Quantitative Structure-Activity Study against *Plasmodium falciparum* of a Series of Derivatives of Azetidine-2-Carbonitriles by the Method of Density Functional Theory

Jean Stéphane N'dri [1], Bafétigué Ouattara [4], Mamadou Guy-Richard Koné [1, 5, *], Ahmont Landry Claude Kablan [3], Georges Stéphane Dembélé [1, 5], Charles Guillaume Kodjo [1, 2] and Nahossé Ziao [1, 5]

[1] Laboratoire de Thermodynamique et de Physico-Chimie du Milieu, UFR SFA, Université Nangui Abrogoua 02 BP 801 Abidjan 02, Côte-d'Ivoire

[2] Laboratoire de Chimie BioOrganique et de Substances Naturelles, Université Nangui Abrogoua, UFR-SFA, 02 B.P. 801 Abidjan 02 Côte-d'Ivoire

[3] UFR des Sciences Biologiques, Université Péléforo Gon Coulibaly de Korhogo, BP 1328 Korhogo, Côte d'Ivoire

[4] Laboratoire de Physique Fondamentale et Appliquée, Université Nangui Abrogoua, UFR-SFA, 02 B.P. 801 Abidjan 02 Côte-d'Ivoire

[5] Groupe Ivoirien de Recherches en Modélisation des Maladies (GIR2M)

Abstract: This work deals with a Quantitative Structure-Activity study against *Plasmodium falciparum* of a series of Azetidine-2-carbonitrile derivatives. Using the MLR and MNLR methods from excel and xlstat software, we have been able to develop two QSAR models based on molecular descriptors and plasmodial activity. Calculation level B3LYP/6-311 G (d, p) was used to determine molecular descriptors. The statistical indicators of the first model obtained by the MLR method are: the regression coefficient found was $\mathbf{R^2}$ = 0.939 with a standard deviation S =0.266, Fischer's coefficient F =82.064, and a cross-validation correlation coefficient $\mathbf{Q_{CV}^2}$ =0.935. The parameters of the second model developed by the MNLR method are: the regression coefficient $\mathbf{R^2}$: de 0.953, a standard deviation S of 0.258, the Fischer's test F of 108.957, and the correlation coefficient of the cross-validation $\mathbf{Q_{CV}^2}$ =0.951. Moreover, these models have shown some interesting statistical performance. The energy of the highest occupied molecular orbital (E_{HOMO}), the dipole moment (μD), and the partition coefficient (log P) are the molecular descriptors responsible for the *Plasmodium falciparum* activity of Azetidine-derivatives 2-carbonitriles. Furthermore, the partition coefficient is the primary descriptor for predicting the biological activity of the studied compounds. From the findings, Eriksson *et al.* and the external validation criteria of Tropsha used to implement the test are verified and accurate.

Keywords: Azetidine-2-carbonitriles; *Plasmodium falciparum*; QSAR; Molecular Descriptors.

1. Introduction

Malaria, also known as "swamp fever", is an infectious disease caused by a unicellular parasite of the genus Plasmodium; it is transmitted by the bite of a female mosquito called Anopheles. With 219 million people ill and 435,000 deaths in 2017, malaria remains the most important and deadly parasitic disease, affecting children under five years of age and pregnant women. 80 % of the cases were recorded in 15 sub-Saharan African countries and India [1-3]. Tropical and subtropical regions are today the main sources of the infection; unfortunately, in sub-Saharan Africa, more than a million lives are lost each year due to this infection [4]. Surveys have shown that an African child loses his or her life every thirty seconds due to

malaria [4]. In Côte d'Ivoire, data from recent studies (Multiple Indicator Cluster Survey (MICS), National Malaria Control Program (NMCP) 2018 reports) do not reveal any particular variations in the distribution of plasmodial species found in the country. *Plasmodium falciparum* remains the predominant species. It is implicated in more than 95% of cases. However, other species are encountered in Côte d'Ivoire in less than 5% of cases. These are *Plasmodium malariae* and *Plasmodium ovale,* and more rarely *Plasmodium vivax* [5]. In humans, malaria is mainly caused by *Plasmodium falciparum* (preponderant in tropical regions*), Plasmodium malariae, Plasmodium ovale* (rarest species, except in West Africa) and *Plasmodium vivax* (least

