*Available online www.jocpr.com*

**Journal of Chemical and Pharmaceutical Research, 2016, 8(9):136-149**

**Research Article**

**ISSN : 0975-7384**
**CODEN(USA) : JCPRC5**

# Application of QSAR Methods on the Study of Bioactive Molecules Derived from Isatin

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

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

## ABSTRACT

*Isatin (1H-indole-2,3-dione) and its derivatives are potent anticancer agents, these compounds inhibit cancer cells proliferation and tumor growth. 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 different physic-chemical parameters (descriptors) of these compounds and its biological activity, using principal components analysis (PCA), multiple linear regression (MLR), multiple non-linear regression (MNLR) and the artificial neural network (ANN). We accordingly propose a quantitative model (non-linear and linear QSAR models), and we interpret the activity of the compounds relying on the multivariate statistical analysis. The topological and the electronic descriptors were computed, respectively, with ACD/ChemSketch and (ChemOffice 8.0; ChemBioOffice 14.0) programs. A good correlation was found between the experimental activity and those obtained by MLR and MNLR respectively such as (R = 0.94 and $R^2$ = 0.88) and (R = 0.96 and $R^2$ = 0.92), 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 MNLR and ANN, we used cross-validation method (CV) such as (R = 0.95 and $R^2$ = 0.90) with the procedure leave-one-out (LOO). This study show that the MLR and MNLR have served to predict activities, but when compared with the results given by an 5-3-1 ANN model we realized that the predictions fulfilled by this latter was more effective and much better than other models. The statistical results indicate that this model is statistically significant and shows very good stability towards data variation in leave-one-out (LOO) cross validation.*

**Keywords**: Anti-cancer; Isatin derivatives; QSAR; PCA; MLR; MNLR; ANN; CV

**Abbreviations**: *QSAR: Quantitative Structure-Activity Relationship; PCA: Principal Component Analysis; MLR: Multiple Linear Regression; MNLR: Multiple Non-Linear Regression; ANN: Artificial Neural Networks; CV: Cross Validation; LOO-CV: Leave One Out Cross-Validation; R: Correlation Coefficient; $R^2$: Coefficient of Determination; $R^2_{aj}$: Adjusted Coefficient of Determination; $q^2$: Coefficient of Prediction; SD: Standard Deviation; MW: Molecular Weight; MR: Molar Refractivity; LogP: Lipophilic; HOMO: Highest Occupied Molecular Orbital; LUMO: Lowest Unoccupied Molecular Orbital; $\eta$: Absolute Hardness; $\chi$: Absolute Electronegativity; NRE: Repulsion Energy; HBA: Hydrogen Bond Acceptor; HBD: Hydrogen Bond Donor; SSE: Sum of Residual (Error) Squares; SSF: Sum of Regression (Factor) Squares; SST: Sum of Total Squares; MSE ($V_E$: Error Variance): Mean Squared Error; MSF ($V_F$: Factor Variance): Mean Squared Factor; F: Fishers F-statistic; F value: Significance level; p-value: Critical Probability*

## INTRODUCTION

Previously, we have reported QSAR studies (MLR, ANN and CV) on a series of molecules derived from isatin such as anti-cancer inhibitors against U937 with electronic and topological descriptors. The MLR method was used to generate statistically significant QSAR models (Boukarai et al., 2015) [1]. Now, in continuation with our earlier work, we report QSAR studies on such a series of isatin with other statistical analysis methods: PCA, MNLR and ANOVA (Principal Components Analysis, Multiple Non-Linear Regression and ANalysis Of VAriance).The present work is an attempt to generate predictive models based on QSAR methods and to find the structural features of the isatins as inhibitor of U937 required for anti-cancer activities to guide the rational synthesis of novel compounds of isatin. QSAR field descriptors i.e. steric, thermodynamic and hydrophobic are useful for the better understanding of molecular modeling studies of this series of compounds in terms of ligand–receptor interactions. In this investigation, a widely used technique viz has been applied for descriptor optimization, and PCA, MLR, MNLR, ANN and CV analysis were applied for QSAR model development. The developed model provides insight into the influence of various interactive fields on the activity and, thus, can help in designing and forecasting the inhibitory activities of the isatins against U937.

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 [2]. 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 [3-7], as anticancer activity, cytotoxic and antineoplastic activites [8,9]. 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 [10].

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 [11,12].

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. Therefore we propose a quantitative model, and we try to interpret the activity of these compounds based on the different multivariate statistical analysis methods include: * The Principal Components Analysis (PCA) has served to classify the compounds according to their activities and to give an estimation of the values of the pertinent descriptors that govern this classification.

*The Multiple Linear Regression (MLR) has served to select the descriptors used as the input parameters for the Multiples Non-Linear Regression (MNLR) and Artificial Neural Network (ANN). * The artificial neural network (ANN) which is a nonlinear method, which allows the prediction of the activities. * Cross-validation (CV) to validate models used with the process leave-one-out (LOO).

## EXPERIMENTAL SECTION

The Biological data used in this study were anti-cancer activity against U937 (inhibition of human monocyte, histiocytic lymphoma cells. ($IC_{50}$)), a set of forty-seven derivatives of isatin. We have studied and analyzed the series of isatin molecule consists of 47 selected derivatives that have been synthesized and evaluated for their anticancer activity in vitro against U937 (in terms of -log ($IC_{50}$)) [13-15]. This in order to determine a quantitative structure-activity relationship between the anticancer activity and the structure of these molecules that are described by their substituents $R_1$, $R_2$, $R_3$, $R_4$, $R_5$ and $R_6$.

The chemical structure of isatin (1H-indol-2,3-dione) is represented in Figure1.

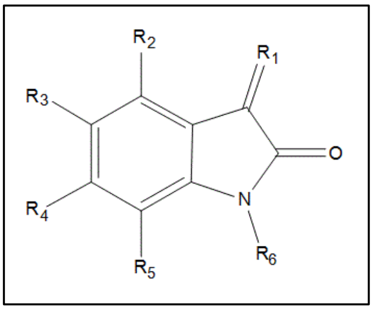

**Figure 1: The general structure of isatin (1H-indole-2,3-dione)**

The chemical structures of 47 compounds of isatin used in this study and their experimental anti-cancer biological activity observed $IC_{50}$ (Cytotoxic concentration required to inhibit the growth of U937 than 50%) are collected from recent publications [13-15]. The observations are converted into logarithmic scale -log ($IC_{50}$) in molar units (M) and are included in Table 1.

