Predictive modeling of the human hepatoma (Huh-7D12) cancer line of a series of bis- (5-arylidene-rhodanine-3-yl) diamine.

Abstract

This work deals with the prediction of the antiproliferative activity of eighteen (18) substances derived from bis-5-arylidene rhodanine against human hepatoma tumor line (Huh-7D12). while applying the functional density theory (DFT) method to the B3LYP / 6-31G (d, p) level, theoretical descriptors were determined and correlated with antiproliferative (Huh-7) activity by linear regression multiple (RML). This correlation has shown that the electron energy, the energy of the lowest vacant molecular orbital (E_{LUMO}) and the molecular volume (VM) are the quantum and geometric descriptors that best influences the antiproliferative activity of the molecules studied. The coefficient of determination R^2 indicates that 97.9% of the molecular descriptors defining this model are taken into account with a standard deviation of 0.015. The significance of the model reflected by the Fischer test is estimated at 123.648. The robustness of the model given by the cross-validation correlation coefficient (Q^2_{CV}) is 97.9%. This model has been validated by Tropsha criteria. The very good correlation between these three descriptors and the Huh-7 activity was confirmed by the nonlinear multiple regression (RNML) method with better statistical data. (R^2 = 0,998 ; Q^2_{CV} = 0,998 ; RMSE = 0,006).

Mots clés : RML, RMNL, Huh-7D12, bis-5-arylidène rhodanine, Molecular Descriptors.

1. Introduction

The liver is an organ of the digestive system that ensures a particular role of purification of the body. It is also a key organ of body to eliminate toxic compounds. Several types of tumors can develop in this organ, the most common form is hepatocellular carcinoma (or hepatocarcinoma). Hepatocellular carcinoma (HCC) is the most common primary liver tumor in the world. The incidence is globally eleven (11) out of one hundred thousand men (100,000) and 1.5 out of 100,000 women [1], and accounts for about 500,000 deaths, the third leading cause of cancer deaths [2]. Surgery, chemotherapy and irradiation are the main therapeutic approaches to cancer, chemotherapy being an important part of the treatment of cancer patients. However, its success is limited due to the lack of selectivity of tumor cells over normal cells, resulting in insufficient drug concentrations in tumors, systemic toxicity, and the appearance of drug-resistant tumor cells [3]. Targeted molecular therapy can cause less damage to normal cells and may have fewer side effects than other types of cancer treatment. It therefore gains importance because of their specificity with respect to cancer cells, while sparing their toxicity for non-targeted cells. It is in this context that Coulibaly et al [4] synthesized a series of bis-5-arylidene rhodanine derivatives to evaluate their potential as anticancer agents. The in vitro antiproliferative activity of synthesized bis-5-arylidene rhodanine has been studied on the human hepatoma (liver) cancer cell line (Huh-7D12). These compounds, which are very active against the Huh-7D12 line, represent a promising starting point for the development of new, more potent anticancer agents in the future. In this context, the study of Quantitative Structure-Activity Relation (QSAR) is well adapted. The remarkable advances known in the development of computer tools and techniques are of considerable help to the use of this science. This study is a highly sought-after technique because it favors the reduction of the number of experiences that are often long, dangerous and costly in terms of time and finance [5–8]. The descriptors are determined by the methods of quantum chemistry. This QSAR study has its origins in the studies carried out by Hansch [9] and by Free and Wilson [10]. Indeed, Hansch has established models relating biological activity with the hydrophobic, electronic and steric properties of molecules. In general, the QSAR model is based on a fifth (1/5) of the initial database. The QSAR model is a mathematical relation that allows to correlate quantitatively the Huh-7D12 line of the series of molecules and their physicochemical properties (descriptors). In this work, the main goal is to apply

QSAR modeling to develop robust and reliable models capable of predicting the antiproliferative activity of a series of twenty (18) bis-5-arylidene rhodanine derivatives against the tumor line of human hepatoma (Huh-7D12).