temperature demanding species). *Plasmodium falciparum* infection is the only potentially fatal infection for humans [6]. The characteristic symptoms in most clinical manifestations of malaria are fevers, headache chills, backache, muscle pain, profuse sweating, nausea, vomiting, diarrhea, and cough. They usually appear within a few weeks after mosquito bites. This endemic infection is responsible for 1.2 million deaths in children under five years of age and 17% of all hospitalizations [7]. Indeed, malaria is a public health problem because of its high prevalence and socioeconomic impact with severe consequences [8]. Several malaria control strategies have been put in place to overcome this deadly parasite, such as chemoprophylaxis and vector control. The administration of drugs characterizes Anti-malarial chemoprophylaxis to target populations such as pregnant women. The aim is to prevent them from developing a severe form of malaria. However, it is now a fact that 20% of deaths in pregnant women are caused by to malaria in sub-Saharan Africa [9]. The skeleton of Azetidin-2-ones has attracted many researchers because of its multiple potentialities against several diseases. Particularly, Azetidin-2-ones have antibacterial properties of cephalosporins and penicillin [10,11]. The Plasmodium falciparum EC50 (μM) activity of a series of thirty (30) Azetidine-2-carbonitrile derivatives were used in our study. Quantitative Structure-Activity Relationship analysis (QSAR) is one of the best and most widely used methods for designing new therapeutic agents [12-14]. This study quantitatively correlates through a mathematical model the structure or properties of compounds with their biological activities. It is increasingly used to reduce the excessive number of experiments, sometimes long and expensive, and drug production costs by pharmaceutical companies [15,16]. QSAR approach has its origins in the studies carried out by Hansch [17] and Free and Wilson [18]. Models relating biological activity to the hydrophobic, electronic, and steric properties of molecules were established by one of QSAR modeling pioneers. Generally, the QSAR model is a function of one-fifth (1/5) of the initial database. By implementing quantum chemistry methods, this work aims to model the observed *Plasmodium falciparum* EC$_{50}$ activities. The molecular descriptors were calculated only from the compounds' chemical structure to predict the Plasmodium falciparum activities of the analogous molecules subsequently. In the QSAR study's specific case, twenty (20) Azetidine-2-carbonitrile derivatives were used for the test set, and ten (10) others from the same series were used for the external validation test. However, statistical methods such as the application-domain, multiple linear regression (RML), and multiple nonlinear regression (RMNL) were employed in this work. This investigation's general objective is to make a descriptive and predictive study of the *Plasmodium falciparum* activity of a series of Azetidine-2-carbonitriles using multivariate statistical analyses.

2. Materials and Methods

2.1. Data sources

Our study's compounds were synthesized and tested by Micah Maetani *et al.* for their *Plasmodium falciparum* [19]. The molecular structures of these compounds are shown in Figure 1 and Table 1.
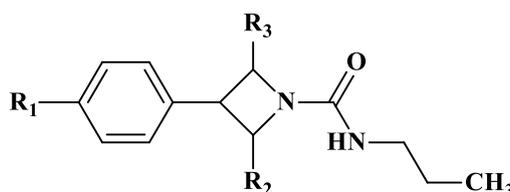

Figure 1. Molecular structure of the compounds studied

2.2. Molecular descriptors

Some theoretical descriptors have been determined to establish a QSAR model. In particular, the energy of the highest occupied molecular orbital (E$_{HOMO}$), the dipole moment(μ_D) and the partition coefficient (log P). The dipole moment (μD) indicates the stability of a molecule in water. Thus, a high dipole moment will reflect the strength of a molecule in water. The Octanol/Water partition coefficient (log P) measures the differential solubility in these two immiscible solvents (water and Octanol) [20]. It is an important measure for identifying drug similarity; according to Lipinski's rule, oral drugs should have log P values greater than or equal to -2 and less than or equal to 5 [21]. This quantity is defined by the expression (1).

$$\log P = \log \frac{[\text{Octanol}]}{[\text{water}]} \tag{1}$$

Where Octanol] and [H$_2$O] are the concentrations of the solute in Octanol and water.

In this work, the partition coefficient was determined using ChemDraw [22]. The energy of the Highest Occupied Molecular Orbital (E$_{HOMO}$) and the dipole moment(μ_D) are obtained in the output file after optimization of molecular structures with Gaussian 09 software.

Table 1. Molecular structures and biological activities of the test sets and validation of the compounds used for the QSAR model.

N°	R1	R2	R3	EC50 (µM)	N°	R1	R2	R3	EC50 (µM)
1				0.010	16				8.375
2				0.249	17				0.005
3*				0.083	18*				0.010
4				0.013	19*				12.630
5				0.016	20				0.139
6*				5.640	21				5.541
7				0.427	22				4.390
8				0.020	23				0.097
9				0.016	24				0.012
10*				0.035	25				0.039
11				0.046	26				0.016
12*				0.019	27				0.015
13				0.051	28*				0.041
14				0.106	29*				1.629
15*				0.019	30*				16.200

* Validation series

2.3. Methodology

The thirty (30) molecules used in this study have Median Effective Concentrations (EC50) ranging from 0.005 to 16.200 µM. This biological activity was expressed by the potential biological pEC50 [23] defined by the following equation (2):

$$pEC_{50} = -\log_{10}(\frac{EC_{50}}{M} * 10^{-3}) \qquad (2)$$

M is the compound's molecular weight in g/mol, and EC50 is the inhibitory concentration in µM.

