Australian Journal of Basic and Applied Sciences
2019 June; 13(6): pages 106-114
DOI: 10.22587/ajbas.2019.13.6.12
Original paper
AENSI Publications
Journal home page: www.ajbasweb.com

Predictive study of human colon (CaCo2) and pulmonary tumor lines (NCI-h727) of a series of bis- (5-arylidene-rhodanine-3-yl) diamine

[a]Koffi Alexis Respect Kouassi, [a, b]Anoubilé Benié, [c]Wacothon Karime Coulibaly, [d]Affoué Lucie Bédé, [a,*]Mamadou Guy-Richard Koné, [d]Adenidji Ganiyou

[a]Laboratoire de Thermodynamique et de Physico-Chimie du Milieu, UFR SFA, Université Nangui Abrogoua, 02 BP 801 Abidjan 02, Côte d'Ivoire,
[b]Laboratoire de Chimie BioOrganique et de Substances Naturelles, UFR-SFA, Université Nangui Abrogoua, 02 B.P. 801 Abidjan 02 Côte-d'Ivoire,
[c]UFR des Sciences Biologiques, Université de Péléforo Gon Coulibaly, Korhogo, BP 1328, Cote d'Ivoire;
[d]Laboratoire de Chimie Organique et de Substances Naturelles, UFR-SSMT, Université Félix Houphouët-Boigny 22 BP 582 Abidjan 22, Côte d'Ivoire

Correspondence Author: Mamadou Guy-Richard Koné, Laboratoire de Thermodynamique et de Physico-Chimie du Milieu, UFR SFA, Université Nangui Abrogoua, 02 BP 801 Abidjan 02, Côte d'Ivoire.
E-mail:- guyrichardkone@gmail.com

Received date: 18 May 2019, Accepted date: 26 June 2019, Online date: 29 June 2019

Abstract

Cancer is a generic term for a large group of diseases characterized by the growth of abnormal cells beyond their normal limits. In order to develop new anticancer drugs, the Quantitative Structure Activity (QSAR) relationship has been applied. to a series of eighteen (18) molecules of bis- (5-arylidene-rhodanine-3-yl) diamine. We have determined the physicochemical descriptors on which observed antiproliferative activity depends, in order to be able to predict biological activities in series of analogous molecules. Two models based on these molecular descriptors and antiproliferative activities of two cell lines (human colon tumor cell line (*CaCo2*) and pulmonary human tumor cell line (*NCI-h727*)) were obtained. The results of calculations showed that energy Gap (ΔE), global softness (S) and molecular volume (VM) are the best descriptors related to the values of antiproliferative activity of the studied molecules. These models obtained give statistically significant results and show good predictability ($R^2 = 0,909$, $Q^2 = 0,909$, F=26.790, S=0.033 for CaCo2 and $R^2 = 0,915$; $Q^2 = 0,915$, F=28.690 and S=0.025 for NCI-h727). The energy descriptor (ΔE) was identified as the priority descriptor in the prediction of human colon tumor (*CaCo2*) line and molecular volume (VM) for human lung tumor line (*NCI-h727*). These models were evaluated using the acceptance criteria of Tropsha et *al.*

Keywords: bis-(5-arylidène-rhodanine-3-yl) diamine, Descriptors, QSAR; DFT; RML

INTRODUCTION

Cancer is a disease characterized by cell proliferation, or malignant tumor, abnormally important formed from mutation transformation or genetic instability of an initially normal cell. This disease is a major problem worldwide and is the leading cause of death in developed countries (Victora *et al.*, 2016). One of the most difficult problems to solve during cancer treatment is the invasion of cancer cells responsible for the spread of tumor cells in the body (Sohn *et al.*, 2010). Despite several efforts in the treatment of cancer, this disease has become a big problem for society's health. The goal is to develop drugs with greater anticancer activity and less toxicity than current drugs (Ghanbari *et al.*, 2014). The search for new drugs, offering broad prospects for evolution and eradication, remains a subject of interest for the scientific, academic and industrial world. In this context, the discovery of selective inhibitors vis-a-vis protein kinases, has become the key point for research of new therapeutic agents. Computational chemistry currently plays an important role in the rational design of drugs (Csizmadia and Enriz, 2000). Quantitative Structure-Activity Relationship (QSAR) is one of the best methods used to design new therapeutic agents (Buha *et al.*, 2013; Tropsha, 2010; Chhabria *et al.*, 2011). It makes it possible to correlate quantitatively through a mathematical model, the structure or the properties of the compounds with their biological activities. It is increasingly used to reduce the excessive number of experiments, sometimes long, expensive and the cost of drug production by pharmaceutical companies (Rekka and Kourounakis, 2008; Oprea *et al.*, 2005; Hansch and Fujita, 1964). This QSAR approach has its origins in the studies carried out by Hansch (Hansch and Fujita, 1964) , Free and Wilson (Free and Wilson, 1964). Hansch has established models linking biological