2. Material and methods

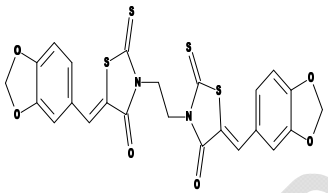
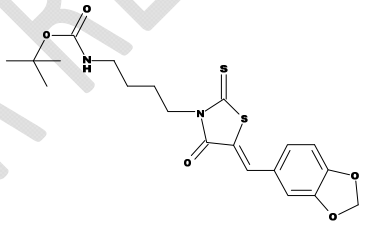
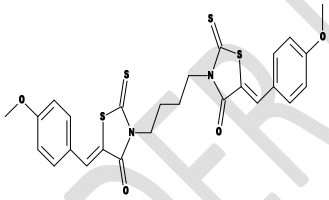
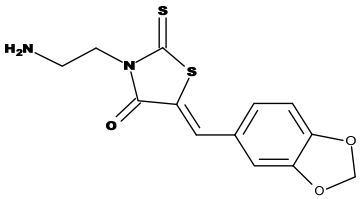
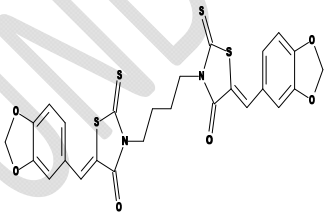
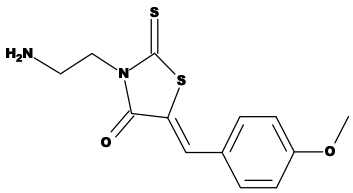
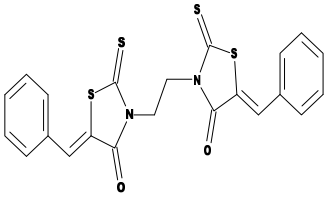
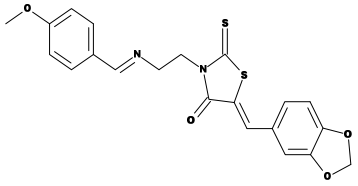
2.1. Materials and Method of Calculation

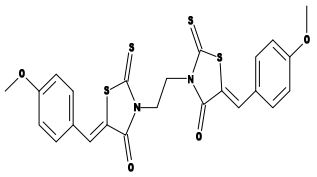
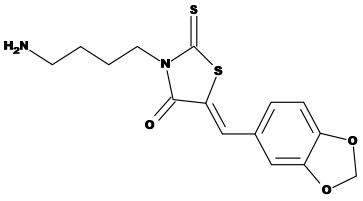
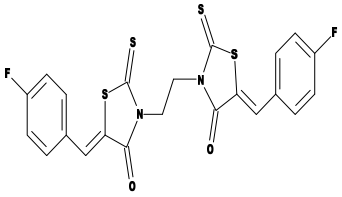
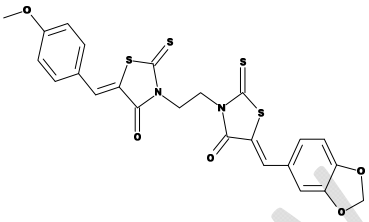
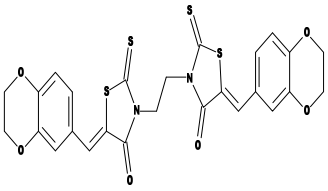
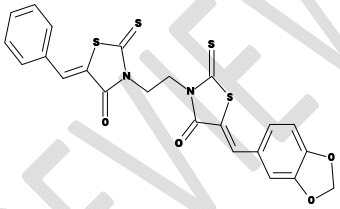
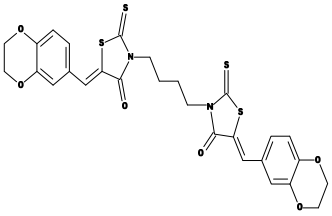
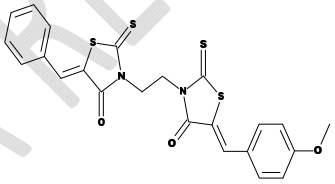
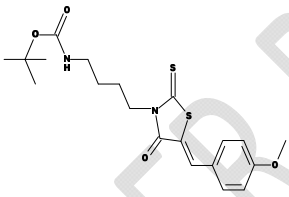
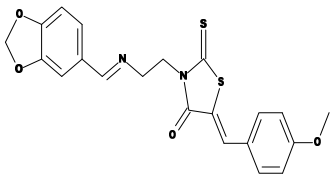
Eighteen (18) molecules of bis-5-arylidene rhodanine derivatives were used in this study (Table 1). Their minimum inhibitory concentration (IC50) varies between 75 and 133 µM. The minimum inhibitory concentration (IC50) is the lowest concentration required to achieve an antiproliferative response. Biological data is usually expressed as the opposite of the log 10 activity base (-log10 (C)) to obtain higher mathematical values when the structures are biologically very efficient [11; 12]. The antiproliferative activity is expressed by the antiproliferative potential pIC50 which is calculated from the following equation (1):

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

Where IC50 represents the median inhibitory concentration of a drug required for 50% inhibition in vitro.

Tableau 1 : molecular structure and antiproliferative activity of the eighteen molecules used.

