Chemical Science International Journal

30(8): 1-12, 2021; Article no.CSIJ.73752
ISSN: 2456-706X
(Past name: American Chemical Science Journal, Past ISSN: 2249-0205)

Elaboration of a Predictive Qsar Model of the Anti-Paludial Activity of a Series of Dihydrothiophenone Molecules at Theory Level B3LYP/ 6-31G (d, p)

Fandia Konate[1], Fatogoma Diarrrassouba[1], Georges Stéphane Dembele[1,2*], Mamadou Guy-Richard Koné[1,2], Bibata Konaté[1,2] and Nahossé Ziao[1,2]

[1]Laboratoire de Thermodynamique et Physico-Chimie du Milieu, UFR SFA, Universite Nangui Abrogoua, 02 BP 801 Abidjan 02, Côte d'Ivoire.
[2]Groupe Ivoirien de Recherches en Modélisation des Maladies (GIR2M), Côte d'Ivoire.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/CSJI/2021/v30i830244
Editor(s):
(1) Prof. Pradip K. Bhowmik, University of Nevada, USA.
(2) Prof. Marcelo Daniel Preite, Pontifical Catholic University of Chile, Chile.
Reviewers:
(1) Moslem Fattahi, Petroleum University of Technology, Iran.
(2) Krishnaraj Somayaji, Manipal Academy of Higher Education, India.
(3) Bhavesh J. Gangani, Bhakat Kavi Narsinh Mehta University, India.
Complete Peer review History: https://www.sdiarticle4.com/review-history/73752

Original Research Article

Received 03 August 2021
Accepted 12 October 2021
Published 16 October 2021

ABSTRACT

The purpose of this study is to develop a QSAR model predictive of the antimalarial activity of a series of Dihydrothiophenone molecules using quantum chemical methods. The molecules were optimized from the B3LYP/6-31G (d, p) level of theory. The extracted descriptors are the vibrational frequency of the carbonyl group ($V_{(C=O)}$), enthalpy of formation (Δ_f H°), the valence angle between the carbon-nitrogen-carbon atoms α($_{C\text{-}N\text{-}C}$) and the ionization potential (I); The application of the RLM method of the XLSTAT program allowed us to develop a regression model. The statistical indicators (R^2=93.50%, S=0.211, F=43.678) of the developed model attest to its robustness and reliability. Internal and external validation parameters (Q^2_{loo} et Q^2_{ext}) reveal that the established model performs well in predicting the antimalarial activity of the series of molecules studied. It can therefore be used to design new HD molecules belonging to its field of applicability at a 95% confidence level.

**Corresponding author: E-mail: 1997sageme@gmail.com;*

Keywords: Antimalarial activity; quantum chemistry; Dihydrothiophenone; RQSA; RLM.

1. INTRODUCTION

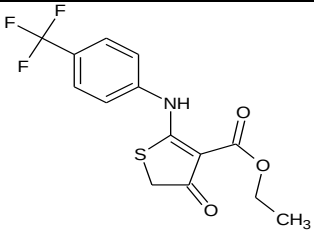
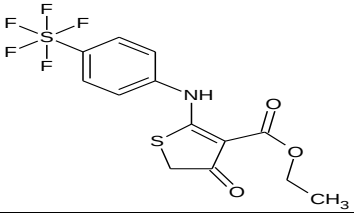
Malaria is an infectious disease caused by a parasitic microorganism of the genus Plasmodium that is transmitted to humans through the bites of Anopheles mosquitoes [1]. According to the World Health Organization (WHO), the number of people infected by this disease is 300 to 500 million, causing approximately 1.5 to 2 million deaths per year. The tropical and subtropical areas of sub-Saharan Africa alone account for 90% of those infected. The majority of deaths occur in children under 5 years of age [2]. The search for new antimalarial drugs remains a major challenge for the pharmaceutical industry in view of the growth of resistant strains. There are widespread parasitic infections of malaria in the world which are difficult to eradicate completely because of the dormant forms of the *plasmodium* genus [3]. However, artemisinin and its derivatives (artesunate, arteether and dihydroartemisinin) have given hope to fight resistant malaria [4, 5]. *Plasmodium* is highly adaptable to its environment and develops numerous resistances, making some of the currently available molecules obsolete in many endemic territories. Although most of these compounds have been known for a long time, their modes of action are not completely elucidated. To tackle the problem of drug resistance, various strategies have been developed to treat malaria [4, 5]. The pharmaceutical industry is moving towards new research methods that consist in predicting the activities of molecules even before they are synthesized. Among these, quantitative structure-activity relationships (QSAR) have become of great interest and are even recommended in the new regulations [6]. They allow the development of mathematical models linking biological activities to molecular structure on the one hand and to explain the origin of these activities and to predict them for molecules for which experimental data are not available on the other hand. It is within this framework that we set ourselves the objective of developing a QSAR model predictive of antimalarial activity from a series of DH molecules using quantum descriptors.