The molecular structures were optimized using Gaussian 09 [24] at the B3LYP/6-311G (d,p) calculation level. The statistical methods of Multilinear

Regression (RML) and Nonlinear (MNLR) implemented in Excel [25] and XLSTAT [26] spreadsheets were used for the modeling.

2.4. Statistical Analyses Metrics

The performance of a model is determined based on various statistical criteria, including the coefficient of determination R^2, the standard deviation S, the correlation coefficients of the cross-validation Q^2_{CV} and Fischer parameters; F. R^2, S, and F is related to the fit of the computed and experimental values. They describe the predictive capacity within the model's limits and estimate the computed values' accuracy [27]. Below, we present expressions 3 to 6 of the statistical indicators used

$$R^2 = \frac{\sum (Y_{i,pred} - \overline{Y}_{exp})^2}{\sum (Y_{i,exp} - \overline{Y}_{exp})^2} \tag{3}$$

$$F = \frac{\sum (y_{i,pred} - y_{i,exp})^2}{\sum (y_{i,exp} - y_{i,pred})^2} * \frac{n-k-1}{k} \tag{4}$$

$$S = \sqrt{\frac{\sum (y_{i,exp} - y_{i,pred})^2}{n-k-1}} \tag{5}$$

$$Q^2_{CV} = 1 - \frac{\sum (y_{i,exp} - y_{i,pred})^2}{\sum (y_{i,exp} - \overline{y}_{exp})^2} \tag{6}$$

where:

$y_{i,exp}$: The experimental value of *Plasmodium falciparum* activity

$y_{i,pred}$: The predicted value of the activity

$\overline{y}_{i,exp}$: The mean value (average) of the experimental values of activity

n : is the sample size

k : is the number of independent variables in the regression equation

As for the cross-validation coefficient 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 the model. However, the correlation coefficient R^2 gives an evaluation of the dispersion of the theoretical values around the experimental ones. The modeling quality is better when the points are close to the fitting line [27].

The performance of a mathematical model, for Eriksson *et al.*[27], is characterized by a value of $Q^2_{CV} > 0.5$ for a satisfactory model when for the excellent model, $Q^2_{CV} > 0.9$. For these authors, a model will be efficient if the acceptance criterion

R^2- $Q^2_{CV} < 0.3$ is respected. In addition, the predictive power of a model can be obtained from five criteria of Tropsha et al. [28,29]. In this case, the model will be considered acceptable if at least three of the criteria are met.

2.5. Applicability Domain

The Domain of Applicability (DA) defines the area in which a compound can be predicted with confidence. It appears necessary, even mandatory, to determine the DA of any QSAR model as recommended by the Organization for Economic Co-operation and Development (OECD) in its five principles for developing a QSAR model [30]. There are several methods for determining the domain, and in this work, the leverage method has been used. DA considers the variation of the dependent variable's standardized residuals with the distance between the values of the descriptors and their mean as an essential variance in the characterization of the lever [31]. The leverage h_{ii} of a compound the following formula defines me:

$$h_{ii} = x_i (X^t X)^{-1} x_i^t \tag{7}$$

x_i is the row vector of the descriptors of compound i, and X is the matrix built on the values of the model descriptors and compounds of the learning set. The exponent t refers to the matrix or the transposed vector. 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} in the space of the predicted values of the explained variable Y_{pred} as described below:

$$Y_{pred} = H Y_{exp} \tag{8}$$

H is defined by the expression (9):

$$H = X (X^t X)^{-1} X^t \tag{9}$$

The range of applicability is delimited by a threshold value of the leverage noted h*. Generally, it is fixed at $3\frac{p+1}{n}$, where n is the number of compounds in the learning set, and p is the number of the model descriptors [32,33]. For standardized residuals, the two limit values generally used are ±3σ, with σ being the standard deviation of the experimental values of the quantity to be explained [34]: that is the "three-sigma rule" [35].

3. Results and Discussion

The values of the molecular descriptors and the values of the biological potentials of the thirty (30) molecules studied are displayed in Table 2. After that, the partial correlation coefficients aij of the descriptors will also be presented in Table 3.

Table 2. Molecular descriptors and biological potentials of the test and validation set.