Code	Molecules	IC_{50} (µM)	Code	Molecules	IC_{50} (µM)
R1		91	**R10**		100
R2		114	**R11**		118
R3		117	**R12**		75
R4		113	**R13**		117

R5		121	**R14**		109
R6		133	**R15**		122
R7		106	**R16**		111
R8		130	**R17**		110
R9		108	**R18**		104

2.2. Calculation Level

The relationship between the values of the biological activity of the studied molecules and their molecular structures was established thanks to the quantum chemistry calculations realized with the Gaussian software 09[13]. Calculations were performed using the Functional Density Theory (DFT) method, which is known to generate a variety of molecular properties [14–17] in QSAR studies that increases predictability, reduces computational time, and influences cost of designing new drugs [11; 18]. The theoretical level of B3LYP / 6-31G (d, p) was used to determine the molecular descriptors. The modeling was carried out using the multilinear regression method implemented in Excel tables [19] and XLSTAT [20].

2.3. Quantum Descriptors

In order to develop a QSAR model, some descriptors of the DFT have been determined. In particular the electronic energy (E) which represents the electronic contribution of all of the atoms of each molecule and the energy of the lowest vacant orbital (E_{LUMO}). These energies were calculated as part of Koopmans' approximation [21]. We have also calculated the molecular volume, which is a geometric

descriptor thanks to the software molinspiration [22]. The molecular volume is the volume occupied by the molecule and is generally expressed in cubic Angstroms (A^3) [23; 24].
For all the descriptors studied, the analysis of the bivariate data, that is to say the calculation of the linear correlation coefficient R between each pair of the set of descriptors, is less than 0.95 ($R < 0.95$), which means that these different descriptors are independent of each other [25; 26; 11].

2.4. Régressions Multiple Linéaires et non Linéaire (RML et RMNL)

The Multiple Linear Regression (RML) statistical method is one of the most popular modeling methods due to its ease of use and ease of interpretation. It has been used to study the relationship between biological activity (dependent variable) and theoretical descriptors (independent variables) [27]. RML minimizes differences between actual and expected values. The advantage of RML is that it is very transparent, since the algorithm is available, and that predictions can be made easily [28]. The RML method is based on the assumption that the property depends linearly on the different variables (the descriptors), according to the relation:

$$\boldsymbol{Y} = a_0 + \sum_{i=1}^{n} a_i X_i \qquad (6)$$

With: Y is the dependent variable (to explain or predict); X_i: the independent (explanatory) variables; n is the number of explanatory variables; a_0 is the constant of the equation of the model; a_i: descriptor coefficients in the model equation.
This method was also used for the selection of molecular descriptors used in multiple nonlinear regression (RMNL). Multiple linear and nonlinear regressions were used to predict the effects on the activity of bis-5-arylidene rhodanine derivatives on Huh-7D12 cancer cells. Multiple nonlinear regression is a nonlinear method (exponential, logarithmic, polynomial, ...) which makes it possible to determine the mathematical model making it possible to explain nonlinearly as well as possible the variability of a property or activity Y according to molecular descriptors X. In all our work we have used the polynomial model based on the descriptors proposed by the linear model which will be raised to the power 2 according to the following equation:

$$\boldsymbol{Y} = a_0 + \sum_{i=1}^{n} a_i X_i + b_i X_i^2 \qquad (7)$$

With: Y is the dependent variable (to explain or predict); X_i: the independent (explanatory) variables; n is the number of explanatory variables; a_0 is the constant of the equation of the model; a_i *and* b_i: descriptor coefficients in the model equation.
RML and RMNL were generated using the XLSTAT software version 2016 [29] to predict the anticancer activity IC_{50}. The equations of the different models were evaluated by the coefficient of determination (R^2) which measures the adequacy of the model and the predictive power of the QSAR model; the Root Mean Square Error (RMSE) which must be less than 10% of the range of the target property value [30]; the Fischer test (F) Test F, for the statistical significance of the model (higher is high, the better is the same set of descriptors and chemicals) [31] and the cross correlation coefficient (Q^2_{CV}) which allows for evaluate the predictive power associated with a QSAR model ($Q^2_{cv} > 0{,}6$ for a satisfactory model while for an excellent model $Q^2_{cv} > 0{,}9$) [32]. These different statistical parameters are given by the following expressions:

$$\boldsymbol{R}^2 = 1 - \frac{\sum\left(y_{i,exp} - \hat{y}_{i,theo}\right)^2}{\sum\left(y_{i,exp} - \bar{y}_{i,exp}\right)^2} \qquad (2)$$

