72	6,40	6,53	5,96
42	6,13	6,38	6,35	5,95
43	5,00	5,28	5,23	5,06
44	5,20	5,70	5,58	5,79
45	5,04	5,69	5,56	5,28
46	5,33	5,18	5,05	5,15
47	5,27	5,19	5,05	5,17

3.3.Artificial Neural Networks (ANN)

In order to increase the probability of good characterization of studied compounds, artificial neural networks (ANN) can be used to generate predictive models of quantitative structure-activity relationships (QSAR) between a set of molecular descriptors obtained from the MLR,and observed activity.The ANN calculated activities model were developed using theproperties of several studied compounds. Some authors [29,30] have proposed a parameter ρ, leading to determine the number of hidden neurons, which plays a major role in determining the best ANN architecture defined as follows:

ρ = (Number of data points in the training set /Sum of the number of connections in the ANN)

In order to avoid over fitting or under fitting, it is recommended that **1.8 < ρ < 2.3**[31].The output layer represents the calculated activity values pIC_{50}. The architecture of the ANN used in this work (5-3-1), **ρ =2.13**.

The values of predicted activities pIC_{50} ANNcalculated using ANN and the observed values are given in Table3. The correlations of predicted and observed activities are illustrated in Figure3.
The correlation between ANN calculated and experimental activities are very significant as illustrated in Figure3 and as indicated by R and R^2 values.

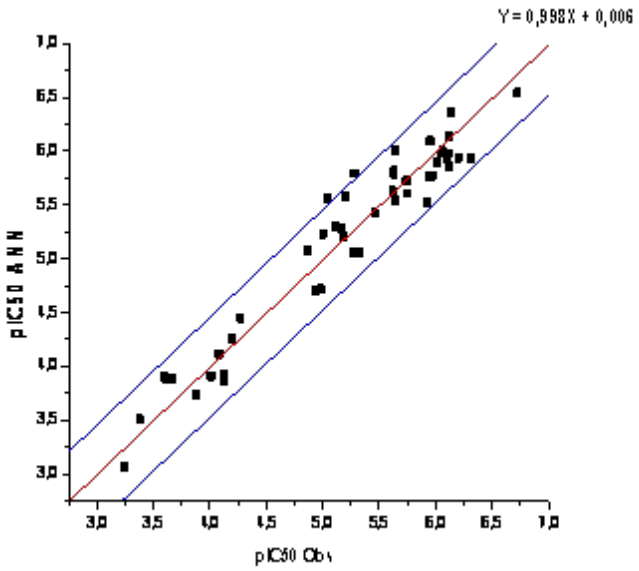

Figure3: Correlations of observed and predicted activities calculated using ANN

N= 47 , R = 0,97 , R^2=0,94

The obtained squared correlation coefficient (R^2) value confirms that the artificial neural network result were the best to build the quantitative structure activity relationship models.
It is important to be able to use ANN to predict the activity of new compounds.To evaluate the predictive ability of the ANN models, '**Leave-one-out**' is an approach particularly well adapted to the estimation of that ability.

3.4.Cross Validation (CV)

To test the performance of the neural network and the validity of our choice of descriptors selected by MLR and trained by ANN, we used cross-validation method (CV) with the procedure leave-one-out (LOO).In this procedure, one compound is removed from the data set, the network is trained with the remaining compounds and used to predict the discarded compound. The process is repeated in turn for each compound in the data set.
In this paper the 'leave-one-out' procedure was used to evaluate the predictive ability of the ANN.
The values of predicted activities pIC_{50} CVcalculated using CV and the observed values are given in Table3. The correlations of predicted and observed activities are illustrated in Figure4.
The correlation between CV calculated and experimental activities are very significant as illustrated in Figure4 and as indicated by R and R2 values.