Compounds	E_{HOMO} (u.a)	μ_D(Debye)	Log P	pEC50
Training Set				
1	-0.2291	6.4163	3.2400	7.5927
2	-0.2267	7.7103	1.9700	6.1506
4	-0.2246	7.6939	3.2400	7.4787
5	-0.2233	7.8824	3.0800	7.3681
7	-0.2304	6.3581	2.1700	5.9430
8	-0.2308	6.1271	3.6400	7.3095
9	-0.2207	8.2133	2.9600	7.4017
11	-0.2504	7.9039	3.8100	6.9344
13	-0.2416	8.7511	2.9800	6.8576
14	-0.2382	7.7888	2.9800	6.5399
16	-0.2539	7.8796	1.1400	4.5137
17	-0.2322	7.4666	4.3500	7.9460
20	-0.2182	7.9527	2.4900	6.4474
21	-0.2516	7.8669	1.4800	4.8010
22	-0.2531	6.3246	1.5200	4.9885
23	-0.2504	7.6185	3.7400	6.6338
24	-0.2472	6.0748	3.7400	7.5414
25	-0.2508	7.8116	3.7400	7.0295
26	-0.2600	9.5610	4.6600	7.4820
27	-0.2189	9.1339	2.6900	7.4210
Validation Set				
3	-0.2261	6.9282	3.2400	6.6736
6	-0.2313	5.8893	1.7500	4.8221
10	-0.2221	8.9437	3.4900	7.0508
12	-0.2358	7.9748	2.8200	7.2646
15	-0.2413	7.5297	2.9800	7.2864
18	-0.2243	9.4994	2.8200	7.6186
19	-0.2083	7.1288	2.4100	4.4879
28	-0.2251	8.3340	3.5500	6.9777
29	-0.2206	5.2004	3.0300	5.4159
30	-0.2218	5.2203	2.8100	4.3884

Table 3. Pearson correlation matrix.

	E_{HOMO}	μ_D	Log P
E_{HOMO}	1.00		
μ_D	0.11	1.00	
Log P	0.07	0.20	1.00

The analysis of the bivariate data of the descriptors studied, i.e., calculating the partial correlation coefficient between each of the descriptors' pairs, is less than 0.70 (aij <0.70), which translates that these different descriptors are independent of each other [27].

3.1. Multiple Linear Regressions (MLR)

The proportionality effect of the evolution between biological activity and this regression equation parameter is reflected by the negative or positive sign of the coefficient of a descriptor of the model translated. Thus, the negative sign indicates that when the descriptor's value is high, the biological activity decreases while the positive sign reflects the opposite effect. The regression equation for the best QSAR model and the statistical indicators as described below:

Model 1: $pEC_{50} = 11.842 + 33.224 \times E_{HOMO} + 0.008 \times \mu_D + 0.908 \times \log P$

$N = 20$ $R^2 = 0.939$ $Q^2_{CV} = 0.935$ $S = 0.266$
      $F = 82.064 > F_{cr} = 3.24$ $R^2 - Q^2_{CV} = 0.004$

The positive signs of the coefficients of dipole moment, highest occupied molecular orbital energy, and partition coefficient indicate that the *Plasmodium falciparum* activity will be enhanced to higher values of these descriptors. The high values of the coefficients of determination (R^2) and cross-validation (Q^2_{CV}) respectively of 0.939 and 0.935 attest to the robustness of model 1. Moreover, this model is permissible because the value of $R^2 - Q^2_{CV}$ is less than 0.3. The Tropsha criteria check for the external validation set is presented in Table 4.

**Table 4**. Analysis of the Tropsha criteria of the external validation set of model 1.

Statistical parameters	Tropsha criteria [28,29]	Model 1
R^2	> 0.7	0.937
Q^2_{CV}	> 0.6	0.933
$\left| R^2 - R_0^2 \right|$	≤ 0.3	0.00
$\dfrac{\left| R^2 - R_0^2 \right|}{R^2}$	< 0.1	0.00
k	$0.85 \leq k \leq 1.15$	1.00
$\dfrac{\left| R^2 - R_0'^2 \right|}{R^2}$	< 0.1	0.00
k'	$0.85 \leq k' \leq 1.15$	0.998

The Tropsha criteria are verified through the external validation set with model 1. Therefore this model is acceptable for predicting the *Plasmodium falciparum* activity of the series of Azetidine-2-carbonitrile derivatives studied. In addition, the external validation criterion according to Roy *et al.* [36] has been verified; the values of the various parameters are listed in Table 5.

Table 5. Verifications of the Roy criteria of the external validation set by Model 1.

