Quantitative structure activity relationship studies of anti-proliferative activity of some indole derivatives combining DFT calculations and statistical results

Halima Hazhazi¹, Hanane Zaki²,³*, Mohamed Benlyas², Youness Filali-Zegzouti², Youcef Boumedjane¹, Mohammed Bouachrine³*

¹Research team  computational & pharmaceutical chemistry, LMCE laboratory, department of chemistry, Faculty of sciences, Biskra University, Biskra, Algeria.
²Research team Biology, Environment & health. Department of biology Faculty of Science and Technics. My Ismail University. Meknes. Morocco
³Research team Matériaux. Environnement & Modélisation. ESTM. My Ismail University. Meknes. Morocco

Abstract—Many studies have focused on indole derivatives mainly their antiproliferative effect. The therapeutic effect of this group of molecule is very important. Quantitative structure-activity relationships (QSAR) have been applied for development relationships between physicochemical properties and their biological activities.

A series of 30 molecules derived from indole is based on the quantitative structure-activity relationship (QSAR). This study was carried out using the principal component analysis (PCA) method, the multiple linear regression method (MLR), non-linear regression (RNLM), the artificial neural network (ANN) and it was validated using cross validation analysis (CV). We accordingly propose a quantitative model and we try to interpret the activity of the compounds relying on the multivariate statistical analyses. A theoretical study of series was studied using density functional theory (DFT) calculations at B3LYP/6-31G(d) level of theory for employing to calculate electronic descriptors when, the topological descriptors were computed with ACD/ChemSketch and ChemDraw 8.0 programs. The best QSAR model was found in agreement with the experimental by ANN (R = 0.99).

Keywords—Breast cancer, anti-proliferative, indole derivatives, QSAR, MLR, MNLR, ANN, CV.

I. INTRODUCTION

Breast cancer is considered as one of the major and widespread reasons that cause death among women all over the world, in case of the late diagnosis [1]. Despite the improvements and the efficiency in early detection, chemotherapy and radiotherapy breast cancer is still at high risks [2-4]. Therefore, it is necessary to find a cure for this disease, a lot of scientific researches were carried out to determine a particular molecule to this treatment [5]. Among the great number of compounds that occur in nature, Indole is the main component. Moreover, indole derivatives have many applications, in the pharmaceutical industry in the treatment of various diseases [6]. Indole derivatives are one of the most promising heterocyclic, which have active sites in treating various diseases [7]. In addition, these compounds have broad spectrum of biological activities involving anticancer, antioxidant, antimicrobial, anticonvulsant, anti-leishmanial, antidepressants, anti-inflammatory activities and they were found to have capabilities of anti-proliferative activity on cancer cells lines [8-14].

![Fig.1: Studied compounds (indole)](image)

On the other hand, Quantitative structure-activity relationship (QSAR) seeks to inquire into the relationship between molecular descriptors which describe the physicochemical properties correlated with biological activity of the set of compounds [15, 16]. The QSAR study is an important step in the development of new drugs. In this paper we have studied a quantitative structure-activity relationship (QSAR) of indole against human breast cancer cells (MCF-7) based on 30 indole derivatives taken from the literature [17-20]. Therefore,
we propose to develop a quantitative model, and we try to predict the activity of these compounds based on the several statistical methods: Principal Component Analysis (PCA), Multiple Non-Linear Regression (MNLR) and Multiple Linear Regression (MLR), Artificial Neural Network (ANN) and Cross Validation analyses (CV). The development of a performant model will help to explain the role of indole derivatives in chemotherapy against breast cancer and also propose other molecules, then predict their anti-cancer activity.