$$\mathbf{RMSE} = \sqrt{\frac{\sum\left(y_{i,exp} - y_{i,theo}\right)^2}{n-k-1}} \quad (3)$$

$$\mathbf{F} = \frac{\sum\left(y_{i,theo} - y_{i,exp}\right)^2}{\sum\left(y_{i,exp} - y_{i,theo}\right)^2} * \frac{n-k-1}{k} \quad (4)$$

$$\boldsymbol{Q^2_{cv}} = \frac{\sum\left(y_{i,theo} - \bar{y}_{i,exp}\right)^2 - \sum\left(y_{i,theo} - y_{i,exp}\right)^2}{\sum\left(y_{i,theo} - \bar{y}_{i,exp}\right)^2} \quad (5)$$

Where:

$y_{i,exp}$: The experimental value of antiproliferative activity on Huh-7D12 cell lines.

$\hat{y}_{i,theo}$: The theoretical value of the antiproliferative activity.

$\bar{y}_{i,exp}$: The mean value of the experimental values of cytotoxicity.

A model is considered efficient according to Eriksson et *al* [33], when $R^2 - Q^2_{CV} < 0{,}3$.

The RML model has been validated by the Tropsha et al criteria defined as follows:

1) $R^2_{Test} > 0{,}7$, **2**) $Q^2_{Cv\ Test} > 0{,}6$, **3**) $\left|R^2_{Test} - R^2_0\right| \leq 0{,}3$,

4) $\frac{\left|R^2_{Test} - R^2_0\right|}{R^2_{Test}} < 0{,}1$ *and* $0{,}85 \leq k \leq 1{,}15$, **5**) $\frac{\left|R^2_{Test} - R'^2_0\right|}{R^2_{Test}} < 0{,}1$ and $0{,}85 \leq k' \leq 1{,}15$

3. Results and Discussions

3.1. Multiple Linear Regression (RLM)

The set of twelve (12) molecules used in the different test sets and the six (6) molecules of the validation set for each model are presented in Table 2. The Pearson correlation matrix between the different physicochemical descriptors are given in Table 3.

Table 2: Molecule Database of Test Set and Validation Set

Molecules	E_{LUMO} (eV)	E (eV)	VM(A^3)	pIC50
Test Set				
R1	-2.826	-44466.076	247.200	4.125
R3	-2.772	-81733.576	427.200	3.914
R4	-2.696	-54906.250	356.120	3.983
R6	-2.686	-88026.892	488.390	3.886
R11	-2.684	-54002.939	377.690	3.967
R12	-2.722	-51863.049	344.080	4.000
R15	-2.826	-44466.076	247.200	4.125
R18	-2.657	-85887.274	459.180	3.975
R7	-2.686	-88026.892	488.390	3.886
R8	-2.684	-54002.939	377.690	3.967
R9	-2.754	-54906.239	356.120	3.932
R13	-2.826	-44466.076	247.200	4.125
Validation Set				
R2	-2.787	-56016.551	376.070	3.928
R5	-2.841	-76603.759	403.270	3.959
R10	-2.657	-85887.274	459.180	3.975
R17	-2.787	-56016.551	376.070	3.928
R14	-2.773	-46605.703	280.800	3.963
R16	-2.860	-78617.369	401.650	3.955

Table 3: Values of the bivariant linear correlation coefficients of the descriptors

	E_{LUMO} (eV)	E (eV)	VM (A^3)

E_{LUMO} (ev)	**1**		
E (ev)	-0.585	**1**	
VM(A^3)	0.812	-0.928	**1**

The linear correlation coefficients R calculated from the series of descriptors are less than 0.95 (R <0.95). This reflects the non-dependence of the descriptors used to develop the models. The correlation between the experimental IC50 inhibition concentrations and the theoretical descriptors of the studied molecules is presented below. Figure 1 represents the correlation between the experimental activities and the theoretical activities predicted by the model. 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 and this parameter of the regression equation. The negative sign indicates that when the value of the descriptor is high, the biological activity decreases. The positive sign reflects the opposite effect. The equation obtained is shown below:

$$\boldsymbol{pIC_{50}^{exp}} = 7.454 + 1.0392 * \mathbf{ELUMO} - 7.4381.10 - 06 * \mathbf{E} - 2.9477.10 - 03 * \mathbf{VM}$$

N=12 $R^2 = 0.979$ $Q^2_{CV} = 0.979$ RMSE = 0.015 F= 123.648 $R^2 - Q^2_{CV} = 0.00$

This model indicates that HOMO energy, electron energy and molecular volume explain to about 98% ($R^2 = 0.979$) the variability of experimental anticancer activity. The negative signs of the coefficients of the electronic energy (E) and the molecular volume (VM), indicate that the anticancer activity will be improved for low values of these descriptors. And the positive sign of the energy of the lowest vacant orbital (E_{LUMO}) also indicates that anticancer activity will be improved for high values of this energy. The meaning of the model is expressed by the Fischer coefficient F = 123.648: the correlation coefficient of the cross validation Q^2_{CV} = 0.979 reflects an excellent robustness of the model (Q^2_{CV}> 0.9). This model is acceptable with $R^2 - Q^2_{CV}$ = 0,979 - 0,979 = 0,000 < 0,3.