Indicators	r_m^2	r'^2_m	$\overline{r_m^2} = \dfrac{(r_m^2 + r'^2_m)}{2}$	$\Delta r_m^2 = \left| r_m^2 - r'^2_m \right|$
Values	1.000	0.9368	0.9368	0.063
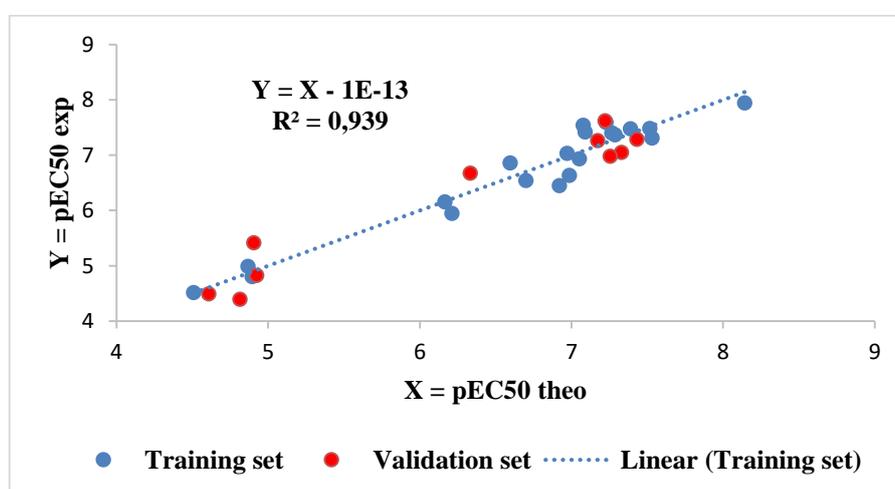
Figure 2. Model 1 regression line

The linear regression curve shows that the values of the predicted activities are close to the experimental ones. The similarity curve (Figure 3) allows a better understanding of the slight difference between the observed and predicted values.
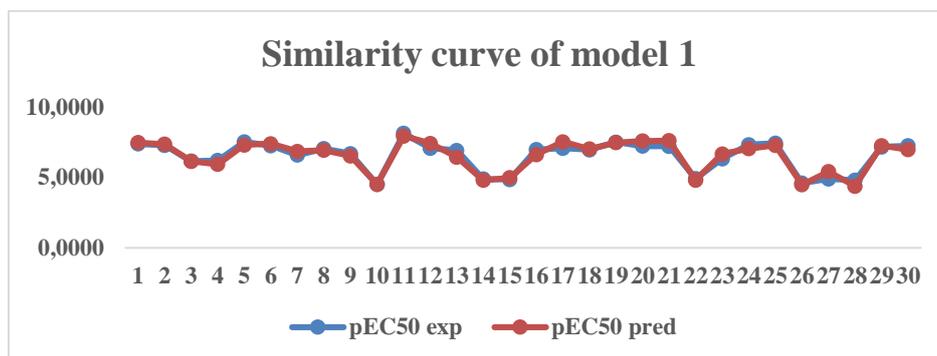

Figure 3. Similarity curve of model 1

Figure 4 below presents a circular diagram illustrating the three (3) descriptors (E_{HOMO}, μ_D, log P) of our model.
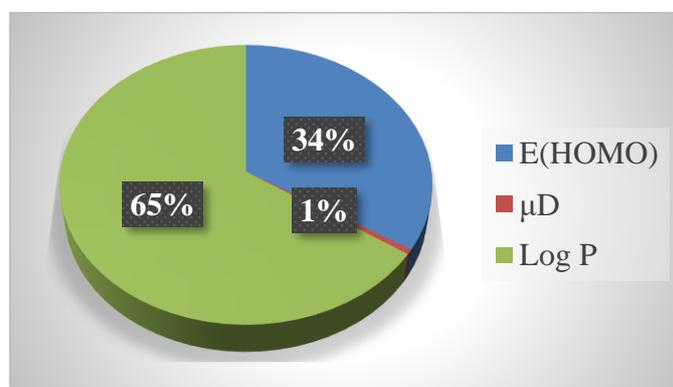

Figure 4. Contribution of the different constituents in Model 1

The partition coefficient has a significant contribution relative to the energy of the highest occupied molecular orbital or dipole moment. Thus, the partition coefficient is the predominant physicochemical descriptor in predicting the *Plasmodium falciparum* strain of the Azetidine-2-carbonitrile derivatives studied.

