

Lipophilicity and QSAR Study of a Series of Makaluvamines by the Method of the Density Functional Theory: B3LYP/6-311++G(d,p)

Sékou DIOMANDÉ, Soleymane KONÉ*

Laboratoire de Chimie Organique et de Substances Naturelles de l'UFR SSMT, Université Félix Houphouët-BOIGNY,
22 BP 582 Abidjan 22, Côte d'Ivoire

*Corresponding author: konesol2003@yahoo.fr

Received October 01, 2019; Revised November 03, 2019; Accepted November 12, 2019

Abstract This work has focused on sixteen (16) Makaluvamines listed in the literature to date. It reveals the importance of lipophilicity and other molecular descriptors in the anticancer activity of these molecules. Indeed, these compounds have a high cytotoxicity several cancer cells including: leukaemia, colon, prostate, breast, ovary..... First, we determined the lipophilicity of ten (10) Makaluvamines for which the experimental values of FC_{50} concentrations on the mentioned cancer cells are known. The lipophilic values are calculated using four methods: Kowwin/LogP, ACD/LogP, A/LogPS and MI/LogP. The experimental values of the FC_{50} concentrations of these Makaluvamines are known. After that, we did a QSAR study of these ten molecules. This study includes the different lipophilic values calculated with other descriptors estimated at the B3LYP/6-311++G(d,p) theory level. Finally, the theoretical FC_{50} concentrations of the other six Makaluvamines were predicted with the four models established from the ten molecules.

Keywords: lipophilicity, Makaluvamines, QSAR, molecular descriptors and anticancer activity

Cite This Article: Sékou DIOMANDÉ, and Soleymane KONÉ, "Lipophilicity and QSAR Study of a Series of Makaluvamines by the Method of the Density Functional Theory: B3LYP/6-311++G(d,p)." *Journal of Materials Physics and Chemistry*, vol. 7, no. 1 (2019): 20-28. doi: 10.12691/jmpc-7-1-3.

1. Introduction

Over the past quarter century, global marine sources have proven to be a rich source of a wide range of new compounds of medical value [1] such as Makaluvamines. These molecules have shown anti-tumor activities against many cancer cells [2]. They have attracted scientists from different disciplines such as organic chemistry, bio-organic chemistry, pharmacology and biology with the aim of developing new drugs. In order to better understand the absorption of drugs into the human body, knowledge of lipophilicity or partition coefficient is essential. Indeed, the water/octanol (LogP) or lipophilic partition coefficient is of crucial importance in predicting pharmacological activity because its transport, its passage through the membranes can be conditioned by its partition between a lipid phase and an aqueous phase [3,4]. The present study aims to determine the lipophilicity of some Makaluvamines in different forms (neutral, protonated and methylated) on the one hand and to establish a quantitative structure-activity relationship (QSAR) on the other hand. This could help to better optimize the process of intercalation of these molecules in topoisomerase II. The lipophilicity calculations are performed using 4 software programs: KowWIN/LogP,

ACD/LogP, A/LogPS and MI/LogP. In addition to lipophilicity, molecular descriptors are used in the research of the QSAR model. These descriptors have been calculated with the DFT method at the B3LYP/6-311++G(d,p) theory level. We will rely on experimental concentration data reducing fluorescence by 50% (FC_{50}) for the research of the QSAR model. This model was developed using XLSTAT software. Each Makaluvamines can be neutral, protonated or methylated, so the models established will be used to predict the FC_{50} concentrations of structures for which experimental values are not available.

1.1. Theoretical Approach to Lipophilicity

The transport, passage through the membranes and pharmacological activity of a molecule can be conditioned by its division between a lipid phase and an aqueous phase, i.e. its hydrophilic character. This can be quantified by the Octanol-Water partition coefficient, noted (logP), which measures the differential solubility of a solute in these two non-miscible solvents [5]. This is an important measure for identifying drug similarity. According to the Lipinski rule, oral drugs must have logP values greater than or equal to -2 and less than or equal to 5[6]. It is defined by formula (1):

$$\text{Log P} = \log \frac{[\text{octanol}]}{[\text{H}_2\text{O}]} \quad (1)$$

[Octanol] and [H₂O] are the concentrations of solute in Octanol and water.

Compounds with $\log P > 0$ values are called lipophilic, and compounds with $\log P < 0$ values are called hydrophilic. If the LogP is positive and very high, it expresses that the molecule is more soluble in Octanol than in water. This reflects its lipophilic character, and inversely, if the LogP is negative it means that the molecule is hydrophilic. A null LogP means that the molecule is as soluble in one solvent as in the other.

2.2. Quantitative Relationship Structure Activity (QSAR)

The principle of a QSAR study is to find a mathematical relationship quantitatively linking a biological activity, measured for a series of compounds having the same structural skeleton under the same experimental conditions, with molecular descriptors (physicochemical properties) using statistical methods [3]. In this context, a QSAR is designed as a statistical model that links a pool of structural descriptors of a chemical compound with its biological activities.

This can be translated by the following equation:

Activity

= f (physico – chemical and / or structural properties)

= f (D₁, D₂, ...).

D_i : molecular descriptor.

The objective of these studies is to analyze the structural data in order to detect the determining factors for the activity studied. The resulting mathematical expression can then be used as a rational way to predict

the studied activity for new molecules or molecules for which experimental data are not available. It is then possible for the pharmacochemist to affirm with some confidence that the properties taken into consideration play an important role in the distribution or even the mechanism of action of the active ingredients studied. Similarly, since physico-chemical properties are quantified by QSARs, the pharmacochemist can calculate by anticipation what should be the biological activity of a new analogue and thus focus his synthesis efforts on the analogues assumed to be more active. This reduces the time, cost of synthesis and the number of compounds to be synthesized becomes reasonable [3,7].

2. Materials and Calculation Methods

2.1. Materials

The study focused on the Makaluvamines in Figure 1. Molecules I and A are each considered in neutral and methylated form, we have ten (10) structures, seven (07) of which were used as a training set and three (03) for external validation. Concentrations reducing fluorescence by 50% (FC₅₀) range from 1.2 to 40 μM [8]. These FC₅₀ values represent the average of tests on about sixty cancer cells. This range of concentrations allowed us to define a quantitative relationship between cancer activity and theoretical descriptors. By convention, negative logarithms, i. e. logarithms of reciprocal molar concentrations (e. g. $\log 1/C$ or pC values), are preferred to obtain higher values for more active compounds (the values in this inverse scale are directly comparable to the pH and pK_a values) [9,10,11].