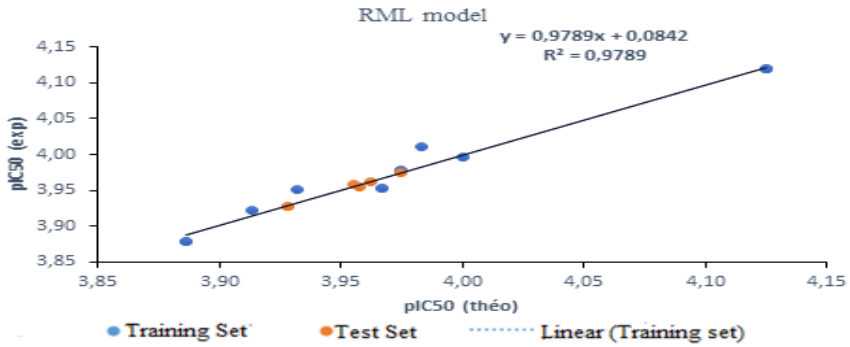


Figure 1: Regression line of the obtained RML model

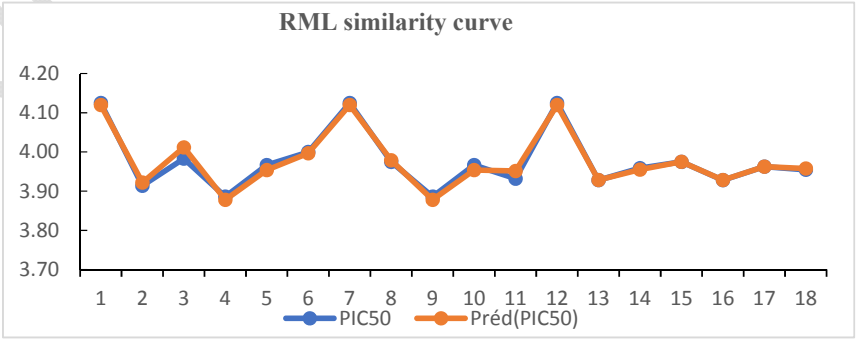


Figure 2: Similarity curve of the experimental and predicted values of the RML model

3.1.1. Verification of Tropsha Criteria

The results of the calculation of the Tropsha criteria of the RML model are as follows:

$R^2_{Test} = 0.987 > 0.7$ $\quad Q^2_{Cv\,Test} = 0.987 > 0.6$ $\quad \left|R^2_{Test} - R^2_0\right| = 0.0128 \leq 0.3$

$\frac{|R^2_{Test} - R^2_0|}{R^2_{Test}} = 0.0130 < 0{,}1$ and $0.85 \leq k = 1.00 \leq 1.15$;

$\frac{|R^2_{Test} - R'^2_0|}{R^2_{Test}} = 0.0130 < 0{,}1$ and $0.85 \leq k' = 1.00 \leq 1.15$

The model is therefore acceptable for predicting Huh7 anticancer activity because it meets the five criteria of Tropsha [34–36].

3.1.2. Analysis of the contribution of the descriptors

The study of the contribution of the descriptors relating to the prediction of the antiproliferative activity of the compounds was carried out for cancer cells of the human liver (Huh-7D12). This contribution of the three descriptors in the prediction of the antiproliferative activity of the bis-5-arylidene rhodanine derivatives was determined from the XLSTAT software version 2016[20]. The different contributions are illustrated in Figure 3.