Citation: Koffi Alexis Respect Kouassi, *et al.*, Predictive study of human colon (CaCo2) and pulmonary tumor lines (NCI-h727) of a series of bis- (5-arylidene-rhodanine-3-yl) diamine. Australian Journal of Basic and Applied Sciences, 13(6): 106-114. DOI: 10.22587/ajbas.2019.13.6.12

activity with the hydrophobic, electronic and steric properties of molecules. This QSAR study makes it possible to correlate quantitatively with a mathematical model, the structure of compounds with their biological activities. A series of rhodanine derivatives were synthesized and evaluated for their *in vitro* antimalarial activities against the resistant strain of *Plasmodium falciparum* by Coulibaly et *al*. (COULIBALY *et al.*, 2012). These molecules showed significant antiproliferative activity on human colon tumor line (CaCo2) and human lung tumor line (NCI-h727). The improvement of the antiproliferative activity of bis-(5-arylidene-rhodanine-3-yl) diamine requires a mastery of the physicochemical properties that govern it. This would help to efficiently orient the synthesis of new molecules based on the structure of bis- (5-arylidene-rhodanine-3-yl) diamine. In this work, the object is to conduct a descriptive and predictive study of the antiproliferative activity of a series of bis- (5-arylidene-rhodanine-3-yl) diamine by applying quantum chemistry methods in order to model the anticancer activities observed.

2. CALCULATION METHODS

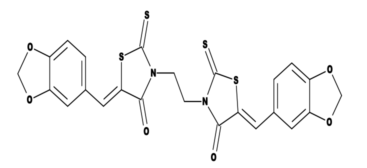
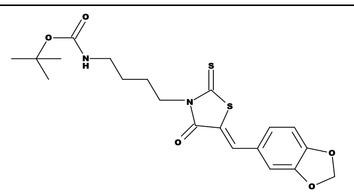
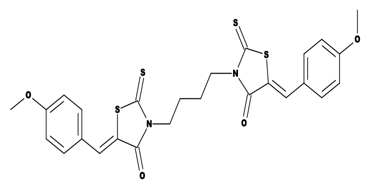
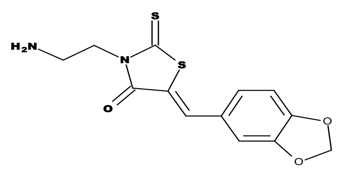
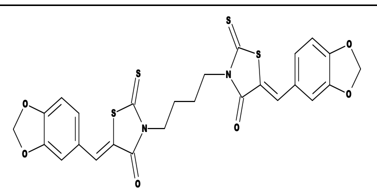
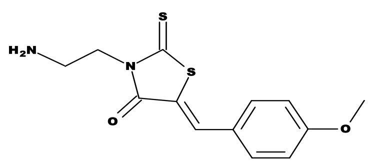
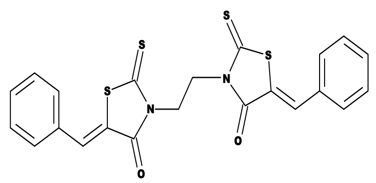
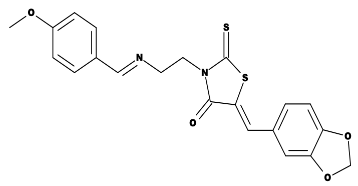
2.1. Experimental data

