



Quantative STRUCTURE Activity Relationship (QSAR) Study of a Series of Molecules Derived from 2,3-dihydro-1H- perimidine having Activity against Protein Tyrosine Phosphatase 1B

Tuo Nanou Tiéba^{1*}, Dembele Georges Stephane¹, Soro Doh¹, Konate Bibata¹, Kodjo Charles Guillaume^{1,2} and ZIAO Nahossé¹

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

²Laboratoire de Chimie BioOrganique et de Substances Naturelles, Université Nangui Abrogoua, UFR-SFA, 02 B.P. 801 Abidjan 02, 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/IRJPAC/2021/v22i830426

Editor(s):

(1) Dr. Farzaneh Mohamadpour, University of Sistan and Baluchestan, Iran.

Reviewers:

(1) F. Solano, University of Murcia, Spain.

(2) Teodorico C. Ramalho, Universidade Federal de Lavras, Brazil.

Complete Peer review History: <http://www.sdiarticle4.com/review-history/76399>

Review Article

Received 14 August 2021
Accepted 28 October 2021
Published 03 November 2021

ABSTRACT

In order to study the Quantitative Structure Activity Relationship (QSAR) against protein tyrosine phosphatase 1B and descriptors, we used a series of fourteen (14) molecules derived from perimidine. The compounds were optimized at the computational level B3LYP / 6-31 G (d, p), to obtain the descriptors of the model. This study was performed using the Linear Multiple Regression (MLR) method. This tool allowed us to obtain a quantitative model from the descriptors that are, the overall softness (S), the energy of the lowest vacant (ELUMO), the bond length I (N-C1). This model has good statistical performance ($R^2 = 0.958$; RMCE = 0.110; F = 43.870). In addition, the external validation test of Tropsha and the domain of applicability from the levers were verified.

*Corresponding author: E-mail: nanoutuo07@gmail.com;

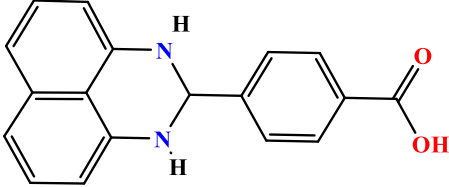
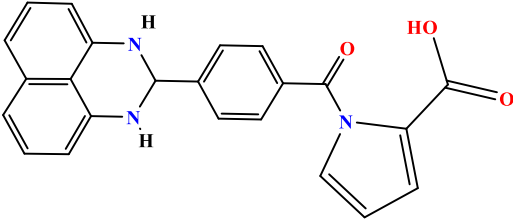
Keywords: QSAR, 2,3-dihydro-1H-perimidine; protein tyrosine phosphatase 1B; MLR and DFT.

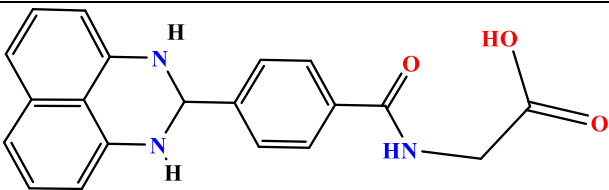
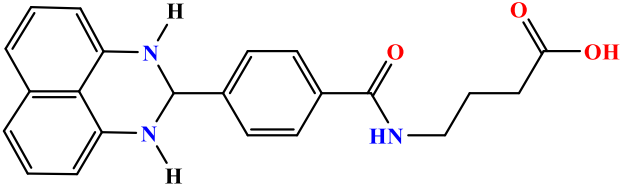
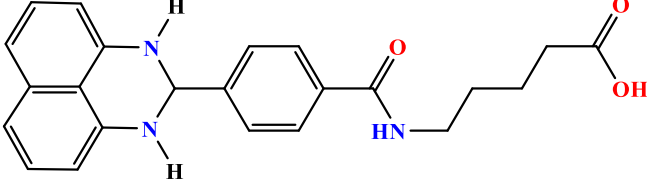
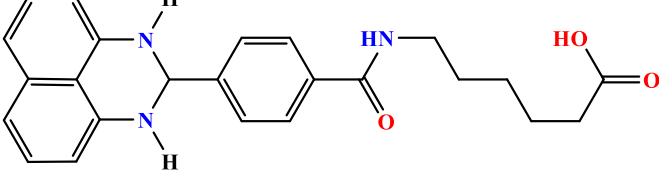
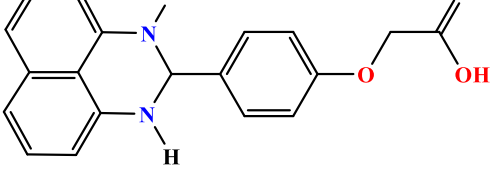
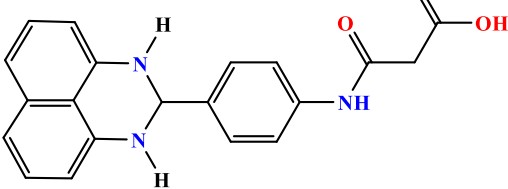
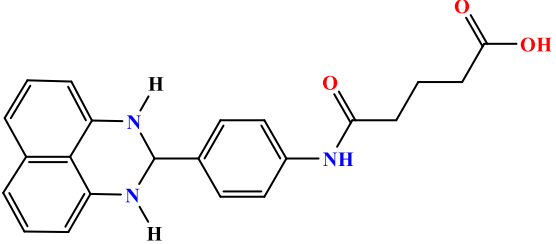
1. INTRODUCTION