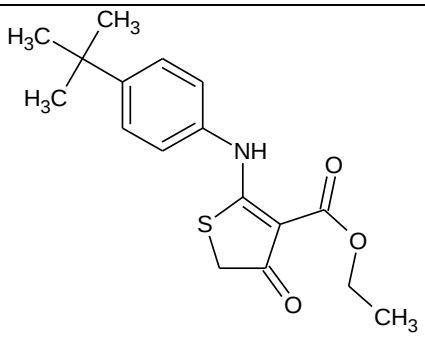
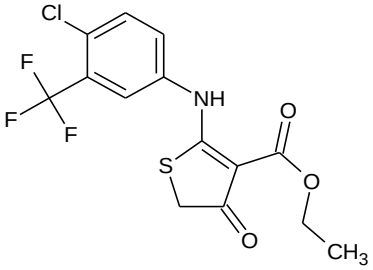
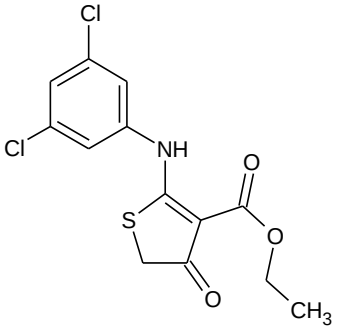
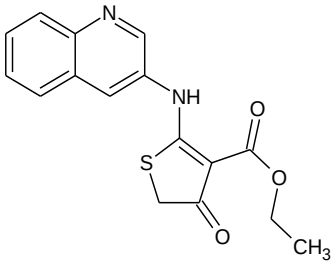
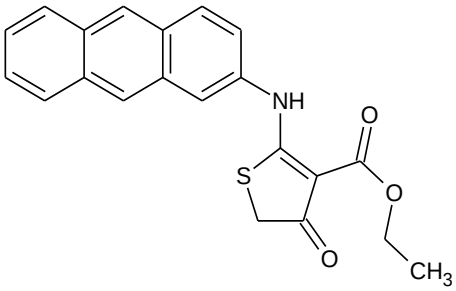
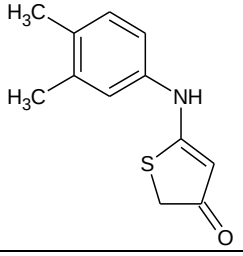
2. COMPUTATIONAL DETAILS

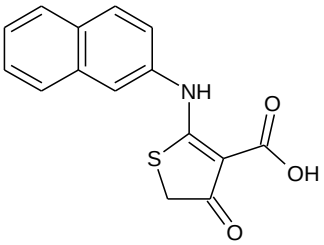
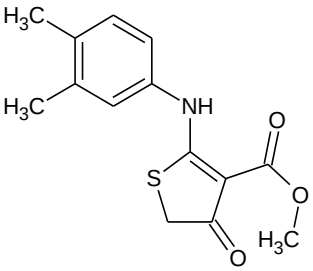
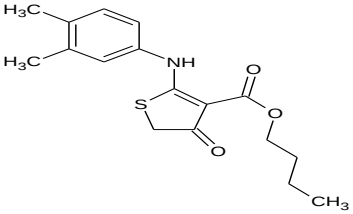
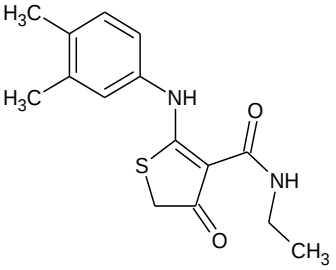
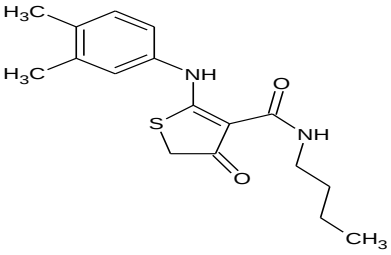
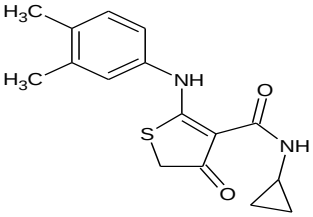
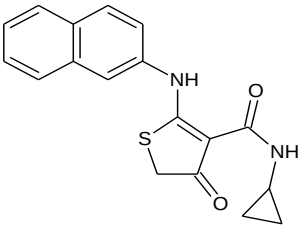
2.1 Traininig Set and Test Set