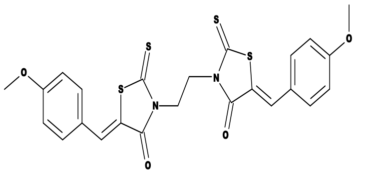
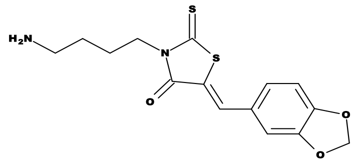
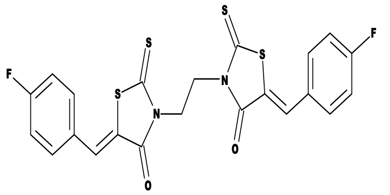
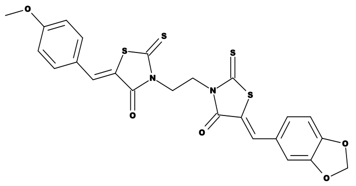
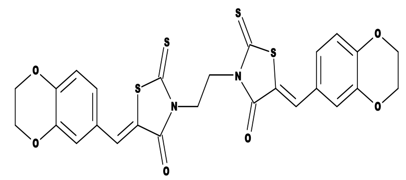
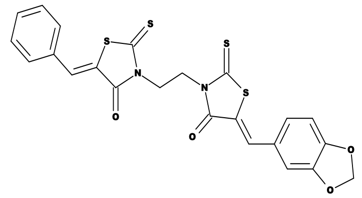
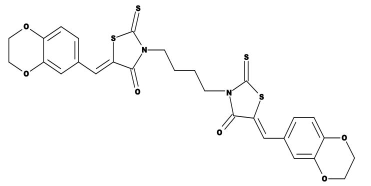
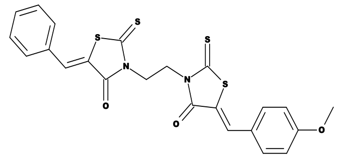
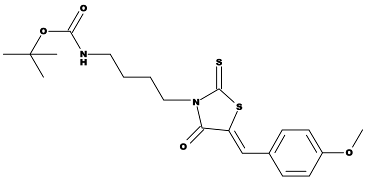
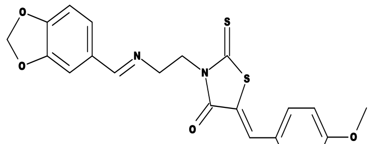
Molecular descriptors were calculated using only the chemical structure of compounds. These descriptors help us predict the inhibitory concentration of similar molecules. This QSAR study is concerned with 18 molecules (**Table 1**) of bis- (5-arylidene-rhodanine-3-yl) diamine synthesized and then tested on two strains of malaria by Coulibaly et *al*. (COULIBALY *et al.*, 2012). To validate our dataset using a QSAR model, the 18 compounds studied were randomly split into two sets. Twelve (12) of them were used for the test game to build the QSAR models and the remaining six (6) for the validation of the obtained models. The antiproliferative activity of bis (5-arylidene-rhodanine-3-yl) diamine studied varies between 67 and 143 µM. This range of concentrations makes it possible to define a quantitative relationship between the antiproliferative activity and theoretical descriptors. Biological data are usually expressed opposite the log 10 base of activity (- log (1/IC50)) to obtain higher mathematical values when the structures are biologically very efficient (Chaltterjee *et al.*, 2011; Chattaraj *et al.*, 1995; Parr *et al.*, 1978). The antiproliferative activity is expressed by the antiproliferative potential pCI50 defined as follows:

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

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

Table 1: Molecular Structure and Antiproliferative Activity of 18 Molecules on CaCo2 and NCI-h727 Cell Lines

Code	Molecules studied	$IC_{50\ Exp}$ (µM) Cell lines		Code	Molecules studied	$IC_{50\ Exp}$ (µM) Cell lines	
		CaCo2	NCI-H727			CaCo2	NCI-H727
M1		67	91	**M10**		92	97
M2		101	84	**M11**		107	74
M3		99	86	**M12**		71	108
M4		106	99	**M13**		105	97

M5		104	133	**M14**		105	116
M6		113	107	**M15**		123	4,013
M7		96	107	**M16**		111	143
M8		96	124	**M17**		106	119
M9		68	110	**M18**		103	87