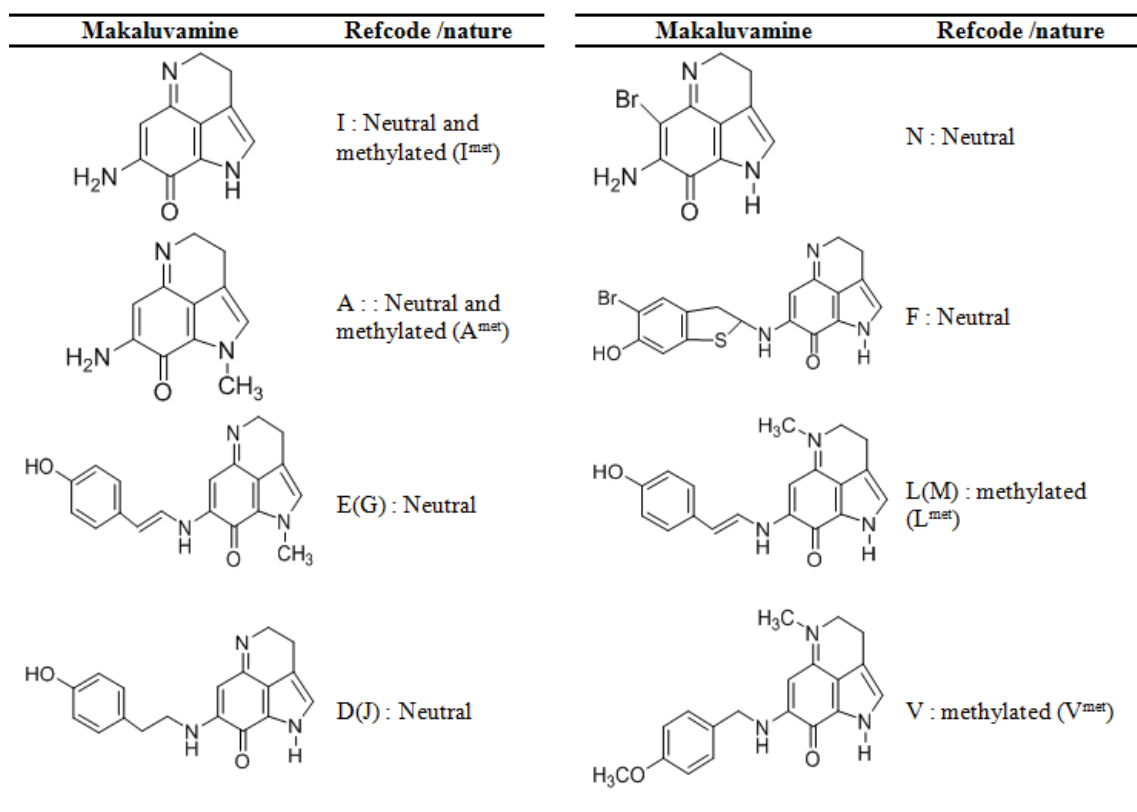


Figure 1. Structures of the studied Makaluvamines (NB: Neutral forms are noted by the refcode X, methylated forms by X^{met})

2.2. Calculation Methods

KowWIN/logP [12], ACD/logP [13] software from ACD/ChemSketch (Freeware) version 2016.2.2, MI/logP [14] and A/logPS [15] were used to calculate the Lipophilicity values. MI/logP and A/logPS are available from vclab.org. A/logPS is a global method, based on electronic state indices while KowWIN/logP, MI/logP and ACD/logP are substructure methods. KowWIN/logP is an atomic method including steric interactions, H-bonding and polar substructure effects. MI/logP is based on functional groups taking into account intramolecular H-bonds and interactions of charges. The ACD/logP is based on the contributions of individual atoms and structural fragments and on intramolecular interactions between the different fragments. The values are given with an uncertainty more or less equal to 0.6 and beyond this value the molecules are considered new for the ACD/logP program database.

2.3. Level of Theory

The relationship between the values of the biological activity of the molecules studied and their molecular structures has been demonstrated by theoretical chemistry calculations with Gaussian 09 [16] software. The density functional theory (DFT) [17] was used at the B3LYP/6-311++G(d,p) level for the calculation of molecular descriptors. Indeed, DFT methods are generally known to generate a variety of molecular properties [18-22] in QSAR studies. These methods increase predictability, reduce calculation time and involve drug conception costs. [23,24]. A molecular geometry optimization calculation is followed by a frequency calculation for each Makaluvamine structure. The global molecular descriptors that we have calculated are as follows: the electronic energy of the molecule (E_{elec}), the energies of the orbital frontiers (the highest occupied: ϵ_{HOMO} and the lowest vacant: ϵ_{LUMO}), first ionization energy (EI), electronic affinity (AE), hardness (η), softness (S), electrophilicity index (ω), chemical potential (μ_{pot}), electronegativity (χ), dipole moment (μ_D), polarizability (α), hydrogen-bond acidity (ϵ_A) and hydrogen-bond basicity (ϵ_B).

The modelling was obtained by the multilinear regression (MLR) method with the software implemented in Excel [25] and XLSTAT [26]. The calculation of the partial correlation coefficient between the descriptors studied must be less than 0.70 ($a_{ij} < 0.70$); this means that these different descriptors are independent from each other [27].

2.4. Estimation of the Predictive Capacity of a QSAR Model

The quality of a model is determined by certain criteria such as the determination coefficient R^2 , the standard deviation S , the correlation coefficients of the cross-validation Q^2_{CV} and the Fischer coefficient F . The statistical indicators R^2 , S and F concern the adjustment of calculated and experimental values. They describe the predictive ability within the limits of the model and allow the accuracy of the values calculated on the learning set to

be estimated [28,29]. The cross-validation coefficient provides information on the predictive power of the model. The R^2 coefficient expresses the dispersion of theoretical values around experimental values. The quality of the model is better when the points are close to the adjustment line [30]. The adjustment of the points to the line can be evaluated by the determination coefficient ϵ

$$R^2 = 1 - \frac{\sum (y_{i,exp} - \hat{y}_{i,theo})^2}{\sum (y_{i,exp} - \bar{y}_{i,exp})^2} \quad (2)$$

Where

$Y_{i,exp}$: Experimental value of anti-cancer activity

$\hat{y}_{i,theo}$: Theoretical value of anti-cancer activity

$\bar{y}_{i,exp}$: The mean value of the experimental values of anti-cancer activity.

