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

**1Jean Stéphane N’DRI,** **4Bafétigué OUATTARA, 1,5\*Mamadou Guy-Richard KONÉ, 3Ahmont Landry Claude KABLAN, 1,5Georges Stéphane DEMBÉLÉ, 1,2Charles Guillaume KODJO, and 1,5Nahossé ZIAO**

*1Laboratoire 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,*

*3UFR des Sciences Biologiques, Université Péléforo Gon Coulibaly de Korhogo, BP 1328 Korhogo, Côte d’Ivoire*

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

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

*\*Corresponding author:* [*guyrichardkone@gmail.com*](mailto:guyrichardkone@gmail.com)*,*

**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, we have been able to develop two QSAR models based on molecular descriptors and antiplasmodial activity. The molecular descriptors were obtained at the calculation level B3LYP/6-311 G (d, p). The statistical indicators of the first model obtained by the MLR method are: the regression coefficient found was **R2** = 0.939 with a standard deviation S = 0.266, Fischer's coefficient F = 82.064 and a cross-validation correlation coefficient = 0.935. The parameters of the second model developed by the MNLR method are: the regression coefficient **R2**: 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 = 0.951. Moreover, these models have shown some interesting statistical performance. Amongst the molecular descriptors; the energy of the highest occupied molecular orbital (EHOMO), the dipole moment (µD), and the partition coefficient (log P) are responsible for the anti-Plasmodium falciparum activity of Azetidine-2-carbonitrile derivatives. Furthermore, the partition coefficient is the primary descriptor for the prediction of the biological activity of the studied compounds. From the findings; the acceptance criteria of 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, which mainly affects children under five years of age and pregnant women. 80 % of the cases were recorded in 15 sub-Saharan African countries and in India [1]–[3]. Tropical and subtropical regions are today the main sources of the infection; unfortunately in sub-Saharan Africa in particular, more than a million lives are lost each year as a result of 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) 2016, 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 its socio-economic impact with serious consequences[8]. In order to overcome this deadly parasite, several malaria control strategies have been put in place such as chemoprophylaxis and vector control. Anti-malarial chemoprophylaxis is characterized by the administration of drugs to target populations such as pregnant women. The aim is to prevent them from developing the 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 the attention of many researchers because of its multiple potentialities against several diseases. Particularly, Azetidin-2-ones have antibacterial properties of cephalosporins and penicillin [10], [11]. In our study, the anti-Plasmodium falciparum EC50 (μM) activity of a series of thirty (30) Azetidine-2-carbonitrile derivatives was used. Quantitative Structure Activity Relationship Analysis (QSAR) is one of the best and most widely used methods for the design of new therapeutic agents [12]–[14]. This study allows to quantitatively correlate 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 the cost of drug production by pharmaceutical companies [15], [16]. This QSAR approach has its origins in the studies carried out by Hansch [17] and Free and Wilson [18]. Indeed, Hansch has successfully established some models relating the biological activity to the hydrophobic, electronic and steric properties of molecules. 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 anti-Plasmodium falciparum EC50 activities, the molecular descriptors being calculated only from the chemical structure of the compounds to subsequently predict the anti-Plasmodium falciparum activities of the analogous molecules. In the specific case of the QSAR study, 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, the field of applicability, Multiple Linear Regression (MLR) and Multiple Non Linear Regression (MNLR) were used in this work. The general objective of this investigation is to make a descriptive and predictive study of the anti-Plasmodium falciparum activity of a series of Azetidine-2-carbonitriles using multivariate statistical analyses.

1. **MATERIALS AND METHODS**
   1. **Data sources**

The compounds in our study were synthesized and tested by Micah Maetani *et al.* for their anti *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

**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

* 1. **Molecular descriptors**

For the development of QSAR models, some theoretical descriptors have been determined. In particular, the energy of the highest occupied molecular orbital (EHOMO), the dipole moment and the partition coefficient (log P). The dipole moment indicates the stability of a molecule in water. Thus, a high dipole moment will indicate low solubility in organic solvents and high solubility in water. The Octanol/Water partition coefficient (log P) measures the differential solubility of a solute in these two immiscible solvents (water and octanol)[20]. It is an important measure for the identification of 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).