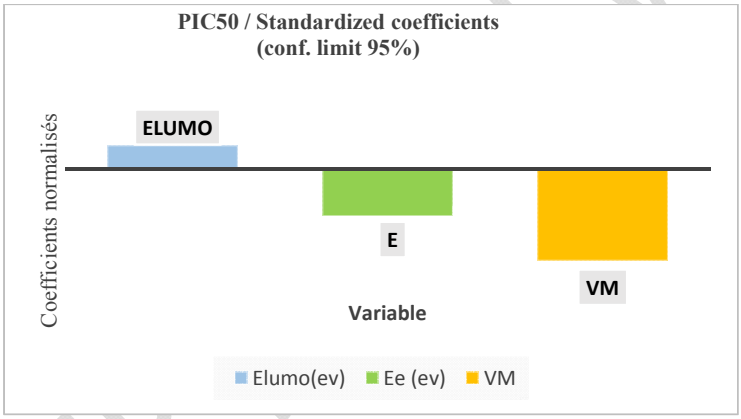


Figure 3: Contribution of descriptors in the RML model

The decreasing order of the contribution of different descriptors in the prediction of the antiproliferative activity of Huh-7D12 is: *VM* > *E* > E_{LUMO}. According to this sequence, the molecular volume is the priority descriptor followed by the electronic energy and finally the energy of the lowest molecular orbital vacant.

3.2. NonLinear Multiple Regression (RMNL)

The statistical nonlinear regression method was used to improve the anticancer activity of the compounds predicted quantitatively. It takes into account the three chosen descriptors ($\mathbf{E_{LUMO}}$, **E**, **VM**). It is the most common tool for studying multidimensional data. This statistical method is applied to the data in Tables 3. The result obtained is the following:

$$\boldsymbol{pIC_{50}^{exp}} = 48.0625 + 30{,}8771*\mathbf{E_{LUMO}} + 5.4404.10^{-05}*\mathbf{E} + 4.5621.10^{-03}*\mathbf{VM} + 5.5441*E_{LUMO}{}^2 + 4.3026.10^{-10}*\mathbf{E}^2 - 8.4391.10^{-06}*\mathbf{VM}^2$$

$N = 12$ $\quad R^2 = 0.998$ $\quad Q^2_{CV} = 0.998$ $\quad RMSE = 0.006$ $\quad R^2 - Q^2_{CV} = 0.00$

In this model, the descriptors (E_{LUMO}, E, VM) used, express the variability of the anticancer activity to a little more than 99%. The correlation coefficient of the cross validation $\boldsymbol{Q^2_{CV}}$**= 0.998** which shows the very good robustness of the model ($\boldsymbol{Q^2_{CV}}$**> 0.9)**. This model is acceptable with $\boldsymbol{R^2 - Q^2_{CV} = 0.998 -}$

$\mathbf{0.998 = 0.000 < 0.3}$. The regression line between the experimental and theoretical anticancer activities of the test set (blue dots) and the test set (red dots) is shown in Figure 4.

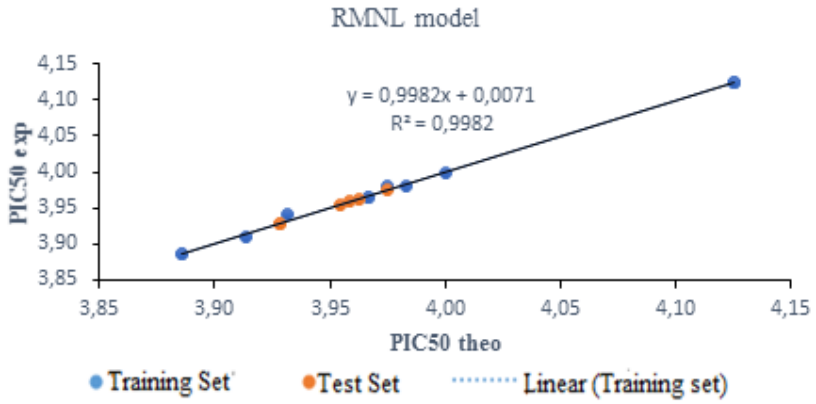


Figure 4: The regression line of the RMNL model

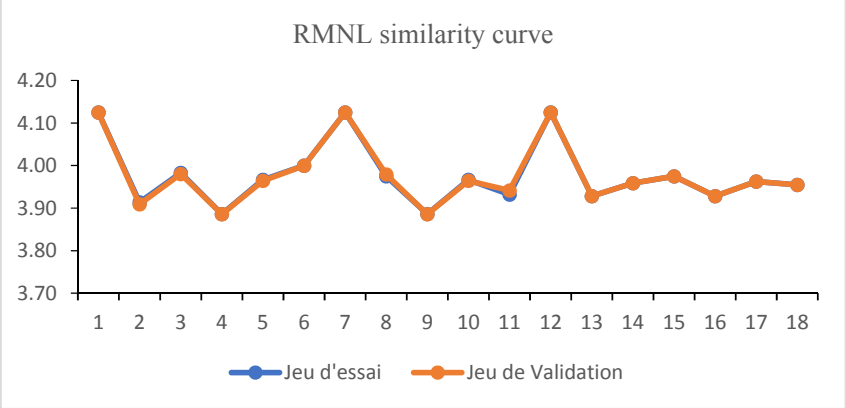


Figure 5: Similarity curve of the RMNL model

For RMNL models, the very low value of the standard error (**RMSE = 0,006**) also demonstrates the good similarity between predicted and experimental values (Figure 5). This curve shows a very good evolution between the experimental values and predicted by the RMNL model of the anticancer activity of the rhodanine derivatives studied.

