Modeling of a series of dihydropyrazole derivatives with antiproliferative activity by quantum chemical methods

ABSTRACT

Cancer is any pathology characterized by the rapid proliferation of abnormal cells within normal tissues of the body. Recent years have been marked by an increase in the number of cases of cancer, particularly that of the prostate. Cancer can affect any part of the body. Despite the efforts made in the fight against cancer in recent years, the objectives set by the WHO and the various supervisory centers in the countries have not been achieved, because the technical strategy for the fight against cancer in no way guarantees a cure without risk of relapse for patients in general, regardless of the continent. Today, we can offer models of molecules for therapeutic purposes using quantum chemical methods such as molecular modeling to predict better activity or by designing new molecules that are more active than existing ones. For this fact our work consists in establishing a mathematical model between the potential of the inhibitory concentration (pIC_{50}) and the descriptors such as the frequency of vibration υ(N-N), the angle of valence α(N-N-C) and the standard entropy of formation ($\Delta_f S^0$). This model is confirmed by very good indicators displaying the following values (**R^2= 0.9108; S= 0.1044; F=20.425**) for the MLR model and **(R^2=0.920; S=0.1393; F=92.823**) for the MNLR model. Moreover, these models allowed us to conclude that the vibration frequency υ(N-N) is the priority descriptor in the prediction of the antiproliferative activities of dihydropyrazole derivatives.

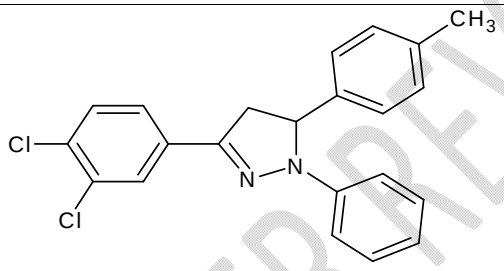
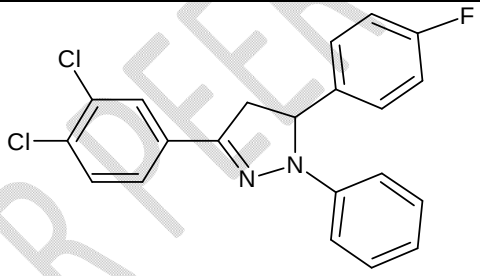
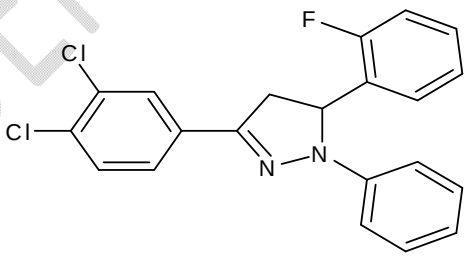
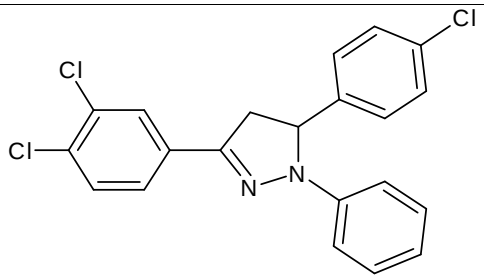
Keywords : QSAR, MNLR, MLR, antiproliferative activity (DU-145), Dihydropyrazole

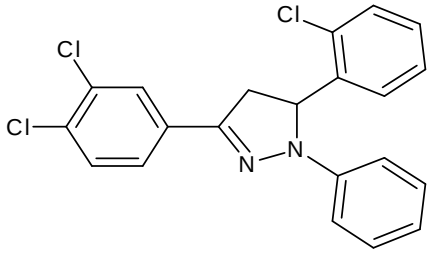
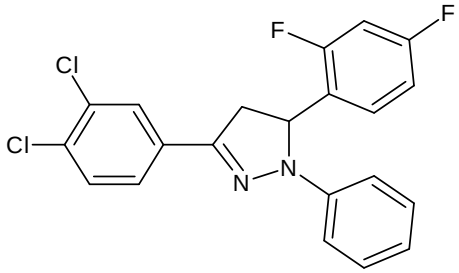
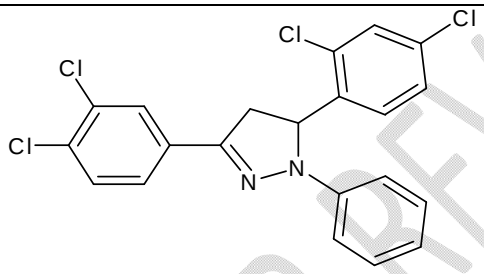
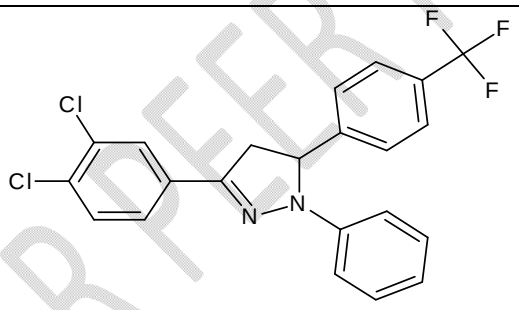
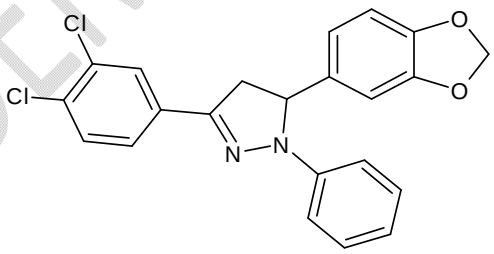
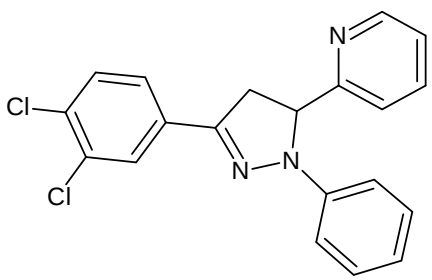
1. INTRODUCTION