II. MATERIALS AND METHODS

Table.1: Observed IC50 of the indole derivatives anti-proliferative agents

| S.N | Structure | pIC50 | S.N | Structure | pIC50 |
|-----|-----------|-------|-----|-----------|-------|
| M1  | ![Structure](image1.png) | 1.988 | M16 | ![Structure](image2.png) | 1.690 |
| M2  | ![Structure](image3.png) | 0.577 | M17 | ![Structure](image4.png) | 1.781 |
| M3  | ![Structure](image5.png) | 1.953 | M18 | ![Structure](image6.png) | 1.274 |
| M4  | ![Structure](image7.png) | 1.273 | M19 | ![Structure](image8.png) | 1.684 |
| M5  | ![Structure](image9.png) | 1.908 | M20 | ![Structure](image10.png) | 1.703 |
| M6  | ![Structure](image11.png) | 1.273 | M21 | ![Structure](image12.png) | 1.674 |

2.1 Chemical data
A dataset of the series of indole Compounds collected from literature [17-20], are listed in table 1. A total of 30 derivatives of indole were studied and analyzed in order to find quantitative structure activity relationship between the anti-proliferative activity and the structure of these molecules. The IC50 values in µM units exhibiting 50% inhibition of cell growth for human breast cancer (MCF 7) were converted in pIC50 by taking logarithm (pIC50 = log10 IC50) for QSAR study.
2.2. Molecular descriptor

The present work is necessary for us to determine several different descriptors to estimate in the QSAR model. The quantum chemical calculations are performed at the B3LYP/6-31G(d) level of theory using GAUSSIAN 03 of programs [21] to calculate some electronic descriptors such as: Frontier molecular orbital’s highest occupied molecular orbital: EHOMO (eV); lowest unoccupied molecular orbital energy :ELUMO (eV) ; The Gap energetic (Gap) (eV), (the difference between EHOMO and ELUMO); Total Energy TE (ua); The absolute electronegativity χ (eV), χ = (EHOMO + ELUMO)/2; the absolute hardness η (eV): η = (EHOMO - ELUMO)/2; The Softness S (eV), it is the reactivity index and defined reciprocal of hardness S = 1/η; The electrophilicity index ω (eV), ω = 2/η η [22] and The dipole moment μ (Debye). On the other side, we have chosen some physico-chemical descriptors, which were computed with Advanced chemistry development’s ACD/ Chem Sketch [23] and ChemDraw Ultra8.0 [24] programs was employed to calculate: Molecular Weight (MW), Torsion energy (TE), Repulsion energy (RE), electronic energy (EE), the octanol/water partition coefficient (log P), Parachor (Pc) and Density (D) Thus 12 descriptors. Data was presented in Table 2.

2.3- Statistical methods

To explain the structure-activity relationship, The 12 quantitative descriptors of the compounds of indole (1 to 30) are studied using different statistical methods:

The principal component analysis (PCA) [25] using the software XLSTAT version 2013 [26]. This is an essentially descriptive statistical method which aims to present, in graphic form. The large information contained in a data, as shown in Table 1. PCA is a helpful statistical technique for summarizing the maximum of information encoded in the structures of compounds. This method is very useful for understanding the distribution of the compounds. The Multiple Linear Regression (MLR) statistical technique is used to study the relation between one dependent variable and several independent variables. It is a mathematic technique that minimizes the differences between actual and predicted values. The multiple linear regression model (MLR) was performed to predict pIC50. and it served to select the used descriptors as the input parameters for (NLMR). MLR and MNLR were generated using the software XLSTAT version 2014. The obtained equations were justified by the determination coefficient (R²) correlation coefficient (R), mean squared error (MSE), Mean Absolute Error (MAE) and Fisher’s criterion (F). [27,28].

The ANN analysis was performed with the use of Matlab software version 2009. A Neural Fitting tool (nftool) toolbox on a data set of the indole compounds [29]. Three components constitute a neural network: the topology of the connections between the nodes, the processing elements or nodes and the learning rule by which new information is encoded in the network. However, there are a number of different ANN models; the most frequent type of ANN in QSAR is the three-layered feed-forward network [30]. In this kind of networks, the neurons are arranged in layers (an input layer, one hidden layer and an output layer). the neurons 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 (CV) 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” [31-33]. For each data set, an input-output model is developed. In this study we used, the Leave-One-Out (LOO) procedure.

III. RESULTS AND DISCUSSION

3.1. Data set for analysis

A QSAR study was performed on 30 indole derivatives as reported previously, in order to identify a quantitative relationship between the structure and anti-proliferative activity against breast cancer cells lines (MCF7). The values of the 12 descriptors (2D and 3D descriptors) are shown in Table 2.