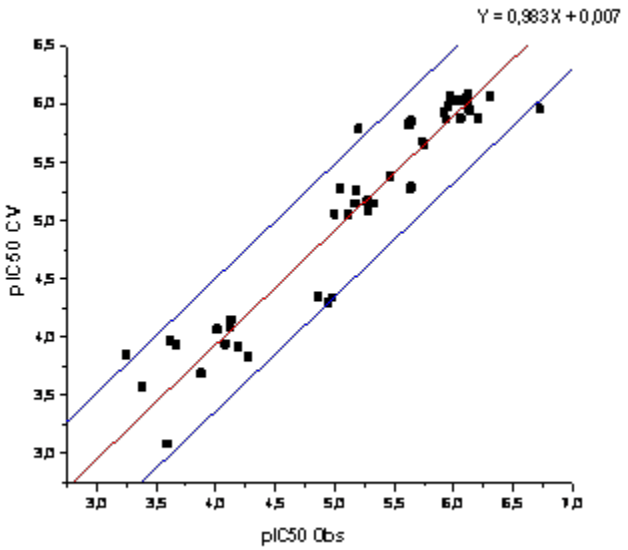

Figure4: Correlations of observed and predicted activities calculated using CV

N= 47 , R = 0,95 , R^2=0,90

The good results obtained with the cross validation, shows that the model proposed in this paper is able to predict activity with a great performance, and that the selected descriptors are pertinent.
The results obtained by MLR and ANN are very sufficient to conclude the performance of the model. Even if it is possible that this good prediction is found by chance we can claim that it is a positive result.So, this model could be applied to all derivatives of isatin accordingly to Table1 and could add further knowledge in the improvement of the search in the domain of anti-cancer agents.
A comparison of the quality of MLR and ANN models shows that the ANN models have substantially better predictive capability because the ANN approach gives better results than MLR.ANN was able to establish a satisfactory relationship between the molecular descriptors and the activity of the studied compounds. A good correlation was obtained with cross validation **Rcv= 0.95**. So the predictive power of this model is very significant.The results obtained in this study, showed that both models MLR and ANN are validated, which means that the prediction of the new compounds is feasible.

4-CONCLUSION

In this study, two different modelling methods, MLR and ANN were used in the construction of a QSAR model for the anti-cancer agents and the resulting models were compared.It was shown the artificial neural network ANN results have substantially better predictive capability than the MLR , yields a regression model with improved predictive power, we have established a relationship between several descriptors and the anticancer activity in satisfactory manners. The good results obtained with the cross validation CV, shows that the model proposed in this paper is able to predict activity with a great performance, and that the selected descriptors are pertinent.
The accuracy and predictability of the proposed models were illustrated by the comparison of key statistical terms like R or R^2 of different models obtained by using different statistical tools and different descriptors has been shown in Table3. It was also shown that the proposed methods are a useful aid for reduction of the time and cost of synthesis and activity determination of anti-cancer agents (compounds based on isatin derivatives).
Furthermore, we can conclude that studied descriptors, which are sufficiently rich in chemical, electronic and topological information to encode the structural featureand have a great influence on the activity may be used with other descriptors for the development of predictive QSAR models.
Previous studies QSAR already performed on the same set of isatin using multiple linear regression, obtained a correlation coefficient **(R = 0.92)** [32]. In this study the correlation coefficient obtained from the MLR **(R = 0.94)**, by using a variety of descriptors, is very important and this coefficient improved by using ANN **(R = 0.97)** so the proposed model is very significant and its performance is tested by cross-validation method CV **(R = 0.95).**

Thus, grace to QSAR studies, especially with the ANN that has allowed us to improve the correlation between the observed biological activity and that predicted, we can enjoy the performance of the predictive power of this model to explore and propose new molecules could be active.

ACKNOWLEDGMENT

We gratefully acknowledge to the group of Applied Chemistry Laboratory.

