

# Nematocidal Activities of Protonated Benzimidazolyl Chalcone Using Quantitative Structure-Activity Relationship

Kone Mamadou Guy-Richard<sup>1,2,3\*</sup>, Dembele Georges Stéphane<sup>1,2</sup>, Ouattara Bafétigué<sup>4</sup>, Karamanis Panagiotis<sup>3</sup>, Ziao Nahossé<sup>1,2</sup>

<sup>1</sup>Laboratoire de Thermodynamique et de Physico-Chimie du Milieu, Université Nangui Abrogoua, Abidjan, Côte-d'Ivoire

<sup>2</sup>Groupe Ivoirien de Recherches en Modélisation des Maladies (GIR2M), Université Nangui Abrogoua, Abidjan, Côte-d'Ivoire

<sup>3</sup>E2S UPPA, CNRS, IPREM, Université de Pau et des Pays de l'Adour, Pau, France

<sup>4</sup>Laboratoire de Physique Fondamentale et Appliquée, UFR SFA, Université Nangui Abrogoua, Abidjan, Côte-d'Ivoire

Email: \*guyrichardkone@gmail.com

**How to cite this paper:** Guy-Richard, K.M., Stéphane, D.G., Bafétigué, O., Panagiotis, K. and Nahossé, Z. (2025) Nematocidal Activities of Protonated Benzimidazolyl Chalcone Using Quantitative Structure-Activity Relationship. *Computational Molecular Bioscience*, 15, 1-17.

<https://doi.org/10.4236/cmb.2025.151001>

**Received:** November 1, 2024

**Accepted:** March 1, 2025

**Published:** March 4, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

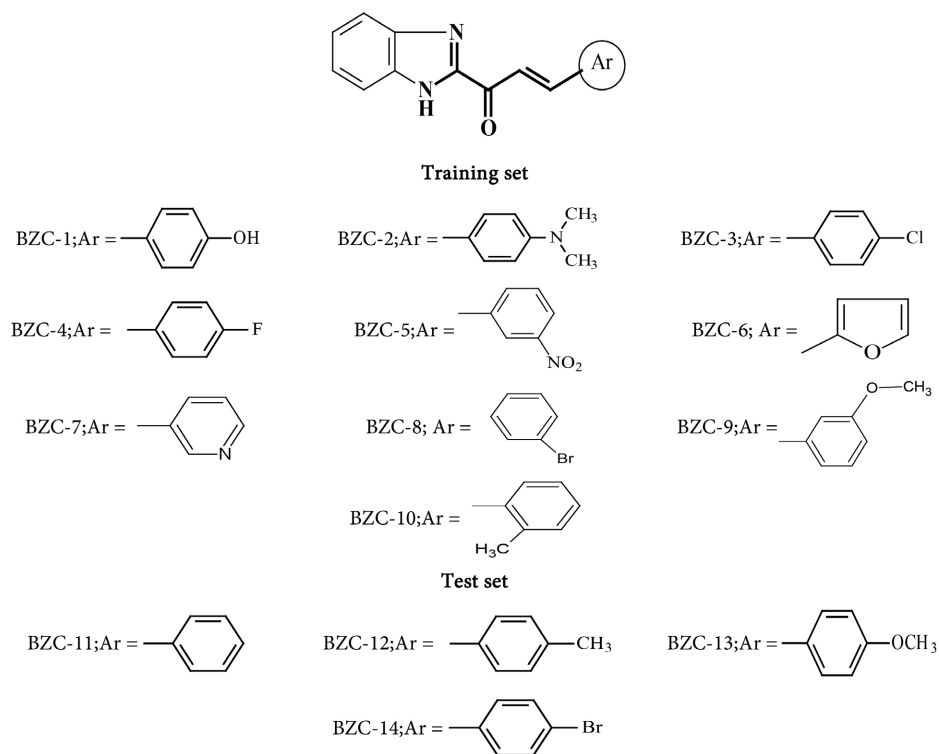
## Abstract

Benzimidazolyl-Chalcones (BZCs) possess nitrogen heteroatoms making them very active molecules when protonated. In this work we will focus on a series of fourteen (14) substituted BZC molecules. These molecules have been synthesised, characterised and tested for their nematocidal activities. By implementing quantum chemical methods, including density functional theory (DFT) at the MPW1PW91/6-311+G (d, p) level, we can achieve accurate predictions of molecular properties. A QSAR study is conducted to determine a quantitative relationship between nematocidal activity and property information of the BZC compound series. The quantum descriptors namely dipole moment ( $\mu$ ), energy gap ( $\Delta E$ ) and mean valence angle ( $\theta_{Nsp}$ ), are all obtained by protonation on the  $sp^2$  nitrogen, the preferred site of protonation in BZCs. These descriptors are the explanatory and predictive parameters of the nematocidal activity of the studied molecules. This study was conducted using the principal component analysis (PCA), the Multiple Linear Regression (MLR) and the non-linear regression (MNLR) methods. The quantitative models were proposed and the nematocidal activity of BZCs was interpreted based on multivariate statistical analysis. This study shows that PCA, MLR and NMR were used to predict the activities, but compared to the statistical indicators of NMR, we realised that the predictions fulfilled by the latter were more effective. The obtained results suggest that the combination of the proposed descriptors ( $\mu$ ,  $\Delta E$ ,  $\theta_{Nsp}$ ) could be useful to predict the nematocidal activity of Benzimidazolyl-Chalcones.

## Keywords

Benzimidazolyl-Chalcone (BZC), Nematocidal Activity, Quantitative Structure-Activity Relationship (QSAR), Quantum Descriptors

## 1. Introduction



**Figure 1.** Molecular structures of training sets and test of Benzimidazolyl-Chalcones used for QSAR models.

Nematodes are a large group of worms found in all living environments. They live as parasites of plants, animals and humans. They are the cause of most parasitic diseases in humans (elephantiasis, filariasis) [1] [2], gastrointestinal infections and loss of productivity in animals [3] [4]. However, for more than two decades, the continued emergence of new, resistant races of nematodes has been of increasing concern to the agricultural, medical and health communities. There is a need to develop an anthelmintic that offers a broad spectrum of action, a high degree of efficacy, a good safety margin and flexibility of use in order to contain the problems of resistance. Several hundred compounds derived from benzimidazoles have been synthesised, of which a few have been selected primarily for their broad-spectrum anthelmintic activity. Benzimidazolyl-chalcone derivatives are of considerable pharmacological interest because of their therapeutic properties in many diseases. Several studies have shown that benzimidazolyl derivatives possess antihistaminic properties [5], antifungals [6], anti-allergic [7], antibacterial [8]-

[10] and antiviral [11]. As these therapeutic properties are linked to the conformation of the molecules and the interactions, they can establish with each other, a QSAR study is used to determine models. These models are increasingly used, due to the growth in computing resources, to explain and/or predict molecular properties in order to limit the excessive number of experiments, which are sometimes long and expensive and to reduce the cost of drug production by pharmaceutical companies [12] [13]. In the specific case of the QSAR study, ten derivatives of Benzimidazolyl-Chalcones were studied and four others from the same series were used for the external validation test (Figure 1). These molecules were synthesised by Ouattara *et al.* [14]. The nematocidity of the 14 BZC derivatives was modelled using several statistical tools, in particular, principal component analysis (PCA), multiple linear regression (MLR) and non-linear regression (MNL). The general objective of this work is to make a descriptive and predictive study of the nematocidal activity of BZCs based on multivariate statistical analyses.