The closer the value of R^2 to 1, the more the theoretical and experimental values are correlated.

Moreover, the variance σ^2 is determined by the relation 1:

$$\sigma^2 = S^2 = \frac{\sum (y_{i,exp} - y_{i,theo})^2}{n - k - 1} \quad (3)$$

Where k is the number of independent variables (descriptors), n is the number of molecules in the test or learning set and $n - k - 1$ is the degree of freedom. The standard deviation S is another statistical indicator used. It evaluates the reliability and precision of a model:

$$S = \sqrt{\frac{\sum (y_{i,exp} - y_{i,theo})^2}{n - k - 1}} \quad (4)$$

The Fischer coefficient F is also used to measure the level of statistical significance of the model, that is to say the quality of the choice of the descriptors constituting the model.

$$F = \frac{\sum (y_{i,theo} - y_{i,exp})^2}{\sum (y_{i,exp} - y_{i,theo})^2} * \frac{n - k - 1}{k} \quad (5)$$

The coefficient of determination of cross-validation Q^2_{CV} , makes it possible to evaluate the accuracy of the prediction on the test set and it is calculated by using the following equation:

$$Q^2_{CV} = \frac{\sum (y_{i,theo} - \bar{y}_{i,exp})^2 - \sum (y_{i,theo} - y_{i,exp})^2}{\sum (y_{i,theo} - \bar{y}_{i,exp})^2} \quad (6)$$

The performance of a model, according to Eriksson and al. [31] is characterized by a value of $Q^2_{CV} > 0.5$ for a satisfactory model and for the excellent model Q^2_{CV} is higher than 0.9. A training set of a model will be performing if the acceptance criterion $R^2 - Q^2_{CV} > 0.3$ is respected. Moreover, the predictive power of a model can

be obtained from the value of the ratio $\frac{\text{Log} \left(\frac{1}{C} \right)_{theo}}{\text{Log} \left(\frac{1}{C} \right)_{exp}}$ for the

validation set. The model is acceptable when the value of the ratio of theoretical and experimental activity tends towards unity.

3. Results et Discussions

3.1. Lipophilicity

The values of lipophilicity are indicated in the following table:

Table 1. Lipophilicity values of Makaluvamines calculated with different methods

Method	KOWWIN/ _{logP}			ACD/ _{logP}			A/ _{logPS}			MI/ _{logP}		
	neutral	protonated	methylated	neutral	protonated	methylated	neutral	protonated	methylated	neutral	protonated	methylated
Makaluvamines												
A	1.05	1.05	-5.1	-0.5±1.4	-0.5±1.4	-	0.58	0.58	0.73	0.84	-2.32	-3.14
I	0.5	0.5	-5.65	-1.0±1.4	-1.0±1.4	-	0.18	0.18	0.67	0.77	-2.39	-3.2
N	1.12	1.12	-5.04	-0.6±1.4	-0.6±1.4	-	0.99	0.99	0.86	1.51	-1.65	-2.47
D(J)	2.47	2.47	-3.69	1.0±1.4	1.0±1.4	-	2.35	2.35	0.98	2.47	-0.69	-1.5
E(G)	3.02	3.02	-3.13	0.6±1.4	0.6±1.4	-	2.68	2.68	1.11	2.75	-0.41	-1.23
F	3.5	3.5	-2.65	2.2±1.5	2.2±1.5	-	3.34	3.34	0.85	3.67	0.51	-0.31
L(M)	2.47	2.47	-3.39	0.2±1.4	0.2±1.4	-	2.43	2.43	0.98	2.68	-0.48	-1.37
V	2.76	2.76	-3.68	1.3±1.4	1.3±1.4	-	2.21	2.21	2.43	2.6	-1.09	-1.3

All values of the partition coefficients of the different forms are less than 5. According to the Lipinski rule, these molecules have good membrane-penetration. The values of lipophilicity obtained by the KowWIN/LogP, ACD/LogP and A/LogP methods are identical for neutral and protonated forms. The MI/LogP method differentiates them. It shows that neutral forms are more lipophilic. In neutral or protonated form, Makaluvamine F is the most lipophilic with a higher partition coefficient value. It can be noted that these molecules all have an uncertainty greater than 0.6 according to the ACD/LogP method. Then, they are new for the ACD program database. The ACD/LogP method does not allow us to access values of lipophilicity concerning methylated forms. The KowWIN/logP and MI/logP methods give negative values of lipophilicity contrary to the A/logP method.

3.2. Molecular Descriptors

The values of the molecular descriptors are grouped in [Table 2](#).

Table 2. Molecular descriptors calculated at level B3LYP/6-311++G(d,p)

Makaluvamine	E_{elec} (kcal/mol)	ϵ_{HOMO} (eV)	ϵ_{LUMO} (eV)	EI (eV)	AE (eV)	η (eV)	S (eV) ⁻¹	ω (eV)	μ_{pot} (eV)	χ (eV)	μ_D (D)	α (u.a)	ϵ_A (eV)	ϵ_B (eV)
A	-417581.09	-5.75	-2.53	5.75	2.53	1.61	0.62	0.81	-1.61	4.14	2.71	164.97	0.06	0.05
A^{met}	-442512.84	-9.92	-6.80	9.92	6.80	1.56	0.64	0.78	-1.55	8.36	3.53	172.53	0.02	0.09
I	-392907.50	-5.83	-2.61	5.83	2.61	1.61	0.62	0.81	-1.61	4.22	2.35	151.55	0.06	0.05
I^{met}	-417836.83	-10.07	-6.96	10.07	6.96	1.55	0.64	0.78	-1.56	8.52	3.43	158.90	0.02	0.09
N	-2007882.86	-5.89	-2.81	5.89	2.81	1.54	0.65	0.77	-1.54	4.35	3.94	172.05	0.06	0.05
D(J)	-634505.12	-5.51	-2.50	5.51	2.50	1.51	0.66	0.75	-1.51	4.00	2.40	256.96	0.06	0.05
E(G)	-658412.08	-5.51	-2.40	5.51	2.40	1.56	0.64	0.78	-1.56	3.96	3.41	268.75	0.06	0.05
F	-2498602.54	-5.95	-2.75	5.95	2.75	1.60	0.63	0.80	-1.60	4.35	5.44	298.60	0.06	0.05
L^{met}(M^{met})	-658675.03	-8.42	-6.46	8.42	6.46	0.98	1.02	0.49	-0.98	7.44	5.23	395.46	0.02	0.08
V^{met}	-659429.87	-8.59	-6.57	8.59	6.57	1.01	0.99	0.51	-1.01	7.58	10.13	275.39	0.02	0.08

The interpretation of these molecular descriptors has already been the subject of an article on Makaluvamines [32]. They are just used in this work to formulate QSAR models.