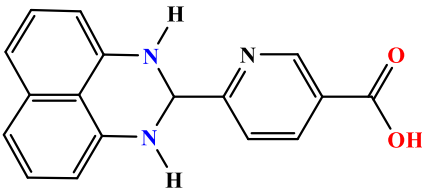
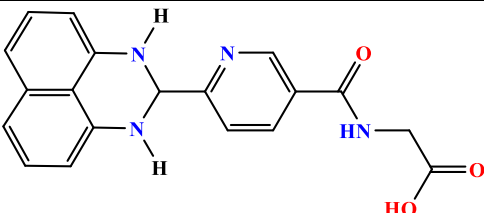
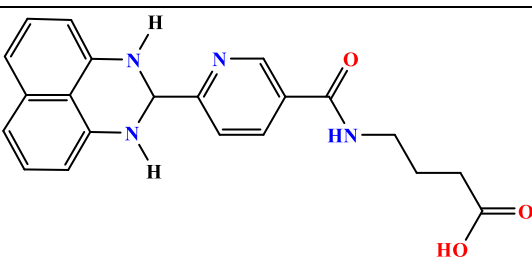
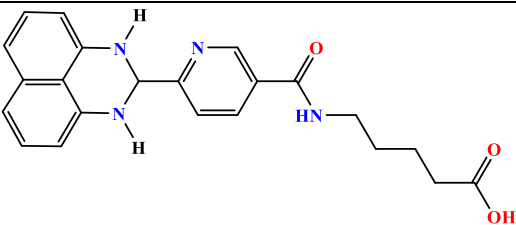
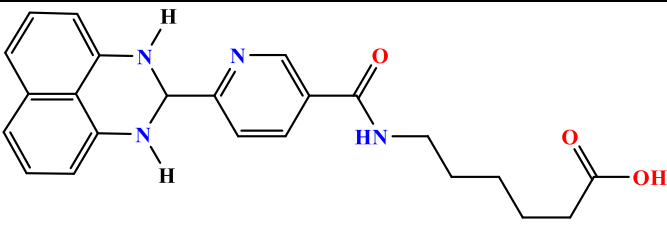
Many diseases such as cancer, diabetes, obesity, infections, autoimmune and neuropsychiatric disorders are linked to the deregulation of protein tyrosine phosphatases (PTP) [1,2]. For the sake of well-being, the scientific community finds interest in finding an inhibitor of this protein. In 2013 Wen-Long Wang et al. [3], show that 2,3-dihydro-1H-perimidine have protein tyrosine phosphatase inhibitory activity. They demonstrate through high throughput screening that 4-(2,3-dihydro-1H-peridin-2-yl) benzoic acid inhibits this protein with an IC₅₀ value of 8.34 ± 1.07 M. 2,3-Dihydro-1H-Perimidines are defined as the product resulting from the condensation of 1,8-diaminonaphthalene (1,8 - DAN) with carbonyl derivatives (Aldehyde). The 1H-perimidines which are heterocyclic compounds constitute a class of organic compounds which has a great structural variety. This allows their use in various scientific fields. They are characterized by either a bond deficit or an excess π bond due to the electron density movements of the nitrogen atoms present in the naphthalene ring [4]. These compounds have a very important interest because of their polyvalent [5], optoelectronic, chemotherapeutic, thermochromic, and photochromic properties. They are still used with

applications [6] in plasma display panels, organic solar cells and optical recording media [4]. 1-H perimidines are used in medical chemistry as antifungal, antimicrobial, antiulcer [7] and antitumor agents. Previous studies have shown that 2,3-dihydro-1H-perimidine have important biological activities and are good inhibitors. With the aim of obtaining new molecules with more interesting biological activities, the study of the quantitative structure-activity relationship (QSAR) is the process by which a molecular structure is correlated with a well-defined effect such as biological activity or chemical reactivity. The development of this type of relationship is expanding rapidly and is becoming essential in pharmaceutical chemistry and drug design [8]. This is why the QSAR methodology, which has emerged over the past few decades [9,10], and which quantitatively links biological activity to molecular structure, has entered into this research process [11,12,13,14,15]. The objective of this work is to develop a QSAR model, by multiple linear regression, capable of predicting the inhibitory activity of the protein tyrosine phosphatases of a series of perimidine Table 1. These compounds were synthesized by the condensation of the 2,3-dihydro-1H-perimidine which are derived from 1,8-diaminonaphthalene and benzoic acid derivatives.