3.2. Multiple Non-Linear Regression (MNLR)

Multiple Non-Linear Regression MNLR is a non-linear method (exponential, logarithmic, polynomial) that allows the mathematical model's determination that best explains the variability of a given property or activity a function of molecular descriptors. To verify the descriptors' efficiency and relevance, we propose to use another statistical method, in this case, a polynomial model of power 2. Model 2 was established from the same test set used for the MLR based on the following formula:

Model 2 :

$$pEC_{50} = -0.442 - 112.754 \times E_{HOMO} - 1.593 \times \mu_D + 1.341 \times \log P - 300.732 \times (E_{HOMO})^2 + 0.108 \times (\mu_D)^2 - 0.084 \times (\log P)^2$$

N = 20 $R^2 = 0.953$ $Q_{CV}^2 = 0.951$ S = 0.258
F = 108.957 > F_{cr} = 3.24

$$R^2 - Q_{CV}^2 = 0.002$$

This model has high values of the R^2 correlation coefficient and the cross-validation correlation coefficient Q_{CV}^2 which are respectively 0.953 and 0.951. Moreover, the result found; $R^2 - Q_{CV}^2 = 0.002 < 0.3$ Model 2 is therefore reliable and acceptable. The regression line of model 2 is represented by Figure 5.
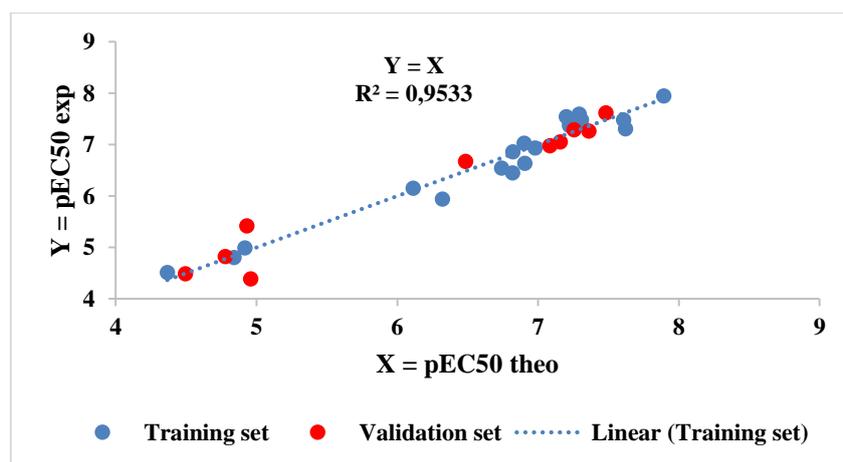
Figure 5. Regression line of Model 2

The regression curve analysis of model 2 shows that all the values of the first twenty-one (21) compounds are almost identical. Whereas, for the values of the remaining nine (9) compounds (between 21-30), there is a slight discrepancy between the predicted and experimental values (Figure 6).
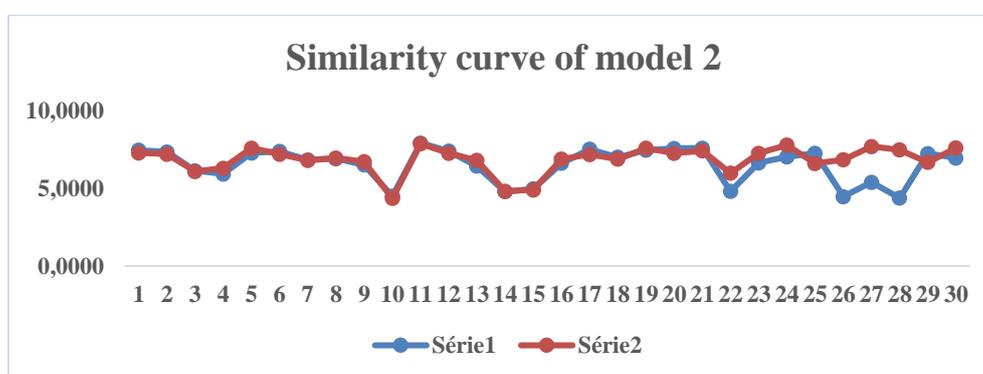

Figure 6. Model 2 similarity curve

3.3 Applicability area

Model 1 cannot be used to predict the biological activity of all molecules. It is, therefore, necessary to define an area in which the formulas can be used.