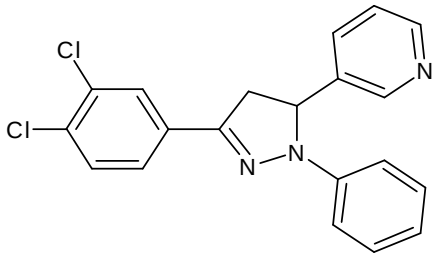
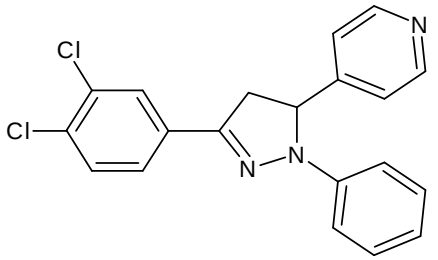
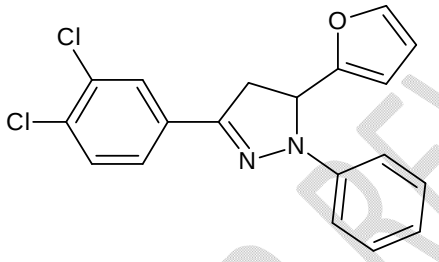
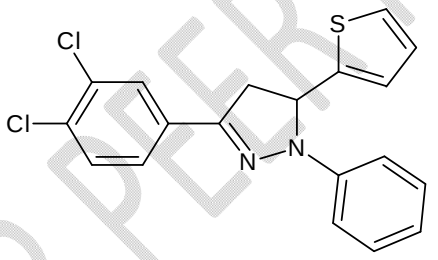
A cancer is a pathology characterized by the presence of one (or more) malignant tumor formed from the transformation by mutations or genetic instability of an initially normal cell [1]. When normal cells are damaged or cannot be cured, they die, which is apoptosis, which is not the case for cancer cells since they continue to grow to reach several cells. There are several types of cancer, in this work we will focus on prostate cancer. Prostate cancer has become the most common cancer and the second leading cause of cancer death in men [2], in developed countries and the first in underdeveloped countries. In 2020 more than 19.3 million cases recorded worldwide [3] and nearly 2747 in Côte d'Ivoire [4]. According to the World Health Organization (WHO), cancer causes nearly 10 million deaths, i.e. almost one in six deaths, cancer is one of the main causes of death in the world [1]. This health problem in men is becoming increasingly important. Today one in nine men are at risk of getting prostate cancer in their lifetime. The incidence of cancer increases considerably with age (60 years and over) due to the increasing accumulation of cancer risk factors and the fact that cell regeneration mechanisms tend to become less effective as aging [5]. Prevention against cancer, especially prostate cancer, is based on physical activity and also a healthy and balanced diet with medication. Surgical operations, chemotherapy are among other treatments used against this cancer. However, the proposal of new Therapeutic Targeted Molecules for the treatment of prostate cancer such as dihydropyrazole derivatives could be a contribution in the drug management of patients. Research has shown that dihydropyrazoles are an important pharmacophore involved in the discovery of new derivatives with therapeutic and biological activities [6]. Indeed, phenyl-substituted dihydropyrazoles exhibit proven biological activities such as antioxidant [2], antibacterial [3], antituberculous [5, 4], antifungal and significant anticancer [7,

8, 9]. Molecular modeling, the basis of theoretical chemistry, helps design and predict the properties of molecules of therapeutic interest using computer techniques. It is used to develop reliable models to predict and explain physico-chemical properties related to the biological activities of molecules or even to propose new molecules with more improved activities. QSAR (Quantitative Structure-Activity Relationship) modeling is one of the techniques that predicts the activities of chemical systems from their molecular structures. Indeed, the development of the QSAR model requires knowledge of the physico-chemical parameters (descriptors) related to the molecular systems studied. The determination of these descriptors is facilitated by the use of methods and tools of quantum chemistry. The compounds studied in our work (**Table 1**) were synthesized, tested and identified as molecules with effective activities against prostate cancer. It is up to us in this work to model, from the derivatives of these different series of molecules, molecules with more effective biological activities against prostate cancer. As general objective for this work, it is for us to contribute to model the antiproliferative profile (prostate cancer) of a series of dihydropyrazoles from their physico-chemical properties.

Table 1: Database molecules and their respective characteristics.

Code	Molecular structures	IC_{50}^{exp} (µM)
DHP1		351
DHP2		119
DHP3		327
DHP4		149

DHP5		134
DHP6		262
DHP7		206
DHP8		257
DHP9		131
DHP10		271

DHP11		211
DHP12		173
DHP13		162
DHP14		84

2. MATERIALS AND METHODS

2.1 Computational Theories Levels and Softwares

In order to predict the anticancer activity of dihydropyrazole derivatives, quantum chemical calculations were performed using Gaussian 09 software [10]. DFT methods are generally known to generate a variety of molecular properties [11-13] in QSAR studies. The latter increase the predictive capacity of QSAR models while shortening the computation time and cost implications in the design of new drugs [14,15]. The B3LYP/6-31+G (d, p) level of theory was used to determine the molecular descriptors. The fourteen (14) molecules used in this study have Inhibitory Concentrations ranging from 84 to 351 µM. The median inhibitory concentration (**IC_{50}**) is a measure of the effectiveness of a given compound in inhibiting a specific biological or biochemical function. Biological data is usually expressed as the opposite of the base-decimal logarithm of activity ($-\log_{10}(\mathbf{IC_{50}})$) to obtain better mathematical values when structures are biologically active [16,17]. The anticancer activity will be expressed by the potential of the inhibitory concentration pIC defined by equation (1):

$$pIC_{50} = -\log_{10}(IC_{50} * 10^{-6}) \quad (1)$$

Where MIC, the minimum inhibitory concentration in µg/mL.

The modeling was developed using three statistical learning methods. Namely the multiple linear regression (MLR) and non-linear (MNLR) methods which are implemented in Excel [18] and XLSTAT [19] spreadsheets.

2.2 Molecular Descriptors Calculation

In this work, to establish the model, we used geometric, spectroscopic and energetic descriptors. The geometric and spectroscopic descriptors used are the valence angle α(N-N-C) in degrees (°) and the nitrogen-nitrogen vibration frequency υ(N-N) in cm-1 (**figure 1**). These descriptors are illustrated in the figure below around the dihydropyrazole core.