**Table 1: Chemical structure and activity observed of isatin derivatives against U937.**

| N° | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | $R_6$ | Experimental $pIC_{50}{}^{a}{}_{Obs}$ |
|---|---|---|---|---|---|---|---|
| **1** | O | H | Br | H | Br | $H_2CCH=CH_2$ | 5,18 |
| **2** | O | H | Br | H | Br | $H_2CCH_2OCH_3$ | 5,46 |
| **3** | O | H | Br | H | Br | $H_2CCH_2CH(CH_3)_2$ | 5,62 |
| **4** | O | H | Br | H | Br | $H_2CC_6H_5$ | 5,94 |
| **5** | O | H | Br | H | Br | $H_2CC_6H_4CH_3{}^{b}$ | 6,31 |
| **6** | O | H | Br | H | Br | $H_2CC_6H_4OCH_3{}^{b}$ | 5,74 |
| **7** | O | H | Br | H | Br | $H_2CC_6H_4OCH_3{}^{c}$ | 5,75 |
| **8** | O | H | Br | H | Br | $H_2CC_6H_4NO_2{}^{b}$ | 6,05 |
| **9** | O | H | Br | H | Br | $H_2CC_6H_4NO_2{}^{d}$ | 5,64 |
| **10** | O | H | Br | H | Br | $H_2CC_6H_4Cl^{b}$ | 6,01 |
| **11** | O | H | Br | H | Br | $H_2CC_6H_4Br^{b}$ | 6,20 |
| **12** | O | H | Br | H | Br | $H_2CC_6H_4I^{b}$ | 5,64 |
| **13** | O | H | Br | H | Br | $H_2CC_6H_4CF_3{}^{b}$ | 6,10 |
| **14** | O | H | H | Br | H | $H_2CC_6H_4CF_3{}^{b}$ | 5,28 |
| **15** | O | H | Br | H | Br | $H_2CC_6H_4COOCH_3{}^{b}$ | 5,92 |
| **16** | O | H | Br | H | Br | $H_2CC_6H_4C(CH_3)_3{}^{b}$ | 5,95 |
| **17** | O | H | Br | H | Br | $H_2CCH=CHC_6H_5$ | 5,63 |
| **18** | O | H | Br | H | Br | $H_2CC_6H_4C_6H_5{}^{b}$ | 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 | $NO_2$ | H | H | H | 3,88 |
| **27** | O | H | $OCH_3$ | 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 | $NO_2$ | H | 3,59 |
| **32** | O | H | Br | Br | Br | H | 5,17 |
| **33** | $N-C_6H_5$ | H | H | H | H | H | 4,12 |
| **34** | $N-C_6H_5$ | H | Br | H | Br | H | 4,86 |
| **35** | O | H | H | H | H | $CH_3$ | 3,62 |
| **36** | O | H | Br | H | Br | $H_2CCH_2C_6H_5$ | 6,11 |
| **37** | O | H | Br | H | Br | $H_2CCH_2C_6H_4Br^{c}$ | 6,11 |
| **38** | O | H | Br | H | Br | $H_2CCH_2C_6H_4Br^{b}$ | 6,06 |
| **39** | O | H | Br | H | Br | $H_2CCH_2C_6H_4OCH_3{}^{c}$ | 5,97 |
| **40** | O | H | Br | H | Br | $H_2CCH_2C_6H_4OCH_3{}^{b}$ | 5,63 |
| **41** | O | H | Br | H | Br | $CH_2C_{10}H_7{}^{e}$ | 6,72 |
| **42** | O | H | Br | H | Br | $CH_2C_{10}H_7{}^{f}$ | 6,13 |
| **43** | O | H | Br | H | Br | $CH_2COC_6H_5$ | 5,00 |
| **44** | O | H | Br | H | H | $CH_2COC_6H_4Br^{c}$ | 5,20 |
| **45** | O | H | Br | H | Br | $CH_2COC_6H_4Br^{b}$ | 5,04 |
| **46** | O | H | Br | H | Br | $CH_2COC_6H_4OCH_3{}^{c}$ | 5,33 |
| **47** | O | H | Br | H | Br | $CH_2COC_6H_4OCH_3{}^{b}$ | 5,27 |

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

[b] Substitutions at para position.

[c] Substitutions at meta position.

[d] Substitutions at ortho position.

[e] 1-naphthylmethyl.

[f] 2-naphthylmethyl.

## Calculation of molecular descriptors

Advanced chemistry development's ACD/ChemSketch program was used to calculate Molar 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)) and Polarizability ($\alpha_e$ ($cm^3$)) [16,17].

Steric, thermodynamic and electronic descriptors are calculated using ChemOffice 8.0 and ChemBioOffice 14.0 [18,19] after optimization of the energy for each compound using the MM2 method (force field method with Gradient setting Root Mean Square (RMS) 0.1 kcal $mol^{-1}$) [20].

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 ($cm^3$)), the lipophilic (LogP), the highest occupied molecular orbital energy ($E_{HOMO}$ (eV)), the lowest unoccupied molecular orbital energy ($E_{LUMO}$ (eV)), the absolute ardness ($\eta$ (eV)), the absolute electronegativity ($\chi$ (eV)), the repulsion energy (NRE (eV)), the hydrogen bond acceptor (HBA) and the hydrogen bond donor (HBD).

$\eta$ and $\chi$ were determined by the following equations [21]:

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

**Statistical analysis**

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

**Table 2: The values of the 10 chemical descriptors**

| | MW | MR | LogP | $E_{HOMO}$ | $E_{LUMO}$ | η | χ | NRE | HBA | HBD |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 3,44,990 | 67,999 | 2,926 | -9,520 | -1,871 | 3,824 | 5,696 | 1,24,93,200 | 2 | 0 |
| **2** | 3,63,005 | 69,880 | 2,078 | -9,456 | -1,946 | 3,754 | 5,701 | 1,45,72,500 | 3 | 0 |
| **3** | 3,75,060 | 77,258 | 3,805 | -9,481 | -1,833 | 3,823 | 5,657 | 1,60,83,100 | 2 | 0 |
| **4** | 3,95,050 | 83,449 | 3,966 | -9,387 | -1,879 | 3,754 | 5,633 | 1,79,47,400 | 2 | 0 |
| **5** | 4,09,077 | 88,490 | 4,453 | -9,270 | -1,862 | 3,704 | 5,566 | 1,96,26,700 | 2 | 0 |
| **6** | 4,25,076 | 89,912 | 3,840 | -9,365 | -1,944 | 3,710 | 5,654 | 2,14,23,800 | 3 | 0 |
| **7** | 4,25,076 | 89,912 | 3,840 | -9,462 | -1,892 | 3,784 | 5,677 | 2,14,46,300 | 3 | 0 |
| **8** | 4,41,054 | 0,0000 | 2,872 | -9,713 | -2,178 | 3,767 | 5,945 | 2,26,92,100 | 3 | 1 |
| **9** | 4,41,054 | 0,0000 | 2,872 | -9,611 | -1,797 | 3,906 | 5,704 | 2,32,12,500 | 3 | 1 |
| **10** | 4,29,492 | 88,254 | 4,524 | -9,485 | -1,978 | 3,753 | 5,732 | 1,94,86,700 | 2 | 0 |
| **11** | 4,73,946 | 91,072 | 4,795 | -9,496 | -1,983 | 3,756 | 5,740 | 1,94,40,900 | 2 | 0 |
| **12** | 5,20,946 | 95,857 | 5,324 | -9,515 | -1,978 | 3,768 | 5,746 | 1,93,82,300 | 2 | 0 |
| **13** | 4,63,048 | 89,423 | 4,887 | -9,649 | -2,061 | 3,793 | 5,855 | 2,52,53,900 | 5 | 0 |
| **14** | 3,84,152 | 81,800 | 4,058 | -9,519 | -1,993 | 3,762 | 5,756 | 2,32,51,500 | 5 | 0 |
| **15** | 4,53,086 | 94,977 | 3,786 | -9,508 | -1,931 | 3,788 | 5,719 | 2,42,16,400 | 3 | 0 |
| **16** | 4,51,158 | 102,11 | 5,671 | -9,251 | -1,838 | 3,706 | 5,545 | 2,52,37,900 | 2 | 0 |
| **17** | 4,21,088 | 93,768 | 4,482 | -9,494 | -1,835 | 3,829 | 5,664 | 2,00,71,800 | 2 | 0 |
| **18** | 4,71,148 | 108,58 | 5,641 | -8,852 | -1,870 | 3,490 | 5,361 | 2,60,81,700 | 2 | 0 |
| **19** | 1,47,133 | 38,694 | 0,016 | -9,425 | -1,651 | 3,887 | 5,538 | 6,41,42,200 | 2 | 1 |
| **20** | 2,26,029 | 46,317 | 1,169 | -9,582 | -1,793 | 3,894 | 5,687 | 7,52,89,700 | 2 | 1 |
| **21** | 2,26,029 | 46,317 | 1,169 | -9,537 | -1,855 | 3,840 | 5,696 | 7,40,08,500 | 2 | 1 |
| **22** | 2,26,029 | 46,317 | 1,169 | -9,633 | -1,851 | 3,891 | 5,742 | 7,39,36,200 | 2 | 1 |
| **23** | 2,26,029 | 46,317 | 1,169 | -9,560 | -1,818 | 3,871 | 5,689 | 7,49,66,300 | 2 | 1 |
| **24** | 1,65,123 | 38,911 | 0,498 | -9,616 | -1,910 | 3,852 | 5,763 | 7,53,91,000 | 3 | 1 |
| **25** | 2,73,029 | 51,102 | 1,697 | -9,579 | -1,848 | 3,865 | 5,713 | 7,35,01,700 | 2 | 1 |
| **26** | 1,93,137 | 0,0000 | -0,024 | -10,02 | -2,435 | 3,795 | 6,230 | 9,56,63,600 | 3 | 2 |
| **27** | 1,77,159 | 45,157 | 0,214 | -9,391 | -1,738 | 3,826 | 5,565 | 8,87,75,800 | 3 | 1 |
| **28** | 3,04,925 | 53,940 | 1,998 | -9,672 | -2,008 | 3,831 | 5,840 | 8,56,84,400 | 2 | 1 |
| **29** | 3,04,925 | 53,940 | 1,998 | -9,693 | -2,020 | 3,836 | 5,856 | 8,51,75,000 | 2 | 1 |
| **30** | 3,98,925 | 63,510 | 3,055 | -9,719 | -1,993 | 3,862 | 5,856 | 8,45,98,600 | 2 | 1 |
| **31** | 2,72,033 | 0,0000 | 0,864 | -9,980 | -2,360 | 3,810 | 6,170 | 1,12,02,200 | 3 | 2 |
| **32** | 3,83,821 | 61,563 | 2,827 | -9,775 | -2,136 | 3,819 | 5,955 | 9,82,84,100 | 2 | 1 |
| **33** | 2,22,247 | 65,426 | 2,461 | -8,811 | -1,019 | 3,895 | 4,915 | 1,28,50,000 | 2 | 1 |
| **34** | 3,80,039 | 80,671 | 4,119 | -9,044 | -1,354 | 3,844 | 5,199 | 1,57,08,700 | 2 | 1 |
| **35** | 1,61,160 | 43,591 | 0,580 | -9,152 | -1,599 | 3,776 | 5,376 | 7,73,01,400 | 2 | 0 |
| **36** | 4,09,077 | 88,204 | 4,246 | -9,335 | -1,882 | 3,726 | 5,609 | 1,89,09,400 | 2 | 0 |
| **37** | 4,87,973 | 95,827 | 5,075 | -9,473 | -1,946 | 3,763 | 5,710 | 2,04,01,000 | 2 | 0 |
| **38** | 4,87,973 | 95,827 | 5,075 | -9,413 | -1,959 | 3,727 | 5,686 | 2,02,80,900 | 2 | 0 |
| **39** | 4,39,103 | 94,667 | 4,120 | -9,390 | -1,922 | 3,734 | 5,656 | 2,24,07,100 | 3 | 0 |
| **40** | 4,39,103 | 94,667 | 4,120 | -9,323 | -1,928 | 3,697 | 5,626 | 2,21,56,900 | 3 | 0 |
| **41** | 4,45,110 | 99,899 | 4,963 | -8,637 | -1,847 | 3,394 | 5,242 | 2,38,29,200 | 2 | 0 |
| **42** | 4,45,110 | 99,899 | 4,963 | -8,663 | -1,842 | 3,410 | 5,252 | 2,34,98,000 | 2 | 0 |
| **43** | 4,23,060 | 88,738 | 3,153 | -9,495 | -1,938 | 3,778 | 5,717 | 2,03,99,500 | 3 | 0 |
| **44** | 5,01,956 | 96,361 | 3,981 | -9,585 | -2,015 | 3,784 | 5,800 | 2,19,05,800 | 3 | 0 |
| **45** | 5,01,956 | 96,361 | 3,981 | -9,592 | -2,005 | 3,793 | 5,799 | 2,18,08,100 | 3 | 0 |
| **46** | 4,53,086 | 95,201 | 3,026 | -9,547 | -1,988 | 3,779 | 5,767 | 2,39,79,700 | 4 | 0 |
| **47** | 4,53,086 | 95,201 | 3,026 | -9,539 | -1,982 | 3,778 | 5,761 | 2,37,38,100 | 4 | 0 |

The study we conducted consists of:
-The principal component analysis (PCA), the multiple linear regressions (MLR), and the non-linear regression (MNLR) available in the XLSTAT and SYSTAT softwares [22,23].
-The Artificial Neural Network (ANN) and the leave-one-out cross validation (CV-LOO) are done on Matlab 7 using a program written in C language.
The structures of the molecules based on isatin derivatives were studied by statistical methods based on the principal component analysis (PCA). PCA is a statistical technique useful for summarizing all the information encoded in the structures of the compounds. It is also very helpful for understanding the distribution of the compounds. This is an essentially descriptive statistical method which aims to present, in graphic form, the maximum of information contained in the data Table 2 and Table 3.