## 2. Materials and Methods

### 2.1. Level of Calculation

In order to establish a descriptive and predictive theory of the biological activities of protonated BZCs, Theoretical Chemistry methods are employed at the MPW1PW91/6-311+G (d, p) level. The modified Perdew-Wang 1 (MPW1) calculations [15] [16] such as mPW1PW91 are hybrid HF-DF models that provide good results for both covalent and non-covalent interactions [17]. In this work, to assess the quantitative structure-activity relationship between nematocidal activity  $LC_{100}$  ( $\mu\text{g/mL}$ ) and protonated BZC descriptors, the quantum chemistry software Gaussian 03 [18] was used. The split-valence and triple-dzeta bases, being sufficiently large and taking into account the diffuse and polarisation functions, are important when dealing with intermolecular interactions. Taking into account these diffuse and polarisation functions gives a quantitative character to the results obtained. The modelling was done using the multilinear regression method implemented in Excel spreadsheets [19] and XLSTAT version 2014 [20].

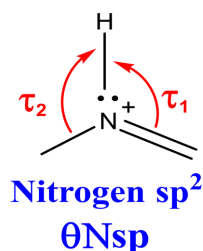
### 2.2. Quantum Descriptors Used

The descriptors linked to the molecules in interaction by protonation were calculated. In particular, on the  $sp^2$  nitrogen, the preferential protonation site in BZCs. For the development of QSAR models, some theoretical descriptors related to the conceptual DFT have been determined such as the energy of the Highest Occupied Molecular Orbital (LUMO), the energy of the Highest Occupied Molecular Orbital (HOMO), the energy gap ( $\Delta E$ ), the dipole moment ( $\mu$ ) and the average valence angle. It should be noted that, the descriptors related to the boundary molecular orbitals have been calculated in a very simple way within the Koopmans approximation [21]. The LUMO energy characterises the sensitivity of the molecule to nucleophilic attack, and the HOMO energy characterises the susceptibility

of a molecule to electrophilic attack. The energy difference  $\Delta E$  between the LUMO and HOMO, is an important parameter that gauges the overall reactivity towards an electron acceptor and the stability of a molecule. The dipole moment ( $\mu$ ) indicates the stability of a molecule in water. Thus, a high dipole moment will reflect low solubility in organic solvents and high solubility in water. The energy gap is calculated from Equation (1):

$$\Delta E = 1/2(\varepsilon_{\text{LUMO}} + \varepsilon_{\text{HOMO}}) \quad (1)$$

The measured valence angle  $\theta_{\text{Nsp}^2}$  is the average of the angles  $\tau_1$  and  $\tau_2$  about the nitrogen  $\text{sp}^2$  (Figure 2).



**Figure 2.** The average valence angle ( $\theta_{\text{Nsp}}$ ) measured on the  $\text{sp}^2$  nitrogen preferential site of protonation.

The descriptors obtained in protonation interaction were characterised by Kone *et al.* [22].

### 2.3. Estimating the Predictive Capacity of a QSAR Model

The BZCs have various nematocidal concentrations ranging from 0.002 to 424.5  $\mu\text{g/mL}$ . This range of concentrations allows a quantitative relationship to be defined between the nematocidal activity and the theoretical descriptors of these molecules. Biological data are generally expressed as the opposite of the base ten logarithm of the activity ( $-\log_{10}(C)$ ) to obtain higher mathematical values where structures are biologically highly effective [23] [24]. The nematocidal activity is expressed by the nematocidal potential  $\text{pLC}_{100}$ . The nematocidal potential is defined from Equation (2):

$$\text{pLC}_{100} = -\log_{10}\left(\frac{\text{LC}_{100}}{M} * 10^{-3}\right) \quad (2)$$

Where  $M$  is the molecular weight ( $\text{g/mol}$ ) and  $\text{LC}_{100}$  is the lethal concentration 100 in worms, it gives the concentration of substance required to destroy 100% of a worm population under the conditions of the experiment ( $\mu\text{g/mL}$ ).

The quality of a model is determined on the basis of different statistical analysis criteria including the coefficient of determination  $R^2$ , standard deviation  $S$ , cross-validation correlation coefficients  $Q_{\text{CV}}^2$  and Fischer  $F$ .  $R^2$ ,  $S$  and  $F$  relate to the fit of calculated and experimental values. They describe the predictive ability within the limits of the model, and allow to estimate the accuracy of the calculated values on the test set [25] [26]. As for the cross-validation coefficient  $Q_{\text{CV}}^2$ , 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 the model. The correlation coefficient  $R^2$  gives an evaluation of the dispersion of the theoretical values around the experimental values. The quality of the modelling is better when the points are close to the adjustment line [27]. The fit of the points to this line can be assessed by the coefficient of determination.

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

Où:

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

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

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

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

Furthermore, the variance  $\sigma^2$  is determined by the relation 4:

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

Where  $k$  is the number of independent variables (descriptors),  $n$  is the number of molecules in the test or training set and  $n - k - 1$  is the degree of freedom.

Another statistical indicator used is the standard deviation  $S$ . It is used to assess the reliability and accuracy of a test set. It is used to assess the reliability and accuracy of a model:

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

The Fisher  $F$  test is also used to measure the statistical significance of the model, *i.e.* the quality of the choice of descriptors making up the model.

$$F = \frac{\sum (y_{i,theo} - y_{i,exp})^2}{\sum (y_{i,exp} - y_{i,theo})^2} * \frac{n - k - 1}{k} \quad (6)$$

The coefficient of determination of the cross-validation,  $Q_{cv}^2$ , is used to evaluate the accuracy of the prediction on the test set. It is calculated using the following relationship:

$$Q_{cv}^2 = \frac{\sum (y_{i,theo} - \bar{y}_{i,exp})^2 - \sum (y_{i,theo} - y_{i,exp})^2}{\sum (y_{i,theo} - \bar{y}_{i,exp})^2} \quad (7)$$

## 2.4. Statistical Analysis

### 2.4.1. Principal Component Analysis (PCA)

The structures of 14 BZC compounds were studied by statistical methods based on Principal Component Analysis (PCA) [28] using the XLSTAT software version

2014 [20]. PCA is a useful statistical technique for summarising all the information encoded in the structures of compounds. It is also very useful for understanding the distribution of compounds [29]. This is an essentially descriptive statistical method which aims to present, in graphic form, the maximum information contained in the data.