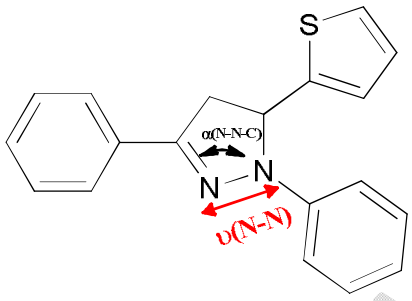


Figure 1: Geometric and spectroscopic descriptors

The standard entropy of compound formation was calculated from the relationship of Otchersky et al [10].

$$\Delta S_f^0(M, 298K) = S_M - \sum_{atoms} x\Delta S(298K) \qquad (2)$$

x : Number of Atoms of X in the Molecule

2.3. Estimation of the Predictive Capacity of a QSAR Model

The quality of a model is determined based on various statistical analysis criteria including the coefficient of determination $\mathbf{R^2}$, the standard deviation (**S**) or the Root of the Mean Square Errors (**RMSE**), the correlation coefficients cross-validation $\mathbf{Q^2_{CV}}$ and Fischer **F**. $\mathbf{R^2}$, **S** and **F** relate to the adjustment of calculated and experimental values. They describe the predictive capacity within the limits of the model, and make it possible to estimate the precision of the values calculated on the test set [20, 21]. As for the cross-validation coefficient $\mathbf{Q^2_{CV}}$, it provides information on the predictive power of the model. This predictive power is said to be "internal" because it is calculated from the structures used to build this model. The coefficient of determination $\mathbf{R^2}$ gives an evaluation of the dispersion of the theoretical values around the experimental values. The quality of the modeling is better when the points are close to the fitted line [22]. The adjustment of the points to this line can be evaluated by the coefficient of determination.

$$R^2 = 1 - \frac{\sum\left(y_{i,exp} - \hat{y}_{i,theo}\right)^2}{\sum\left(y_{i,exp} - \bar{y}_{i,exp}\right)^2} \qquad (3)$$

where:

$y_{i,exp}$: Experimental value of anticancer activity

$\hat{y}_{i,theo}$: Theoretical value of anticancer activity

$\bar{y}_{i,exp}$: Average value of experimental values of anticancer activity.

The closer the value of $\mathbf{R^2}$ is to 1, the more the theoretical and experimental values are correlated

The Root Mean Square RMSE Error is another statistical indicator used. It is used to assess the reliability and accuracy of a model:

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

The Fisher F test is also used to measure the level of statistical significance of the model, ie the quality of the choice of descriptors constituting the model.

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

The coefficient of determination of the cross-validation $\mathbf{Q^2_{CV}}$ makes it possible to evaluate the accuracy of the prediction on the learning set. It is calculated using the following relationship:

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

2.4. *Model Acceptance Criterion*

The performance of a mathematical model, for Eriksson et al. [23], is characterized by a value of $Q^2_{cv} > 0.5$. for a satisfactory model, while for the excellent model $Q^2_{cv} > 0.9$. According to these authors, given a test set, a model will perform well if the acceptance criterion $R^2 - Q^2_{cv} < 0.3$ is respected. In addition, the randomization test is used by the modeler to affirm that the good correlations between the descriptors and the activity presented by the QSAR model are not due to chance. To do this, the observations are disorganized, for example ten times, by randomly changing the activity column while keeping the descriptor columns fixed. We then obtain ten models with specific statistical characteristics. For an acceptable QSAR model, the average correlation coefficient (R_r) of the randomized models must be lower than the correlation coefficient (R) of the non-randomized model. The parameter R^2_{p} [24] developed by Roy K. is used to judge a QSAR model. This parameter can be calculated using the following equation:

$$R^2_p = R^2 * \sqrt{R^2 - R_r} \quad (7)$$

The value of R^2_p should be greater than 0.5 for an acceptable model. If the randomization of observations leads to weak prediction models, this means that the predictive capacities of the QSAR model built are not due to chance [25].

2.5. *Statistical Methods*

Multiple Linear and Non-Linear Regressions (MLR and MNLR)

The statistical technique of Multiple Linear Regression (MLR) is used to study the relationship between a dependent variable (Property) and several independent variables (descriptors). This statistical method minimizes the differences between actual and predicted values. It also made it possible to select the descriptors used as input parameters in the Multiple NonLinear Regression (MNLR). As for the analysis of Multiple Nonlinear Regression (MNLR), it also allows to improve the structure-property relationship in order to quantitatively evaluate the property. It is the most common tool for studying multidimensional data. It is based on the following pre-programmed XLSTAT functions:

$$y = a + (bx_1 + cx_2 + dx_3 + ex_4) + (fx_{12} + gx_{22} + hx_{32} + ix_{42}) \quad (8)$$

Where a, b, c, d,... represent the parameters and, x1, x2, x3, x4,... represent the variables.

2.6. Applicability Area

The domain of applicability of a QSAR model is the physico-chemical, structural or biological space, in which the model equation is applicable to make predictions for new compounds [26]. It corresponds to the region of the chemical space including the compounds of the training set and the similar compounds, which are close in this same space [27]. Indeed, the model, which is built on the basis of a limited number of compounds, by relevant descriptors, chosen from among many others, cannot be a universal tool for predicting the activity of any other molecule with trust. It appears necessary, even mandatory, to determine the DA of any QSAR model. This is what the Organization for Economic Co-operation and Development (OECD) recommends in the development of a QSAR model [28]. There are several methods for determining the applicability domain of a model [27]. Among these, the approach used in this work is that of leverage. This method is based on the variation of the standardized residuals of the dependent variable with the distance between the values of the descriptors and their mean, called leverage [29]. The h_{ii} are the diagonal elements of a matrix H called the hat matrix. H is the projection matrix of the experimental values of the explained variable Y_{exp} in the space of values of the predicted explained variable Y_{pred} such that:

$$Y_{pred} = HY_{exp} \quad (9)$$

H is defined by the expression (10):

$$H = X(X^tX)^{-1}X^t \quad (10)$$

The domain of applicability is delimited by a leverage threshold value denoted by h*. In general, it is set to $3\frac{p+1}{n}$, where n is the number of compounds in the training set, and p is the number of descriptors in the model [30, 31]. For standardized residuals, the two limit values generally used are ±3σ, σ being the standard deviation of the experimental values of the quantity to be explained [32]: this is "the three sigma rule" [33].

3. RESULTS AND DISCUSSION

This QSAR study was conducted using a series of fourteen molecules (14) molecules derived from dihydropyrazoles. These compounds were synthesized and tested on prostate cancer cell lines (DU-145). It should be noted that this study was carried out on 14 molecules divided into two categories, ten (10) for the learning game and four (4) for the validation game. The objective of this part of the work is to model the antiproliferative activity of dihydropyrazole derivatives from the descriptors. The values of the descriptors as well as those of the experimental biological activities of the molecules are listed in **table 2**.

Table 1:Physico-chemical descriptor and experimental pIC_{50} of learning and validation sets.