Table 1. Molecular structure, code and inhibitory concentration (IC₅₀) of the fourteen (14) derivatives of perimidine

Codes	Structures	IC ₅₀ (μ M)
A1		8.34
A2		20.23

Codes	Structures	IC ₅₀ (μ M)
A3		27.75
A4		22.21
A5		5.53
A6		7.82
A7		5.88
A8		6.45
A9		6.91

Codes	Structures	IC ₅₀ (μ M)
A10		10.82
A11		0.66
A12		15.24
A13		3.56
A14		0.59

2. MATERIALS AND METHODS

2.1 Computational Theory Level

Prediction of activity against protein tyrosine phosphatase 1B was performed using Gaussian 09 software [16]. DFT methods are generally known to generate a variety of molecular properties [17-19] in QSAR studies. The latter increase the predictability of QSAR models while reducing calculation time and cost implications in the design of new drugs [20,21]. The theory level

B3LYP / 6-31 G (d, p) was used to determine molecular descriptors. The fourteen (14) molecules used in this study have Inhibitory Concentrations (ICs) varying from 0.59 to 27.75 μ M. Biological data are generally expressed as the opposite of the decimal logarithm of activity ($-\log_{10}(C)$) to obtain better mathematical values when structures are biologically active [22,23]. The antimalarial activity will be expressed by the potential of the inhibitory concentration pIC defined by equation (1):

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

IC50, the inhibitory concentration in μM .

The modeling was developed using namely the linear multiple regression (LMR) method implemented in Excel [24] and XLSTAT [25] spreadsheets.

2.2 Molecular Descriptors Used

In order to build our QSAR model, several molecular descriptors were calculated. In particular, the energy of the vacant lowest orbital, softness and bond length (N-C1). The energy of the lowest vacant provides information on the electron acceptor character of the molecule. ELUMO indicates the ability of a molecule to accept electrons, the lower the energy of BV, the higher the probability that the molecule will accept electrons. Overall softness is the ability of an atom or molecule to retain an acquired charge. The lower the overall softness of a system, the more resistant it is to electron transfer and therefore the more stable it is expression is given by the relation:

$$S = \frac{1}{\eta} \quad (2)$$

Avec :

$$\eta = \frac{I-AE}{2} = \frac{1}{2}(E_{\text{LUMO}} - E_{\text{HOMO}}) \quad (3)$$

The geometric descriptor used is the bond length l (N-C1) in Armstrong (\AA) (Fig. 1). This descriptor is illustrated by the figure below around the nucleus of perimidine.

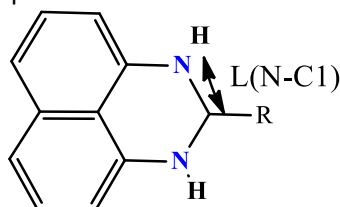


Fig. 1. Geometric descriptor of the perimidine derivatives used: bond length (l (N-C1)) in Armstrong (\AA)

2.3 Estimation of the Predictive Capacity of a QSAR Model

quality of a model is determined based on various statistical analysis criteria including the coefficient of determination R^2 , the standard deviation (S) or the Root of the Mean Squares of

Errors (RMCE), the correlation coefficients cross-validation Q_{CV}^2 and Fischer $F.R^2$, S and F relate to the fit of the calculated and experimental values. They describe the predictive capacity within the limits of the model, and make it possible to estimate the precision of the values calculated on the test set [26,27]. As for the cross-validation coefficient Q_{CV}^2 , it provides information on the predictive power of the model. This forecasting power is said to be "internal" because it is calculated from the structures used to build this model. The coefficient of determination R^2 gives an evaluation of the dispersion of the theoretical values around the experimental values. The quality of the modeling is better when the points are close to the line of fit [28]. The fit of the points to this line can be assessed by the coefficient of determination.

$$R^2 = 1 - \frac{\sum(y_{i,exp} - \hat{y}_{i,theo})^2}{\sum(y_{i,exp} - \bar{y}_{i,exp})^2} \quad (4)$$

Où :

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

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

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

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

Another statistical indicator used is the Root of the Mean Squares of RMCE Errors. It allows the reliability and accuracy of a model to be assessed:

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

The Fisher F test is also used to measure the level of statistical significance of the model, that is to say the quality of the choice of descriptors constituting 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 makes it possible to evaluate the accuracy of the prediction on the training set. It is calculated using the following relation:

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

2.4 Acceptance Criteria for a Model

The performance of a mathematical model, for Eriksson et al. [29], is characterized by a value of $Q_{cv}^2 > 0.5$ for a satisfactory model, while for the excellent model $Q_{cv}^2 > 0.9$. According to these authors, given a test set, a model will perform well if the acceptance criterion $R^2 - Q_{cv}^2 < 0.3$ is met.