Table 2: Dataset used for QSAR analysis of series of indole derivatives

| molecules | E_HOMO | E_LUMO | ΔE  | μ   | χ    | TotE | Log P | RE   | TE   | Kow  | MW   | D   |
|-----------|--------|--------|-----|-----|------|------|-------|------|------|------|------|-----|
| M1        | -4.907 | -0.846 | 4.061 | 2.855 | 2.030 | -879.519 | 1.744 | 16622.8 | 7.5280 | 2.860 | 262.31 | 1.214 |
| M2        | -4.905 | -0.064 | 4.841 | 2.796 | 2.484 | -996.254 | 2.498 | 20936.1 | 14.036 | 4.053 | 302.37 | 1.216 |
| M3        | -5.395 | -1.403 | 3.992 | 2.423 | 3.399 | -940.636 | 3.300 | 17066.4 | 16.621 | 5.171 | 287.36 | 1.166 |
| M4        | -5.229 | -1.297 | 3.932 | 1.239 | 3.263 | -1055.11 | 3.174 | 20431.2 | 0.9840 | 5.137 | 317.38 | 1.174 |
| M5        | -5.171 | -1.158 | 4.013 | 3.782 | 3.164 | -899.114 | 2.022 | 15064.2 | 24.456 | 3.545 | 263.29 | 1.192 |
3.2. Data Modeling.

3.2.1 Principal component analysis

The 12 descriptors (variables) describing the 30 molecules were submitted to Principal Components Analysis (PCA). The first two principal axes are sufficient to describe the information provided by the data matrix. Figure 2 presents the percentages of variance: F1 = 40.50%, F2 = 21.23% and the total information is estimated on 61.74%.

![Scree plot](image)

*Fig. 2. The principal components and their variances*

The principal component analysis (PCA) was carried out to have an idea about the link between the different variables.

The obtained matrix (Table 3) summarizes the correlations between the 12 descriptors and provides information on the negative or positive correlation between variables. Figure 3 shows these descriptors in a correlation circle. In general, the correlation matrix shows a low interrelationship between most of the descriptors. Good co-linearity ($r > 0.5$) was observed between some of the variables. High interrelationship was observed between $E_{HOMO}$ and $\chi$ ($r = -0.979$), $E_{HOMO}$ and $\Delta E$ ($R = -0.96$), $E_{LUMO}$ and $\chi$ ($R = -0.91$) and $\Delta E$ and $\chi$ ($R = 0.90$), the variables $\Delta E$ and $\chi$ are removed to decrease the correlations.
### Table 3: The correlation matrix (pearson (r)) between different obtained descriptors

| Variables | EHOMO | ELUMO | ΔE | μ | χ | TotE | Log P | RE | TE | Kow | MW | D |
|-----------|-------|-------|----|---|---|------|-------|----|---|-----|----|---|
| EHOMO     |  1    |       |    |   |   |      |       |    |   |     |    |   |
| ELUMO     | 0.845 |       | 1  |   |   |      |       |    |   |     |    |   |
| ΔE        | -0.969| -0.688| 1  |   |   |      |       |    |   |     |    |   |
| μ         | -0.347| -0.536| 0.22| 1 |   |      |       |    |   |     |    |   |
| χ         | -0.979| -0.918| 0.90| 0.419| 1  |      |       |    |   |     |    |   |
| TotE      | 0.106 | 0.142| 0.020| -0.132| 1  |      |       |    |   |     |    |   |
| Log P     | 0.194 | 0.018|   | 0.149| -0.114| -0.378| 1  |    |   |     |    |   |
| RE        | 0.538 | 0.559| -0.317| -0.549| -0.186| 0.541| 1  |    |   |     |    |   |
| TE        | -0.106| -0.093| 0.219| 0.125| -0.090| 0.042| 0.142| 1  |   |     |    |   |
| Kow       | 0.428 | 0.436| -0.380| -0.413| -0.257| 0.564| 0.583| -0.312| 1  |     |    |   |
| MW        | 0.288 | 0.225| -0.119| -0.259| -0.312| 0.509| 0.576| 0.049| 0.428| 1  |    |   |
| D         | -0.040| -0.293| 0.267| 0.128| -0.790| 0.326| -0.042| -0.028| -0.009| 0.178| 1  |   |

The correlation circle (Figure 3) which shows that the F1 axis (40.50% of the variance) appears to represent the Density (D) and the Total energy (TE). The F2 axis (21.24% of the variance) seems to represent the EHOMO and gap Energy (ΔE).