Molecules	α(N-N-C)(°)	υ(N-N)(cm^{-1})	Δ_fS^0(kcal/molK)	pIC_{50}
learning set				
DHP1	110.3072	1163.1700	-1.2055	3.4547
DHP4	110.3639	1162.4800	-1.1312	3.8268
DHP5	110.3326	1164.5500	-1.1304	3.8729
DHP6	110.3338	1163.9800	-1.1344	3.5817
DHP7	110.1449	1167.7300	-1.1354	3.6861
DHP8	110.2819	1166.8900	-1.2220	3.5901
DHP9	110.3088	1170.8500	-1.2184	3.8827
DHP11	110.3291	1163.9000	-1.1030	3.6757
DHP13	110.5349	1156.7100	-1.0439	3.7905

DHP14	110.3062	1173.4800	-1.0428	4.4949
Validation Set				
DHP2	110.3661	1162.1800	-1.1324	3.9245
DHP3	110.4924	1158.7400	-1.1338	3.4855
DHP10	110.2262	1164.9300	-1.1019	3.5670
DHP12	110.0557	1179.0400	-1.1017	3.7620

In order to highlight the interdependence of the descriptors retained, the values of the bivariant linear correlation coefficients aij between the pairs of descriptors are less than 0.70 ((a_{ij} < 0.70) (**Table 3**). Moreover, two descriptors are said to be interdependent when (a_{ij} < 0.70.

Table 3: Correlation matrix between the different physico-chemical descriptors.

Variables	α(N-N-C)	υ(N-N)	ΔfS
α(N-N-C)	**1**	-0.6635	0.4388
υ(N-N)	-0.6635	**1**	-0.1987
ΔfS	0.4388	-0.1987	**1**

In the equation of a model, the negative or positive sign of the coefficient of a descriptor translates the effect of proportionality between the evolution of the median inhibitory concentration IC_{50} and the physico-chemical parameters of the regression equation. Thus, the negative sign indicates that when the value of the descriptor is high, the inhibitory concentration IC_{50} decreases while the positive sign reflects the opposite effect. In this work two tools of statistical analysis were used, they are among others the Multiple Linear Regression (**MLR**) and the Multiple Non Linear Regression (**MNLR**).

*3.1. Multiple Linear Regression (***MLR***)*

The equation of the QSAR model relating the inhibitory potential, the vibration frequency **υ(N-N)**, the valence angle **α(N-N-C)** and the standard entropy of formation ($\Delta_f S^0$) are presented below monitoring of statistical indicators (**Table 4**).

$$\boldsymbol{pIC_{50} = -226.87 + 1.48 * \alpha(N-N-C) + 0.06 * \upsilon(N-N) + 2.47 * \Delta_f S^0}$$

The model equation was obtained using descriptors obtained after molecule (**α(N-N-C)**, **υ(N-N)**, $\Delta_f S^0$). The positive sign of the various descriptors shows that the antiproliferative activity $\mathbf{pIC_{50}}$ evolves in the same direction. The study of the significance of this model is carried out by evaluating the statistical indicators and the validation criteria.

Table 2: Statistical analysis report of the $\mathbf{pIC_{50}}$ inhibitory potential of dihydropyrazole derivatives of the MLR model

Number of observations N	10
Coefficient of determination R^2	0.9108
Standard deviation RMSE S	0.1044
Fischer F-test F	20.425 > F_{cr} = 3.48
Cross Validation Correlation Coefficient $\boldsymbol{Q^2_{CV}}$	0.9108
R^2 - $\boldsymbol{Q^2_{CV}}$	0.000

The value of the coefficient of determination $\boldsymbol{R^2}$ which is 0.9108, shows that the estimated values of the $\mathbf{pIC_{50}}$ contain **91.08%** of the experimental values. The value of the Fisher test (**F = 20.425**) is relatively high compared to the critical value of the Fisher Snedecor table $\mathbf{F_{cr} = 3.48}$ [34]. This value 20.425 of the Fisher test, higher than the critical value, shows that the error made is less than what the

model explains. The standard deviation (**S=0.1044**) expresses the small variation of the predicted values compared to the experimental mean. For this model, the cross-validation correlation coefficient Q^2_{cv} is equal to $Q^2_{cv} = 0.9108$. This value, greater than 0.9, reflects a model said to be excellent according to Erikson et al. [35]. This model is acceptable, because it is in agreement with the acceptance criterion of these authors $R^2 - Q^2_{cv} = 0.9108 - 0.9108 = 0.000 < 0.3$. All these statistical indicators show that the model developed explains the antiproliferative activity in a statistically significant and satisfactory way.

The regression plot of the MLR model showing theoretical anticancer activity as a function of experimental activity is shown in **Figure 2.**

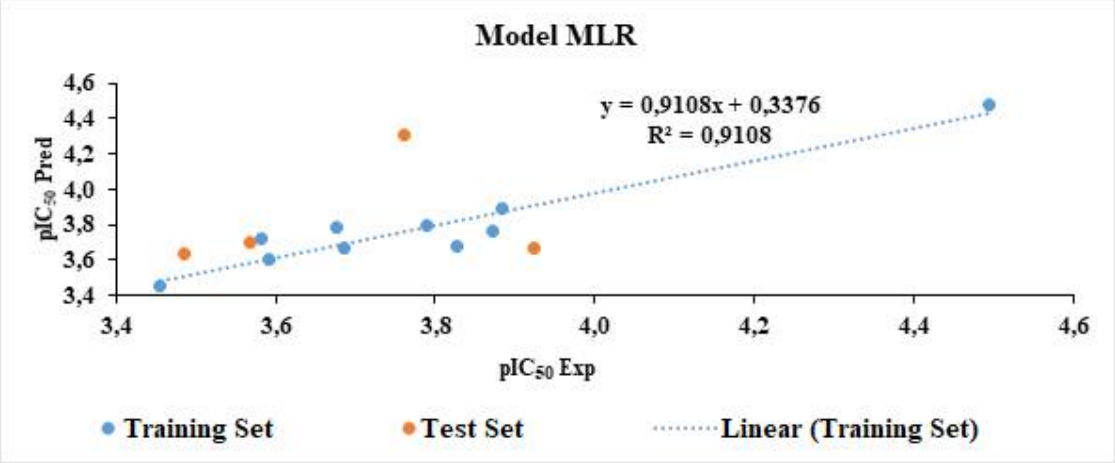


Figure 2: Regression line of the MLR model.