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

| Variables | MW | η | MR | HBA | HBD | LogP | NRE | $pIC_{50}$ |
|---|---|---|---|---|---|---|---|---|
| **MW** | **1** | | | | | | | |
| **η** | -0,468 | **1** | | | | | | |
| **MR** | 0,680 | -0,513 | **1** | | | | | |
| **HBA** | 0,190 | 0,047 | 0,019 | **1** | | | | |
| **HBD** | -0,683 | 0,491 | **-0,856** | -0,127 | **1** | | | |
| **LogP** | **0,903** | -0,574 | **0,799** | 0,002 | **-0,745** | **1** | | |
| **NRE** | **0,870** | -0,575 | 0,631 | 0,406 | -0,705 | **0,831** | **1** | |
| **$pIC_{50}$** | 0,880 | -0,603 | 0,649 | 0,050 | -0,707 | 0,900 | 0,815 | **1** |

The multiple linear regression statistic technique is used to study the relation between one dependent variable and several independent variables. It is a mathematic technique that minimizes differences between actual and predicted values. It has served also to select the descriptors used as the input parameters in the multiple non-linear regression (MNLR) and artificial neural network (ANN).
The (MLR) and the (MNLR) were generated to predict cytotoxic effects $IC_{50}$ activities of isatin derivatives. Equations were justified by the correlation coefficient (R), the coefficient of determination ($R^2$), the mean squared error (MSE), the Fishers F-statistic (F) and the significance level (F value) [24-26].
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 [27]. 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” [28-30]. For each data set, an input-output model is developed. In this study we used, the leave-one-out (LOO) procedure.

## RESULTS AND DISCUSSION

### Data set for analysis

The QSAR analysis was performed using the -log ($IC_{50}$) of the 47 selected molecules that have been synthesized and evaluated for their anticancer activity in vitro against U937 (experimental values) **[13-15]**. 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 and the values of the 10 chemical descriptors as shown in Table 2.
The principle is to perform in the first time, a main component analysis (PCA), which allows us to eliminate descriptors that are highly correlated (dependent), then perform a decreasing study of MLR based on the elimination of descriptors aberrant until a valid model (including the critical probability: $p\text{-value} < 0.05$ 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.

### Principal Components Analysis (PCA)

The totality of the 7 descriptors (variables) coding the 47 molecules was submitted to a principal components analysis (PCA). 8 principal components were obtained (Figure 2).
The first two axes F1 and F2 contributing respectively 66.45% and 14.37% to the total variance, the total information is estimated to a percentage of 80.83%.

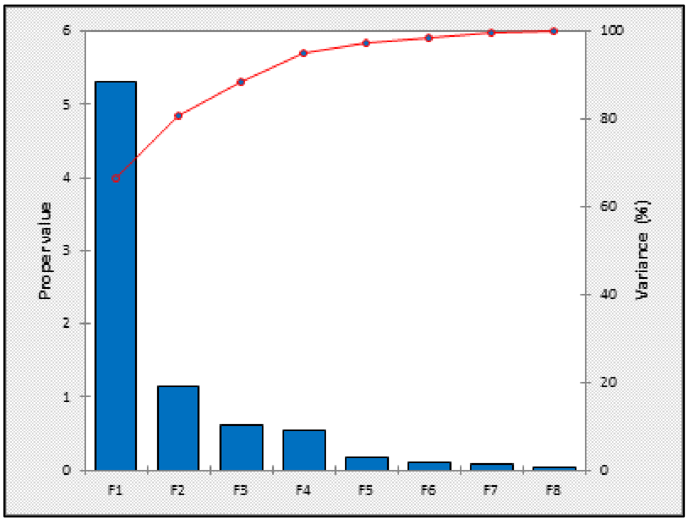

**Figure 2: The principal components and there variances**

The Pearson correlation coefficients are summarized in the above Table3. The obtained matrix provides information on the negative or positive correlation between variables. The principal component analysis (PCA) was conducted to identify the link between the different variables. Correlations between the 7 descriptors are shown in Table3 as a correlation matrix and in Figure 3 these descriptors are represented in a correlation circle.

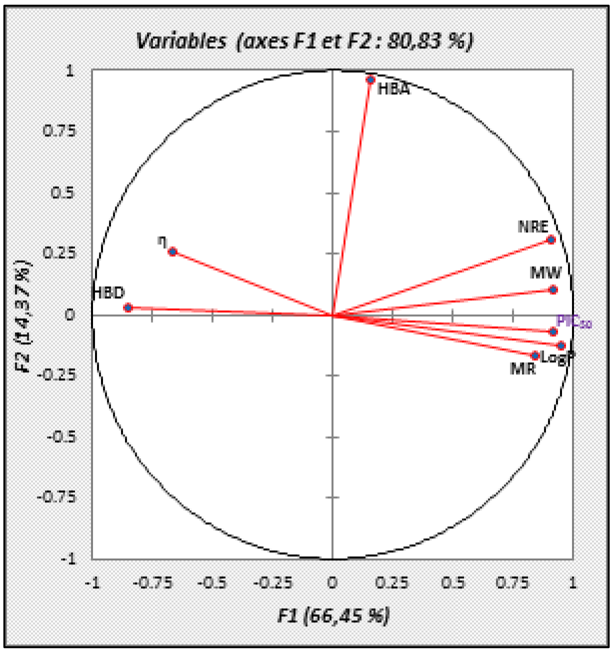

**Figure 3: Correlation circle**

HBD, MR and LogP are negatively correlated (r (HBD, MR) = -0,856; r (HBD, LogP) = -0,745).
LogP and MR are correlated (r (LogP, MR) = 0,799).
NRE, LogP and MW are highly correlated (r (NRE, LogP) = 0,831; r (NRE, MW) = 0,870) and (r (LogP, MW) = 0,903). The following variable is then removed (NRE).

**Multiple Linear Regressions (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, $R^2$, $R^2_{aj}$, $q^2$, SD, MSE, MSF and the F-values to select the best regression performance. Where R is the correlation coefficient; $R^2$ is the coefficient of determination; $R^2_{aj}$ is the adjusted coefficient of determination; $q^2$ is the coefficient of prediction; SD is the standard deviation; MSE is the mean squared error; MSF is the mean squared factor; F is the Fisher F-statistic.

Treatment with multiple linear regressions 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:

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

(Equation 1)

N = 47 R = 0.940 R2 = 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 the 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 Table 4.

The correlations of predicted and observed activities are illustrated in Figure 4. The descriptors proposed in equation1 by MLR were, therefore, used as the input parameters in the multiples non-linear regression (MNLR) and artificial neural network (ANN).