From other side the analysis of diagrams according to the planes F1 and F2 (of the total variance) of the studied series are presented in Figure 4 we can discern three groups of molecules:

- Group 1: contains the molecules: M8, M9, M11, M10, and M13. (Green color)
- Group 2: contains the molecules M27, M24, M16, M28 and M12. (Red color)
- Group 3: contains the rest of the molecules. (Blue color)

When we return to the structures of molecules M8, M9, M10 and M11 (group 1), we note that all these molecules are alike in their structures, and have as basic structure compound 1- Aryl-1H-1,2,3-Triazol-4-yl methyl 1H indole-2-carboxylate. The molecules M24, M27 and M28 (group 2) have the same basic derivative which is 5-(3-indolyl)-2-Substituted-1,3,4-thiadiazoles. Group 3 is the most important of the groups because it contains a large number of molecules (20 molecules) which have the same behavior.
3.2.2 Multiple linear regressions MLR

Our work is based on the development of the best QSAR model to clarify the correlation between the different descriptors and the biological activities pIC50 values of the indole derivatives. This method utilised several coefficients: R is the correlation coefficient, $R^2$ is the coefficient of determination, MSE mean squared error, MAE Mean Absolute Error and F is the Fisher F-statistic those coefficients adopt the best regression performance.

The obtained relationship in this model by the linear combination of the essential descriptors: $E_{HOMO}$, $E_{LUMO}$, $\mu$, TE, Log P, RE, TE, Kow, MW, D.

The QSAR models using multiple linear regressions method is represented by the following equations:

$$pIC50 = -2.91 - 0.45 * E_{HOMO} + 1.15 * E_{LUMO} - 6.44 \text{ E-02} * \mu + 8.74 \text{ E-04} * \text{TE} + 0.20 * \text{LogP} - 4.12 \text{ E-06} * \text{RE} - 9.31 \text{ E-03} * \text{TE} + 0.22 * \text{Kow} + 3.03 \text{ E-04} * \text{MW} + 2.84 \text{ D}.$$  

N= 24     N_{test} = 6    R= 0.80     R²= 0.641

MSE = 0.48     MAE = 0.44     F = 2.326

The model shows a good correlation coefficient ($R$ =0.800) between ten descriptors and the anti-proliferative activity. This equation shows that the anti-proliferative activity of the indole derivatives depends on the electronic and the topological side of the molecule. Anti-proliferative activity increases by increasing the topological properties, Log P, RE, Kow, MW, D and by diminishing the electronic properties $E_{HOMO}$, $\mu$, TE, RE, TE. Figure 5 presents the graphical representations of graphical calculated and observed pIC50 by MLR.

3.2.3 Multiple nonlinear regressions MNLR

We have utilized the technique of nonlinear regression model to improve the structure activity relationship in a quantitative way, the selected descriptors from the MLR model are used like data base matrix for the MNLR. The resulting equation is:

$$pIC50 = 120.18 + 2.44 * E_{HOMO} + 1.01 * E_{LUMO} + 0.75 * \mu + 6.63 \text{ E-03} * \text{TE} - 4.56 \text{ E-05} * \text{RE} - 4.73 \text{ E-02} * \text{TE} - 1.44 * \text{Kow} + 0.16 * \text{MW} - 203.25 * \text{D} + 0.44 * E_{HOMO}^2 + 0.43 * E_{LUMO}^2 - 0.11 * \mu^2 + 6.27 \text{ E07} * \text{TE}^2 + 0.74 * \text{LogP}^2 - 5.9 \text{ E-09} * \text{RE}^2 + 1.13 \text{ E-03} * \text{TE}^2 + 0.12 * \text{Kow}^2 - 2.15 \text{ E-04} * \text{MW}^2 + 79.02 * \text{D}^2.$$  

N= 24     N_{test} = 6    R= 0.95     R²= 0.90     MSE = 0.13

MAE = 0.24

As illustrated in Figure 5, the correlation between calculated and experimental activities is very remarkable.
The obtained correlation coefficient was significant \( R = 0.95 \). Figure 6 shows a regular distribution of the PIC50 observed values depend on the experimental values.