3.3. QSAR Models

We have developed four (04) QSAR models. Indeed, this is due to the fact that the lipophilicity of each molecule was calculated using four (04) different methods. The results of each method are combined with the other molecular descriptors to search a model. For each model, seven (07) molecules are used for the training set and three (3) for the validation set. The experimental data of FC₅₀ concentrations are relative to the activities of the molecules on various sources of cancer cells of leukaemia, colon, prostate, breast, ovary, renal....[8].

3.3.1. Model 1

This model was established with lipophilicity obtained from the KOWWIN/LogP software and other molecular descriptors.

3.3.1.1. Model 1 Training and Validation Set

The data used to build this model are presented in [Table 3](#). As for the partial correlation coefficients between the descriptors, they are summarized in [Table 4](#).

Analysis of the data in [Table 4](#) shows that the value of a_{ij} between two descriptors is less than 0.70. This shows that the descriptors used in this model are independent.

Table 3. Concentrations reducing fluorescence by 50% (FC₅₀) [8] and molecular descriptors of the training and validation set

Makaluvamines	FC ₅₀ (μM)	Log(1/C)	LogP	α (u.a)	E_{elec} (kcal/mol)	μ_D (D)
I	17	4.77	0.50	151.55	-392907.50	2.35
I^{met}	16	4.80	-5.65	158.90	-417836.83	3.43
A	20	4.70	1.50	164.97	-417581.09	2.71
A^{met}	16	4.80	-5.10	172.53	-442512.84	3.54
D	10	5.00	2.47	256.96	-634505.12	2.40
E	1.6	5.80	3.02	268.75	-658412.08	3.41
F	12	4.92	3.50	298.60	-2498602.54	5.44
N	18	4.75	1.12	172.05	-2007882.86	3.94
L^{met}	1.2	5.92	-3.39	395.46	-658675.03	5.23
V^{met}	40	4.40	-3.68	275.39	-659429.87	10.13

Table 4. Values of partial correlation coefficients (α_{ij}) of the descriptors

	LogP	α (u.a)	E_{elec} (kcal/mol)	μ_D (D)
LogP	1.00	0.24	-0.57	-0.52
α (u.a)	0.24	1.00	0.03	0.50
E_{elec} (kcal/mol)	-0.57	0.03	1.00	-0.38
μ_D (D)	-0.52	0.50	-0.38	1.00

3.3.1.2. Validation of Model 1

The contribution of a descriptor to anticancer activity in correlation with other descriptors in the regression equation depends not only on the sign of its coefficient but also on its own sign. When the descriptor and its coefficient are the same sign, it improves biological activity. On the other hand, if they have opposite signs, the descriptor reduces the activity. Equation (7) below represents the best model based on the data in Table 3.

$$\log\left(\frac{1}{C}\right) = 4.52 - 2.48 \cdot 10^{-2} \cdot \text{LogP} + 2.86 \cdot 10^{-3} \cdot \alpha - 8.35 \cdot 10^{-8} \cdot E_{elec} - 0.10 \cdot \mu_D \quad (7)$$

With:

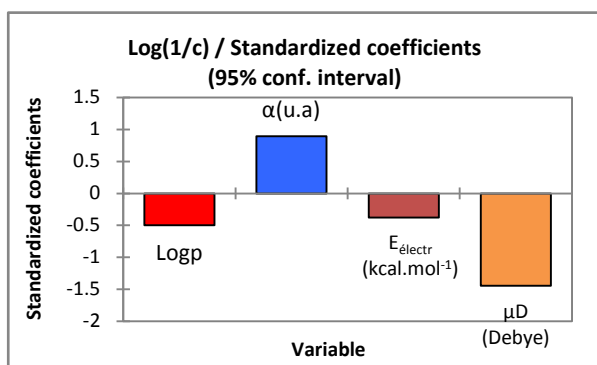
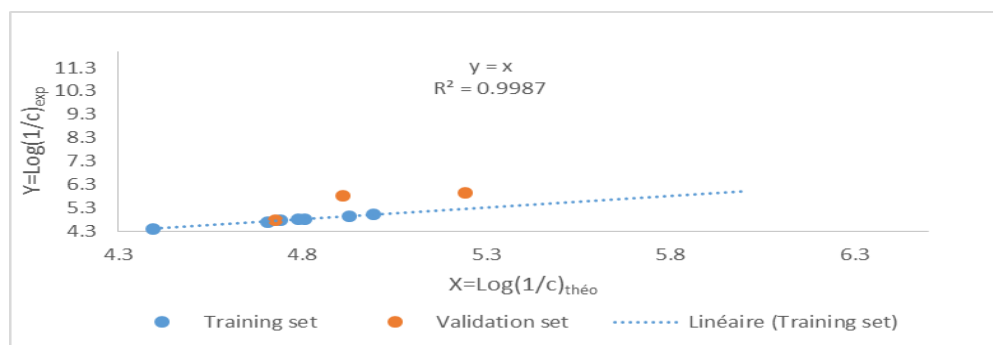
$$\alpha \text{ (expressed in u.a)}, E_{elec} \text{ (expressed in } \frac{\text{kcal}}{\text{mol}} \text{)}$$

$$\text{et } \mu_D \text{ (expressed in D)}$$

$$R^2 = 0.999; Q_{CV}^2 = 0.936; F = 385.981; S = 0.789; N = 7;$$

$$R^2 - Q_{CV}^2 = 0.063$$

Model 1 exalts four (04) descriptors: lipophilicity, polarizability, electronic energy and dipole moment. Figure 2 below shows the coefficients (sign and importance) assigned to these descriptors.

**Figure 2.** Coefficients defining the contributions of the four descriptors to anticancer activity in model 1**Figure 3.** Mean regression of experimental anticancer activity in relation with theoretical anticancer activity for all molecules of the training set and the validation set of model 1

According to this model, low lipophilicity ($\text{LogP} < 0$) contributes to improving the anticancer activity of a molecule in the series. This is the case for the Makaluvamines A^{met} , I^{met} , L^{met} and V^{met} . These molecules are all in charged form. The model predicts that high polarizability (α) will improve the anticancer activity of any molecule in the series. Makaluvamine L^{met} has the highest polarizability of the series. Since the electronic energy is always negative, equation (7) shows that the lower the energy of a molecule in the series, the more it improves the activity. The model predicts a decrease in anticancer activity of one tenth of the dipole moment value of any molecule in the series.