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

In this work the partition coefficient was determined using ChemDraw[22].

* 1. **Methodology**

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

Where M is the molecular weight of the compound 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 modeling was done using the multilinear regression method implemented in Excel [25] and *XLSTAT*[26] spreadsheets.

* 1. **Statistical analyses metrics**

The quality of a model is determined on the basis of various statistical criteria including the determination coefficient **R2**, the standard deviation S, the correlation coefficients of the cross-validation and Fischer parameters; **F**. **R2**, **S** and **F** is related to the fit of the computed and experimental values. They describe the predictive capacity within the limits of the model, and allow to estimate the accuracy of the computed values [27]. As for the cross-validation coefficient , 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 R² gives an evaluation of the dispersion of the theoretical values around the experimental ones. The quality of the modeling 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 > 0.5 for a satisfactory model when for the excellent model, > 0.9. According to their studies~~,~~ a model will perform successfully and accurately if the acceptance criterion **R2**- < 0.3. Moreover, the predictive power of a model can be obtained using five criteria highlighted by 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 the development of a QSAR model [30]. There are several methods for determining the domain, and in this work the leverage method has been used. This method is based on the variation of standardized residuals of the dependent variable with the distance between the values of the descriptors and their mean, called leverage [31]. The leverage hii of a compound i is defined by the following formula:

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 are the diagonal elements of a matrix H called hat matrix. H is the projection matrix of the experimental values of the explained variable Yexp in the space of the predicted values of the explained variable as described below:

H is defined by the expression (5):

The range of applicability is delimited by a threshold value of the leverage noted h\*. Generally, it is fixed at , 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].

1. **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**. Thereafter, the values of the partial correlation coefficients aij of the descriptors are also presented in **Table 3**.

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Compounds** | **EHOMO (u.a)** | **(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**: Values of the partial correlation coefficients of the descriptors.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **EHOMO** |  | **Log P** |
| **EHOMO** | 1.00 |  |  |
|  | 0.11 | 1.00 |  |
| **Log P** | 0.07 | 0.20 | 1.00 |

For all the descriptors studied, the analysis of the bivariate data such as the calculation of the partial correlation coefficient between each of the pairs and the descriptors; when this value is less than 0.70 (aij < 0.70), it means that the different descriptors are independent ~~of~~ from each other [27] .

* 1. **Multiple Linear Regressions (MLR)**

It should be noted that the negative or positive sign of the coefficient of a model descriptor reflects the proportionality effect between the evolution of the biological activity and the regression equation. Thus, the negative sign indicates that when the value of the descriptor 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:

N = 20 S = 0.266 F = 82.064 > Fcr = 3.24

The positive signs of the coefficients of dipole moment, highest occupied molecular orbital energy and partition coefficient indicate that the anti-Plasmodium falciparum activity will be enhanced to higher values of these descriptors. The significance of model 1 is translated by high values of the correlation coefficient R2: 0.939 and the cross-validation correlation coefficient : 0.935. This model is acceptable because the value of is less than 0.3. The Tropsha criteria check for the external validation set are presented in **Table 4**.

**Table 4:** Verification of Tropsha criteria for the external validation set for Model 1

|  |  |  |
| --- | --- | --- |
| Statistical parameters | Tropsha criteria [28,29] | Model 1 |
|  |  | 0.937 |
|  |  | 0.933 |
|  |  | 0.00 |
|  |  | 0.00 |
|  |  | 1.00 |
|  |  | 0.00 |
|  |  | 0.998 |

All Tropsha criteria are verified by the external validation set of model 1. This model is therefore acceptable for the prediction of the anti-Plasmodium falciparum activity of the series of Azetidine-2-carbonitrile derivatives studied. Also the external validation criterion according to Roy *et al*.[36] has been verified and the values of the different parameters are listed in **table 5**.