2.2. Computational methods

The geometries of bis- (5-arylidene-rhodanine-3-yl) diamine have been optimized with the DFT method, which provides a large number of molecular properties in QSAR studies (Chattaraj *et al.*, 1995; Parr *et al.*, 1978). The calculation of the electronic descriptors was carried out using Gaussian software 09 (Frisch *et al.*, 2013) using the theoretical level B3LYP / 6-31G (d, p). This Hybrid functional gives better energies and it is in agreement with the high-level ab initio methods (Kapp *et al.*, 1996; Johnson *et al.*, 1993). As for the split-valence and double-dzeta base (6-31G (d, p)), it is sufficiently extensive and taking into account of the diffuse and polarization functions is important for the explanation of the free doublets of the heteroatoms. All geometrical optimizations of the molecules were carried out beforehand in order to obtain structure in its initial state. In addition, this stable configuration was confirmed by the frequency analysis which revealed the absence of imaginary frequency. 2D structures of all molecules were constructed using Gauss View 5 (Dennington *et al.*, 2009). Modeling was performed using the multilinear regression method implemented in Excel (Thomas *et al.*) and XLSTAT version 2016 (Zanuncio *et al.*, 2016).

2.3. Calculation of the molecular descriptors

In order to develop QSAR models, theoretical descriptors related to the conceptual DFT were determined such as the lowest unoccupied molecular orbit energy (E_{LUMO}), the highest occupied molecular orbit energy. (E_{HOMO}). These descriptors are all determined from the optimized molecules. The descriptors for the boundary molecular orbitals were calculated as part of the Koopmans approximation (Koopmans, 1934). E_{LUMO} energy characterizes the sensitivity of the molecule to nucleophilic attack, and the EHOMO energy characterizes the sensitivity of a molecule to an electrophilic attack (SIMON, 1998). These descriptors were used to calculate the energy gap ($\Delta E = E_{LUMO} - E_{HOMO}$). Other descriptors have also been used: molecular volume (*VM*), which is a fundamental physical property of molecules, very important for understanding their structure, function and interactions (Connolly, 1985)[28]. This molecular descriptor was determined from the Free Molinspiration Software cheminformatics (*Chemoinformatics*, 2004; *molinspiration cheminformatics software - Recherche Google*).

2.4. Multiple Linear Regression (MLR)

The Multiple Linear Regression Statistical Technique (MLR) is used to study the relationship between a dependent variable (Biological activity) and several independent variables (descriptors). This statistical method minimizes the differences between the actual and predicted values. The MLR was generated using the XLSTAT software version 2016 (Zanuncio *et al.*, 2016) to predict the antiproliferative activity of two cell lines ((CaCo2) and (NCI-h727)). The equations of both models were evaluated by the coefficient of determination (R^2), the mean squared error (S), the Fischer test (F) and the cross-correlation coefficient (Q^2_{CV}) (Rücker *et al.*, 2007b).

2.5. Statistical analysis

The modeling not only provides a model adjusted to the experimental data but also makes it possible to predict the biological activity of new molecules not yet synthesized (Esposito *et al.*, 2004). To achieve this objective, several validation methods are used to estimate the reliability of the model including the analysis of the coefficient of determination R^2, the standard deviation S, the cross-validation correlation coefficients Q^2_{CV} and the coefficient of Fischer F. Adjustment of calculated and experimental values given by statistical indicators R^2, S and F describe the predictive ability within the limits of the model and make it possible to estimate the accuracy of the values calculated on the formation game (Agrawal *et al.*, 2002; Mattioni and Jurs, 2003). The predictive power information of the model is provided by the cross-validation coefficient Q^2_{CV}. R^2 is interpreted as the proportion of the variability of the dependent variable (here the antiproliferative activity) explained by the model. The closer the R^2 is to 1, the better the model (Kamchonwongpaisan *et al.*, 2004). Its value is defined by :

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

With:
$y_{i,exp}$: The experimental value of the antiproliferative activity on 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 the toxicity.
Moreover, the variance $\boldsymbol{\sigma}^2$is determined by the relation 3:

$$\boldsymbol{\sigma}^2 = \boldsymbol{s}^2 = \frac{\sum\left(y_{i,exp} - y_{i,theo}\right)^2}{n-k-1} \qquad (3)$$

With k being the number of independent variables (descriptors), n is the number of molecules in the test or learning set and n-k-1 the degree of freedom. The standard deviation or standard deviation S is another statistical indicator used. It is used to evaluate the reliability and accuracy of a model:

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

The Fisher F test is also used to measure the level of statistical significance of the model, i.e. the quality of the choice of the descriptors that make up the model.

$$\mathbf{F} = \frac{\sum\left(y_{i,theo} - y_{i,exp}\right)^2}{\sum\left(y_{i,exp} - y_{i,theo}\right)^2} * \frac{n-k-1}{k} \qquad (5)$$