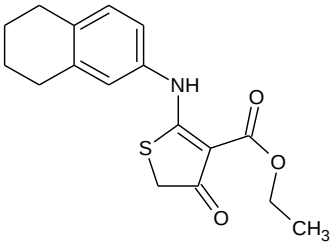
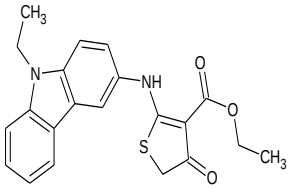
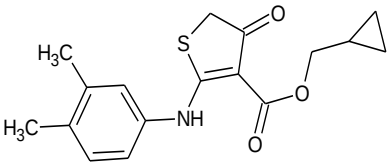
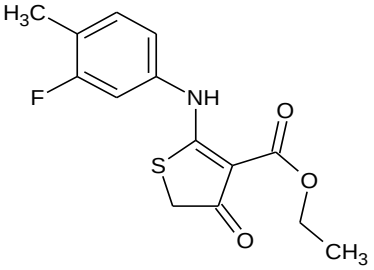
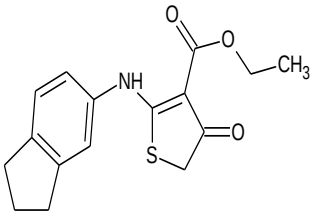
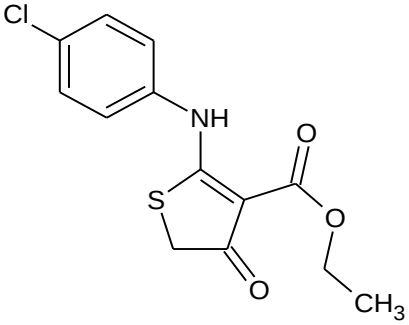
In the development of the QSAR model predicting the antimalarial activity of Dihydrothiophenone, a database of twenty-two DH compounds [7] was considered. These molecules were synthesized by Xu et *al*. [7]. They demonstrated the in vitro inhibitory capacity of these compounds against the enzyme as well as chloroquine-sensitive (*Pf3D7*) and -resistant (*PfDd2*) strains. The choice of these molecules is due to the availability of their experimental activities. They constituted our database. Table 1 presents these different molecules.

Table 1. Series of molecules studied

CODE	STRUCTURE	pIC_{50}
	Training Set	
DH2		**6.435**
DH3		**6.252**

DH4		**6.643**
DH5		**6.124**
DH7		**5.925**
DH11		**6.454**
DH14		**6.705**
DH16		**4.941**

DH17		**5.907**
DH18		**5.827**
DH20		**5.628**
DH22		**4.403**
DH24		**5.241**
DH25		**4.718**
DH30		**4.879**
DH31		**4.919**

CODE	STRUCTURE	pIC_{50}
Test Set		
DH10		**7.036**
DH15		**6.406**
DH21		**5.381**
DH9		**6.271**
DH13		**6.070**
DH1		**5.953**

2.2 Computational Theory Level and Software

GaussView 5.0 software [8] was used to represent the 3D structure and visualize the studied molecules. Then, the software Gaussian 09 [9] was used for geometric optimization and frequency calculation. The method used is the Density Functional Theory (DFT) with the level of theory B3LYP/6-31G (d, p). As for the 2D structures, they have been represented with Chemsketck [10]. EXCEL software [11] and XLSTAT [12] were used respectively for the graphical representation and the establishment of the models.

2.3 Statistical Analysis

To develop a QSAR model, a data analysis method is needed. This method makes it possible to quantify the relationship between the activity studied and the molecular structure (descriptors). There are several methods to build a model and analyze the statistical data of the latter. But the one used in our study is the multiple linear regression (MLR). In general, the MLR method is based on the hypothesis that the activity depends linearly on the different variables (descriptors) X_1, X_2……X_i , according to the relation: $Y=a_0+\sum_{i=1}^{n} aiXi$

With:
Y: the dependent variable (to be explained or predicted);
X_i : independent (explanatory) variables ;
a_0 : the constant of the model equation;
a_i : the coefficients of descriptors in the model equation.

Descriptor selection is a crucial step in QSAR modeling. In this study, descriptor selection was based on the following criterion:

The descriptors must be independent of each other two by two. To do so, the partial correlation coefficient a_{ij} between descriptors i and j must be lower than 0.70 ($a_{ij} < 0.70$) [13] . For a multilinear regression, the coefficients R et a_{ij} are expressed as follows:

The coefficient of determination R^2 [14] is given by the following relation:
$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$ avec $R = \sqrt{R^2}$

- **Standard deviation** [15]

It is an indicator of dispersion. It informs about the way the distribution of the data is spread around the mean. The closer its value is to 0, the better the fit and the higher the reliability of the prediction.

$$s = \sqrt{\frac{RSS}{n-p-1}}$$

$$s_{PRESSS} = \sqrt{\frac{PRESS}{n-p-1}}$$

TSS: Total Sum of Squares; ESS: Extended Sum of Squares; RSS: Residual Sum of Squares

$$ESS = \sum (Y_{i,cal} - \overline{Y}_{exp})^2$$
$$TSS = \sum (Y_{i,exp} - \overline{Y}_{exp})^2$$
$$RSS = \sum (Y_{i,exp} - Y_{i,cal})^2$$

Other statistical parameters were also determined:

- Paramètre de Roy et *al*. R_P^2, [16]

Allows to know if the model is due to chance correlations or not. If this parameter is greater than 0.5, the model is due to chance.

$$R_P^2 = R\sqrt{R^2 - R_r^2}$$

With R_r^2 , the average value of R_{ri}^2 models obtained with the randomized property.