The correlation between MLR calculated and experimental activities are very significant as illustrated in Figure4 and as indicated by R and $R^2$ values.

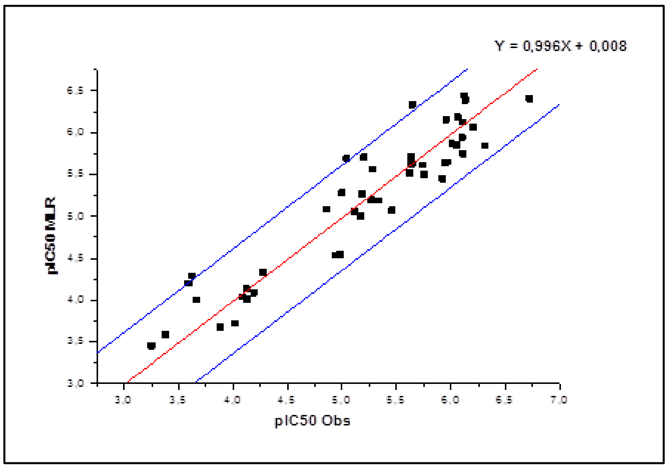

**Figure 4: Correlations of observed and predicted activities calculated using MLR**

### Validation criteria of the MLR model (ANOVA: ANalysis Of VAriance)

To validate the correlation equation provided by the statistical method of multiple linear regression (MLR), different criteria may be used.

### Overall assessment of the regression

Table 5 summarizes the variances, the degrees of freedom (df), the sums of squares (SS), Fisher's F value ($F_{exp}$) and overall p-value value of the model.

**Table 4: The observed, the predicted activities ($pIC_{50}$), according to different methods MLR, MNLR, ANN and CV for the 47 derivatives of isatin**

| N° | $pIC_{50\ Obs}$ | $pIC_{50\ MLR}$ | $pIC_{50\ MNLR}$ | $pIC_{50\ ANN}$ | $pIC_{50\ CV}$ |
|---|---|---|---|---|---|
| **1** | 5,18 | 5,27 | 5,34 | 5,20 | 5,26 |
| **2** | 5,46 | 5,06 | 5,34 | 5,42 | 5,38 |
| **3** | 5,62 | 5,51 | 5,56 | 5,62 | 5,83 |
| **4** | 5,94 | 5,63 | 5,76 | 5,75 | 5,87 |
| **5** | 6,31 | 5,84 | 5,99 | 5,93 | 6,07 |
| **6** | 5,74 | 5,61 | 5,76 | 5,72 | 5,68 |
| **7** | 5,75 | 5,49 | 5,56 | 5,60 | 5,66 |
| **8** | 6,05 | 5,85 | 6,06 | 5,98 | 6,03 |
| **9** | 5,64 | 5,62 | 5,58 | 5,54 | 5,86 |
| **10** | 6,01 | 5,86 | 5,90 | 5,90 | 6,04 |
| **11** | 6,20 | 6,06 | 5,90 | 5,94 | 5,88 |
| **12** | 5,64 | 6,33 | 5,87 | 6,01 | 5,29 |
| **13** | 6,10 | 5,94 | 5,87 | 5,93 | 6,06 |
| **14** | 5,28 | 5,56 | 5,78 | 5,79 | 5,09 |
| **15** | 5,92 | 5,45 | 5,44 | 5,52 | 5,93 |
| **16** | 5,95 | 6,15 | 6,21 | 6,09 | 5,98 |
| **17** | 5,63 | 5,62 | 5,57 | 5,78 | 5,28 |
| **18** | 6,12 | 6,44 | 6,40 | 6,14 | 6,08 |
| **19** | 3,25 | 3,44 | 3,15 | 3,06 | 3,85 |
| **20** | 3,67 | 4,00 | 3,98 | 3,89 | 3,93 |
| **21** | 4,19 | 4,08 | 4,18 | 4,25 | 3,92 |
| **22** | 4,13 | 4,00 | 3,99 | 3,92 | 4,15 |
| **23** | 4,08 | 4,03 | 4,07 | 4,10 | 3,94 |
| **24** | 4,01 | 3,72 | 3,56 | 3,91 | 4,07 |
| **25** | 4,27 | 4,33 | 4,46 | 4,45 | 3,83 |
| **26** | 3,88 | 3,67 | 3,38 | 3,73 | 3,69 |
| **27** | 3,38 | 3,59 | 3,63 | 3,50 | 3,57 |
| **28** | 4,98 | 4,53 | 4,78 | 4,72 | 4,34 |
| **29** | 4,94 | 4,53 | 4,77 | 4,70 | 4,29 |
| **30** | 5,11 | 5,06 | 5,11 | 5,30 | 5,05 |
| **31** | 3,59 | 4,20 | 4,08 | 3,90 | 3,08 |
| **32** | 5,17 | 4,99 | 5,19 | 5,28 | 5,15 |
| **33** | 4,12 | 4,13 | 4,03 | 3,86 | 4,08 |
| **34** | 4,86 | 5,08 | 5,24 | 5,08 | 4,35 |
| **35** | 3,62 | 4,28 | 4,04 | 3,88 | 3,97 |
| **36** | 6,11 | 5,74 | 5,87 | 5,85 | 6,04 |
| **37** | 6,11 | 6,12 | 5,88 | 5,97 | 6,03 |
| **38** | 6,06 | 6,18 | 5,98 | 6,00 | 5,88 |
| **39** | 5,97 | 5,64 | 5,72 | 5,77 | 6,07 |
| **40** | 5,63 | 5,70 | 5,81 | 5,81 | 5,83 |
| **41** | 6,72 | 6,40 | 6,35 | 6,53 | 5,96 |
| **42** | 6,13 | 6,38 | 6,35 | 6,35 | 5,95 |
| **43** | 5,00 | 5,28 | 5,39 | 5,23 | 5,06 |
| **44** | 5,20 | 5,70 | 5,41 | 5,58 | 5,79 |
| **45** | 5,04 | 5,69 | 5,38 | 5,56 | 5,28 |
| **46** | 5,33 | 5,18 | 5,24 | 5,05 | 5,15 |
| **47** | 5,27 | 5,19 | 5,24 | 5,05 | 5,17 |

**Table 5: Variance analysis**

| **Source** | **SS** | **df** | **Variance** | **F-exp** | **p-Value** |
|---|---|---|---|---|---|
| Regression (Factor) | 33.73 | 7 | 4.819 | 42.588 | 0 |
| Residual (Error) | 4.413 | 39 | 0.113 | - | - |
| Total | 38.143 | 46 | 4.932 | - | - |

-The variability not explained by the model is the sum of residual squares **SSE** = **4.413** with a degree of freedom equal to **39** (**N-p-1= 47-7-1**).