#### 2.4.2. 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 (biological activity) and several independent variables (descriptors). This statistical method minimises the differences between the actual and predicted values. It was also used to select the descriptors used as input parameters in the multiple non-linear regression (MNLR). Multiple Non-Linear Regression (MNLR) analysis is a technique for improving the structure-activity relationship in order to quantitatively assess biological activity. It takes into account several parameters. 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  $a, b, c, d, \dots$  represent the parameters and  $x_1, x_2, x_3, x_4, \dots$  represent the variables.

The (MLR) and (MNLR) were generated using the XLSTAT software version 2014 [20], to predict LC<sub>100</sub> nematocidal activity. The equations of the different models were evaluated by coefficient of determination ( $R^2$ ), root mean square error ( $S$ ), Fischer test ( $F$ ) and cross correlation coefficient ( $Q_{CV}^2$ ) [30].

All descriptor values for the ten (10) BZC molecules in the test set and the other four (4) molecules in the validation set are presented in **Table 1**.

**Table 1.** Quantum descriptors and experimental potentials of the test and validation sets.

| Compounds    | $\Delta E$ (eV) | $\mu$ (D) | $\theta_{Nsp}$ (°) | pLC <sub>100exp</sub> |
|--------------|-----------------|-----------|--------------------|-----------------------|
| Training Set |                 |           |                    |                       |
| BZC-1        | -3.150          | 4.293     | 125.13             | 3.095                 |
| BZC-2        | -2.645          | 4.695     | 125.13             | 2.837                 |
| BZC-3        | -3.255          | 9.512     | 125.13             | 4.372                 |
| BZC-4        | -3.422          | 7.943     | 125.13             | 4.346                 |
| BZC-5        | -3.712          | 15.726    | 125.14             | 7.503                 |
| BZC-6        | -3.250          | 4.781     | 125.14             | 4.919                 |
| BZC-7        | -3.642          | 8.259     | 124.84             | 8.096                 |
| BZC-8        | -3.089          | 14.292    | 125.14             | 6.311                 |
| BZC-9        | -2.811          | 7.724     | 125.14             | 2.817                 |
| BZC-10       | -3.377          | 5.991     | 125.14             | 4.34                  |

## Continued

|        |        | Test set |        |       |
|--------|--------|----------|--------|-------|
| BZC-11 | -3.498 | 5.273    | 125.13 | 8.094 |
| BZC-12 | -3.314 | 5.812    | 125.13 | 6.215 |
| BZC-13 | -3.054 | 5.028    | 125.13 | 5.612 |
| BZC-14 | -3.133 | 12.953   | 125.14 | 2.887 |

## 2.5. Acceptance Criteria for a QSAR Model

The performance of a mathematical model, for Eriksson *et al.* [31], is characterized by a value of  $Q_{cv}^2 > 0.5$  for a satisfactory model when for the excellent model  $Q_{cv}^2 > 0.9$ . According to them, given a test set, a model will perform well if the acceptance criterion  $R^2 - Q_{cv}^2 < 0.3$  is respected.

According to Tropsha *et al.* [32] [33], For the external validation set, the predictive power of a model can be obtained from five criteria. These criteria are:

- 1)  $R_{Test}^2 > 0.7$ , 2)  $Q_{Cv Test}^2 > 0.6$ , 3)  $|R_{Test}^2 - R_0^2| \leq 0.3$
- 4)  $\frac{|R_{Test}^2 - R_0^2|}{R_{Test}^2} < 0.1$ . et  $0.85 \leq k \leq 1.15$ , 5)  $\frac{|R_{Test}^2 - R_0'^2|}{R_{Test}^2} < 0.1$  et

$0.85 \leq k' \leq 1.15$ .

In addition, Roy and Roy [34], have further refined the predictive capacity of a QSAR model. They have developed quantities  $r_m^2$  et  $\Delta r_m^2$ , called metric values,  $r_m^2$  determines how close the observed activity is to the prediction. The metric values  $r_m^2$  et  $\Delta r_m^2$  are calculated from the observed and predicted activities. Currently, these two different variants  $r_m^2$  et  $\Delta r_m^2$ , can be calculated for t test set (internal validation) or for the test set (external validation). A QSAR model is acceptable to these authors, if both criteria are met.

$$\overline{r_m^2} = \frac{r_m^2 + r_m'^2}{2} > 0.5$$

$$\Delta r_m^2 = |r_m^2 - r_m'^2| > 0.2$$

Where  $r_m^2 = r^2 * \left(1 - \sqrt{r^2 - r_0^2}\right)$  and  $r_m'^2 = r^2 * \left(1 - \sqrt{r^2 - r_0'^2}\right)$ .

## 2.6. Domain of Applicability

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 [35]. It corresponds to the region of the chemical space including the compounds of the training set and similar compounds, which are close in the same space [36]. Indeed, the model, which is built on the basis of a limited number of compounds, by relevant descriptors, chosen among many oers, cannot ba universal tool to predict the activity of any other molecule with confidence. It appears necessary, even mandatory, to determine the DA of any QSAR model. This is recommended by the Organisation for Economic Co-operation and

Development (OECD) in the development of a QSAR model [37]. There are several methods for determining the domain of applicability of a model [36]. Among these, the approach used in this work is the leverage approach. This method is based on the variation of the standardised residuals of the dependent variable with the distance between the values of the descriptors and their mean, called leverage [38]. The  $h_{ii}$  are the diagonal elements of a matrix  $H$  called hat matrix.  $H$  is the projection matrix of the experimental values of the explained variable  $Y_{exp}$  into the space of the predicted values of the explained variable  $Y_{pred}$  such that:

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

$H$  is defined by the expression (10):

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

The area of applicability is delimited by a threshold value of the lever noted  $h$ . In general, it is set at  $3\frac{p+1}{n}$ , where  $n$  is the number of compounds in the training set, and  $p$  is the number of descriptors in the model [39] [40]. For standardised residuals, the two limit values generally used are  $\pm 3\sigma$ ,  $\sigma$  being the standard deviation of the experimental values of the quantity to be explained [41]: this is the “three sigma rule” [42].