Table 4: Values of the theoretical activity / experimental activity ratio of the validation set of the two models.

Model RMNL			
Compounds	$pCI_{50}exp$	$pCI_{50}pred$	$pCI_{50}pred/ pCI_{50}exp$
R2	3.928	3.928	1.00
R5	3.959	3.959	1.00
R10	3.975	3.975	1.00
R17	3.928	3.928	1.00
R14	3.963	3.963	1.00
R16	3.955	3.955	1.00

All values in the pCI50pred / pCI50exp report tend to 1 (Table 4). This indicates a good correlation between the theoretical and experimental toxicity of the rhodanines studied. This model is acceptable for predicting the toxicity of rhodanines on the human hepatoma line (Huh-7D12).

4. Conclusion

The electron energy, the highest occupied orbital energy (E_{HOMO}) and the molecular volume (VM) have been used to describe and predict the activity of 18 molecules derived from bis-5-arylidene rhodanine

against the cancer line. of human hepatoma (Huh-7D12). Multiple linear regression was used to quantify the relationships between molecular descriptors and the properties of the antiproliferative activity of bis-5-arylidene rhodanine derivatives. This study revealed a strong correlation between the experimental antiproliferative activities and the theoretical descriptors calculated by DFT. In addition, the good correlation between the Huh-7D12 activity and these three descriptors was confirmed by the nonlinear multiple regression method. The molecular volume appears as the priority descriptor.

References

[1] L. Fartoux, C. Desbois-Mouthon and O. Rosmorduc, Carcinome hépatocellulaire, EMC-Hépatologie (2009), 7–38.

[2] D.M. Parkin, F. Bray, J. Ferlay and P. Pisani, Global cancer statistics, 2002, CA: a cancer journal for clinicians, 55 (2) (2005), 74–108.

[3] G. Xu and H.L. McLeod, Strategies for enzyme/prodrug cancer therapy, Clinical Cancer Research, 7 (11) (2001), 3314–3324.

[4] W. Coulibaly, L. Paquin, A. Bénié, Y.-A. Bekro, E. Durieux, L. Meijer, R. Le Guével, A. Corlu and J.-P. Bazureau, Synthesis of New N, N'-Bis (5-arylidene-4-oxo-4, 5-dihydrothiazolin-2-yl) piperazine Derivatives Under Microwave Irradiation and Preliminary Biological Evaluation, Scientia pharmaceutica, 80 (4) (2012), 825–836.

[5] D. Soro, L. Ekou, M.G.-R. Koné, T. Ekou, S.T. Affi, L. Ouattara and N. Ziao, Prediction of the Inhibitory Concentration of Hydroxamic Acids by DFT-QSAR Models on Histone Deacetylase 1, International Research Journal of Pure and Applied Chemistry (2018), 1–13.

[6] A. Tropsha, Best practices for QSAR model development, validation, and exploitation, Molecular informatics, 29 (6-7) (2010), 476–488.

[7] M.T. Chhabria, B.M. Mahajan and P.S. Brahmkshatriya, QSAR study of a series of acyl coenzyme A (CoA), Medicinal Chemistry Research, 20 (9) (2011), 1573–1580.

[8] V.M. Buha, D.N. Rana, M.T. Chhabria, K.H. Chikhalia, B.M. Mahajan, P.S. Brahmkshatriya and N.K. Shah, Synthesis, biological evaluation and QSAR study of a series of substituted quinazolines as antimicrobial agents, Medicinal Chemistry Research, 22 (9) (2013), 4096–4109.

[9] C. Hansch and T. Fujita, p -σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure, J. Am. Chem. Soc., 86 (8) (1964), 1616–1626.

[10] S.M. Free and J.W. Wilson, A Mathematical Contribution to Structure-Activity Studies, J. Med. Chem., 7 (4) (1964), 395–399.

[11] N.J.-B. Kangah, M.G.-R. Koné, A.L.C. Kablan, S.A. Yéo and N. Ziao, Antibacterial Activity of Schiff Bases Derived from Ortho-Diaminocyclohexane, Meta-Phenylenediamine and 1, 6-Diaminohexane, International Journal of Pharmaceutical Science Invention, 6 (13) (2017), 38–43.