- The variability explained by the model is the sum of regression squares **SSF** = **33.730** with a degree of freedom equal to **7** (**N-(N-p-1)-1= p = 47-39-1**).

- The results seem excellent and the model is significant because we achieved good results for **F-exp** Fisher (**42,588**) and lower overall p-value at **α (F value) = 0.05** level (**p-value <0.05**).

**Test for significance**

-The first test that comes to mind is the significance of the correlation i.e. the correlation coefficient **R** is it significantly different from (**0**)?

-The test is: $\mathbf{H_{0:}}$ R = 0

$\mathbf{H_1}$ : R ≠ 0

-If the correlation coefficient is zero, we reject the hypothesis $\mathbf{H_0}$ (null hypothesis) and accept $\mathbf{H_1}$ (not null hypothesis). So the model is significant.

**Confidence Interval (CI)**

-The confidence interval (**CI**) **1-α** is a range of values that has a chance of **1-α** to contain the true value of the estimated parameter.

-If the p-value value exceeds (**0.05**), we reject $H_1$ and $H_0$ is accepted. So the model is not significant.

- If α > p-value, reject $H_0$ ($H_1$ acceptance).

- If α < p-value, $H_0$ acceptance (reject $H_1$).

**Student test**

-The Student law with (**N-p-1**) degree of freedom $\boldsymbol{t_{calc}}$ is written:

$$\boldsymbol{t_{calc=}}\left(\frac{\boldsymbol{R}}{\sqrt{\frac{\mathbf{1}-\boldsymbol{R^2}}{\boldsymbol{N}-\boldsymbol{p}-\mathbf{1}}}}\right)$$

-$H_0$ is rejected (null hypothesis) where: $\boldsymbol{t_{calc}} > \boldsymbol{t}_{\left(\mathbf{1}-\frac{\alpha}{2}\right),(\mathbf{N-p-1})}$

Where $\boldsymbol{t}_{\left(\mathbf{1}-\frac{\alpha}{2}\right),(\mathbf{N-p-1})}$ is the value of the Student law for $(\mathbf{N-p-1})$ degree of freedom, a probability$\left(\mathbf{1}-\frac{\alpha}{\mathbf{2}}\right)$.

-In our case we have **N = 47** and **R = 0.94**. This corresponds to $\boldsymbol{t_{calc}}$ **= 16.94**, one rejects $H_0$ (null hypothesis) where: $\boldsymbol{t_{calc}} > \boldsymbol{t}_{\left(\mathbf{1}-\frac{\alpha}{2}\right),(\mathbf{N-p-1})}$.

-According to the Student table $\left(\mathbf{1}-\frac{\alpha}{\mathbf{2}}\right)$ **= 0.975** and **N = 47** is obtained $\boldsymbol{t}_{(\mathbf{0.975,39})}$ **= 2.0227**.

$\boldsymbol{t_{calc}} > \boldsymbol{t}_{(\mathbf{0.975,39})}$ . Then we reject the null hypothesis $H_0$.

**Fisher test**

Analysis of variance (**V**) was used to test the equality of means, is called the **F statistic of Fisher**.

-Hypothesis $\mathbf{H_0}$ : SSF = SSE ($V_F = V_E$) Where (Error Variance) $\mathbf{V_E}$ **= MSE**

-against hypothesis $\mathbf{H_1}$ : SSF > SSE ($V_F > V_E$) Where (Factor Variance) $\mathbf{V_F}$ **= MSF**

-The Fisher F is calculated according to the following equation:

$$\mathbf{F_{exp}} = \frac{VF}{VE} = \frac{MSF}{MSE} = \frac{SSF/p}{SSE/N-p-1}$$

-To a threshold of (**0.05**) comparing $\boldsymbol{F_{exp}}$ obtained by the theoretical calculation and that obtained from Fisher's table $\boldsymbol{F}_{(\boldsymbol{p},\boldsymbol{N-p-1})}$ for one degree of freedom (**p, N-p-1**) with **p = 7** and **N = 47**, such as **(N-p-1) = 39**.

-We Accept $H_1$ if $\boldsymbol{F_{exp}} > \boldsymbol{F}_{(\mathbf{7,39})}$ .

-We then find $\boldsymbol{F}_{(\mathbf{7,39})}$ **= 2.255** and $\boldsymbol{F_{exp}}$ **= 42.588**, so we accept $H_1$ and $H_0$ is rejected.

**Correlation Coefficient: R**

This coefficient determines the variance of the target activity is explained by the model of QSAR i.e. by the regression of target activity based on the initial activity.

$$\mathbf{R} = \sqrt{\mathbf{1}-\frac{\boldsymbol{SSE}}{\boldsymbol{SST}}}$$

-A good correlation between the target activity and initial activity if **R** is closer to **1**.

-A non-linear correlation between the target activity and initial activity if **R** is closer to **0**.

-In our case we have **R = 0.94**, so a good correlation was shown between the observed activity and that obtained by **MLR**.

### Coefficient of Determination: $R^2$

The coefficient of determination $\mathbf{R^2}$, gives the rate of explanation or percentage of the variation of **Y** (endogenous variables) explained by the variation in **X** (exogenous variable).

$$\mathbf{R^2 = \frac{SSF}{SST}}$$

-In our case we have $\mathbf{R^2 = 0.88}$, this figure means that **88%** of the variable Y (activity) is attributable to the variation in the variable X (descriptors), which indicates that this model is statistically explanatory.

### Adjusted Coefficient of Determination: $R^2_{aj}$

The overall quality of the linear regression is measured by the coefficient of determination ($\mathbf{R^2_{aj}}$) "adjusted" taking into account the degree of freedom.

$$\mathbf{R^2_{aj} = 1- \frac{N-1}{N-p-1}(1- R^2)}$$

-With: N = 47, p = 7 and $R^2 = 0.88$.

-In our case we have $\mathbf{R^2_{aj} = 0.86}$, so the overall quality of the MLR is best. This indicates that this model is statistically significant.

### Coefficient of Prediction: $q^2$

The $\mathbf{q^2}$ value is used as the determining factor in selection of optimal models. The coefficient of prediction ($\mathbf{q^2}$) was calculated using:

$$\mathbf{q^2 = 1- \frac{VE}{SST} = 1- \frac{MSE}{SST}}$$

-SST: sum of total squares.

-In our case we have $\mathbf{q^2 = 0.99 > 0.6}$, So the predictive power of this model is very significant, which shows that the model proposed in this paper is able to predict activity with a great performance, and that the selected descriptors are pertinent. This means that the prediction of the new compounds is feasible.

-we can enjoy the performance of the predictive power of this model to explore and propose new molecules could be active.

### Standard Deviation: SD

The standard deviation (**SD**) measures the variation in the target activity is not explained by the QSAR model. In particular, over the standard deviation is small, the correlation is best.

$$\boldsymbol{SD = \sqrt{\frac{SSE}{N-p-1}} = \sqrt{VE} = \sqrt{MSE}}$$

-N: (**N = 47**) number of data points considered.