### 3. Results and Discussion

#### 3.1. Principal Component Analysis (PCA)

Three descriptors for the 14 compounds are subjected to PCA analysis. The two main axes are sufficient to describe the information provided by the data matrix. Indeed, the percentages of variance are 53.73% and 25.49% for the F1 and F2 axes, respectively. The total information is estimated at 79.23%. Principal component analysis (PCA) [29] was conducted to identify the relationship between the different descriptors. The bold values are different from 0 at a significance level of  $p = 0.05$ . The correlations between the three descriptors are presented in **Table 2** as a correlation matrix and in **Figure 2** where these descriptors are represented in a correlation circle. The Pearson correlation coefficients are summarised in **Table 2**. The resulting matrix provides information on the negative or positive correlation between the variables.

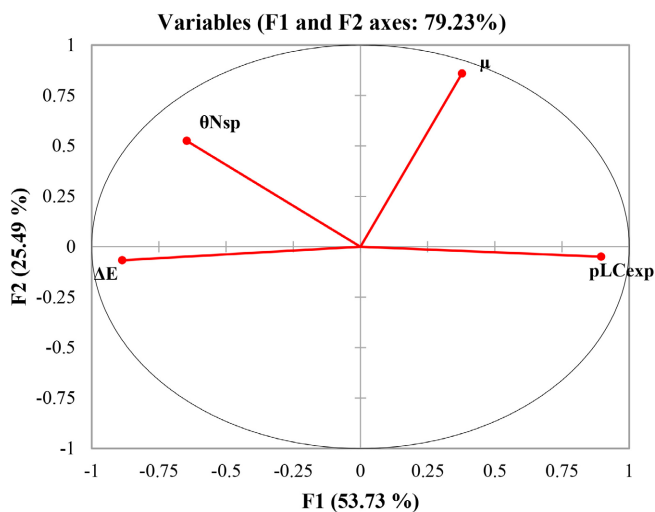
**Table 2.** Correlation matrix (Pearson(n)) between the different descriptors.

| Variables             | $\Delta E$ (eV) | $\mu$ (D) | $\theta_{Nsp}$ (°) | pLC <sub>100exp</sub> |
|-----------------------|-----------------|-----------|--------------------|-----------------------|
| $\Delta E$ (eV)       | 1               | -0.2821   | 0.3893             | -0.7386               |
| $\mu$ (D)             | -0.2821         | 1         | 0.0165             | 0.2172                |
| $\theta_{Nsp}$ (°)    | 0.3893          | 0.0165    | 1                  | -0.4514               |
| pLC <sub>100exp</sub> | -0.7386         | 0.2172    | -0.4514            | 1                     |

Bold values are different from 0 at a significant level for  $p < 0.05$ . Very significant at  $p < 0.01$ . Very significant at  $p < 0.001$ .

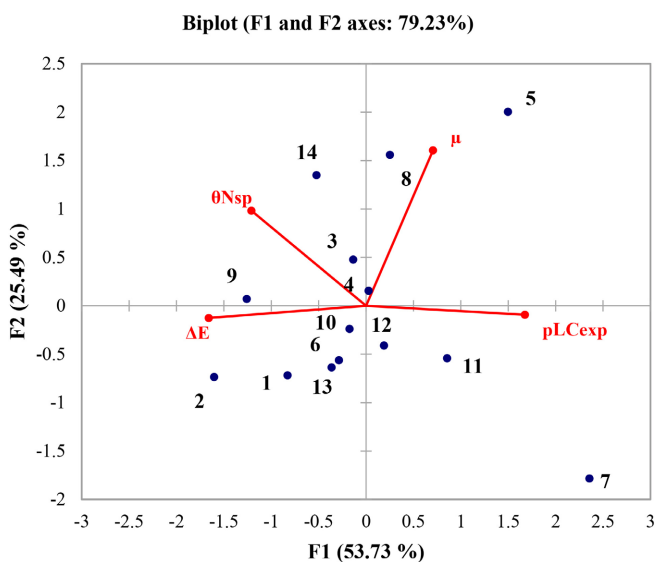


The Pearson correlation coefficients are summarised in the following **Table 2**. The matrix obtained provides information on the negative or positive correlation between the variables. The energy gap ( $\Delta E$ ) is negatively correlated with the nematocidal activity pLC<sub>100</sub> ( $r = -0.7386$  and  $p < 0.05$ ) at a significant level.



**Figure 3.** Circle of correlation.

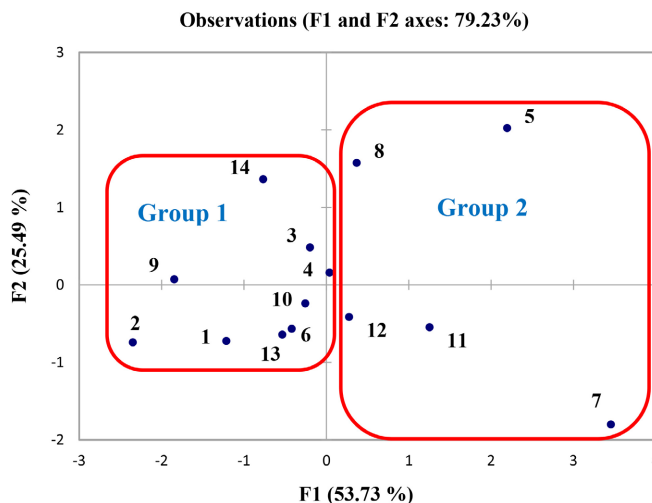
The correlation circle was performed to detect the connection between the different descriptors. Principal component analysis from the correlation circle (**Figure 3**) revealed that the F1 axis (53.73% of the variance) appears to represent the mean valence angle ( $\theta$ Nsp) and the energy gap ( $\Delta E$ ), and the F2 axis (25.49% of the variance) appears to represent the dipole moment ( $\mu$ ).



**Figure 4.** Cartesian diagram according to F1 and F2: correlation between the descriptors used and the BZCs.

The Cartesian diagram in **Figure 4** linking the 14 BZCs to the three descriptors

studied, shows that there is a connection between the compounds BZC-5 and BZC-8 and the dipole moment ( $\mu$ ).



**Figure 5.** Cartesian diagram according to F1 and F2: Separation between group 1 ( $pLC_{100} < 6.00$ ) and group 2 ( $pLC_{100} > 6.00$ ).