[12] T.N.P. Huynh, Synthèse et études des relations structure/activité quantitatives (QSAR/2D) d'analyse benzo [c] phénanthridiniques (Université d'Angers2007).

[13] Gaussian 09, Revision D.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2013, - Google Search.

[14] P. Ayers, Parr RG, J. Am. Chem. Soc., 2000 (2000), 122–2010.

[15] F. de Proft and P. Geerlings, Conceptual and computational DFT in the study of aromaticity, Chemical reviews, 101 (5) (2001), 1451–1464.

[16] F. de Proft, J.M.L. Martin and P. Geerlings, On the performance of density functional methods for describing atomic populations, dipole moments and infrared intensities, Chemical Physics Letters, 250 (3-4) (1996), 393–401.

[17] J.R. Pliego Jr, Thermodynamic cycles and the calculation of pKa, Chemical Physics Letters, 367 (1-2) (2003), 145–149.

[18] R. Franke, Theoretical drug design methods (Elsevier Science Ltd1984).

[19] K.N. N'guessan, M.G.-R. Koné, K. Bamba, O.W. Patrice and N. Ziao, Quantitative structure anti-cancer activity relationship (QSAR) of a series of ruthenium complex azopyridine by the density functional theory (DFT) method, Computational Molecular Bioscience, 7 (02) (2017), 19.
[20] XLSTAT version 2016.5.03- Google Search, Copyright Addinsoft 1995-2014 XLSTAT and Addinsoft are Registered Trademarks of Addinsoft. 2016, https://www.xlstat.com
[21] T. Koopmans, Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den einzelnen Elektronen eines Atoms, Physica, 1 (1-6) (1934), 104–113.
[22] Logiciel libre Molinspiration Cheminformatics,*http://www.molinspiration.com*. Accession en (02 May 2019).
[23] B. Lee and F.M. Richards, The interpretation of protein structures, Journal of molecular biology, 55 (3) (1971), 379-IN4.
[24] A. Shrake and J.A. Rupley, Environment and exposure to solvent of protein atoms. Lysozyme and insulin, Journal of molecular biology, 79 (2) (1973), 351–371.
[25] V.V. Nalimov, The application of mathematical statistics to chemical analysis (Elsevier2014).
[26] A.R. Katritzky, V.S. Lobanov and M. Karelson, CODESSA, University of Florida, Gainesville, FL (1994).
[27] K. Roy, S. Kar and R.N. Das, Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment (Academic press2015).
[28] M. Larif, A. Adad, R. Hmammouchi, A.I. Taghki, A. Soulaymani, A. Elmidaoui, M. Bouachrine and T. Lakhlifi, Biological activities of triazine derivatives. Combining DFT and QSAR results, Arabian Journal of Chemistry, 10 (2017), S946-S955.
[29] S. Karabulut, N. Sizochenko, A. Orhan and J. Leszczynski, A DFT-based QSAR study on inhibition of human dihydrofolate reductase, Journal of molecular graphics and modelling, 70 (2016), 23–29.
[30] G.A.F. Seber and A.J. Lee, Linear regression analysis (John Wiley & Sons2012).
[31] K. Asgaonkar, G. Mote and T. Chitre, QSAR and molecular docking studies of oxadiazole-ligated pyrrole derivatives as enoyl-ACP (CoA) reductase inhibitors, Scientia pharmaceutica, 82 (1) (2013), 71–86.
[32] C. Rücker, G. Rücker and M. Meringer, y-Randomization and its variants in QSPR/QSAR, Journal of chemical information and modeling, 47 (6) (2007), 2345–2357.
[33] L. Eriksson, J. Jaworska, A.P. Worth, M.T.D. Cronin, R.M. McDowell and P. Gramatica, Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs, Environmental health perspectives, 111 (10) (2003), 1361–1375.
[34] A. Tropsha, P. Gramatica and V.K. Gombar, The importance of being earnest, QSAR & Combinatorial Science, 22 (1) (2003), 69–77.
[35] O. Ouattara, A. Thomas Sopi, M.G.-R. Koné, K. Bamba and N. Ziao, Can Empirical Descriptors Reliably Predict Molecular Lipophilicity? A QSPR Study Investigation, Int. Journal of Engineering Research and Application, 7 (15) (2017), 50–56.
[36] A. Golbraikh and A. Tropsha, Beware of q2!, Journal of molecular graphics and modelling, 20 (4) (2002), 269–276.