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Indicators |  |  |  |  |
| Values |  |  | 0.9368 | 0.063 |

The analysis of this table shows that, the is greater than 0.5 and the is less than 0.2. We can therefore state that the model is robust and has good predictive power.

The regression line between the experimental and theoretical activities of *Plasmodium falciparum* control activities of the test set (blue dots) and the validation set (red dots) is shown in **Figure 2**.

**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 small difference between the experimental and predicted values.

**Figure 3:** Similarity curve of model 1.

The result of the study of the relative contribution of the descriptors in predicting the anti-Plasmodium falciparum activity of compounds is presented in **Figure 4.**

**Figure 4**: Contribution of the different constituents in Model 1

The partition coefficient has a large contribution relative to the energy of the highest occupied molecular orbital or dipole moment. Thus, the partition coefficient proves to be the priority descriptor in the prediction of the anti-Plasmodium falciparum activity of the Azetidine-2-carbonitrile derivatives studied.

**3.2 Multiple Non Linear Regression (MNLR)**

Multiple Non Linear Regression MNLR is a nonlinear method (exponential, logarithmic, polynomial) that allows to determine the mathematical model that best explains the variability of a given property or activity as a function of molecular descriptors. In this work, we used the polynomial model based on the descriptors proposed by the linear model which will be raised to the power 2. The model 2 was established from the same test set used for the MLR based on the following formula:

Model 2 :

N = 20 S = 0.258 F = 108.957 > Fcr = 3.24

This model has high values of the R2 correlation coefficient and the cross-validation correlation coefficient which are respectively 0.953 and 0.951. Moreover, the result found; 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 analysis of the regression curve 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**

The models obtained 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 (EHOMO), the dipole moment (µD) and the partition coefficient (log P) allowed us to predict the anti-Plasmodium falciparum activity of the Azetidine-2-carbonitrile derivatives studied. In the final step of our study, we found out some strong correlations between the calculated and experimental values of the biological potential. Two QSAR models were obtained from the MLR and MNLR methods, interestingly; these proposed models have revealed that the partition coefficient is the major descriptor for improving anti-Plasmodium falciparum activity. The outcomes of this work are promising and represent a real compass for the design of 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. Based on the results obtained; the current work will play an important role in understanding the relationship between the physicochemical parameters of the structure and the biological activity.

**REFERENCES**

[1] J. SCHANTZ-DUNN et N. NOUR, « Malaria and Pregnancy: A Global Health Perspective. », *Rev. Obstet. Gynecol.*, p. 186‑192.

[2] WHO, « World malaria report 2013 », Geneva, Switzerland, 2013.

[3] WHO, « World malaria report 2014 », Geneva, Switzerland, 2014.

[4] Roll Back Malaria, « Données, chiffres et stratégies clés. Plan d’action mondialcontre 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. GOLVAN, *Parasitologie médicale*, 4e éd. Paris: Flammarion, 1983.

[8] M. GENTILLINI, *Médecine Tropicale*, 4e éd. Paris: Flammarion, 1986.

[9] WHO, « World malaria report 2008 », Geneva, Switzerland, 2008.

[10] S. F. Ayad, S. M. Mahmood, et 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.*, vol. 2, no 14, p. 67‑78, 2014.

[11] M. Noolvi, S. Agrawal, H. Patel, A. Badiger, M. Gaba, et A. Zambre, « Synthesis, antimicrobial and cytotoxic activity of novel azetidine-2-one derivatives of 1Hbenzimidazole », *Arab. J Chem*, no 17, p. 219‑226, 2014.