**Figure 5** shows a distribution of the BZCs into two groups: group 1 containing compounds with  $pLC_{100} < 6.00$  and group 2 containing compounds with  $pLC_{100} > 6.00$ .

### 3.2. Multiple Linear Regression (MLR)

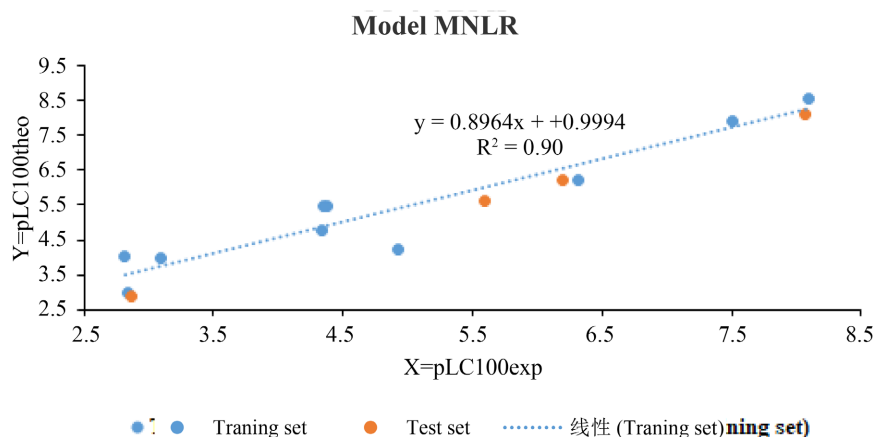
The equation of the QSAR model with statistical data is presented below. **Figure 6** shows the correlation between the experimental and theoretical nematocidal potentials of the test set (blue dots) and the validation set (red dots). The resulting model relates the nematocidal activity to the theoretical descriptors of the protonated molecules. The negative or positive sign of the coefficient of a descriptor in the model reflects the proportionality effect between the evolution of biological activity and this parameter in the regression equation. The negative sign indicates that when the value of the descriptor is high, biological activity decreases. The positive sign indicates the opposite effect. The resulting equation is presented below:

$$pLC_{100} = 1096 - 2.13190 * \Delta E + 0.24723 * \mu - 8.78931 * \theta Nsp \quad (11)$$

$$N = 10, \quad R^2 = 0.90, \quad Q_{CV}^2 = 0.82, \quad S = 0.64, \quad F = 69.21$$

The negative signs of the coefficients of energy gap ( $\Delta E$ ) and mean valence angle ( $\theta Nsp$ ) reflect that nematocidal activity will be enhanced for low values of these descriptors. And the positive sign of the dipole moment ( $\mu$ ) also reflects that the nematocidal activity will be improved for high values of the dipole moment. The significance of the model is reflected by the Fischer coefficient  $F = 69.21$ : the correlation coefficient of the cross-validation  $Q_{CV}^2 = 0.82$ . This model is acceptable with  $R^2 - Q_{CV}^2 = 0.90 - 0.82 = 0.008 < 0.3$ . The regression line between the

experimental and theoretical nematocidal potentials of the test set and the validation set is shown in **Figure 6**.



**Figure 6.** The regression line of the MLR model.

### Verification of Tropsha's Criteria

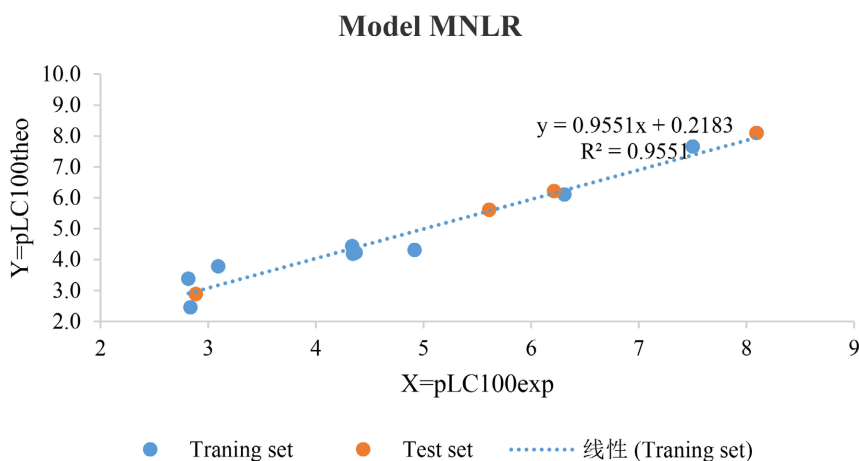
$$R_{Test}^2 = 1 > 0.7, \quad Q_{Cv\ Test}^2 = 1 > 0.6, \quad |R_{Test}^2 - R_0^2| = 0.007 \leq 0.3$$

$$\frac{|R_{Test}^2 - R_0^2|}{R_{Test}^2} = 0.007 < 0.1 \quad \text{and} \quad 0.85 \leq k = 0.996 \leq 1.15;$$

$$\frac{|R_{Test}^2 - R_0'^2|}{R_{Test}^2} = 0.007 < 0.1 \quad \text{and} \quad 0.85 \leq k' = 1.004 \leq 1.15.$$

As *mt* the Tropsha criteria, so the model is acceptable for predicting nematocidal activity.

### 3.3. Multiple Non-Linear Regression (MNLr)



**Figure 7.** The regression line of the MNLr model.

The non-linear regression statistical method was used to improve the predicted  $L_{00}$  nematocidal activity in a quantitative way. It takes into account the three

selected descriptors ( $\Delta E$ ,  $\mu$ ,  $\theta_{\text{Nsp}}$ ). It is the most common tool for the study of multidimensional data. This statistical method is applied to the data in **Table 1** containing 14 molecules associated with the three descriptors. The resulting equation is:

$$\text{pLC}_{100} = 2788144 - 10.43862 * \Delta E - 0.39036 * \mu - 44603 * \theta_{\text{Nsp}} - 1.36271 * \Delta E^2 + 0.03197 * \mu^2 + 178.38332 * \theta_{\text{Nsp}}^2 \quad (12)$$

$$N = 10, R^2 = 0.955, Q_{\text{CV}}^2 = 0.955, S = 0.422, F = 170.228$$

The significance of the model is expressed by the Fischer coefficient  $F = 17228$ : the correlation coefficient of the cross-validation  $Q_{\text{CV}}^2 = 0.955$ . This model is acceptable with  $R^2 - Q_{\text{CV}}^2 = 0.955 - 0.955 = 0.000 < 0.3$ . The regression line between the experimental and theoretical nematocidal potentials of the test set (blue points) and the test set (red points) is shown in **Figure 7**.