- **Adjusted coefficient of determination** $R^2_{ajusté}$ [17]

It allows to measure the robustness of a model unlike R^2 . This coefficient is used in multiple regression because it takes into account the number of parameters (descriptors) of the model.

$$R^2_{ajusté} = 1 - \frac{(n - intercept)(1 - R^2)}{n - p}$$

- **Fisher-Snedecor coefficient F** [18]

It is used to test the overall significance of the linear regression. The Fisher-Snedecor coefficient is related to the coefficient of determination by the following relationship:

$$F = \frac{R^2}{1 - R^2} \frac{n - p - 1}{p}$$

- **Cross-validation criterion PRESS [19]**

The sum of squared prediction errors Prediction Sum of Squares (PRESS) is defined by the relation:

$$\mathrm{PRESS} = \sum (y_{i,exp} - y_{i,préd})^2$$

This criterion allows to select the models with a good predictive power. (We always look for the smallest PRESS).Cross-validation coefficient Q^2_{LOO} [20] It measures the accuracy of the prediction on the training set data.

$$Q^2_{LOO} = 1 - \frac{\sum (y_{i,exp} - y_{i,préd})^2}{\sum (y_{i,exp} - \bar{y}_{exp})^2} = 1 - \frac{PRESS}{TSS}$$

- **External validation coefficient Q^2_{ext} [19]**

It measures the accuracy of the prediction on the test set data.

$$Q^2_{ext} = 1 - \frac{n}{n_{ext}} \frac{PRESS(test)}{TSS}$$

- **Leverage h_{ii} [20]**

The leverage is a kind of distance to the barycenter of points in the space of explanatory variables. It identifies observations that are abnormally far from the others. For observation i

$$h_i = x_i (X^T X)^{-1} x^T \qquad (i=1,.., n)$$

Where xi is the row vector of the descriptors of compound i and X is the model matrix deduced from the values of the descriptors in the training set. The index T refers to the transposed matrix/vector. The critical value of the lever h* is, in general, fixed at $\frac{3\,(k+1)}{N}$ [21], where N is the number of compounds in the training set, and k is the number of descriptors in the model. If a compound has a residual and leverage that exceeds the critical value h*, that compound is considered outside the applicability domain of the developed model.

- **"External validation criteria" or "Tropsha criteria** [19, 22].
-

There are five criteria:

- Criteria 1: $R^2_{ext} > 0.70$
- Criteria 2: $Q^2_{ext} > 0.60$
- Criteria 3 : $\frac{|R^2_{ext} - R^2_0|}{R^2_{ext}} < 0.1$ et k=0.9475 avec 0,85<k<1.15
- Criteria 4: $\frac{|R^2_{ext} - R'^2_0|}{R^2_{ext}} < 0.1$ et k'= 1.0521 avec 0,85<k'<1.15
- Criteria 5 : $|R^2_{ext} - R^2_0| < 0.3$ with

$\mathbf{R^2_{ext}}$: Coefficient of determination for the molecules in the test series; $\mathbf{R^2_0}$: Coefficient of determination of the regression between predicted and experimental values for the test series ; $\mathbf{R'^2_0}$: Coefficient of determination of the regression between experimental and predicted values for the test series; **k**: slope of the correlation line (predicted values versus experimental values); **k'**: slope of the correlation line (experimental values versus predicted values).

- **Normality test**

The verification of the normality test conditions the quality of the confidence intervals around the parameters and the predictions. The normality of the residuals can be verified by analyzing some graphs or by using a normality test. The independence of the residuals can be verified by using the Durbin-Watson test [23]. For this purpose, the XLSTAT software is used.

This test is developed to detect autocorrelation between the residuals of a linear regression.

The XLSTAT software displays the following values as output:

U=c
p-value=d
alpha= 0,05
where c and d are reals.

Interpretation of the test:

H_0 : The residues are not aotocorrelated;
H_a : The residues describe an aotocorrelation process.

If the calculated p-value(d) is greater than the significance level alpha=0.05, then the null hypothesis H_0 cannot be rejected. The risk of rejecting the null hypothesis when it is true is 100*b in %. This test can be confirmed by representing the standardized residuals according to the values predicted by the established model. If we have a random distribution of the cloud of points obtained, we conclude that there is no autocorrelation of the residues.