According to Tropsha et al. [30,31,32], for the external validation set, the predictive power of a model can be obtained from five criteria. These criteria are as follows:

- 1) $R_{Test}^2 > 0.7$, 2) $Q_{cvTest}^2 > 0.6$,
- 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 $0.85 \leq k' \leq 1.15$

2.5 Area of Applicability (DA)

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 [33]. It corresponds to the region of chemical space including compounds from the learning game and similar compounds, which are closely related in the same space [34]. It appears necessary, even obligatory, to determine the DA of any QSAR model. This is what the Organization for Economic Cooperation and Development (OECD) recommends in developing a QSAR model [35]. There are several methods for determining the domain of applicability of a model [34]. Among these, the approach used in this work is that of leverage. This method is based on the variation of the standardized residuals of the dependent variable with the

distance between the values of the descriptors and their mean, called leverage [36]. The h_{ii} are the diagonal elements of an **H** matrix called the hat matrix. H is the projection matrix of the experimental values of the explained variable $Y_{expé}$ in the space of the values of the predicted explained variable $Y_{préd}$ such that:

$$Y_{préd} = HY_{expé} \quad (8)$$

H is defined by the expression (21) :

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

The field of applicability is delimited by a threshold value of the leverage denoted h^* . In general, it is fixed 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 [37,38]. For standardized residues, the two limit values generally used are $\pm 3\sigma$, σ being the standard deviation of the experimental values of the quantity to be explained [39]: this is "the rule of three sigmas" [40].

3. RESULTS AND DISCUSSION

This QSAR modeling work was conducted using a series of fourteen (14) perimidine derivatives. The molecules were divided into two groups, nine (9) were used for the learning game and five (5) for the validation game. The values of the descriptors as well as those of the experimental biological activities of the molecules are listed in Table 2.

The table 3 presents the values of the partial correlation coefficients a_{ij} of these descriptors.

The calculation of the partial correlation coefficient between each of the pairs of all the descriptors is less than 0.95 ($a_{ij} < 0.95$), which means that these different descriptors are independent of each other [41,8].

Table 2. Quantum descriptors and experimental inhibitors of protein tyrosine phosphatase 1B (PTP1B) activities of the Training Set test and validation set

Codes	E _{LUMO} (ev)	s	LN-C1	IC ₅₀	pIC ₅₀
Training Set					
A1	-1.597	0.605	1.463	8.34	1.685
A2	-1.958	0.683	1.462	20.23	1.300
A4	1.179	0.335	1.460	22.21	1.259
A5	-1.056	0.537	1.463	5.53	1.863
A6	-1.213	0.526	1.457	7.82	1.713
A8	-0.736	0.502	1.464	6.45	1.796

Codes	E _{LUMO} (ev)	s	LN-C1	IC ₅₀	pIC ₅₀
A10	-1.891	0.659	1.464	10.82	1.572
A11	-1.560	0.411	1.463	0.66	2.786
A12	-1.197	0.565	1.461	15.24	1.423
Validation Set					
A3	-1.276	0.561	1.463	27.75	1.163
A7	-0.547	0.475	1.463	5.88	1.836
A9	-0.670	0.493	1.464	6.91	1.766
A13	-1.451	0.580	1.465	3.56	2.054
A14	-1.320	0.577	1.486	0.59	2.835

Table 3. Values of the bivariate linear correlation coefficients of the descriptors

Variables	E _{LUMO} (ev)	s(eV ⁻¹)	N-C1(Å)
E _{LUMO} (ev)	1		
s(eV ⁻¹)	-0.803	1	
N-C1(Å)	-0.242	0.169	1

3.1 Quantitative Structure Activity Relationship (QSAR) Model and Contribution of Descriptors

It should be noted that the negative or positive sign of the coefficient of a descriptor of the model reflects the effect of proportionality between the evolution of the biological activity of interest and this parameter of 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 translates the opposite effect. The model equation obtained using the theoretical descriptors related to the optimized molecules and the statistical indicators are presented below:

$$pIC_{50} = -61,812 - 0,679 * E_{LUMO}(ev) - 6,163 * s(eV^{-1}) + 45,200 * LN-C1(\text{Å})$$

The negative signs of the different coefficients of the model parameters indicate that the inhibitors of protein tyrosine phosphatase 1B (PTP1B) activity (pIC₅₀) evolves inversely with (E_{LUMO} and s).

The correlation coefficient indicates that 95.6% of the molecular descriptors that define this model are taken into account at a standard deviation of 0.110. The significance of the model is reflected by the Fischer coefficient F = 43.870. Fisher's value shows that the error made is less than what the model explains [42][42]. For this model,

the cross-validation correlation coefficient Q_{cv}^2 is equal to $Q_{cv}^2 = 0.958$. This value, greater than 0.9, reflects a model said to be excellent according to Erikson et al. [43]. This model is acceptable because it agrees with the acceptance criteria of these authors $R^2 - Q_{cv}^2 = 0.958 - 0.958 = 0.000 < 0.3$. All these statistical indicators clearly show that the model developed explains the antimalarial activity in a statistically significant and satisfactory manner.

The regression graph of the RML model showing theoretical antimalarial activity as a function of experimental activity is shown in Fig. 2.

Analysis of the regression curve of the RML model shows that all points are around the regression line. This result indicates that there is a small difference (RMCE = 0.110) between the values of pIC₅₀^{exp} and pIC₅₀th, so a good similarity at the level of these values, apart from some difference. This similarity is illustrated in Fig. 3.

The checks of the Tropsha criteria for the validation sets are presented in Table 5.

Analysis of the table shows that out of the five (5) Tropsha criteria there is one (1) criterion that is not met. However Ouattara et al [32] have shown that if 3/4 of the criteria are met then the model is acceptable for the prediction.