Figure 7 shows the range of applicability of the models using the Leverage method.
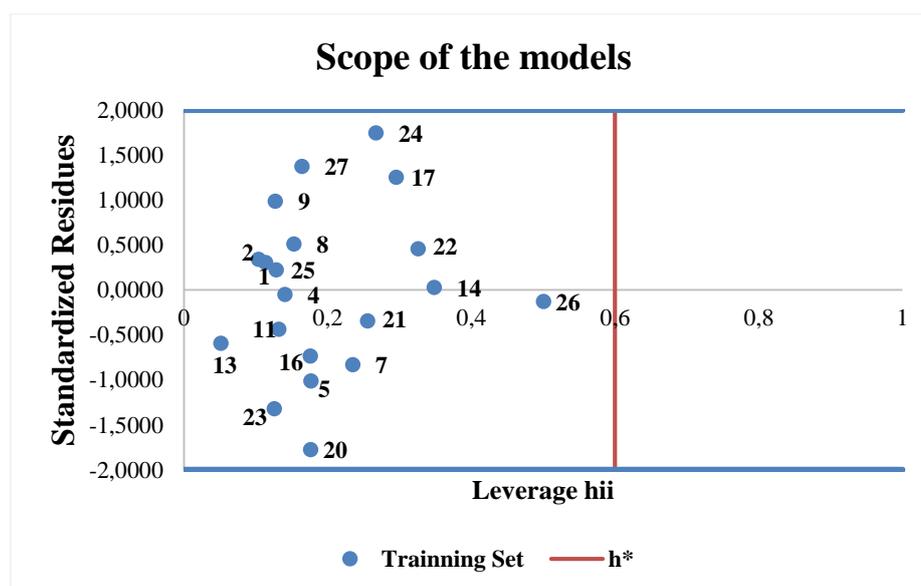

Figure 7. William's Plot of the MLR and MNLR Models

The graph shows that for all molecules of the test set and descriptors of the MLR and MNLR models, the limit value of the lever h* is 0.6. The maximum values of the standardized residues are ±2 according to the "three-sigma rule". These different values delimit the range of applicability of the models, as shown in the graph in Figure 7. All molecules have some values h < h*. Compounds that will fall within this range will have a reliable and predictable biological activity.

4. Conclusion

The highest occupied molecular orbital energy (E_{HOMO}), the dipole moment (μD), and the partition coefficient (log P) allowed us to predict the *Plasmodium falciparum* activity of the Azetidine-2-carbonitrile derivatives studied. In the final step of our study, we found out some. At the end of our work, we observe an excellent correlation between the predicted and experimental values of the effective potential of the median concentration pEC_{50}. Two QSAR models were obtained from the MLR and MNLR methods; these proposed models have revealed that the partition coefficient is the primary descriptor for improving *Plasmodium falciparum* activity. This work's outcomes are promising and represent an accurate compass for designing new and highly effective molecules of the Azetidine-2-carbonitrile family against *Plasmodium falciparum*. The significance of these models was verified using a test set of ten molecules. These results are very encouraging and promising, as they could serve as a solid basis for understanding the quantitative relationship between the physicochemical descriptors of the molecular structure of Azetidine-2-carbonitriles and the *Plasmodium falciparum* strain. Although the set objectives have been achieved for the rest of this work, we plan to use the two QSAR models constructed to screen available Azetidine-2-carbonitrile derivatives libraries to discover those which possess Plasmodium falciparum activities.

References

1- J. Schantz-Dunn, M. N. NOUR, Malaria and Pregnancy: a Global Health Perspective, *Rev. Obstet. Gynecol.*, **2009**, 2, 186-192.

2- World Health Organization, Report of the sixteenth WHOPES working group meeting, Geneva, Switzerland, **2013**.

3- World Health Organization, World malaria report 2014, Geneva, Switzerland, **2014**.

4- Organisation Mondiale de la Santé, Plan d'action mondial contre le paludisme (GMAP), **2005**.

5- M. TANOH, « évaluations des besoins, Draft 1, **2020**.

6- WHO, Grands lignes du plan d'action de l'OMS pour la lutte contre le paludisme. 1993 - 2000, Amsterdam, Conférence ministérielle sur le paludisme, **1992**.

7- Y. J. Golvan, Eléments de parasitologie médicale, 4 édition, Paris, Flammarion, **1983**.

8- M. Gentillini, B. Duflo, Médecine Tropicale, 4 édition, Paris, Flammarion, **1986**.

9- World Health Organization, World malaria report 2014, Geneva, Switzerland, **2008**.

10- S. F. Ayad, S. M. Mahmood, A. A. Mohamed, Synthesis and characterization of mono/bis β-lactams by using [2+2] cycloaddition reaction and study antihyperglycemic activity, *Best Int. J. Humanit. Med. Sci.*, **2014**, 2, 67-78.

11- M. Noolvi, S. Agrawal, H. Patel, A. Badiger, M. Gaba, A. Zambre, Synthesis, antimicrobial and cytotoxic activity of novel azetidine-2-one derivatives of 1Hbenzimidazole, *Arab. J Chem.*, **2014**, 17, 219-226.

12- M. T. Chhabria, B. M. Mahajan, P. S. Brahmkshatriya, QSAR Study of a Series of Acyl Coenzyme A (CoA): Cholesterol Acyltransferase Inhibitors Using Genetic Function Approximation, *Med. Chem. Res.*, **2011**, 20, 1573-1580.