The low value of the standard deviation (S) which is 0.1044 for this model attests to the good regression between the predicted and experimental values. The curves (**Figure 3**) reflect the similar evolution of the data from this model for the **pIC$_{50}$** prediction of the 14 dihydropyrazole derivatives despite some deviations recorded.

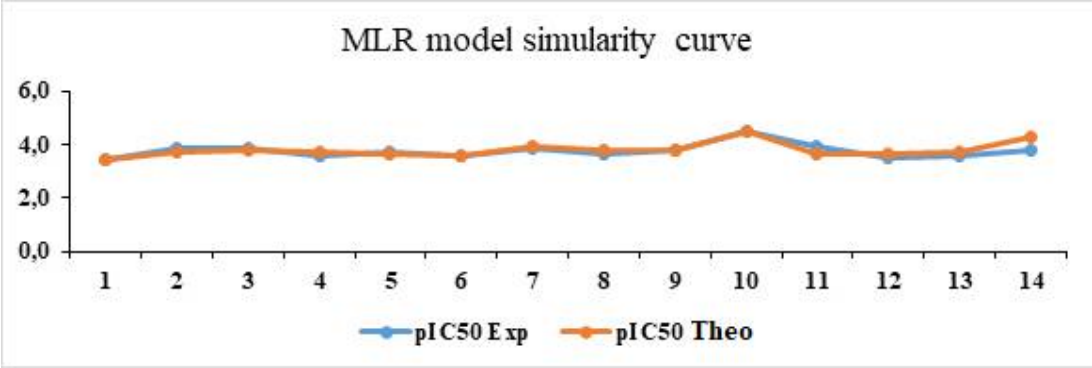


Figure 3 : Evolution curves of the experimental and predicted **pIC$_{50}$** values of the MLR model.

3.1.1. Internal Validation of the MLR Model

The randomization test of this model is only for the compounds in the test set, because the model relies on them. We limited ourselves to ten iterations. The randomized determination coefficients ($\mathbf{R^2_r}$) of each iteration are given in **Table 5**:

Table 5: Randomized determination coefficients

Interactions	1	2	3	4	5	6	7	8	9	10
R^2_r	0.015	0.101	0.112	0.123	0.137	0.168	0.214	0.280	0.362	0.184

The Roy parameter (R^2_p = 0.77) being lower than the R^2 coefficient of the model (0.9108), we can therefore confirm that the established model is not due to chance.

3.1.2. External Validation of the MLR Model

The external validation of the model was carried out respectively with the derivatives of dihydropyrazoles (DHP2, DHP3, DHP10, DHP12). In addition, the values of the external validation test were verified by calculating the pIC_{50}^{exp}/ pIC_{50}^{pred} ratio (**Table 6**).

Table 6: Values of the ratio between predicted and experimental antimalarial activities from the database.

CODE	$\mathbf{pIC_{50}^{exp}}$	$\mathbf{pIC_{50}^{pred}}$	$\mathbf{pIC_{50}^{exp}/\ pIC_{50}^{pred}}$
DHP2	3.925	3.664	1.07
DHP3	3.486	3.638	0.958
DHP10	3.567	3.699	0.964
DHP12	3.762	4.305	0.874

The values of the ratio tend towards 1. This reflects the good correlation between the predicted and experimental potentials of the dihydropyrazole molecules of the validation set.

After the verification of the various properties of external validation, we can affirm that the model is exploitable for the prediction of the antiproliferative activity of dihydropyrazoles.

3.2. Multiple Nonlinear Regression (MNLR)

We propose another MNLR model based on the vibration frequency **υ(N-N)**, the valence angle **α(N-N-C)** and the standard entropy of formation ($\Delta_f S^0$). This **MNLR** model is presented below with the corresponding statistical indicators in **Table 7**.

$$\boldsymbol{pIC_{50} = -18014 + 437,16947 * \alpha(N-N-C) - 10,65210 * \upsilon(N-N) - 40,93001 * \Delta_f S^0 - 1,97663 * \alpha(N-N-C)^2 + 0,00459 * \upsilon(N-N)^2 - 18,73594 * (\Delta_f S^0)^2}$$

Table 7: Statistical analysis report of the pIC_{50} inhibitory potential of the dihydropyrazole derivatives of the MNLR model.

Number of observations N	10
Coefficient of determination R^2	0.9207
Standard deviation RMSE S	0.1393
Fischer F-test F	92.823 > F_{cr} = 3.48
Cross Validation Correlation Coefficient $\boldsymbol{Q^2_{CV}}$	0.9207
R^2 - $\boldsymbol{Q^2_{CV}}$	0.000

The coefficient of determination R^2 = **0.9207** (close to **1**: ideal case) justifies the good correlation between the predicted values and the observed values. The model therefore has excellent explanatory power. The low value of the standard deviation (**RMSE=0.1393**) shows the good statistical fit of the model and therefore its strong predictive power. The significance of the model descriptor is expressed by the Ficher-Snedecor coefficient whose value is: **92.823**. This high value, well above the threshold value **(= 3.48**) indicates that there is a strong relationship between the antiproliferative activity and the descriptors of this model. The cross-validation correlation coefficient (Q^2_{cv}) is equal to 0.9207 > 0.9

reflecting an excellent model according to Erikson *et al.* . $R^2 - Q^2_{cv}$=**0.9207 -0.9207 = 0.000 < 0.3** means that the model is acceptable. All these statistical indicators show that the model developed explains the antiproliferative activity in a significant and satisfactory way. It can be used to predict the antiproliferative activity of other molecules.

The regression line of the **MNLR** model between the experimental and predicted antidiabetic activities of the learning game and the validation game is shown in **Figure 4**.