Table 4. Rapport d'analyse statistique du modèle RML

R ²	Q _{cv} ²	RMCE	Fisher(F)	α
0.958	0.958	0.110	43.870	> 95 %

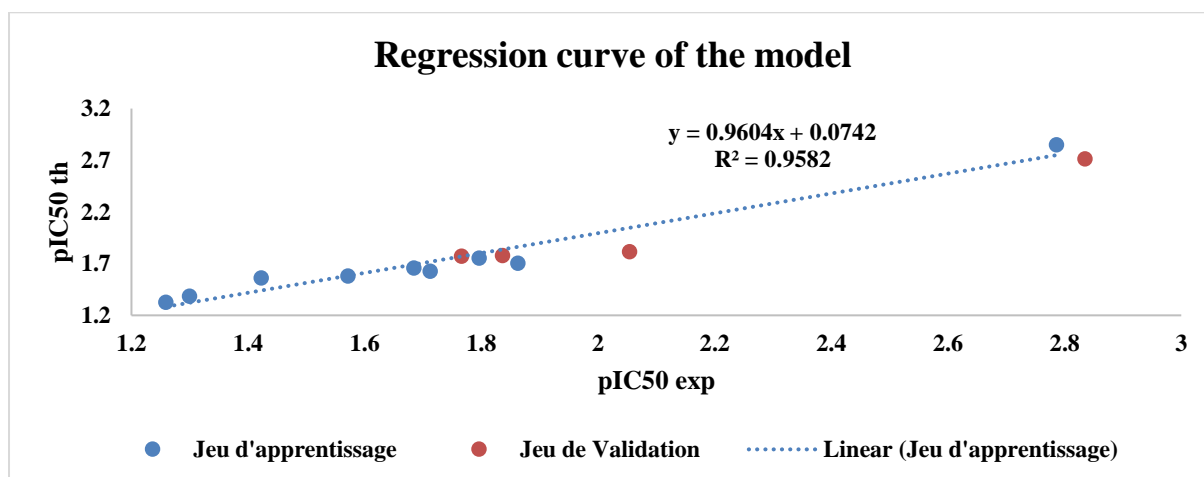


Fig. 2. The regression line of the RML model

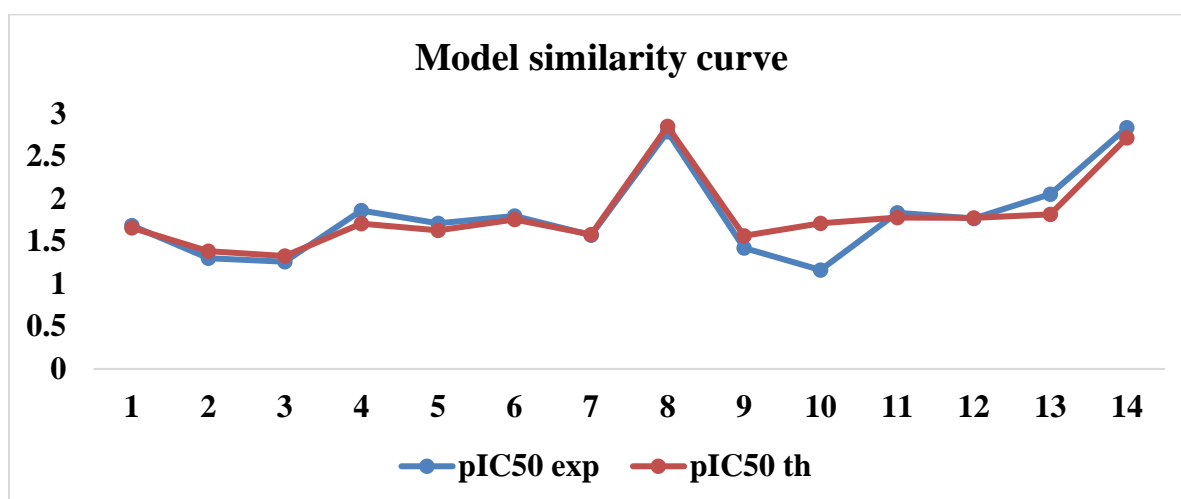


Fig. 3. Similarity curve of the experimental and predicted values of the RML model

Table 5. Verifications of the tropsha criteria of the external validation set of the RML model

R^2_{ext}	Q^2_{cv}	$R^2 - R^2_0$	$\frac{R^2 - R^2_0}{R^2}$	k	$\frac{R^2 - R^2_0'}{R^2}$	k'
0.776	0.742	0.296	0.381	0.986	0.044	0.995

The contributions of the three (3) physicochemical descriptors in the prediction of activity against the protein tyrosine phosphatase 1B of perimidine derivatives were illustrated by the normalized coefficients and shown in Fig. 4.

According to the contribution of these descriptors, the softness (s) displays the highest normalized coefficient (-1.506) followed by the energy of the lowest vacant (E_{LUMO}) with -1.423 and the bond length L (N-C1) has the lowest coefficient (0.197) compared to the other

descriptors. Note that softness (s) is the most influential physicochemical descriptor. Thus, to improve the inhibitory concentration IC_{50} in the synthesis of new derivatives of perimidine, it is necessary to play to the maximum on the softness (s).