13- V. M. Buha, D. N. Rana, M. T. Chhabria, K. H. Chikhalia, B. M. Mahajan, P. S. Brahmkshatriya, N. K. Shah, Synthesis, Biological Evaluation and QSAR Study of a Series of Substituted Quinazolines as Antimicrobial Agents, *Med. Chem. Res.*, **2013**, 22, 4096-4109.

14- A. Tropsha, Best Practices for QSAR Model Development, Validation, and Exploitation, *Mol. Inform.*, **2010**, 29, 476-488.

15- T. I. Oprea, Chemoinformatics in Drug Discovery, WILEY-VCH Verlag, Allemagne, **2005**.

16- E. A. Rekka, P. N. Kourounakis, Chemistry and Molecular Aspects of Drug Design and Action, CRC Press, **2008**.

17- C. Hansch, T. Fujita, Additions and Corrections-$\rho - \sigma - \pi$ Analysis. A Method for correlation of biological activity and chemical structure, *J. Am. Chem. Soc.*, **1964**, 86, 1616-1626.

18- S. M. Free, J. W. Wilson, A Mathematical Contribution to Structure-Activity Studies, *J Med Chem.*, **1964**, 7, 395-399.

19- M. Maetani, N. Kato, V. A. Jabor, F. A. Calil, M. C. Nonato, C. A. Scherer, S. L. Schreiber, Discovery of Antimalarial Azetidine-2-carbonitriles That Inhibit P. falciparum Dihydroorotate Dehydrogenase, *ACS Med. Chem. Lett.*, **2017**, 4, 438-442.

20- P. J. Taylor, Hydrophobic Properties of Drugs, In Quantitative Drug DesignII, Pergamon Press, Oxford (UK), **1990**, 4, 241-294.

21- C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv. Drug Deliv. Rev.*, **1997**, 6, 3-25.

22- ChemDraw Ultra, *CambridgeSoft©Mulder 875-317589-4732*. 1986.

23- M. G. Koné, J. S. N'dri, C. G. Kodjo, Combining of DFT and QSAR results to predict the

antibacterial activity of a series of azetidinones derived from dapsone as inhibitors of Bacillus Subtilis and Pseudomonas aeruginosa, *SDRP J. Comput. Chem. Mol. Model.*, **2018**, 2, 1-8.

24- M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, D. J. Fox, Gaussian 09 (Wallingford, CT: Gaussian Inc), **2009**.

25- Microsoft Excel, *Microsoft Office Professionnel.* 2016.

26- XLSTAT Version 2014.5.03 Copyright Addinsoft 1995-2014 (2014) XLSTAT and Addinsoft are Registered Trademarks of Addinsoft.

27- J. S. N'dri, A. L. Kablan, B. Ouattara, M. G. Koné, L. P. Ouattara, C. G. Kodjo, N. Ziao, QSAR studies of the Antifungal activities of α-diaminophosphonates Derived from Dapsone by DFT Method, *J. Mater. Phys. Chem.*, **2019**, 7, 1-7.

28- A. Golbraikh, A. Tropsha, Beware of q², *J Mol Graph Model.*, **2002**, 20, 269-276.

29- A. Tropsha, Best Practices for QSAR Model Development, Validation, and Exploitation, *Mol. Inform.*, **2010**, 29, 476-488.

30- K. Roy, S. Kar, R. N. Das, Statistical Methods in QSAR/QSPR, A Primer on QSAR/QSPR Modeling, *Springer cham.*, **2015**, 37-59.

31- J. Jaworska, N. N. Jeliazkova, T. Aldenberg, QSAR Applicability Domain Estimation by Projection of the Training Set in Descriptor Space: A Review, *Alternatives to laboratory animals,* **2005**, 33, 445-459.

32- M. Ghamali, S. Chtita, M. Bouachrine, T. Lakhlifi, Méthodologie générale d'une étude RQSA/RQSP, *Rev. Interdiscip.*, **2016**.

33- S. Chtita, M. Ghamali, R. Hmamouchi, B. Elidrissi, M. Bourass, M. Larif, T. Lakhlifi, Investigation of Antileishmanial Activities of Acridines Derivatives against Promastigotes and Amastigotes Form of Parasites Using QSAR Analysis, *Adv. Phys. Chem.*, **2016**, 1-16.

34- T. Asadollahi, S. Dadfarnia, A. Shabani, J. Ghasemi, 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*, **2011**, 16, 1928-1955.

35- S. Chtita, M. Larif, M. Ghamali, M. Bouachrine, 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, *J. Taibah Univ. Sci.*, **2015**, 9, 143-154.

36- P. P. Roy, K. Roy, On some aspects of variable selection for partial least squares regression models, *QSAR Comb. Sci.*, **2008**, 27, 302-313.