The values of the Fischer coefficient (F), the cross-validation coefficient (Q_{CV}^2) et de $R^2 - Q_{CV}^2$ show that the model of equation (7) is acceptable.

The external validation of the model was performed with molecules I, E and L^{met} . The values of the ratio

$$\frac{\log\left(\frac{1}{C}\right)_{theo}}{\log\left(\frac{1}{C}\right)_{exp}}$$

of the external validation set are reported in Table 5.

Table 5. Experimental and theoretical activities, and their relationship to the molecules in the validation set.

Molécules	$\text{Log}\left(\frac{1}{C}\right)_{exp}$	$\text{Log}\left(\frac{1}{C}\right)_{theo}$	$\frac{\log\left(\frac{1}{C}\right)_{theo}}{\log\left(\frac{1}{C}\right)_{exp}}$
Maka I	4.77	4.73	1.00
Maka E	57.95	4.91	0.80
Maka L^{met}	5.92	5.24	0.90

The values of the reports $\frac{\log\left(\frac{1}{C}\right)_{theo}}{\log\left(\frac{1}{C}\right)_{exp}}$ are close to 1.

These results indicate a good correlation between the experimental and theoretical values of the anticancer activity of the molecules studied. Figure 3 shows the regression line between the experimental and theoretical anticancer activity of the training set (blue dots) and the test set (red dots).

Table 6. Concentrations reducing fluorescence by 50% (FC₅₀) [8] and molecular descriptors of the training and validation set of models 2, 3 and 4

Makaluvamines	FC ₅₀ (μ M)	Log(1/C)	Model 2			Model 3		Model 4			
			LogP	α (u.a)	E_{el} ($\frac{kcal}{mol}$)	LogP	α (u.a)	LogP	EI (ev)	η (ev)	μ_D (D)
I	17	4.77	-0.95	151.55	-392907.50	0.18	151.55	0.77	5.83	1.61	2.35
I^{met}	16	4.80	-0.95	158.90	-417836.83	0.67	158.90	-3.20	10.07	1.55	3.43
A	20	4.70	-0.49	164.97	-417581.09	0.58	164.97	0.84	5.75	1.61	2.71
A^{met}	16	4.80	-0.49	172.53	-442512.84	0.73	172.53	-3.14	9.92	1.56	3.54
D	10	5.00	0.99	256.96	-634505.12	2.35	256.96	2.47	5.51	1.51	2.40
E	1.6	5.80	0.63	268.75	-658412.08	2.68	268.75	2.75	5.51	1.56	3.41
F	12	4.92	2.23	298.60	-2498602.54	3.34	298.60	3.67	5.95	1.60	5.44
N	18	4.75	-0.59	172.05	-2007882.86	0.99	172.05	1.51	5.89	1.54	3.94
L^{met}	1.2	5.92	0.17	395.46	-658675.03	0.98	395.46	-1.37	8.42	0.98	5.23
V^{met}	40	4.40	1.32	275.39	-659429.87	2.43	275.39	-1.30	8.59	1.01	10.13

Table 7. Expressions of models 2, 3, 4 and their different coefficients

Model	Expression of the model	R ²	Q _{CV} ²	F	S	R ² - Q _{CV} ²	N
2	$Log(\frac{1}{C}) = 3.72 - 0.18 * LogP + 5.71 * 10^{-3} * \alpha + 3.60 * 10^{-8} * E_{elec}$	0.998	0.844	501.388	0.535	0.154	7
3	$Log(\frac{1}{C}) = 4.02 - 0.19 * LogP + 5.25 * 10^{-3} * \alpha$	0.984	0.894	123.841	0.534	0.09	7
4	$Log(\frac{1}{C}) = 4.49 + 0.16 * LogP + 0.19 * EI - 0.45 * \eta - 0.10 * \mu_D$	0.998	0.926	213.791	0.968	0.072	7

Table 8. Values of partial correlation coefficients (a_{ij}) of the descriptors of models 2, 3 and 4.

	Model 2			Model 3		Model 4			
	LogP	α (u.a)	E_{el} ($\frac{kcal}{mol}$)	LogP	α (u.a)	LogP	EI (ev)	η (ev)	μ_D (D)
LogP	1.00	0.67	-0.50	1.00	0.56	1.00	-0.94	0.36	0.50
α (u.a)	0.67	1.00	0.21	0.56	1.00				
E_{el} ($\frac{kcal}{mol}$)	-0.50	0.21	1.00						
EI (ev)						-0.94	1.00	0.27	0.49
η (ev)						0.36	0.27	1.00	-0.76
μ_D (D)						0.50	0.49	-0.76	1.00

We used values of lipophilicity obtained with the three other methods ACD/LogP, A/LogPS and MI/Log and other molecular descriptors in Table 2 to establish three QSAR models 2, 3 and 4 respectively. Table 6 contains, in addition to the experimental data, the descriptors that are used in the expressions of these models. Table 7 contains the mathematical expressions of the said models and the values of the different coefficients: determination R², standard deviation S, correlation of cross-validation Q_{CV}² and Fischer F.

For these three models as for model 1, R² is close to 1. The experimental and theoretical values correlate well. All values of Q_{CV}² > 0.5, R² - Q_{CV}² < 0.3. The training set of each model is efficient. The high values of the F coefficient also reflect a good statistical significance of the model and the quality of the choice of the descriptors that make up the model.

The partial correlation coefficients between the descriptors of models 2, 3 and 4 are grouped in Table 8.

Analysis of the data in this table shows the a_{ij} coefficient between two descriptors of a given model is less than 0.70. This shows that these descriptors in the model are independent.