The graph of the standardized residuals according to the levers h_{ii} in Fig. 5, makes it possible to visualize the domain of applicability of the model.

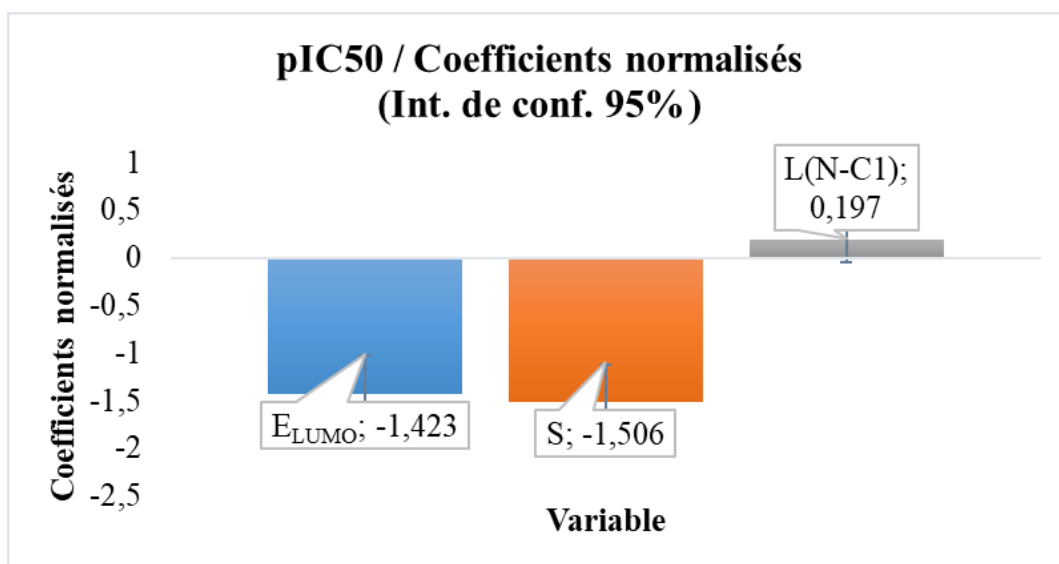


Fig. 4. Contribution of descriptors in the model

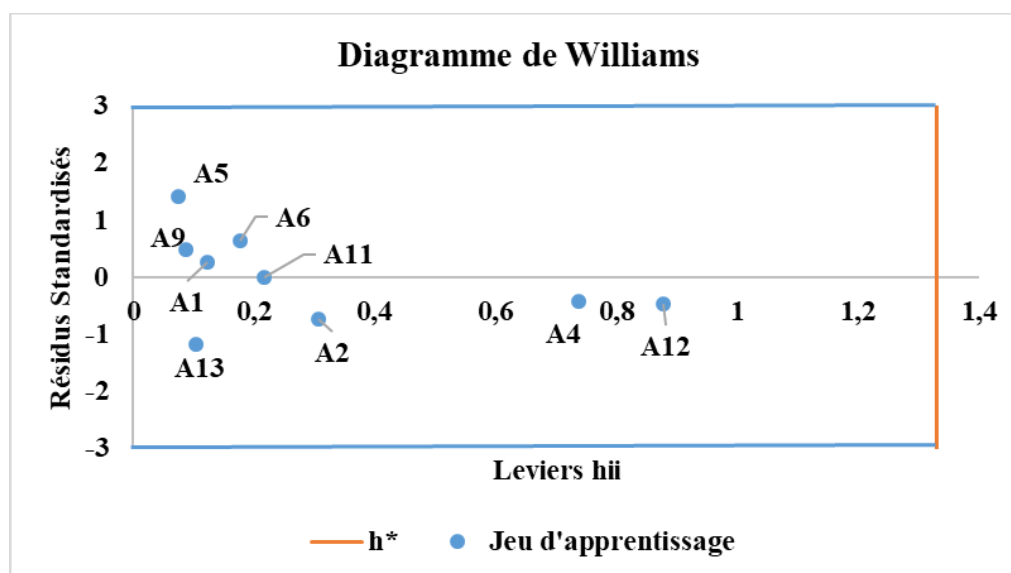


Fig. 5. Scope of the model

For the 9 molecules of the learning set and the 3 descriptors of the model, the threshold value of the levers h^* is 1.33. The extreme values of the standardized residuals are ± 3 according to the "three sigma rule" [40]. These different values delimit the range of applicability [44] of the model as shown in the graph in Fig. 5. All compounds are within the range, resulting in the reliability of the model.