### 3.2.3 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 the observed activity. The calculated activities model was developed using the properties of several studied compounds. Some authors \cite{34, 35} have proposed a parameter \( \rho \), leading to the number of hidden neurons, which play a major role in determining the best ANN architecture. These are defined as follows:

\[
\rho = \frac{\text{Number of data points in the training set}}{\text{Sum of the number of connections in the ANN}}
\]

The values of predicted activities (pIC50) using ANN and the observed values are given in Table 4. The correlation between calculated ANN and experimental anti-proliferative values is very significant as indicated by \( R \) and \( R^2 \) values illustrated in figure 8.

| Molecules | Obs (pIC50) | Pred (pIC50) | Pred (pIC50)/Obs (pIC50) |
|-----------|-------------|--------------|--------------------------|
| M1        | 1.988       | 2.0078       |                          |
| M2        | 0.577       | 0.5947       |                          |
| M3        | 1.953       | 2.0063       |                          |
| M4        | 1.273       | 1.2653       |                          |
| M5        | 1.908       | 1.7715       |                          |
| M6        | 1.273       | 1.4139       |                          |
| M7        | 1.621       | 1.5753       |                          |
| M8        | 1.105       | 1.0249       |                          |
| M9        | 1.209       | 1.2162       |                          |
| M10       | 1.484       | 1.4810       |                          |
| M11       | 1.459       | 1.5133       |                          |
| M12       | 1.233       | 1.2304       |                          |
| M13       | 1.588       | 1.5925       |                          |
| M14       | 1.218       | 1.3461       |                          |
| M15       | 1.727       | 1.7242       |                          |
| M16       | 1.690       | 1.7099       |                          |
| M17       | 1.781       | 1.7874       |                          |
| M18       | 1.274       | 1.2230       |                          |
| M19       | 1.684       | 1.7271       |                          |
| M20       | 1.703       | 1.5881       |                          |

Table.4: The observed and ANN predicted activities

The obtained correlation coefficient \( R \) value confirms that the artificial neural network result was the best to build the quantitative structure activity relationship models. A comparison of the quality of MLR, MNLR and ANN models table 5 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.

| Molecules | Obs (pIC50) | (pIC50) RLM | (pIC50) RNLM | (pIC50) ANN | (pIC50) CV |
|-----------|-------------|-------------|--------------|-------------|------------|
| M1        | 1.989       | 1.791       | 1.944        | 2.0078      | 1.78       |
| M2        | 1.909       | 1.470       | 0.605        | 0.5947      | 0.61       |

Table.5: observed, predicted activities according to different used methods
3.2.4 Cross Validation

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 which is well adapted to the estimation of that ability. A good correlation was obtained with cross validation RCV = 0.74. So, the predictive power of this model is very significant. The results obtained showed that models MLR, MNLR and ANN are validated, which means that the prediction of the new compounds is feasible.

In this study, three different modelling methods, MLR, MNLR and ANN were used in the construction of a QSAR model for 30 derivatives of indole and the resulting models were compared (Table 5 - Table 6). It was shown that the artificial neural network ANN results have better predictive capability than the MLR and MNLR. We established a relationship between several descriptors and the anti-proliferative activity IC_{50} in satisfactory manners. The good results obtained with the cross validation (CV) shows that the model proposed in this paper are able to predict activity with a good performance, and that the selected descriptors are pertinent.

|     | RLM | RNLM | ANN | CV  |
|-----|-----|------|-----|-----|
| R   | 0.80| 0.95 | 0.99| 0.74|
| MSE | 0.48| 0.13 | 0.003| 0.08|
| MAE | 0.44| 0.24 | 0.03| 0.18|