3.3.1.3. Validation of Models 2, 3 and 4

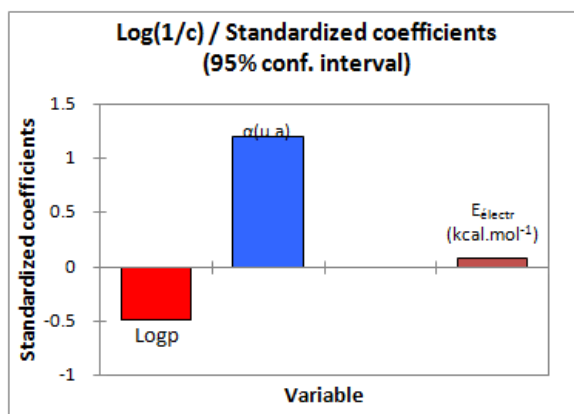
Model 2 presents the contribution of three (03) descriptors to the anticancer activity of the series of compounds. These are lipophilicity, polarizability and electronic energy. As for model 3, two descriptors,

lipophilicity and polarizability, contribute to the anticancer activity. Model 4 correlates four (04) descriptors to predict the anticancer activity of the molecules studied. These are lipophilicity, first ionization energy, hardness, and dipole moment.

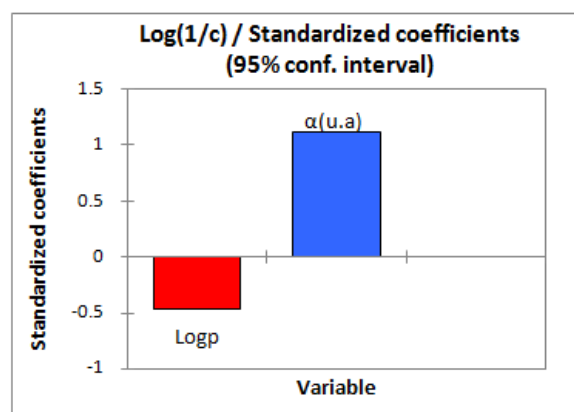
Figure 4 shows the coefficients assigned to the different descriptors of models 2 (4a), 3 (4b) and 4 (4c).

Models 2 and 3 show that negative lipophilicity (LogP < 0) improves the activity of a Makaluvamine. However, if LogP > 0, activity decreases. A high value of polarizability contributes to increasing the activity of the molecules in the series studied. According to model 2, the high values of electron energy, which are lower in absolute values, should allow a better anticancer activity of these molecules. Model 4 shows that a high value of lipophilicity and also the first ionization energy strongly contributes to improve the anticancer activity of a molecule in the series studied. For a good contribution to this activity in such a molecule, the hardness and dipole moment must be low. The molecule must be slightly polar and soft. External validation was performed with molecules I, E and V^{met} and for models 2 and 3 then with molecules I, E and L^{met} for

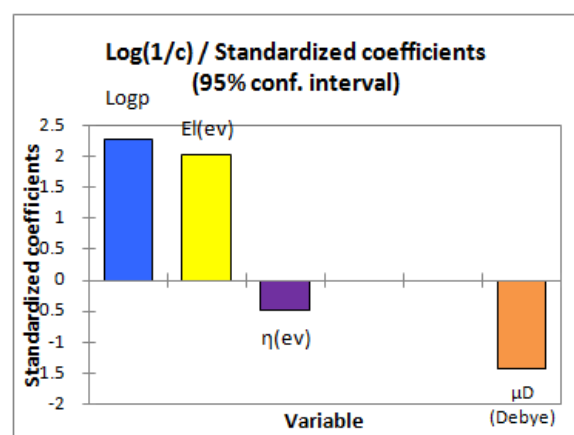
model 4. The values of the ratio $\frac{\log(\frac{1}{C})_{theo}}{\log(\frac{1}{C})_{exp}}$ of the external validation set are shown in Table 9.



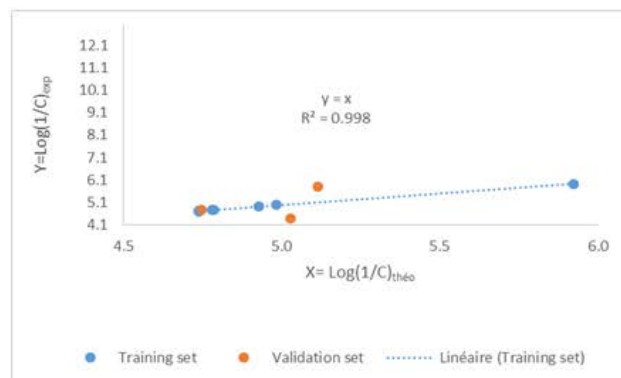
a



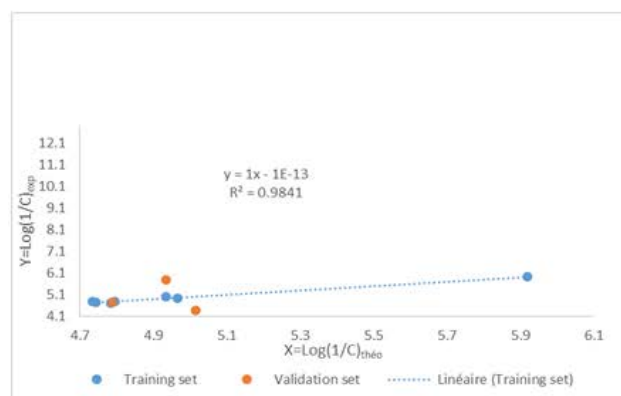
b



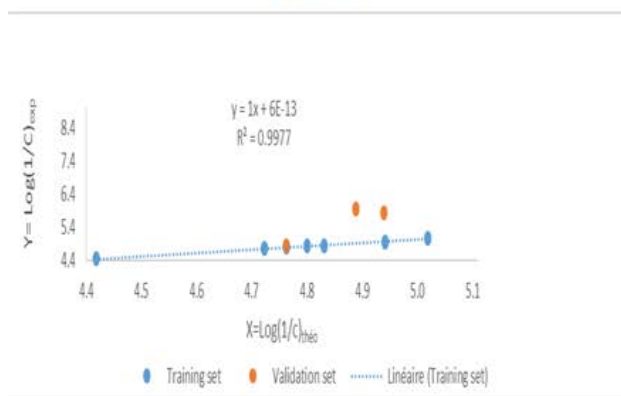
c



Model 2



Model 3



Model 4

Figure 4. Coefficients attributed to the different descriptors of models 2, 3 and 4

Figure 5. Mean regression of experimental anticancer activity in relation with theoretical anticancer activity for all molecules in the training set and the validation set of three models 2, 3 and 4

Table 9. Experimental and theoretical activities, and their ratio for the molecules in the validation set