4. CONCLUSION

In this work we established a relationship between the potential of the inhibitory concentration pIC₅₀ and the physicochemical

descriptors of pyrimidine derivatives. Overall softness (S), lowest vacant energy (E_{LUMO}), I-bond length (N-C1) are the descriptors that best reflect activity against protein tyrosine phosphatase 1B of pyrimidine derivatives. The statistical indicators ($R^2 = 0.958$; RMCE = 0.110; $F = 43.870$) of the model show that they are acceptable, robust and have good predictive power. The external validation test showed a good correlation between the theoretical and experimental pIC₅₀ inhibitory potential observations. This shows that these models are therefore acceptable for the prediction of the inhibitory IC₅₀ concentration of pyrimidine

derivatives. Also, we note in terms of the contribution of descriptors that a decrease in softness (s) could promote a significant improvement in the inhibitory concentration IC₅₀ of perimidine analogues against the protein tyrosine phosphatase 1B. In addition, the domain of applicability of this model determined from the levers shows that a prediction of pIC₅₀ of new perimidine derivatives is acceptable when its leverage value is less than 1.33.

DISCLAIMER

The products used for this research are commonly and predominantly use products in our area of research and country. There is absolutely no conflict of interest between the authors and producers of the products because we do not intend to use these products as an avenue for any litigation but for the advancement of knowledge. Also, the research was not funded by the producing company rather it was funded by personal efforts of the authors.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

- Zhang Z. Protein tyrosine phosphatases: Prospects for therapeutics, *Curr. Opin. Chem. Biol.* 2001;5:416–423.
- Julien S, Dube N, Hardy S, et M. Tremblay, Inside the human cancer tyrosine phosphatome, *Nat. Rev. Cancer.* 2011;11:35–49.
- Wang WL, Yang DL, Gao LX, Tang CL, Ma WP, Ye HH, ET AL. 1H-2,3-Dihydroperimidine Derivatives: A New Class of Potent Protein Tyrosine Phosphatase 1B Inhibitors, *Molecules.* 2014;19:102-121,.
- Harry NA, Radhika S, Neetha MG, Anilkumar. A novel catalyst-free mechanochemical protocol for the synthesis of 2,3-dihydro-1H-perimidines, *J. Heterocycl. Chem.* 2020;13:2–8.
- Giani AM. Fluorescence studies on 2-(het)aryl perimidone derivatives, *J. Lumin.* 2016;179:384–392.
- Cucciolito ME, Panunzi B, Ruffo FT. Preparation, structure, and metal coordination of 2-(2-methyl-2,3-dihydro-1H-perimidin-2-yl)benzene-1,3-diol, *Tetrahedron Lett.* 2013;54(112):1503–1506.
- Kazumi Maruyama MI, Nobuhara Y, Yamada T, Okabe S. Synthesis and cytoprotective Antiulcer Activity of 2- or 4-(1H-Pyrazol-1-yl)pyrimidine Derivatives Related to Mepirizole and Dulcerozine, *Chem. Pharm. Bull.* 2002;143:2091.
- N'dri J, Koné MG, KODJO C, AFFI S, KABLAN A, OUATTARA O, Soro D. Quantitative Activity Structure Relationship (QSAR) of a Series of Azetidinones Derived from Dap-sone by the Method of Density Functional Theory (DFT), *IRA International Journal of Applied Sciences (ISSN 2455-4499).* 2017;8(12):55-62.
- De HV, Waterbeemd S, Rose. Quantitative Approaches to Structure-Activity Relationships, *Wermuth's the Practice of Medicinal Chemistry.* 2008;23:491-513.
- Shahlaei M. Descriptor Selection Methods in Quantitative Structure-Activity Relationship Studies: A Review Study, *American Chemical Society.* 2013;113(110):8093-8103.
- N'dri JS, Koné MGR, Kodjo CG, Affi ST, Kablan ALC, Ouattara ZA, Ziao N. Quantitative Structure antifungal Activity Relationship (QSAR) study of a series of Schiff bases derivatives from 4aminobenzene-sulphonamide By DFT method, *IOSR: Journal of Pharmacy.* 2017;7(14):27-33.
- N'guessan KN, Koné MGR, Bamba K, Ouattara WP, Ziao N. Quantitative Structure Anti-Cancer Activity Relationship (QSAR) of a Series of Ruthenium Complex Azopyridine by the Density Functional Theory (DFT) Method, *Computational Molecular Bioscience.* 2017;7:19-31.
- Kpidi H, Yapo O, Koné MGR, Gadji G, Gnagne A, N'dri JZ. Monitoring and Modeling of Chlorophyll-a Dynamics in a Eutrophic Lake: M'koa Lake (Jacqueville, Ivory Coast), *American Journal of Environmental Protection.* 2018;6(11):1-9.
- N'dri JS, Koné MGR, Kodjo CG, Kablan ALC, Ouattara L, Ouattara O, Ziao N. Combining of DFT and QSAR results to predict the antibacterial activity of a series of azetidiones derived from dapson as inhibitors of *Bacillus Subtilis* and *Pseudomonas aeruginosa.*, *SDRP Journal of Computational Chemistry & Molecular Modelling.* 2018;2(12):1-9.
- Soro D, Ekou L, Koné MGR, Ekou T, Affi ST, Ouattara LN, Ziao. Prediction of the