Correlation coefficient (R), Mean squared error (MSE), Mean Absolute Error (MAE)
In this paper we have used different statistical methods: MLR, MNLR, ANN, cross validation CV and various electronic and topologic descriptors for construction of QSAR model for the anti-proliferative activity of indole derivatives, also, were compared the statistical terms R, R², MAE, MSE Resulting models. Moreover, the neural network ANN results (R= 99, MAE= 0.03 MSE= 0.003) have better predictive capability than the MLR and MNLR. A good correlation was obtained with cross validation RCV = 0.74 that confirms the great ability of our model to predict the activity. we established a relationship between several descriptors and inhibition values pIC50 of several organic compounds based on substituted indole in satisfactory manners. That studied model which is sufficiently rich in chemical, electronic and topologic information may be utilized for predicting and developing new molecules with better effect. Thus, thanks to QSAR studies, especially with the ANN that 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 that could be active in experiment.

REFERENCES

[1] Siegel R, Naishadham D and Jemal A: Cancer statistics, 2013. CA cancer J clin 63: 11-30, 2013.
[2] Aliaabadi HM, Maranchuk R, Kucharaki C, Mahdipoor P, Hugh J, Uludag H. Effective response of doxorubicin- sensitive and resistant breast cancer cells to combinational siRNA therapy. J Contr Release Official: J Contr Release Soc 2013; 172(1): 219-228.
[3] Hong SE, Kim EK, Jin HO, Kim HA, Lee JK, Koh JS, et al S6K1 inhibition enhances tamoxifen-induced cell death in MCF-7 cells through translational inhibition of Mcl-1 and survivin. Cell Biol Toxicol 2013: 29(4): 273-282.
[4] Choi HN, Jin HO, Kim JH, Hong SE, Kim HA, Kim EK, et al. Inhibition of S6K1 enhances glucose deprivation-induced cell death via downregulation of anti-apoptotic proteins in MCF-7 breast cancer cells. Biochem Biophys Res Comm 2013; 432(1): 123-128
[5] Hao Wang, Yi-Feng Ye, Effect of survivin siRNA on biological behaviour of breast cancer MCF7 cells. J Tropical Medicine (2015)225-228.
[6] T.M.T. Carvalho, L.M.P. Amaral, V.M.F. Morais, M.D.M. Ribeiro da Silva, Thermodynamic properties of alkyl 1H-indole carboxyate derivatives: a combined experimental and computational study, J. Chem. Thermodynamics (2016).
[7] Sirassu Narisimha, Nukala Satheesh Kumar, Kumara Swamy Battula, Vasudeva Reddy Nagavelli .S. K. Althaf Hussain , M. Srinivasa Rao. Indole-2-carboxylic acid derived mono and bis 1,4 disubstituted 1,2,3-triazoles: Synthesis, characterization and evaluation of anticancer, antibacterial, and DNA-cleavage activities. Bioorg. Med.Chem. Lett 26 (2016) 1639–1644
[8] Daly, S; Hayden, K.; Mulik, I.; Porch, N.; Tang, H.; Rogelj, S.; Frolova, L. V.; Lepthien, K.; Kornienko, A.; Magedov, I.V. Bioorg Med. Chem. Lett. 2011, 21, 4720
[9] Holla, B. S.; Mahalinga, M.; Karthikeyan, M. S.; Poojary, B.; Akberali, P. M.; Kumari, N. S. Eur. J. Med. Chem. 2005, 40, 1173.
[10] Mady, M. F.; Awad, G. E. A.; Jørgensen, K. B. Eur. J. Med. Chem. 2014, 84, 433.
[11] Rao, P. S.; Kurumurthy, C.; Veeraswamy, B.; Kumar, G. S.; Poornachandra, Y.; Kumar, C. G.; Babu, V. S.; Kotamraju, S.; Narsaiah, B. Eur. J. Med. Chem. 2014, 80, 184.