-p: (**p = 7**) number of restrictions on the degrees of freedom (equal to the number of parameters).

-In our case we have **SD = 0.33**, so the correlation between the observed activity and that obtained by MLR is best.

### Multiples Non-Linear Regression (MNLR)

We have used also the technique of nonlinear regression model to improve the structure-activity relationship to quantitatively evaluate the effect of substituent. We have applied to the data matrix constituted obviously from the descriptors proposed by MLR corresponding to the 47 molecules. The coefficients R, $R^2$, and the F-values are used to select the best regression performance. We used a pre-programmed function of XLSTAT following:

$$\mathbf{Y = a + (bX_1 + cX_2 + dX_3 + eX_4 \ldots) + (fX_1^2 + gX_2^2 + hX_3^2 + iX_4^2 \ldots)}$$

Where a, b, c, d,…: represent the parameters and $X_1$, $X_2$, $X_3$, $X_4$,…: represent the variables. The resulting equations:

$$\mathbf{pIC_{50\ MNLR}} = - 42{,}404 + 0{,}015\ \mathbf{MW} + 26{,}617\ \boldsymbol{\eta} - 0{,}008\ \mathbf{MR} - 0{,}016\ \mathbf{HBD} + 0{,}191\ \mathbf{LogP}$$

$$-1{,}820E\text{-}05\ (\mathbf{MW})^2 - 3{,}912\ (\boldsymbol{\eta})^2 - 3{,}569E\text{-}05\ (\mathbf{MR})^2 - 0{,}293\ (\mathbf{HBD})^2 + 0{,}016\ (\mathbf{LogP})^2$$

(**Equation 2**)

N = 47 R = 0.960 R2 = 0.920 MSE = 0.085

With the optimal MNLR model, the values of predicted activities $pIC_{50\ MNLR}$ calculated from equation2 and the observed values are given in Table 4. The correlations of predicted and observed activities are illustrated in Figure 5.

The correlation between MNLR calculated and experimental activities are very significant as illustrated in Figure5 and as indicated by R and $R^2$ values.

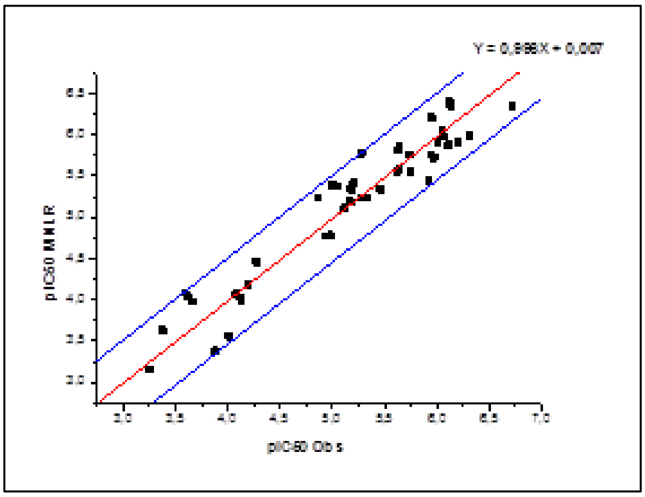

**Figure 5: Correlations of observed and predicted activities calculated using MNLR**

## Artificial Neural Networks (ANN)

In order to increase the probability of good characterization of studied compounds, artificial neural networks (ANN) can be used to generate predictive models of quantitative structure-activity relationships (QSAR) between a set of molecular descriptors obtained from the MLR, and observed activity. The ANN calculated activities model were developed using the properties of several studied compounds. Some authors **[31,32]** 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 $1.8 < \rho < 2.3$ [33]. 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 $pIC_{50\ ANN}$ calculated using ANN and the observed values are given in Table 4. The correlations of predicted and observed activities are illustrated in Figure 6.

The correlation between ANN calculated and experimental activities are very significant as illustrated in Figure6 and as indicated by R and $R^2$ values.

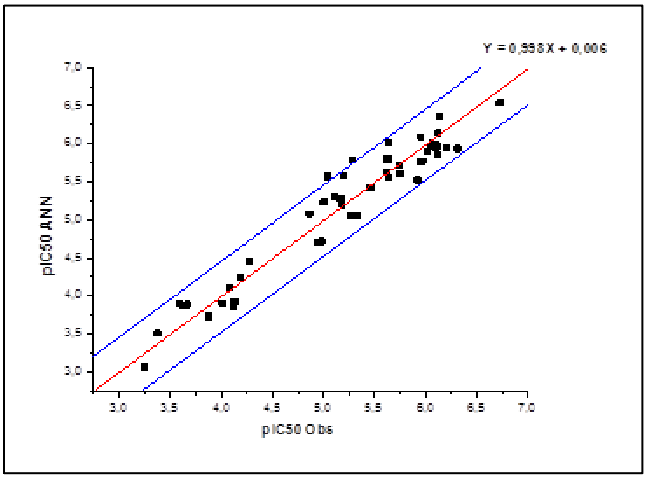

**Figure 6: Correlations of observed and predicted activities calculated using ANN**

N = 47 R = 0.97 R2 = 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.

**Cross Validation (CV)**

To test the performance of the neural network and the validity of our choice of descriptors selected by MLR and trained by MNLR and ANN, we used cross-validation method (CV) with the procedure leave-one-out (LOO). In this procedure, one compound is removed from the data set, the network is trained with the remaining compounds and used to predict the discarded compound. The process is repeated in turn for each compound in the data set.

In this paper the 'leave-one-out' procedure was used to evaluate the predictive ability of the ANN.

The values of predicted activities $pIC_{50\ CV}$ calculated using CV and the observed values are given in Table4. The correlations of predicted and observed activities are illustrated in Figure 7.

The correlation between CV calculated and experimental activities are very significant as illustrated in Figure7 and as indicated by R and $R^2$ values.

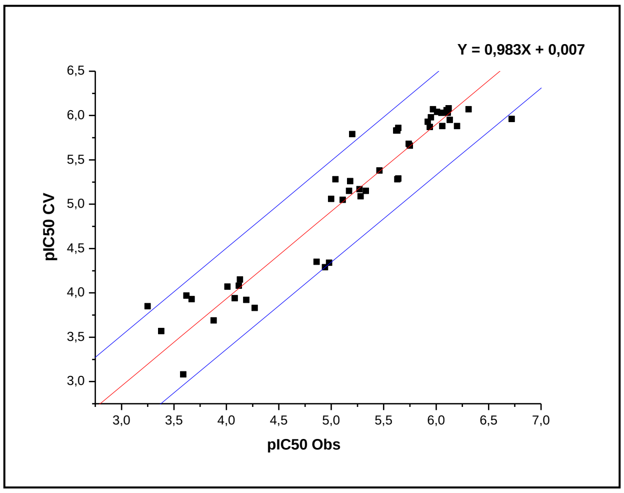

**Figure 7: Correlations of observed and predicted activities calculated using CV**

N = 47 R = 0.95 R2 = 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 MNLR are very sufficient to conclude the performance of the model. Even if it is possible that this good prediction is found by chance we can claim that it is a positive result. So, this model could be applied to all derivatives of 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, MNLR and ANN models shows that the ANN models have substantially better predictive capability because the ANN approach gives better results than MLR and MNLR. ANN was able to establish a satisfactory relationship between the molecular descriptors and the activity of the studied compounds. A good correlation was obtained with cross validation $\mathbf{R_{CV} = 0.95}$. So the predictive power of this model is very significant. The results obtained in this study, showed that models MLR, MNLR and ANN are validated, which means that the prediction of the new compounds is feasible.