- Inhibitory Concentration of Hydroxamic Acids by DFT-QSAR Models on Histone Deacetylase 1, *International Research Journal of Pure & Applied Chemistry*. 2018;16(12):1-13.
16. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE. Gaussian 09, Revision A.02, Gaussian, Inc., Wallingford CT; 2009.
 17. Chattaraj PK, Cedillo A, Parr RG. *J. Phys. Chem.* 1991;103:7645.
 18. Ayers PW, Parr RG. *J. Am Chem. Soc.* 2010;122:2000.
 19. De Proft F, Martin JML, Geerlings P. *Chem. Phys Let.* 1996;250:393.
 20. Hansch C, Sammes PG, Taylor JB. in: *Comprehensive Medicinal Chemistry, Computers and the Medicinal Chemist*. 1990;4:33-58.
 21. Franke R. *Theoretical Drug Design Methods*, Elsevier; 1984.
 22. Chatterjee S, Hadi A, Price B. *Regression Analysis by Examples*, Wiley VCH: New York; 2000.
 23. Phuong H. *Synthèse et étude des relations structure/activité quantitatives (QSAR/2D) d'analogues Benzo[c]phénanthridiniques*, France; 2007.
 24. M. Excel; 2016.
 25. XV. 2. C. Addinsoft, XLSTAT and Addinsoft are Registered Trademarks of Addinsoft. 2014;1995-2014.
 26. Snedecor GW, Cochran WG. *Methods, Statistical*, Oxford and IBH: New Delhi, India. 1967;381.
 27. Kangah NJB, Koné MGR, Kodjo CG, N'guessan BR, Kablan ALC, Yéo SA, et al. Antibacterial activity of schiff bases derived from ortho diaminocyclohexane, meta-phenylenediamine and 1,6-Diaminohexane: Qsar Study with Quantum Descriptors, *International Journal of Pharmaceutical Science Invention*. 2017; 6(13):38-43.
 28. Esposito EX, Hopfinger AJ, Madura JD. *Methods for Applying the Quantitative Structure-Activity Relationship Paradigm, Methods in Molecular Biology*. 2004;275:131-213.
 29. Eriksson L, Jaworska J, Worth A, Cronin MD, Mc Dowell RM, Gramatica P. *Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and Regression-Based QSARs, Environmental Health Perspectives*. 2003;111(110):1361-1375.
 30. Golbraikh , Tropsha A. Beware of qsar, *J. Mol. Graph. Model.* 2002;20:269-276.
 31. Tropsha A, Gramatica P, Gombar VK. The importance of being earnest, validation is the absolute essential for successful application and interpretation of QSPR models, *QSAR Comb. Sci.* 2003;22:69-77.
 32. Ouattara O, Affi TS, Koné MGR, Bamba K, Ziao N. Can Empirical Descriptors Reliably Predict Molecular Lipophilicity ? A QSPR Study Investigation, *Int. Journal of Engineering Research and Application*. 2017;7(15):50-56.
 33. Jeliaskova NN, Jaworska J. An Approach to Determining Applicability Domains for QSAR Group Contribution Models: An Analysis of SRC KOWWIN, *ATLA*. 2005;3(3):461–470.
 34. Sahigara F, Mansouri K, Ballabio D, Mauri A, Todeschini VCR. Comparison of Different Approaches to Define the Applicability Domain of QSAR Models, *Molecules*. 2012;17:4791-4810.
 35. Roy K, et al. A Primer on QSAR/QSPR Modeling Chapter 2 Statistical Methods in QSAR/QSPR, *Springer Briefs in Molecular Science*. 2015;37-59.
 36. Jaworska J, Jeliaskova NN, Aldenberg T. QSAR Applicability Domain Estimation by Projection of the Training Set in Descriptor Space: A Review, *ATLA*. 2005;33:445–459.
 37. Ghamali M, Chtita S, Bouachrine M, Lakhlifi T. Méthodologie générale d'une étude RQSA/RQSP, *Revue Interdisciplinaire*. 2016;1(11).
 38. Chtita S, Ghamali M, Hmamouchi R, Eldrissi B, Bourass M, Larif M, et al. Investigation of Antileishmanial Activities of Acridines Derivatives against Promastigotes and Amastigotes form of Parasites Using QSAR Analysis, *Advances in Physical Chemistry*. 2016;1-16.
 39. Asadollahi T, Dadfarnia S, Shabani A, Ghasemi J, Sarkhosh M. 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.
 40. Chtita S, Larif M, Ghamali M, Bouachrine M, Lakhlifi T. 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. 2015;9:143-154.
41. Vessereau A. Méthodes statistiques en biologie et en agronomie, Paris: Lavoisier (Tec & Doc). 1988;538.
 42. Fortuné A. Techniques de Modélisation Moléculaire appliquées à l'Etude et à l'Optimisation de Molécules Immunogènes et de Modulateurs de la Chimiorésistance. Médicaments; 2006.
 43. Eriksson L, Jaworska J, Worth A, Cronin MD, Mc Dowell RM, Gramatica P. Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs, Environmental Health Perspectives. 2003;111(110):1361-1375.
 44. Gramatica P. Principles of QSAR models validation: internal and external., QSAR Comb. Sci. 2007;26(15):694–701.

© 2021 Tiéba et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
The peer review history for this paper can be accessed here:
<http://www.sdiarticle4.com/review-history/76399>