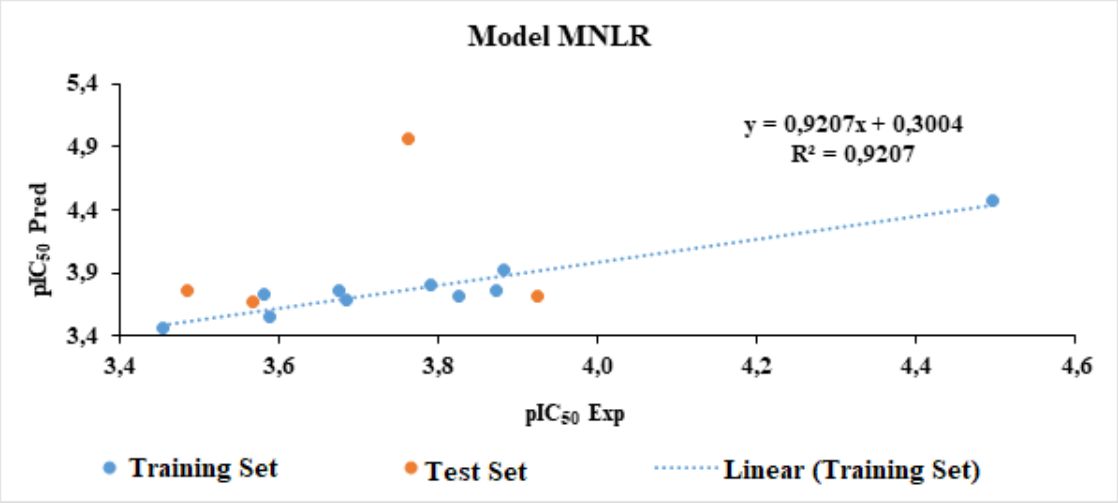


Figure 4: Regression line of the MNLR model.

The low value of the standard deviation (**S**) which is 0.1393 for this model attests to the good regression between the predicted and experimental values. The curves (**Figure 5**) reflect the similar evolution of the data from this model for the **pIC_{50}** prediction of the 14 dihydropyrazole derivatives despite some deviations recorded.

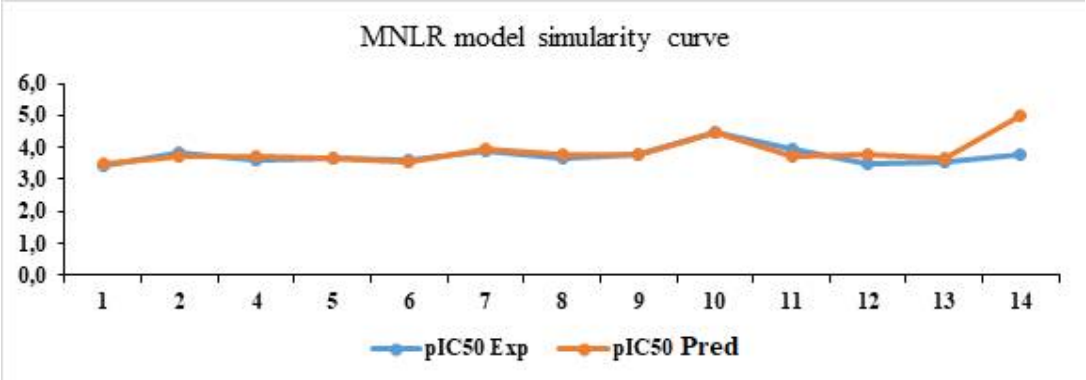


Figure 5: Evolution curves of the experimental and predicted **pIC_{50}** values of the MNLR model.

3.2.1. Internal Validation of the MNLR Model

The randomization test of this model is only for the compounds in the test set, because the model relies on them. We limited ourselves to ten iterations. The randomized determination coefficients (**R^2_r**) of each iteration are given in **Table 8**:

Table 8 : Randomized determination coefficients

interactions	1	2	3	4	5	6	7	8	9	10
R^2_r	0.318	0.329	0.380	0.421	0.442	0.446	0.460	0.472	0.506	0.544

Using the values in the table, we note that Roy's parameter (($\mathbf{R^2_p}$ **= 0.65**) is lower than the R^2 coefficient of the model (0.9207), so we can confirm that the established model is not due to chance.

3.2.2. External Validation of the MNLR Model

The external validation of the model was carried out respectively with the derivatives of dihydropyrazoles (DHP2, DHP3, DHP10, DHP12). In addition, the values of the external validation test were verified by calculating the pIC_{50}^{exp}/ pIC_{50}^{pred} ratio (**Table 9**).

Table 9: Values of the ratio between predicted and experimental antimalarial activities from the database.

CODE	pIC_{50}^{exp}	pIC_{50}^{pred}	pIC_{50}^{exp}/ pIC_{50}^{pred}
DHP2	3.925	3.709	1.058
DHP3	3.486	3.761	0.927
DHP10	3.567	3.674	0.971
DHP12	3.762	4.965	0.758

The values of the ratio tend towards 1. This reflects the good correlation between the predicted and experimental potentials of the dihydropyrazole molecules of the validation set.

3.3. Analysis of Model Descriptors

The analysis of the contribution of each of the three descriptors in the model for the prediction of antiproliferative activity was carried out using XLSTAT software [19]. The different contributions are illustrated in **Figure 6**.

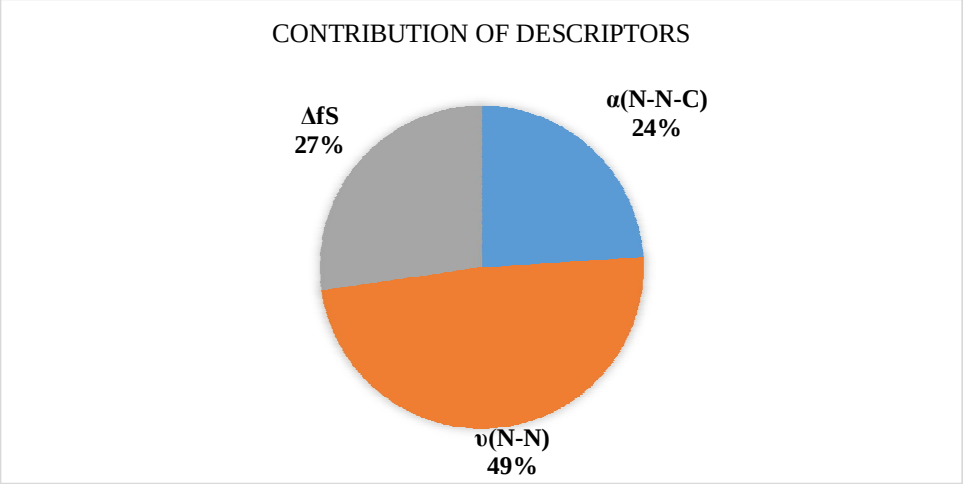


Figure 6: Contribution of the different constituents in the model.

Figure 6 shows us that the vibration frequency **υ(N-N)** is the descriptor which has the greatest contribution compared to the other descriptors. Thus, the vibration frequency **υ(N-N)** is the priority descriptor in the prediction of anticancer activity.

3.4. Applicability Domains and Model

The graph of standardized residuals as a function of the levers h_{ii}in **figure 7**, makes it possible to visualize the domain of applicability of the model.