2.4 Molecular Descriptor Values

Table 2. Calculated molecular descriptors

CODE	$\dot{V}_{(C=O)}$ (cm^{-1})	$\Delta_f H°$ (kcal/mol)	$\alpha_{(C-N-C)}$ (°C)	I(ev)
Training Set				
DH2	1776.47	-937.114	129.867	6.316
DH14	1771.9	-1038.931	129.89	5.394
DH16	1776.59	-485.32	129.985	5.614
DH17	1779.69	-790.924	129.873	5.93
DH18	1773.29	-780.256	129.524	5.885
DH20	1771.13	-929.722	128.764	5.857
DH22	1789.04	-591.683	129.995	5.741
DH24	1786.09	-689.737	130.291	5.735
DH30	1788.75	-651.42	129.825	5.778
DH5	1775.62	-934.599	129.49	6.362
DH25	1783.82	-598.466	129.846	5.744
DH4	1772.27	-926.335	129.735	5.923
DH3	1777.91	-967.288	129.685	6.461
DH11	1775.92	-695.372	130.236	6.108
DH31	1788.53	-644.786	129.926	5.758
DH7	1776.25	-729.063	129.335	6.292
Test Set				
DH10	1771.9	-915.744	129.68	5.861
DH15	1771.91	-920.022	129.843	5.634
DH21	1769.85	-886.285	128.018	5.971
DH9	1773.69	-829.523	129.442	6.049
DH13	1771.22	-860.393	129.404	5.868
DH1	1776.25	-730.541	129.335	6.292

3. RESULTS AND DISCUSSION

3.1 Study of the Developed Model

The four descriptors of the developed model allowed us to establish the regression equation which is written:

$$pIC_{50\ PfDHODH} = -0.06827*\dot{V}_{(C=O)} - 0.00255*\Delta_f H° + 0.93880*\alpha_{(C-N-C)} + 0.56459*I$$

$$N = 16\ ;\ R = 0.968\ ;\ R^2 = 0.935;\ s = 0{,}211\ ;\ F = 43.678;$$
$$p - value < 0.0001;\ TSS = 8.329;\ ESS = 7.795$$

Examination of the above parameters shows that the correlation coefficient is very high (R=0.968). This high value reflects that there is a strong linear correlation between the antimalarial activity and the calculated molecular descriptors. The coefficient of determination $R^2 = 0.935$ shows that 93.50% of the experimental variance in malaria activity is explained by the model descriptors. Moreover, the standard deviation (s=0.211) is low, indicating a good fit and high reliability of the prediction. The p-value is less than 0.0001 and therefore at α=0.05 (5% risk). It is therefore clear that the regression equation of the model is significant for the prediction of the antimalarial activity of the series of molecules studied. This significance is confirmed by the high Fischer value (F=43.678). It should be noted that a QSAR model can be obtained in a hazardous way. Therefore, one must always ensure its stability. To do so, two validation methods were used: internal and external validation.

3.2 Internal Model Validation

For internal validation, Leave-One-Out (LOO) cross-validation and the Y-randomization test. were used.

- **Leave-One-Out cross-validation**

Table 3. Statistical parameters of the cross-validation of the LOO model of the model

N	Press	Q^2_{LOO}	S_{press}
16	0.819	0.876	0.088

Table 3 shows that the value of $Q^2_{LOO} = 0.876$ is satisfactory because $Q^2_{LOO} > 0.50$. Moreover, about 88% of the molecules in the training set have their activities predicted by the model. The model therefore has a high predictive capacity. This result shows that the model is not very sensitive to this operation of putting aside a molecule and putting it back in the training set (Leave-One-Out). This justifies the stability of this model. Moreover, to ensure that the model is not due to chance, the randomization test was performed. Ten iterations were performed.

- **Y-randomization test**

The average values of the randomization parameters are recorded in Table 4.

The TODESHNI parameter $^{c}R_p^2$=0.669>0,5 shows that our model really exists and is not due to chance correlations [24]. Based on the results of the internal validation, it can be concluded that the established model is robust and not due to chance.

3.3 External Validation of the Model

Since internal validation is necessary but not sufficient, it is imperative to proceed with external validation.

Verification of the Tropsha criteria

Criteria 1 : $R^2_{ext} = 0.799 > 0.70$
Criteria 2: Q^2_{ext}=0.738>0.60
Criteria 3 : $\frac{|R^2_{ext} - R^2_0|}{R^2_{ext}} = 0.0090 < 0,1$ et k=0,912 avec 0.85<k<1.15
Criteria 4 : $\frac{|R^2_{ext} - R'^2_0|}{R^2_{ext}} = 0.00078 < 0,1$ et k'= 0.9533 avec 0.85<k'<1.15
Criteria 5: $|R^2_{ext} - R^2_0|$=0.008 <0.3

After the calculation of the different parameters of the Tropsha criteria, we notice that the five (5) criteria are verified. This result shows that the model obtained has an acceptable predictive power.