### Checking Roy's Criteria

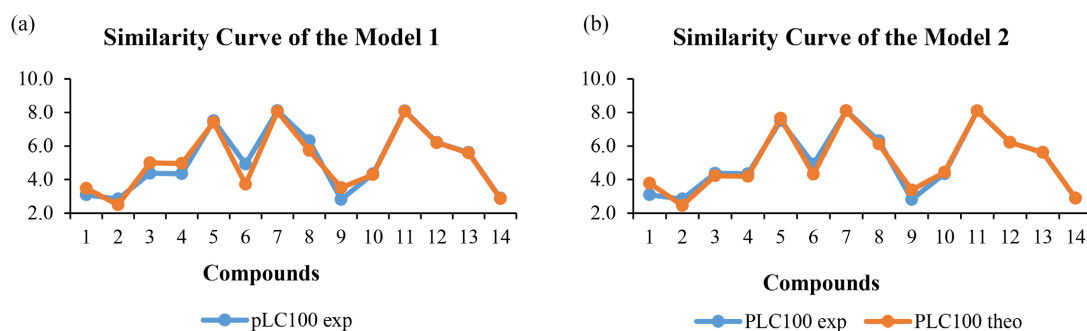
$$\overline{r_m^2} = \frac{r_m^2 + r_m'^2}{2} = 0.919 > 0.5; \quad \Delta r_m^2 = |r_m^2 - r_m'^2| = 0.000 < 0.2$$

$$\text{Avec } r_m^2 = 0.919 \text{ et } r_m'^2 = 0.919$$

All values meet Tropsha's criteria, so the model is acceptable for predicting nematocidal activity.

Of the two models, the model obtained by the statistical method MNLR has a significantly better predictive ability than the MLR approach.

The low standard deviation values are 0.64 and 0.422 in models 1 and 2, respectively, showing that the predicted and experimental values are very similar (**Figure 8**). These curves show similar evolution of these values in both models of BZC series derived from benzimidazoles, despite some recorded differences.

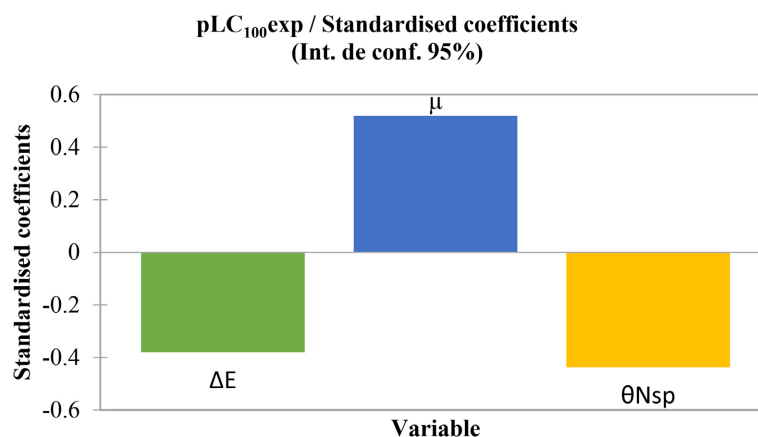


**Figure 8.** Similarity curve of experimental and predicted values for models 1 and 2.

However, as this model is a function of three theoretical descriptors, it is essential to determine the contribution of each one in the prediction of the nematocidal activity for this series of molecules. Indeed, the knowledge of this contribution makes it possible to establish the order of priority of the various descriptors and to define the choice of the parameters to be optimised for the achievement of a better activity.

### 3.4. Analysis of the Contribution of the Descriptors

The contribution of the three descriptors of this model in predicting the nematocidal activity of Benzimidazolyl-Chalcones was determined. The various contributions are illustrated in **Figure 9**.

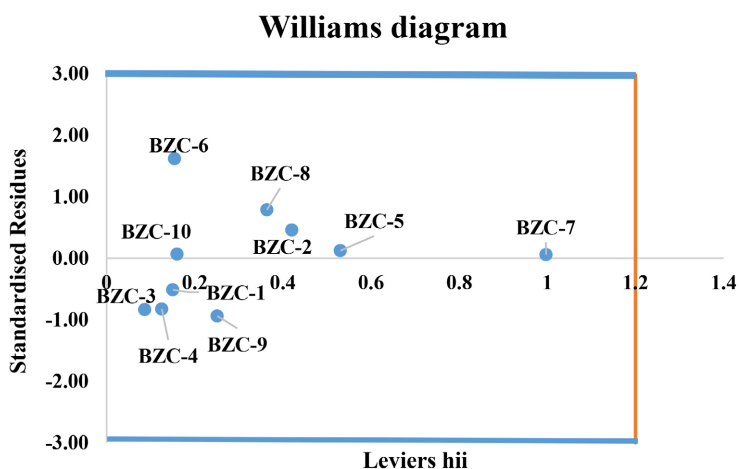


**Figure 9.** Contribution of descriptors in the models.

The dipole moment ( $\mu$ ) shows a large proportion followed by the mean valence angle ( $\theta N_{sp}$ ) and finally the energy gap ( $\Delta E$ ). Thus it is noted that to improve nematocidal activity, a high value of dipole moment ( $\mu$ ) is required. This claim is evidenced by **Figure 4** where compounds BZC-5 and BZC-8 which have high values of dipole moment ( $\mu$ ) correlate well with  $pLC_{100}$  activity.

### 3.5. Domain of Applicability Analysis

The range of applicability of the MLR and MNLR models was determined by the lever method. The values of the  $h_{ii}$  levers and the standardised residues of the molecules were used to plot the standardised residues against the  $h_{ii}$  levers in **Figure 10**.



**Figure 10.** Graph of Standardised Residues according to the levers of the MLR and MNLR models.

For the 10 molecules of the training set and the 3 descriptors of the model, the threshold value of the levers  $h^*$  is 1.2. The extreme values of the standardised residuals are  $\pm 3$  according to the “three sigma rule” [42]. These different values delimit the area of applicability [43] of the model as shown on the graph in **Figure 10**. This figure shows that all molecules have levers below the threshold lever ( $h^* = 1.2$ ) and standardised residue values between +3 and -3. This result means that all molecules belong to the applicability domain.