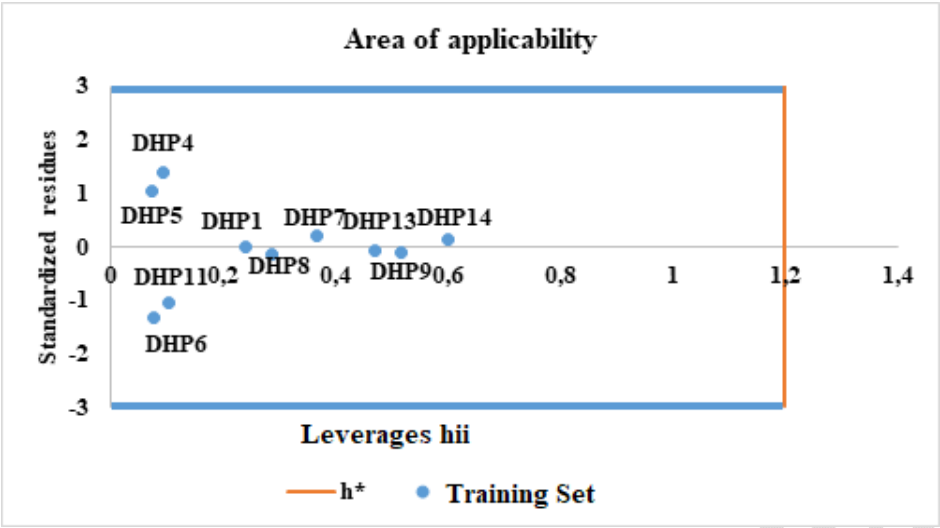


Figure 7: Graph of standardized residuals of antimalarial activity according to the levers of the models.

The 10 compounds of the training set and the three (3) descriptors of the models have a threshold value of the levers ***h**** = **1.2**. The extreme values of the standardized residuals are **±3** according to the "three sigma rule". These different values delimit the domain of applicability of the model as shown in the graph of **Figure 7**.

4.CONCLUSION

Dihydropyrazoles are molecules that offer an exemplary pharmacological profile. In particular, their dynamic biological activities make them an interesting research subject. The synthesis of the most effective molecules derived from dihydropyrazoles requires knowledge and understanding of the physicochemical properties in relation to its various biological activities, in particular the antiproliferative activity. In this study, a QSAR model was developed on fourteen molecules derived from dihydropyrazoles in order to relate the antiproliferative activity (***DU-145***) with quantum descriptors (of the vibration frequency **υ(N-N)**, the valence angle **α(N-N-C)** and the standard entropy of formation ($\Delta_f S^0$)) in order to predict that of new molecules or not. Multiple linear regression (MLR) was used to quantify the relationship between molecular descriptors of vibrational frequency **υ(N-N),** valence angle **α(N-N-C)** and standard entropy of formation ($\Delta_f S^0$) and potential the inhibitory concentration (pIC_{50}) of molecules derived from dihydropyrazoles. The strong correlation observed between the calculated and experimental values of this antiproliferative activity indicates that the vibration frequency υ(N-N), the valence angle **α(N-N-C)** and the standard entropy of formation ($\Delta_f S^0$) are closely related to the potential inhibitory concentration of dihydropyrazoles. The statistical indicators of this model display the following values: R^2 = **0.9108**, **S** = **0.1044**, **F** = **20.425**. The validation techniques commonly used, namely the standard statistical indicators of the model, the external validation and the domain of applicability, were used. For the RMNL model we have: R^2 = 0.920**, S** = 0.1393, **F** = 92.823 > 2.9, Q^2_{CV} = 0.920). It appears from the two models that the vibration frequency **υ(N-N)**, turns out to be the priority descriptor in the prediction of the antiproliferative activity of cancer of the derivatives of dihydropyrales studied. The two models obtained were validated using a validation set comprising four (4) molecules by applying the Tropsha criteria. Indeed, the similarity curve illustrates this perfectly by the almost perfect alignment of the points.

The model developed is in accordance with the five principles of the OECD for the validation of QSAR models. This QSAR model obtained therefore makes it possible to explain and predict the behavior of the inhibitory concentration of molecules and also to design other new molecules similar to dihydropyrazoles that are more active against prostate cancer by identifying priority descriptors

which are the vibration frequency **υ(N-N)**, the valence angle **α(N-N-C)** and the standard entropy of formation ($\Delta_f S^0$).

This model is able to describe **91.08%** of the total variance of the antiproliferative activity of experimental prostate cancer (***DU-145***) cell lines and could be used effectively to estimate the antiproliferative activity of dihydropyrazole derivatives for which data experiments are not yet available.

References

[1] Futura Santé, «futura-sciences,» 2022. [En ligne]. Available: https://www.futura-sciences.com/sante/definitions/medecine-cancer-108/. [Accès le 16 07 16/07/2022].

[2] A. Shaikh, R. Bhandare, K. Palleapati, S. Nissankararao, V. Kancharlapalli et S. Shaik, « Antimicrobial, antioxidant, and anticancer activities of some novel isoxazole ring containing chalcone and dihydropyrazole derivatives.,» *Molecules* , n° %125, p. 1047, 2020.

[3] A. Shaik, R. Yejella et S. Shaik, «Synthesis, Antimicrobial, and Computational Evaluation of Novel Isobutylchalcones as Antimicrobial Agents.,» *Int. J. Med. Chem*, p. 1–14, 2017.

[4] B. Lokesh, Y. Prasad et A. Shaik, «Synthesis, Biological evaluation and molecular docking studies of new pyrazolines as an antitubercular and cytotoxic agents. Infect,» *Infectious Disorders - Drug Targets*, n° %119, p. 310, 2019.

[5] V. Kancharlapalli, A. Shaikh et K. Palleapati, «Antitubercular evaluation of isoxazole appended 1-carboxamido-4,5-dihydro-1H-pyrazoles.,» *J. Res. Pharm.*, n° %123, p. 156–163, 2019.

[6] L. S. Goodman et A. Gilman, The Pharmacological Basis of Therpeutics, New York: Macmillan, 1980.

[7] P. Kishor, V. Ramana et A. Shaikh, «Antitubercular Evaluation of Isoxazolyl Chalcones,» *Res. J. Pharm. Biol. Chem. Sci.*, n° %18, p. 730–735, 2017.