3.4 Correlation between the Values Predicted by the Model and the Experimental Values

In Fig. 1, all points tend to be close to the regression line. This shows a strong linear correlation between the predicted values of the antimalarial activity by the model and the experimental values. This graph confirms that the model is validated and is very efficient in the prediction of the studied activity..

Table 4. First 10 iterations of Y-randomization

ITERATION	1	2	3	4	5	6	7	8	9	10
R^2ri	0.670	0.601	0.662	0.514	0.261	0.282	0.169	0.190	0.266	0.477

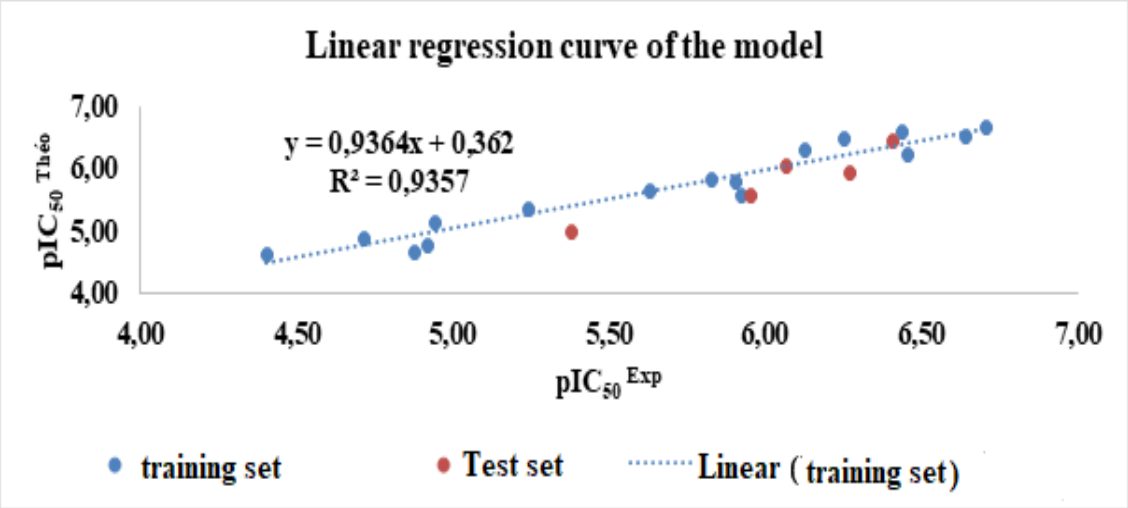


Fig. 1. Correlation between experimental and theoretical values

3.5 Normality Tests of the Model

- **Shapiro-Wilk test ($E_{théo}$) [25]**

Table 5. Parameter values of the Shapiro-Wilk test

W	p-value	alpha
0.9218	0.0828	0.05

The data in Table 5 shows that the calculated p-value is greater than alpha= 0.05 (5% threshold). Thus, the theoretical values of the first reduction potential obtained from the model follow a normal distribution. This normal distribution is confirmed by the distribution of the scatterplot along the first bisector in Fig. 2.

- **Durbin-Watson test [23]**

Table 6. Parameter values for the Durbin-Watson test

U	p-value	alpha
1.4524	0.1329	0.05

The values in Table 6 show that the calculated p-value is greater than alpha=0.05 (5% threshold). It is therefore clear that the residuals are not self-correlated (zero correlation). Under these conditions, these residuals do not contain information capable of influencing the model's prediction of the first reduction potential. This interpretation is confirmed by the random distribution of the scatterplot in Fig. 3.