## CONCLUSION

In this study, three different modelling methods, MLR, MNLR and ANN were used in the construction of a QSAR model for the anti-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 and MNLR, 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 Table4. It was also shown that the proposed methods are a useful aid for reduction of the time and cost of synthesis and activity determination of anti-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 feature and have a great influence on the activity may be used with other descriptors for the development of predictive QSAR models.

Previous studies QSAR already performed on the same set of isatin using multiple linear regression, obtained a correlation coefficient (R = 0.92) [34]. In this study the correlation coefficient obtained from the MLR ($R_{MLR}$ = 0.94), by using a variety of descriptors, is very important and this coefficient improved by using MNLR and ANN respectively ($R_{MNLR}$ = 0.96) and ($R_{ANN}$ = 0.97) so the proposed model is very significant and its performance is tested by cross-validation method CV ($R_{CV}$ = 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] Y Boukarai, F. Khalil, M. Bouachrine, Int J Sci Engineer Res **2015**.
[2] JB Gibbs. *Science*, **2000**, 287, 1969-1973.
[3] SN Pandeya, S Smitha, M Jyoti, SK Sridhar, *Acta Pharm*. **2005** 5527–5546.
[4] VM Sharma, P Prasanna, VA Seshu, B Renuka, VL Rao, GS Kumar, CP Narasimhulu, PA Babu, RC Puranik, D Subramanyam, A Venkateswarlu, S Rajagopal, KBS Kumar, CS Rao, NVSR Mamidi, DS Deevi, R Ajaykumar, R Rajagopalan, *Bioorg. Med. Chem. Lett*. **2002,** 12, 2303–2307.
[5] MJ Moon, SK Lee, JW Lee, WK Song, SW Kim, JI Kim, C Cho, SJ Choi, YC Kim, *Bioorg Med Chem* **2006,** 14, 237–246.
[6] AH Abadi, SM Abou-Seri, DE Abdel-Rahman, C Klein, O Lozach, L Meijer, *Eur J Med Chem* **2006**, 41 296–305.
[7] A Gursoy, N Karali, *Eur J Med Chem* 38, **2003**, 633–643.
[8] A Cane, M C Tournaire, D Barritault, M Crumeyrolle-Arias Biochem Biophysics Res Commun, **2000**, 276, 379-384.
[9] K L Vine, J M Locke, M Ranson, K Benkendorff, S G Pyne, J B Bremner *Bioorg Med Chem*, **2007**, 15, 931-938.
[10] G Bacher, B Nickel, P Emig, U Vanhoefer, S Seeber, A Shandra, Klenner, T Beckers, Cancer Res, **2001**, 61, 392-399.
[11] C Nantasenamat, C Isarankura-Na-Ayudhya, T Naenna, V Prachayasittikul, J Excli **2009**, 8 74-88.
[12] C Nantasenamat, C Isarankura-Na-Ayudhya & V Prachayasittikul, *J Expert Opin Drug Discov*, **2010**, 5 (7) 633-654.
[13] KL Vine, JM Locke, M Ranson, K Benkendorff, SG Pyne, JB Bremner, *Bioorg Med Chem* **2007**, 15 931–938.
[14] KL Vine, JM Locke, M Ranson, SG Pyne, JB Bremner, *J Med Chem* **2007,** 50 5109–5117.
[15] L Matesic, JM Locke, JB Bremner, SG Pyne, D Skropeta, M Ranson, KL Vine, *Bioorg Med Chem*, **2008**, 16, 3118–3124.
[16] Advanced Chemistry Development Inc, Toronto, Canada, (**2009**).
[17] ACD/ChemSketch Version 45 for Microsoft Windows User’s Guide.
[18] ACD/Labs Extension for ChemOffice Version 80 for Microsoft Windows User’s Guide.
[19] ACD/Labs Extension for ChemBioOffice Version 140 for Microsoft Windows User’s Guide.
[20] A, N L Conformational Analysis 130 MM2 A Hydrocarbon Force Field Utilizing V1 and V2 Torsional Terms, *J Am Chem Soc* **1977**, 99, 8127-8134.
[21] U Sakar, R Parthasarathi, V Subramanian, PK Chattaraji, *J Mol Des IECMD*, **2004,** 1-24.
[22] XLSTAT 2015 Add-in software (XLSTAT Company) www.xlstat.com
[23] SYSTAT 13 Add-in software (SYSTAT Company) www.systat.com
[24] M Larif, A Adad, R Hmammouchi, AI Taghki, ASoulaymani, A Elmidaoui, M Bouachrine, T Lakhlifi, Arabian J Chem, **2013.**
[25] M Ghamali, S Chtita, A Adad, R Hmamouchi, M Bouachrine, T Lakhlifi, Int J Adv Res Comput Sci Software Engineer, **2014.**
[26] Y Boukarai, F Khalil, M Bouachrine, *J. Chem. Pharm. Res* **2016.**

[27] VJ Zupan & J Gasteiger, Neural Networks for Chemists - An Introduction, VCH Verlagsgesellschaft, Weinheim/VCH Publishers, New York 106 (12) (**1993**) 1367-1368.
[28] B Efron, *J AmStat Assoc* **1983**, 78, 316-331.
[29] MA Efroymson, Multiple regression analysis, In Mathematical Methods for Digital Computers, Ralston, A, Wilf, HS, Eds,Wiley NewYork, **1960.**
[30] DW Osten, *J Chemom*, 1998 2, 39-48.
[31] S-S So & WG Richards, *J Med Chem*, **1992**, 35, 3201-3207.
[32]TA Andrea & H Kalayeh, *J Med Chem*, **1991**, 34, 2824–2836
[33] M Elhallaoui, Modélisatrice moléculaire et étude QSAR d'antagonistes non compétitifs durécepteur NMDA par les méthodes statistiques et le réseau de neurones, **2002**, 106.
[34] R Sabet, M Mohammadpoura, A Sadeghi, Fassihi, Europ J Med Chem, **2010.**