Molecules	Model 2			Model 3			Model 4		
	$\text{Log}(1/C)_{\text{exp}}$	$\text{Log}(1/C)_{\text{theo}}$	$\frac{\text{Log}(1/C)_{\text{theo}}}{\text{Log}(1/C)_{\text{exp}}}$	$\text{Log}(1/C)_{\text{exp}}$	$\text{Log}(1/C)_{\text{theo}}$	$\frac{\text{Log}(1/C)_{\text{theo}}}{\text{Log}(1/C)_{\text{exp}}}$	$\text{Log}(1/C)_{\text{exp}}$	$\text{Log}(1/C)_{\text{theo}}$	$\frac{\text{Log}(1/C)_{\text{theo}}}{\text{Log}(1/C)_{\text{exp}}}$
I	4.77	4.75	1.00	4.77	4.79	1.00	4.77	4.74	1.00
E	5.80	5.12	0.90	5.80	4.93	0.90	5.80	4.92	0.80
V ^{met}	4.40	5.03	1.10	4.40	5.02	1.10			
L ^{met}							5.92	4.87	0.80

The ratio values $\frac{\text{Log}(1/C)_{\text{theo}}}{\text{Log}(1/C)_{\text{exp}}}$ are close to 1 for the three models. These results reflect a good correlation between the experimental and theoretical values of the anticancer activity of the molecules studied. The regression lines between the experimental and theoretical anticancer activity of the training set (blue points) and the test set (red points) for these three models are contained in Figure 5.

These plots show a good correlation between the experimental data and the results of theoretical calculations of the anticancer activity of the molecules studied.

All four established models show the contribution of lipophilicity to anticancer activity, three of them show the contribution of polarizability. The dipole moment and electronic energy appear in two of the models. By referring to the quality parameters, models 2 and 3 are the best. They confirm the importance of two descriptors in predicting and improving anticancer activities: lipophilicity and polarizability. It should also be noted that according to the Lipinski rule, lipophilicity (LogP) is important for the identification of similarity with a drug [27]. Polarizability is also one of the parameters that reflects the molecular properties related to hydrophobicity and therefore to biological activities [5].

3.3.2. Prediction of Anticancer Activity: Theoretical Determination of FC_{50} concentrations

For all forms (neutral, protonated and methylated) of each Makaluvamine C, H, K(P) and O in Figure 6 and

those of the compounds in Figure 1 for which experimental FC_{50} concentration values are not available, we made a theoretical estimate. The four models established were applied.

The four values of lipophilicity (KowWIN/logP, ACD/LogP, A/LogP and MI/LogP) of the different forms of these Makaluvamines and the different molecular descriptors that are exalted in the models have been calculated. The application of the models led to the results in Table 10.

These data cannot be compared due to the lack of experimental values. These are now new data to be exploited. Despite the differences between the neutral, protonated and methylated forms, it is models 2 and 3 that give the best agreement with the experimental values of FC_{50} of Makaluvamines in the training and validation set of the established models. The experimental FC_{50} concentrations of Makaluvamines L^{met} and E show that these molecules are the most effective. It is important to note that the theoretical values, derived from our calculations, which are generally low for forms L, LH⁺, EH⁺ and E^{met} always reflect their efficacy.

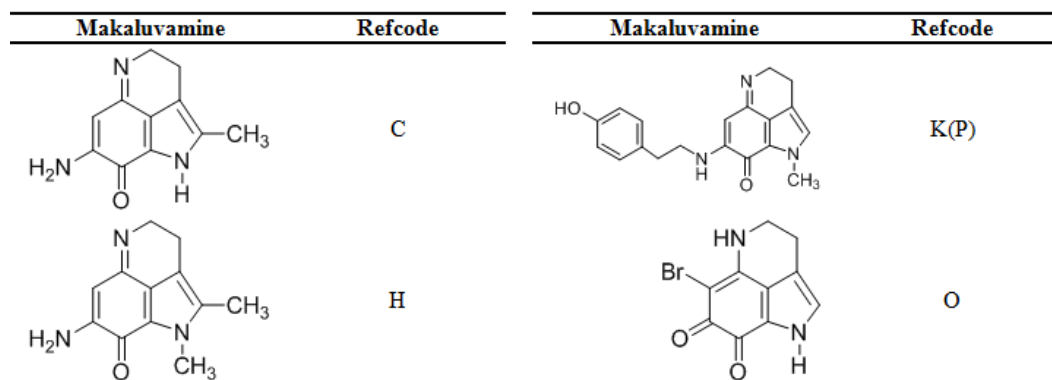


Figure 6. Structures of Makaluvamines without experimental FC_{50} values

Table 10. Theoretical values of FC_{50} (μ M) concentrations of the anticancer activity of a series of Makaluvamines calculated from models 1 (M1), 2 (M2), 3 (M3) and 4 (M4)

Molecules	M ₁ : FC_{50}	M ₂ : FC_{50}	M ₃ : FC_{50}	M ₄ : FC_{50}
IH ⁺	33.1	49.0	20.9	3.6
AH ⁺	31.6	52.5	22.4	3.3
DH ⁺	46.8	30.9	12.6	5.5
D ^{met}	55.0	2.1	0.7	95.5
EH ⁺	9.8	6.2	2.9	2.6
E ^{met}	8.9	0.4	0.2	35.5
FH ⁺	316.2	89.1	8.7	44.7
F ^{met}	177.8	7.4	0.6	371.5
NH ⁺	24.6	186.2	21.4	3.0
N ^{met}	16.6	12.0	1.2	32.4
L	13.8	30.9	12.9	11.0
LH ⁺	10.2	4.9	2.3	3.1
VH ⁺	81.3	32.4	13.5	8.5
V	11.8	31.6	13.2	8.7
C	20.4	18.2	16.2	20.0
CH ⁺	25.1	19.5	17.8	11.0
C ^{met}	15.5	16.2	15.5	14.5
H	20.4	18.6	14.8	20.9
HH ⁺	24.5	20.0	15.9	11.5
H ^{met}	15.8	16.6	13.8	15.9
K(P)	97.7	10.5	11.5	10.5
KH ⁺ (PH ⁺)	40.7	10.7	11.8	19.5
K ^{met} (P ^{met})	17.8	9.1	5.1	20.9
O	70.8	23.4	27.5	70.8
OH ⁺	144.5	25.7	3ces 0.2	67.6
O ^{met}	114.8	22.4	19.1	66.1