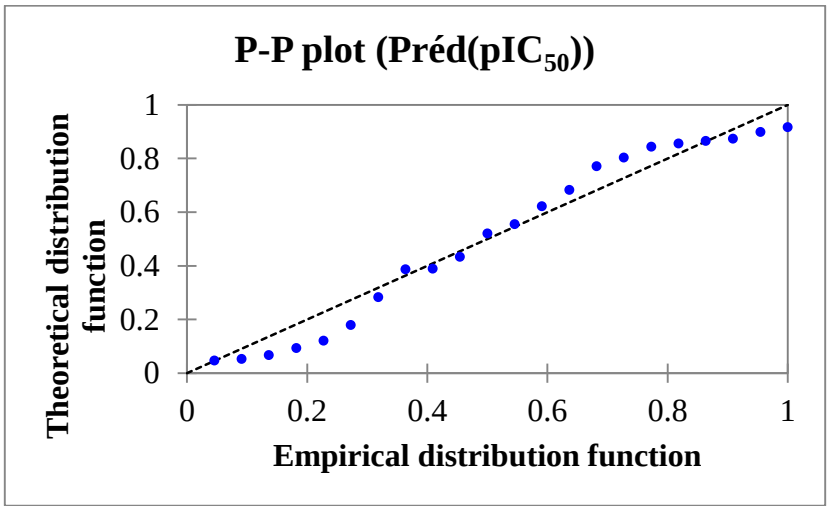


Fig. 2. P-P plot of the model

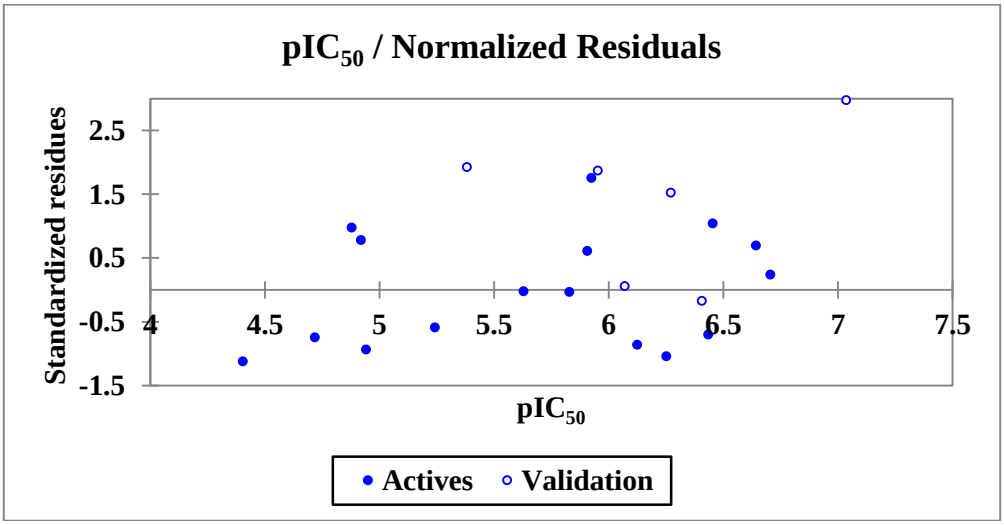


Fig. 3. Graph of the normalized residuals of the model

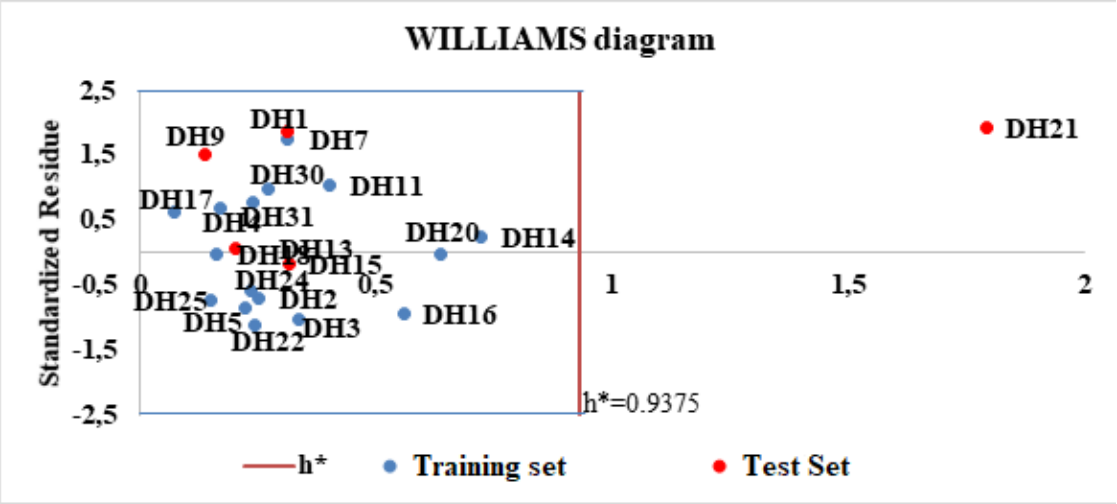


Fig. 4. Williams diagram of the model

3.6 Area of Applicability of the Model

The Domain of Applicability (DA) was determined by analysis of the Williams diagram in Fig. 4.

Examination of the Williams plot shows that for the training and test sets, all observations have their standardized residuals within ±3 standard deviation units (±3σ) [26]**.** This justifies the absence of outliers. The choice "3 standard deviation units" was made because our data follow a normal distribution. Through the diagram, we can notice that for the twenty-two (22) compounds, all the standardized residues are between -3 δ and + 3δ [26]**.** However, compound DH21 in the validation set that has a leverage value hii=1.070 greater than the threshold value h*=0.9375 displays a standardized residue of less than 3 δ. As a result, our established QSAR model can be used to predict the antimalarial activity of other molecules listed in its DA.