REFERENCES

[1] J. B. Gibbs. Science, (2000), 287, 1969-1973.
[2] S.N. Pandeya, S. Smitha, M. Jyoti, S.K. Sridhar, Acta Pharm. 55 (2005) 27–46.
[3] V.M. Sharma, P. Prasanna, V.A. Seshu, B. Renuka, V.L. Rao, G.S. Kumar, C.P. Narasimhulu, P.A. Babu, R.C. Puranik, D. Subramanyam, A. Venkateswarlu, S. Rajagopal, K.B.S. Kumar, C.S. Rao, N.V.S.R. Mamidi, D.S. Deevi, R. Ajaykumar, R. Rajagopalan, Bioorg. Med. Chem. Lett. 12 (2002) 2303–2307.
[4] M.J. Moon, S.K. Lee, J.W. Lee, W.K. Song, S.W. Kim, J.I. Kim, C. Cho, S.J. Choi, Y.C. Kim, Bioorg. Med. Chem. 14 (2006) 237–246.
[5] A.H. Abadi, S.M. Abou-Seri, D.E. Abdel-Rahman, C. Klein, O. Lozach, L. Meijer, Eur. J. Med. Chem. 41 (2006) 296–305.
[6] A. Gursoy, N. Karali, Eur. J. Med. Chem. 38 (2003) 633–643.
[7] A. Cane, M. C. Tournaire, D. Barritault, M. Crumeyrolle-Arias. Biochemistry and Biophysics Research Commuication, (2000), 276, 379-384.
[8] K. L. Vine, J. M. Locke, M. Ranson, K. Benkendorff, S. G. Pyne, J. B. Bremner. Bioorganic and Medicinal Chemistry, (2007), 15, 931-938.
[9] G. Bacher, B. Nickel, P. Emig, U. Vanhoefer, S. Seeber, A. Shandra, Klenner, T. Beckers, Cancer Research, (2001), 61, 392-399.
[10] C. Nantasenamat, C. Isarankura-Na-Ayudhya, T. Naenna& V. Prachayasittikul, A practical, overview of quantitative structure-activity relationship, J. Excli. 8 (2009) 74-88.
[11] C. Nantasenamat, C. Isarankura-Na-Ayudhya& V. Prachayasittikul, Advances in computational methods to predict the biological activity of compounds, J. Expert Opin. DrugDiscov. 5(7) (2010) 633-654.
[12] K.L. Vine, J.M. Locke, M. Ranson, K. Benkendorff, S.G. Pyne, J.B. Bremner, Bioorg., Med. Chem. 15 (2007) 931–938.
[13] K.L. Vine, J.M. Locke, M. Ranson, S.G. Pyne, J.B. Bremner, J. Med. Chem. 50 (2007) 5109–5117.
[14] L. Matesic, J.M. Locke, J.B. Bremner, S.G. Pyne, D. Skropeta, M. Ranson, K.L. Vine, Bioorg. Med. Chem. 16 (2008) 3118–3124.
[15] Advanced Chemistry Development Inc., Toronto, Canada. (2009). http://www.acdlabs.com/resources/freeware/chemsketch.
[16] ACD/ChemSketch Version 4.5 for Microsoft Windows User's Guide.
[17] ACD/Labs Extension for ChemDraw Version 8.0 for Microsoft Windows User's Guide.
[18] ACD/Labs Extension for ChemBioDraw Version 14.0 for Microsoft Windows User's Guide.
[19] A, N. L. Conformational Analysis 130. MM2. A Hydrocarbon Force Field Utilizing V1 andV2Torsional Terms, J. Am. Chem. Soc. Vol. 99, pp.8127-8134, (1977).
[20] U. Sakar, R. Parthasarathi, V. Subramanian & P.K. Chattaraji, Toxicity analysis of polychlorinated dibenzofurans through global, J. Mol. Des. IECMD, (2004) 1-24.
[21] SYSTAT 13 Add-in software (SYSTAT Company). www.systat.com.
[22] 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, Arabian Journal of Chemistry. (2013).
[23] M. Ghamali, S. Chtita, A. Adad, R. Hmamouchi, M. Bouachrine, T. Lakhlifi, Biological activity of molecules based on benzylpiperidine inhibitors of human acetylcholinesterase(HuAChE). Predicting by Combining DFT and QSAR calculations, International Journal of Advanced Research in Computer Science and Software Engineering. (2014).
[24] S. Chtita, R. Hmamouchi, M. Larif, M. Ghamali, M. Bouachrine, T. Lakhlifi, QSPR studies of 9-aniliioacridine derivatives for their DNA drug binding properties based on density functional theory using statistical methods: Model, validation and influencing factors, J. of Taibah Univ. for Sci. (2015),http://dx.doi.org/10.1016/j.jtusci.2015.04.007.
[25] V.J. Zupan& J. Gasteiger, Neural Networks for Chemists - An Introduction, VCH, Verlagsgesellschaft, Weinheim/VCH Publishers, New York. 106(12) (1993) 1367-1368.
[26] B. Efron, Estimating the error rates of a predictive rule: improvement on cross-validation, J. Am.Stat. Assoc. 78 (1983) 316-331.
[27] M.A. Efroymson, Multiple regression analysis, In Mathematical Methods for DigitalComputers, Ralston, A., Wilf, H.S., Eds,WileyNewYork, (1960).
[28] D.W. Osten, Selection of optimal regression models via cross-validation, J. Chemom. 2(1998) 39-48.
[29] S-S. So & W.G. Richards, Application of neural networks: quantitative structure-activity relationships of the derivatives of 2, 4 diamino (substituted-benzyl) pyrimidines as DHFR inhibitors, J. Med. Chem. 35 (1992) 3201-3207.
[30] T.A. Andrea & H. Kalayeh, Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors, J. Med. Chem. 34 (1991) 2824–2836.
[31] M. Elhallaoui, Modélisatrice moléculaire et étude QSAR d'antagonistes non compétitifs du récepteur NMDA par les méthodes statistiques et le réseau de neurones, Thesis of Doctorat, Fes,Morocco. (2002) 106.
[32] R.Sabet, M. Mohammadpoura, A. Sadeghi, Fassihi'' QSAR study of isatine analogues as in vitro anti-cancer agents''. (2010).