4. Conclusion

The QSAR study carried out on a series of Makaluvamines made it possible to identify major descriptors in the prediction of the anticancer activity of the family of the said molecules. These include lipophilicity (LogP) and polarizability (α). In addition to these two descriptors, the dipole moment can also be decisive. When it appears as a descriptor in a model, the molecule must be low polar. The prediction of the theoretical anticancer activity of molecules without experimental FC₅₀ has demonstrated the efficacy of L, LH⁺, EH⁺ and E^{met} molecules in Models 2 and 3. Thus, this study opens new ways in the conception of new molecules with improved biological activities in the effective fight against certain cancers

References

- [1] Blunt, J.W.; Copp, B.R.; Munro, M.H.G.; Northcote, P.T.; Prinsep, M.R. *Nat. Prod. Rep.* 2005, 22, 15.
- [2] Dijoux, M.-G.; Schnabel, P. C.; Hallock, Y. F.; Boswell, J. L.; Johnson, T. R.; Wilson, J. A.; Ireland, C. M.; Van Soest, R.; Boyd, M. R.; Barrows, L. R.; Cardellina, J. H. *Bioorg. Med. Chem.* 2005, 13, 6035-6044.
- [3] Chtita S. Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSAR. Recherche de nouveaux médicaments [thèse]. [Meknès]: Moulay Ismail; 2017.
- [4] Numbury Surendra Babu, Didugu Jayaprakash. Global and Reactivity Descriptors Studies of Cyanuric Acid Tautomers in Different Solvents by using of Density Functional Theory (DFT). *Int J Sci Res IJSR.* 2015; 4(6) :615-20.
- [5] P.J. Taylor, —Hydrophobic Properties of Drugs, In *Quantitative Drug Design*, Pergamon Press, Oxford (UK), 4, 1990, 241-294.
- [6] C.A. Lipinski, F. Lombardo, B.W. Dominy, and P.J. Feeney, —Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Advanced Drug Delivery Reviews*, 6(1-3), 1997, 3-25.
- [7] Patrick GL. Chimie pharmaceutique. Paris : De Boeck; 2003.
- [8] Marie-Geneviève Dijoux, Peter C. Schnabel, Yali F. Hallock, Jamie L. Boswell, Tanya R. Johnson, Jennifer A. Wilson, Chris M. Ireland, Rob Van Soest, Michael R. Boyd, Louis R. Barrows and John H. Cardellina, II, Antitumor activity and distribution of pyrroiminoquinones in the sponge genus *Zyzzya*, *Bioorganic and Medicinal Chemistry*, 13, 2005, 6035-6044.
- [9] S.Chalteejee, A.Hadi, B. Price, *Regression Analysis by Examples; Wiley VCH: New York, USA, 2000.*
- [10] Phuong HTN, thèse de doctorat « Synthèse et étude des relations structure/activité quantitatives (QSAR/2D) d'analogues Benzo[c]phénanthridiniques », Université d'Angers, (France), 2007.
- [11] Kubinyi Hugo, *QSAR: Hansch analysis and related approaches*, Weinheim; New York; Basel; Cambridge; Tokyo; VCH, 1993.
- [12] T.Cheng, Y.Zhao, X.Li, F. Lin, Y.Xu, X. Zhang, Y.Li, R. Wang, L.Lai, *J.Chem.Inf.Model.*; 47, 2007, 2140-2148.
- [13] Petrauskas A, Kolovanov EA. *Perspect Drug Discov Des* 18, 2000, 19-38.
- [14] *molinspiration*, <http://www.molinspiration.com>.(Septembre 2018)
- [15] I.V.Tetko,V.Y.Tanchuk, *J.Chem.Inf.Comput.Sci.* 42, 2002, 1136-1145.
- [16] Gaussian 09, Revision A.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria,
- [17] M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2009.
- [18] Hohenberg, P.; Kohn, W. Inhomogeneous electron gas, 1964. *Phys.Rev.* 136, B864.
- [19] P.K. Chattaraj, A.Cedillo, and R.G Parr, *J. Phys.Chem.*, 1991, 103: 7645.
- [20] P.W.Ayers, and R.G.Parr, *J.Am Chem., Soc.*, 2000, 122:2010.
- [21] F. De Proft, J.M.L.Martin, and P. Geerlings, *Chem. Phys. Let.*, 1996, 250:393
- [22] P.Geerlings, F. De Proft, J.M.L.Martin, In *Theoretical and Computational Chemistry*; Seminario, J., Ed.,; Elsevier; Amsterdam, 1996, 4 (Recent Developments in Density Functional Theory): 773.
- [23] F.De Proft, J.M.L. Martin, and P. Geerlings, *Chem. Phys.Let.*, 1996, 256: 400.
- [24] C.Hansch, P.G. Sammes, and J.B. Taylor, *Computers and the medicinal chemist*; in: *Comprehensive Medicinal Chemistry*, 1990, 4, Eds. Pergamon Press, Oxford: 33-58.
- [25] R. Franke, *Theoretical Drug Design Methods*, Elsevier, Amsterdam, 1984.
- [26] Microsoft® Excel® 2013 (15.0.4420.1017) MSO (15.0.4420.1017) 64 Bits (2013) Partie de Microsoft Office Professionnel Plus.
- [27] XLSTAT Version 19.5.47062 (64 bit) Copyright 1995-2018 (2018) XLSTAT and Addinsoftware Registrered Trademarks of Addinsoft. <https://www.xlstat.com>.
- [28] A.Vessereau, *Méthodes statistiques en biologie et en agronomie*. Lavoisier (Tec and Doc). Paris 1988 : 538
- [29] G.W. Snedecor, W.G. Cochran, *Statistical Methods*; Oxford and IBH: New Delhi, India; 1967: 381.
- [30] M.V.Diudea, *QSAR/QSAR Studies for Molecular Descriptors*; Nova Science: Huntingdon, New York, USA, 2000.
- [31] E.X.Esposito, A.J.Hopfinger, J.D.Madura, *Methods in Molecular Biology*, 2004, 275: 131-213.
- [32] L.Eriksson, J. Jaworska, A. Worth, M.T.D. Cronin, R.M.Mc Dowell, P.Gramatica, *Methods for Reliability and Uncertainly Assessment and for Applicability Evaluations of Classification and Regression-Based QSARs*, *Environmental Health Perspectives*, 2003, 111(10):1361-1375.
- [33] Sékou DIOMANDÉ, Affoué Lucie BÉDÉ, Soleymane KONÉ and El-Hadji Sawaliho BAMBBA, Determination of protonation and methylation sites of neutral makaluvamines, relative stability and reactivity potential of the charged forms, *International Journal of Innovation and Applied Studies*, Vol. 25 No. 1 Dec. 2018, pp. 516-527.