4. CONCLUSION

In this study, the aim was to develop predictive QSAR model relating antimalarial activity from a series of dihydrothiophenone molecules analogous to quantum descriptors from conceptual density functional theory. The study revealed that the vibrational frequency ($V_{C=O}$), the nitrogen-carbon-nitrogen valence angle $\alpha_{(C\text{-}N\text{-}C)}$, enthalpy of formation ($\Delta_f H°$) and the ionization potential (I) are the priority descriptors in the prediction of antimalarial activity. Regarding the parameters of the internal and external validations, they revealed that the model is validated and very efficient in predicting the first reduction potential. The standard deviations are much lower than 0.50 indicating a good fit and a high reliability of the prediction. Its applicability range was defined to detect outliers and influential compounds. The study showed that the compound DH21 is influential in predicting antimalarial activity. This model can be used to predict the activity of new molecules on the one hand, and on the other hand, to identify descriptors that improve antimalarial activity, thus giving orientations to design new more active molecules.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Pully E. Ouvrage du Collège des Universitaires de Maladies Infectieuses et Tropicales., 2eme éd. 2012;99.
2. Gentillini M. Médecine Tropicale, 4e éd., Paris: Flammarion; 1986.
3. OMS, «World malaria report ,» Geneva; 2018.
4. Hein TT. «An overview of the clinical use of artemisinin and its derivatives in the treatment of falciparum malaria in Vietnam, Tran. R. Roc. Trop.Med. Hyg. 1994;88:7-8.
5. Harinasuta T, Karbwang J. et al. Qinghaosu, «a promising antimalarial ,» Journal of American Medical Association. 1994;3.

6. Margossian N, Le règlement REACH - La règlementation européenne sur les produits chimiques, L. Nouvelle, Éd., Paris: Dunod; 2008.
7. Xu M. Novel selective and potent inhibitors of malaria parasite dihydroorotate dehydrogenase: Discovery and optimization of dihydrothiophénone derivatives,» J. Med. Chem. 2013;56, n° %120:7911 7924.
8. Dennington R, Keith T, et J. Millam, «GaussView; 2009.
9. G. W. T. H. B. S. G. E. S. M. A. R. J. R. C. G. S. V. B. B. M. G. A. P. H. N. M. C. X. L. H. P. H. A. F. I. J. B. G. Z. J. L. S. M. H. M. J. Frisch, «Gaussian 09 Revision D.01,» Gaussian, Inc., Wallingford CT, 2016.
10. Acdlabs, Advanced Chemistry Development/ Chemskecht; 1994-2010.
11. Microsoft Excel, «(15.0.4420.1017) MSO (15.0.4420.1017) 64 Bits,» Microsoft Office Professionnel; 2016.
12. XLSTAT, « XLSTAT and Addinsoft are Registered Trademarks of Addinsoft,» Copyright Addinsoft , 2014.
13. Vessereau A. Méthodes statistiques en biologie et en agronomie, Paris: Lavoisier (Tec & Doc). 1988; 538.
14. Chatterje S, et al. Hadi, Regression Analysis by Example, 4 éd., I. H. John Wiley & Sonc, Éd., 2006.
15. Siegel AF. Practical Business Statistics, 3 éd., IRWIN; 1997.
16. Roy P, Paul S, Mitra I, et al. K. Roy, «On Two Novel Parameters for Validation of Predictive QSAR Models,» Molecules. 2009;14, n° %15:1660-1701.
17. Besse P, Pratique de la modélisation statistique, Publications du laboratoire de statistique et Probabilité; 2003.
18. Cook RD, et al. S. Weisberg, «An introduction to Regression Graphics,» Wiley Series in Probability and Statistics; 1994.
19. Golbraikh A, et al. Tropsha, «Beware of qsar, J. Mol. Graph. Model. 2002;20:269.
20. Gramatica P. Principles of QSAR models validation: internal and external.,» QSAR Comb. Sci. 2007; 26,n° %15:694 – 701.
21. Netzeva T, Worth A, Aldenberg T, Benigni R, Cronin M, Gramatica P, et al. J. Van De Sandt, «Current status of methods for defining the applicability domain of (quantitative) structure–activity relationships,» Alternatives to Laboratory Animals. 2005;33, n° %12:155–173.
22. Tropsha A, Gramatica P, et al. V. Gombar, «The importance of being earnest, validation is the absolute essential for successful application and interpretation of QSPR models,» QSAR Comb. Sci. 2003;22:69.
23. Durbin J, Watson GS. Testing for serial correlation in Least Squares Regression,II,» Biometrika. 1951;38, n° %11 2:159 179.
24. Roy K. A Primer on QSAR/QSPR Modeling, Chapter 2 Statistical Methods in QSAR/QSPR,» Springer Briefs in Molecular Science. 2015;37-59.
25. Shapiro SS, et Wilk MB. «An analysis of variance test for normality (Complete Samples),» Biometrika. 1965;52, n° %13-4:591-611.
26. Touhami I, Mokrani K, Messadi D. «Modèles QSRR hybrides algorithme génétique-régression linéaire multiple des indices de rétention de pyrazines en chromatographie gazeuse,» Lebanese Science Journal. 2012;13, n° %11.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle4.com/review-history/73752