[12] H.N. Dogan, A. Duran, S. Rollas, G. Sener, M.K. Uysal, D. Gül, Bioorg. Med. Chem. 10 (2002) 2893e2898.
[13] F. Poorrajab, S.K. Ardestani, S. Emami, M. Behrouzi-Fardmoghodam, A. Shafiee, A. Foroumadi, Eur. J. Med. Chem. 44 (2009) 1758e1762.
[14] M. Yusuf, R.A. Khan, B. Ahmed, Bioorg. Med. Chem. 17 (2008) 8029e8034.
[15] Hansch C., Muir R.M., Fujita T., Maloney P.P., Geiger F., Streich M., J. Am. Chem. Soc. 85 (1963) 2817–2825.
[16] Bodor N., Current Medicinal Chemistry 5 (1988) 353–380. From book: Biochemistry of Redox Reactions, by Bernard Testa, editor: London [u, a], Acad. Press (1995).
[17] Xing-yue Ji , Si-tu Xue, Yue-chen Zhan, Jia-jia Shen, Lin-tao Wu, Jie Jia, Zhen Wang, Wu-ying Li , J. Med. Chem. 83 (2014) 409-418.
[18] S. K. Althaf Hussain, M. Srinivasa Rao. Bioorg. Med. Chem , 26 (2016) 1639-1644.
[19] Dalip Kumar, N. Maruthi Kumar, Kuei-Hua Chang , Kavita Shah, Eur. J. Med. Chem. 45 (2010) 4664-4668.
[20] E. Rajanarendar, K. Govardhan Reddy, S. Ramakrishna, M. Nagi Reddy, B. Shireesha, G. Durgaiah, Y. N. Reddy. Bioorg. Med. Chem. 22 (2012) 6677–6680.
[22] M.J. Frisch & al, Gaussian 03, Revision B.01, Gaussian, Inc., Pittsburgh, PA, (2003).
[23] U. Sakar, R. Parthasarathi, V. Subramanian & P.K. Chattaraj, Toxicity analysis of polychlorinated dibenzo furans through global, J. Mol. Des. IECMD, (2004) 1-24.
[24] Advanced Chemistry Development Inc., Toronto, Canada, (2009), (www.acdlabs.com/resources/freeware/chemsketch/).
[25] ACD/Labs Extension for ChemDraw Version 8.0 for Microsoft Windows User’s Guide.
[26] M Larif, A Adad, R Hmamouchi, A I Taghki, A Soulaymani, A Elmidiaoui, M Bouachrine, T Lakhlifi article in press in Arabian Journal of Chemistry 2013.
[27] SYSTAT 13 Add-in software (SYSTAT Company). www.systat.com.
[28] Y. Boukarai, F. Khalil, M. Bouachrine, QSAR study of isatin (1H-indole-2-3-dione) analogues as in vitro anticancer agents using the statistical analysis methods, International Journal of Scientific & Engineering Research (2015)
[29] Youness Boukarai, Fouad Khalil and Mohamed Bouachrine, QSAR study of 5,6-bicyclic heterocycles analogues as anti-Alzheimer’s agents using the statistical analysis methods. J. Chem. Pharm. Res., 2016, 8(3):1000-1013
[30] Adad A., Hmamouchi R., Taghki A. I., Abdellaoui A., Bouachrine M., Lakhlifi T., Atmospheric half-lives of persistent organic pollutants (POPs) study combining DFT and QSPR results, J. of Chemical and Pharmaceutical Research, Vol. 5, Issue 7 (2013) 28-41.
[31] V.J. Zupan, J. Gasteiger, Neural Networks for Chemists - An Introduction, VCH Verlagsgesellschaft, Weinheim/VCH Publishers, New York. 106(12) (1993) 1367-1368.
[32] B. Efron, Estimating the error rates of a predictive rule: improvement on crossvalidation, J. Am.Stat. Assoc. 78 (1983) 316-331
[33] M.A. Efroymsion, Multiple regression analysis, In Mathematical Methods for DigitalComputers, Ralston, A., Wilf, H.S., Eds,WileyNewYork, (1960).
[34] D.W. Osten, Selection of optimal regression models via cross-validation, J. Chemom. 2(1998) 39-48.
[35] S-S. So & W.G. Richards, Application of neural networks: quantitative structure-activity relationships of dihydrofolate reductase inhibitors, J. Med. Chem. 35 (1992) 3201-3207.
[36] T.A. Andrea & H. Kalayeh, Applications of neural networks in quantitative structure-activity