## 4. Conclusion

This study has made it possible to highlight relationships between the nematocidal activity  $LC_{100}$  ( $\mu\text{g/mL}$ ), which is an interaction quantity by its size, and the descriptors of the interacting molecules by protonation. From a chemical point of view, these descriptors obtained by molecule interaction will be able to guide the experimenter in the synthesis of new more active molecules. The descriptors of the protonated molecules ( $\Delta E$ ,  $\mu$ ,  $\theta_{\text{Nsp}}$ ) can explain and predict the nematocidal activity of BZCs because there are strong correlations between the calculated and experimental values of nematocidal activity. Statistical methods such as principal component analysis (PCA), multilinear and non-linear regression were used. The robustness study of the two models (MLR and MNLR) constructed shows good stability and excellent predictive power. Moreover, compared to the MLR model, the MNLR model ( $R^2 = 0.955$ ,  $S = 0.422$ ,  $F = 170.228$ ) is better and constitutes an efficient tool to predict the nematocidal activity of the best studied BZC analogues called “leads”. Furthermore, the study of the contribution of the descriptors showed that the dipole moment ( $\mu$ ) is the priority descriptor in predicting the nematocidal activity of the studied protonated BZCs. Furthermore, the positive sign of the dipole moment coefficient ( $\mu$ ) in the MLR model equation reflects that high values of dipole moment ( $\mu$ ) could enhance the nematocidal activity of BZCs.

## Acknowledgements

Our thanks go to the entire team of the Institute of Analytical Sciences and Physico-Chemistry for the Environment and Materials (IPREM), the University of Pau and the Pays de l'Adour. Particularly Dr. Panaghiotis Karamanis, CNRS-IPREM-UPPA/ECP Technopole Helioparc Research Officer who facilitated the collaboration with the French computing center.

## Disclaimer (Artificial Intelligence)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during writing or editing of this manuscript.

## Conflicts of Interest

Authors have declared that no competing interests exist.

## References

- [1] Horton, J. (2003) Global Anthelmintic Chemotherapy Programs: Learning from History. *Trends in Parasitology*, **19**, 405-409. [https://doi.org/10.1016/s1471-4922\(03\)00171-5](https://doi.org/10.1016/s1471-4922(03)00171-5)
- [2] Hotez, P.J., Brindley, P.J., Bethony, J.M., King, C.H., Pearce, E.J. and Jacobson, J. (2008) Helminth Infections: The Great Neglected Tropical Diseases. *Journal of Clinical Investigation*, **118**, 1311-1321. <https://doi.org/10.1172/jci34261>
- [3] Geerts, S. and Gryseels, B. (2000) Drug Resistance in Human Helminths: Current Situation and Lessons from Livestock. *Clinical Microbiology Reviews*, **13**, 207-222. <https://doi.org/10.1128/cmr.13.2.207>
- [4] Waller, P.J. (2003) The Future of Anthelmintics in Sustainable Parasite Control Programs for Livestock. *Helminthologia*, **40**, 97-102.
- [5] Souness, J.E., Aldous, D. and Sargent, C. (2000) Immunosuppressive and Anti-Inflammatory Effects of Cyclic AMP Phosphodiesterase (PDE) Type 4 Inhibitors. *Immunopharmacology*, **47**, 127-162. [https://doi.org/10.1016/s0162-3109\(00\)00185-5](https://doi.org/10.1016/s0162-3109(00)00185-5)
- [6] Podunavac-Kuzmanovic, S. and Cvetkovic, D. (2011) Lipophilicity and Antifungal Activity of Some 2-Substituted Benzimidazole Derivatives. *Chemical Industry and Chemical Engineering Quarterly*, **17**, 9-15. <https://doi.org/10.2298/ciceq100329044p>
- [7] Kus, C., Ayhan-Kilcigil, G., Eke, B.C. and iScan, M. (2004) Synthesis and Antioxidant Properties of Some Novel Benzimidazole Derivatives on Lipid Peroxidation in the Rat Liver. *Archives of Pharmacol Research*, **27**, 156-163. <https://doi.org/10.1007/bf02980099>
- [8] Carta, A., Loriga, M., Zanetti, S. and Sechi, L.A. (2003) Quinoxalin-2-Ones: Part 5. Synthesis and Antimicrobial Evaluation of 3-Alkyl-, 3-Halomethyl- and 3-Carboxyethylquinoxaline-2-Ones Variously Substituted on the Benzo-Moiety. *Il Farmaco*, **58**, 1251-1255. [https://doi.org/10.1016/S0014-827X\(03\)00198-8](https://doi.org/10.1016/S0014-827X(03)00198-8)
- [9] Andriole, V.T. (2000) The Quinolones. In: Andriole, V.T., Ed., *The Quinolones*, Elsevier, 477-495. <https://doi.org/10.1016/b978-012059517-4/50017-9>
- [10] Emami, S., Shafiee, A. and Foroumadi, A. (2005) Quinolones: Recent Structural and Clinical Developments. *Iranian Journal of Pharmaceutical Research*, **3**, 123-136.
- [11] Mertens, A., Mueller-Beckmann, B., Kampe, W., Hoelck, J.P. and Von der Saal, W. (1987) Nonsteroidal Cardiotonics. 1.2-Pyridyl-6, 7-Dihydro-3h, 5h-Pyrrolo[2, 3-F]benzimidazol-6-Ones, a Novel Class of Cardiotonic Agents. *Journal of Medicinal Chemistry*, **30**, 1279-1287. <https://doi.org/10.1021/jm00391a004>
- [12] Oprea, T.I. (2005) Chemoinformatics in Drug Discovery. WILEY-VCH Verlag.
- [13] Rekka, E.A. and Kourounakis, P.N. (2008) Chemistry and Molecular Aspects of Drug Design and Action. Taylor & Francis Group
- [14] Ouattara, M., Sissouma, D., Koné, M., Menan, H., Touré, S. and Ouattara, L. (2011) Synthesis and Anthelmintic Activity of Some Hybrid Benzimidazolyl-Chalcone Derivatives. *Tropical Journal of Pharmaceutical Research*, **10**, 767-775. <https://doi.org/10.4314/tjpr.v10i6.10>
- [15] Zhao, Y., Tishchenko, O. and Truhlar, D.G. (2005) How Well Can Density Functional Methods Describe Hydrogen Bonds to  $\pi$  Acceptors? *The Journal of Physical Chemistry B*, **109**, 19046-19051. <https://doi.org/10.1021/jp0534434>
- [16] Adamo, C. and Barone, V. (1998) Exchange Functionals with Improved Long-Range Behavior and Adiabatic Connection Methods without Adjustable Parameters: The mPW and mPW1PW Models. *The Journal of Chemical Physics*, **108**, 664-675. <https://doi.org/10.1063/1.475428>