[12] M. Chhabria, B. Mahajan, et P. Brahmkshatriya, « QSAR Study of a Series of Acyl Coenzyme A (CoA): Cholesterol Acyltransferase Inhibitors Using Genetic Function Approximation », *Med. Chem. Res.*, vol. 20, p. 1573‑1580, 2011.

[13] V. Buha *et al.*, « Synthesis, Biological Evaluation and QSAR Study of a Series of Substituted Quinazolines as Antimicrobial Agents », *Med. Chem. Res.*, vol. 22, no 120, p. 4096‑4109, 2013.

[14] A. Tropsha, « Best Practices for QSAR Model Development, Validation, and Exploitation », *Mol. Inform.*, vol. 29, p. 476‑488, 2010.

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

[16] E. A. Rekka et P. N. Kourounakis, *Chemistry and Molecular Aspects of Drug Design and Action*. LLC. Etats Unies: Taylor & Francis Group, 2008.

[17] C. Hansch et T. Fujita, « ρ – σ – π,analysis: method for correlation of biological activity and chemical structure », *J Am Chem Soc*, vol. 86, p. 1616‑1626, 1964.

[18] S. M. Free et J. W. Wilson, « A Mathematical Contribution to Structure-Activity Studies », *J Med Chem*, vol. 7, p. 395‑399, 1964.

[19] M. Maetani *et al.*, « Discovery of Antimalarial Azetidine-2-carbonitriles That Inhibit P. falciparum Dihydroorotate Dehydrogenase. », *ACS Med. Chem. Lett.*, vol. 4, no 18, p. 438‑442, 2017.

[20] P. J. Taylor, *Hydrophobic Properties of Drugs, In Quantitative Drug DesignII*, vol. 4. Oxford (UK): Pergamon Press, 1990.

[21] C. A. Lipinski, F. Lombardo, B. Dominy, et P. Feeney, « Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings », *Adv. Drug Deliv. Rev.*, vol. 6, no 11‑3, p. 3‑25, 1997.

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

[23] J. S. N’dri *et al.*, « 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.*, vol. 2, no 12, p. 1‑9, 2018.

[24] M. J. Frisch *et al.*, *Gaussian 09*. Wallingford CT: Gaussian, Inc, 2009.

[25] Microsoft Excel, *Microsoft Office Professionnel*. 2016.

[26] XLSTAT, *XLSTAT and Addinsoft are Registered Trademarks of Addinsoft*. 2014.

[27] J. S. N’dri *et al.*, « QSAR studies of the Antifungal activities of α-diaminophosphonates Derived from Dapsone by DFT Method », *J. Mater. Phys. Chem.*, vol. 7, no 11, p. 1‑7, 2019.

[28] A. Golbraikh et A. Tropsha, « Beware of qsar », *J Mol Graph Model*, vol. 20, p. 269, 2002.

[29] A. Tropsha, « Best Practices for QSAR Model Development, Validation, and Exploitation », *Mol. Inform.*, vol. 29, p. 476‑488, 2010.

[30] K. Roy, « A Primer on QSAR/QSPR Modeling, Chapter 2 Statistical Methods in QSAR/QSPR », *Springer Briefs Mol. Sci.*, p. 37‑59, 2015.

[31] 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.

[32] M. Ghamali, S. Chtita, M. Bouachrine, et T. Lakhlifi, « Méthodologie générale d’une étude RQSA/RQSP », *Rev. Interdiscip.*, vol. 1, no 11, 2016.

[33] S. Chtita *et al.*, « Investigation of Antileishmanial Activities of Acridines Derivatives against Promastigotes and Amastigotes Form of Parasites Using QSAR Analysis », *Adv. Phys. Chem.*, p. 1‑16, 2016.

[34] 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, p. 1928‑1955, 2011.

[35] 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 », *J. Taibah Univ. Sci.*, vol. 9, p. 143‑154, 2015.

[36] P. P. Roy et K. Roy, « on some aspects of variable selection for partial least squares regression models », *QSAR Comb Sci*, vol. 27, p. 302‑313, 2008.