[8] A. Shaikh, K. Lohitha, S. Vani, P. Basu, A. Shaik, K. Supraja et S. Harish, «Synthesis and screening of novel lipophilic diarylpropeones as prospective antitubercular, antibacterial and antifungal agents,» *Biointerface Res. Appl. Chem.*, n° %19, 2019.

[9] A. Shaikh, Y. Prasad et S. Shaik, «Design, Facile Synthesis, Characterization and Computational Evaluation of Novel Isobutylchalcones as Cytotoxic Agents,» *FABAD J. Pharm. Sci.*, Vols. %1 sur %2Part-A, n° %140, p. 7–22, 2015.

[10] M. J. Frisch, G. W. Trucks, H. B. Schlegel et G. E. Scuseria, «Gaussian 09, Revision A.02,» Gaussian, Inc., Wallingford CT, 2009.

[11] P. K. Chattaraj, A. Cedillo et R. G. Parr, *J. Phys. Chem.*, vol. 103, p. 7645, 1991.

[12] P. W. Ayers et R. G. Parr, *J. Am Chem. Soc*, vol. 122, p. 2000, 2010.

[13] F. De Proft, J. M. L. Martin, P. Geerlings et :. :, *Chem. Phys Let.*, vol. 250, p. 393, 1996.

[14] C. Hansch, P. G. Sammes et J. B. Taylor, «in:Comprehensive Medicinal Chemistry,» *Computers and the medicinal chemist*, vol. 4, pp. 33-58, 1990.

[15] R. Franke, «Theoretical Drug Design Methods,» *Elsevier*, 1984.

[16] S. Chaltterjee, A. Hadi et B. Price, «Regression Analysis by Examples,» *Wiley VCH: New York*, 2000.

[17] H. Phuong, «Synthèse et étude des relations structure/activité quantitatives (QSAR/2D) d'analoguesBenzo[c]phénanthridiniques,» France, 2007.

[18] M. Excel, 2016.

[19] Addinsoft, XLSTAT Version 2014.5.03 Copyright, *XLSTAT and Addinsoft are Registered Trademarks of Addinsoft.*, 2014, pp. 1995-2014 .

[20] G. W. Snedecor et W. G. Cochran, « Methods, Statistical,» *Oxford and IBH: New Delhi, India;* , p. 381, 1967.

[21] N. J.-B. Kangah, M. G.-R. Koné, C. G. Kodjo, B. R. N'guessan, A. L. C. Kablan, S. A. Yéo et N. Ziao, «Antibacterial Activity of Schiff Bases Derived from Ortho Diaminocyclohexane, Meta-Phenylenediamine and 1,6-Diaminohexane: Qsar Study with Quantum Descriptors,,» *International Journal of Pharmaceutical Science Invention,* vol. 6, n° %13, pp. 38-43, 2017.

[22] E. X. Esposito, A. J. Hopfinger et J. D. Madura, «Methods for Applying the Quantitative Structure-Activity Relationship Paradigm,,» *Methods in Molecular Biology,* vol. 275, pp. 131-213., 2004.

[23] L. Eriksson, J. Jaworska, A. Worth, M. D. Cronin, R. M. Mc Dowell et P. Gramatica, «Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs,» *Environmental Health Perspectives,* vol. 111, n° %110, pp. 1361-1375, 2003.

[24] K. Roy et S. Paul, «Exploring 2D and 3D QSARs of 2,4-Diphenyl-1,3-oxazolines for Ovicidal Activity Against Tetranychus urticae,» *QSAR Comb. Sci.,* vol. 28, n° %14, p. 406–425, 2009.

[25] P. Gramatica, «Principles of QSAR models validation: internal and external.,» *QSAR Comb. Sci.,* vol. 26, n° %15, p. 694 – 701, 2007.

[26] N. N.-Jeliazkova et J. Jaworska, « An Approach to Determining Applicability Domains for QSAR Group Contribution Models: An Analysis of SRC KOWWIN,» *ATLA 33,* p. 461–470, 2005.

[27] F. Sahigara, K. Mansouri, D. Ballabio, A. Mauri et V. C. a. R. Todeschini, «Comparison of Different Approaches to Define the Applicability Domain of QSAR Models,» *Molecules,* vol. 17, pp. 4791-4810, 2012.

[28] K. Roy et e. al, «A Primer on QSAR/QSPR Modeling Chapter 2 Statistical Methods in QSAR/QSPR,,» *Springer Briefs in Molecular Science,* pp. 37-59, 2015.

[29] J. Jaworska, N. N. Jeliazkova et T. Aldenberg, «QSAR Applicability Domain Estimation by Projection of the Training Set in Descriptor Space: A Review,» *ATLA 33,* p. 445–459, 2005.

[30] M. Ghamali, S. Chtita, M. Bouachrine et T. Lakhlifi, «Méthodologie générale d'une étude RQSA/RQSP, Revue Interdisciplinaire,» vol. 1, n° %11, 2016.

[31] S. Chtita, M. Ghamali, R. Hmamouchi, B. Elidrissi, M. Bourass, M. Larif, M. Bouachrine et T. Lakhlifi, «Investigation of Antileishmanial Activities of Acridines Derivatives against Promastigotes and Amastigotes Form of Parasites Using QSAR Analysis,» *Advances in Physical Chemistry*, pp. 1-16, 2016.

[32] T. Asadollahi, S. Dadfarnia, A. Shabani, J. Ghasemi et M. Sarkhosh, «QSAR Models for CXCR2 Receptor Antagonists Based on the Genetic Algorithm for Data Preprocessing Prior to Application of the PLS Linear Regression Method and Design of the New Compounds Using In Silico Virtual Screening,» *Molecules,* vol. 16, pp. 1928-1955, 2011.

[33] S. Chtita, M. Larif, M. Ghamali, M. Bouachrine et T. Lakhlifi, «Quantitative structure–activity relationship studies of dibenzo[a,d]cycloalkenimine derivatives for non-competitive antagonists of N-methyl-D-aspartate based on density functional theory with electronic and topological descriptors,» *Journal of Taibah University for Science* , vol. 9, pp. 143-154, 2015.

[34] R. D. Cook et S. Weisberg, «Wiley Series in Probability and Statistics,» *An introduction to Regression Graphics,* 1994.

[35] L. Eriksson, J. Jaworska, A. Worth, M. T. D. Dowell et P. Gramatica, «Methods for Reliability and Uncertainty Assessment and for applicability Evaluations of classification and Regression-Based QSARs,» *Environmental Health Perspectives,* vol. 111, n° %110, pp. pp. 1361-1375, 2003.