- [17] Lynch, B.J., Fast, P.L., Harris, M. and Truhlar, D.G. (2000) Adiabatic Connection for Kinetics. *The Journal of Physical Chemistry A*, **104**, 4811-4815. <https://doi.org/10.1021/jp000497z>
- [18] Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., *et al.* (2004) Gaussian 03, Revision E.01. Gaussian, Inc.
- [19] (2016) Microsoft Excel, "(15.0.4420.1017) MSO (15.0.4420.1017) 64 Bits". Microsoft Office Professionnel.
- [20] XLSTAT (2014) XLSTAT and Addinsoft are Registered Trademarks of Addinsoft. Copyright Addinsoft.
- [21] Koopmans, T. (1934) Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms. *Physica*, **1**, 104-113. [https://doi.org/10.1016/s0031-8914\(34\)90011-2](https://doi.org/10.1016/s0031-8914(34)90011-2)
- [22] Kone, M.G., Affi, S.T., Ziao, N., Bamba, K. and Assanvo, E.F. (2016) Protonation Sites in Benzimidazolyl-Chalcones Molecules: An Ab Initio and DFT Investigation. *Computational Chemistry*, **4**, 65-72. <https://doi.org/10.4236/cc.2016.43006>
- [23] Chatterjee, S., Hadi, A. and Price, B. (2000) Regression Analysis by Examples. Wiley VCH.
- [24] Huynh, T.N.P. (2007) Synthèse et étude des relations structure/activité quantitatives (QSAR/2D) d'analogues Benzo[c]phénanthridiniques. <https://theses.hal.science/tel-00346332v1/preview/TheseHUYNH.pdf>
- [25] Snedecor, G.W. and Cochran, W.G. (1967) Statistical Methods. Oxford and IBH, 381.
- [26] Diudea, M. (2000) QSPR/QSAR Studies for Molecular Descriptors, New York. Nova Science.
- [27] Esposito, E.X., Hopfinger, A.J. and Madura, J.D. (2004) Methods for Applying the Quantitative Structure-Activity Relationship Paradigm. In: Bajorath, J., Ed., *Chemoinformatics*, Humana Press, 131-213. <https://doi.org/10.1385/1-59259-802-1:131>
- [28] Larif, M., Chtita, S., Adad, A., Hmamouchi, R., Bouachrine, M. and Lakhli, T. (2013) Predicting Biological Activity of Anticancer Molecules 3-Ary 1-4-Hydroxyquinolin-2-(1H)-One by DFT-QSAR Models. *Int J*, **3**, 32-42.
- [29] Hmamouchi, R., Taghki, A., Larif, M., Adad, A., Abdellaoui, A., Bouachrine, M. and Lakhli, T. (2013) The Inhibitory Activity of Aldose Reductase of Flavonoid Compounds: Combining DFT and QSAR Calculations. *Journal of Chemical and Pharmaceutical Research*, **5**, 198-209.
- [30] Rücker, C., Rücker, G. and Meringer, M. (2007) Y-Randomization and Its Variants in QSPR/QSAR. *Journal of Chemical Information and Modeling*, **47**, 2345-2357. <https://doi.org/10.1021/ci700157b>
- [31] Eriksson, L., Jaworska, J., Worth, A.P., Cronin, M.T.D., McDowell, R.M. and Gramatica, P. (2003) Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs. *Environmental Health Perspectives*, **111**, 1361-1375. <https://doi.org/10.1289/ehp.5758>
- [32] Golbraikh, A. and Tropsha, A. (2002) Beware of  $q^2$ ! *Journal of Molecular Graphics and Modelling*, **20**, 269-276. [https://doi.org/10.1016/s1093-3263\(01\)00123-1](https://doi.org/10.1016/s1093-3263(01)00123-1)
- [33] Tropsha, A., Gramatica, P. and Gombar, V.K. (2003) The Importance of Being Earnest: Validation Is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR & Combinatorial Science*, **22**, 69-77. <https://doi.org/10.1002/qsar.200390007>
- [34] Roy, P.P. and Roy, K. (2008) On Some Aspects of Variable Selection for Partial Least Squares Regression Models. *QSAR & Combinatorial Science*, **27**, 302-313.



- <https://doi.org/10.1002/qsar.200710043>
- [35] Nikolova-Jeliazkova, N. and Jaworska, J. (2005) An Approach to Determining Applicability Domains for QSAR Group Contribution Models: An Analysis of SRC Kowwin. *Alternatives to Laboratory Animals*, **33**, 461-470.  
<https://doi.org/10.1177/026119290503300510>
- [36] Sahigara, F., Mansouri, K., Ballabio, D., Mauri, A., Consonni, V. and Todeschini, R. (2012) Comparison of Different Approaches to Define the Applicability Domain of QSAR Models. *Molecules*, **17**, 4791-4810.  
<https://doi.org/10.3390/molecules17054791>
- [37] Roy, K., Kar, S. and Das, R.N. (2015) Statistical Methods in QSAR/QSPR. In: Roy, K., Kar, S. and Das, R.N., Eds., *A Primer on QSAR/QSPR Modeling*, Springer International Publishing, 37-59. [https://doi.org/10.1007/978-3-319-17281-1\\_2](https://doi.org/10.1007/978-3-319-17281-1_2)
- [38] Jaworska, J., Nikolova-Jeliazkova, N. and Aldenberg, T. (2005) QSAR Applicability Domain Estimation by Projection of the Training Set in Descriptor Space: A Review. *Alternatives to Laboratory Animals*, **33**, 445-459.  
<https://doi.org/10.1177/026119290503300508>
- [39] Ghamali, M., Chtita, S., Bouachrine, M. and Lakhli, T. (2016) Méthodologie générale d'une étude RQSA/RQSP. *Revue Interdisciplinaire*, **1**, 4-5.
- [40] Chtita, S., Ghamali, M., Hmamouchi, R., Elidrissi, B., Bourass, M., Larif, M., *et al.* (2016) Investigation of Antileishmanial Activities of Acridines Derivatives against Promastigotes and Amastigotes Form of Parasites Using Quantitative Structure Activity Relationship Analysis. *Advances in Physical Chemistry*, **2016**, Article ID: 5137289. <https://doi.org/10.1155/2016/5137289>
- [41] Asadollahi, T., Dadfarnia, S., Shabani, A.M.H., Ghasemi, J.B. and Sarkhosh, M. (2011) 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*, **16**, 1928-1955. <https://doi.org/10.3390/molecules16031928>
- [42] Chtita, S., Larif, M., Ghamali, M., Bouachrine, M. and Lakhli, T. (2015) 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*, **9**, 143-154. <https://doi.org/10.1016/j.jtusci.2014.10.006>
- [43] Gramatica, P. (2007) Principles of QSAR Models Validation: Internal and External. *QSAR & Combinatorial Science*, **26**, 694-701.  
<https://doi.org/10.1002/qsar.200610151>