The coefficient for determining the cross-validation Q^2_{cv}, allows to evaluate the accuracy of the prediction on the test set is calculated using the following relationship:

$$\boldsymbol{Q}^2_{\boldsymbol{cv}} = \frac{\sum\left(y_{i,theo} - \bar{y}_{i,exp}\right)^2 - \sum\left(y_{i,theo} - y_{i,exp}\right)^2}{\sum\left(y_{i,theo} - \bar{y}_{i,exp}\right)^2} \qquad (6)$$

The performance of a mathematical model, for Eriksson et *al*, is characterized by a value of $Q^2_{cv} > 0{,}6$ for a satisfactory model while for the excellent model $Q^2_{cv} > 0{,}9$ (Eriksson *et al.*, 2003). According to them, given a test set, a model will perform well if the acceptance criterion $R^2 - Q^2_{cv} < 0{,}3$ is met (Eriksson *et al.*, 2003; Rücker *et al.*, 2007a). According to Tropsha et *al (Tropsha, 2010; OUATTARA et al., 2017; Golbraikh and Tropsha, 2002)*, for the external validation set, the predictive power of a model can be obtained from five criteria. These criteria are as follows:

1) $R^2_{Test} > 0{,}7$, **2)** $Q^2_{Cv\,Test} > 0{,}6$, **3)** $|R^2_{Test} - R^2_0| \leq 0{,}3$, **4)** $\frac{|R^2_{Test} - R^2_0|}{R^2_{Test}} < 0{,}1 \; and \; 0{,}85 \leq k \leq 1{,}15$,

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

3. RESULTS AND DISCUSSION

Citation: Koffi Alexis Respect Kouassi, *et al.*, Predictive study of human colon (CaCo2) and pulmonary tumor lines (NCI-h727) of a series of bis- (5-arylidene-rhodanine-3-yl) diamine. Australian Journal of Basic and Applied Sciences, 13(6): 106-114. DOI: 10.22587/ajbas.2019.13.6.12

3.1. Data set for analysis

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 **Tables 2** and **4**. The Pearson correlation matrix between the different physico-chemical descriptors is given in **Tables 3** and **5**.

Table 2: Descriptive values and expected antiproliferative activity of rhodanine derivatives for the Caco2 line.

Molecules	VM (A^3)	ΔE (eV)	S (eV^{-1})	PIC50
M1	428,810	0,150	13,316	3,876
M5	459,180	0,214	9,363	3,971
M7	488,390	0,197	10,152	3,907
M8	344,080	0,067	29,638	4,027
M10	376,070	0,056	35,855	4,013
M11	247,200	0,091	21,876	4,131
M12	280,800	0,094	21,305	4,013
M14	427,200	0,066	30,499	3,936
M15	401,650	0,109	18,330	3,845
M16	403,270	0,114	17,543	3,924
M17	356,120	0,170	11,761	4,060
M18	247,200	0,091	21,876	4,131
M2	462,420	0,153	13,079	4,076
M3	459,180	0,074	27,023	4,066
M4	377,720	0,230	8,688	4,004
M6	387,580	0,273	7,321	3,971
M9	377,690	0,054	37,123	3,959
M13	356,120	0,098	20,418	3,967

Table 3: Pearson correlation matrix between the different physico-chemical descriptors

	VM	ΔE	S
VM	**1**		
ΔE	0,548	**1**	
S	-0,357	-0,930	**1**

Table 4: Descriptive values and expected antiproliferative activity of rhodanine derivatives for line NCI-h727.

Molecules	VM (A^3)	ΔE (eV)	S (eV^{-1})	PIC50
M1	377,690	0,054	37,123	3,959
M2	376,070	0,056	35,855	4,013
M3	247,200	0,091	21,876	4,131
M4	356,120	0,098	20,418	3,967
M5	427,200	0,066	30,499	3,936
M6	403,270	0,114	17,543	3,924
M9	459,180	0,214	9,363	3,971
M11	488,390	0,197	10,152	3,907
M12	377,690	0,054	37,123	3,959
M13	344,080	0,067	29,638	4,027
M15	376,070	0,056	35,855	4,013
M17	247,200	0,091	21,876	4,131
M7	459,180	0,214	9,363	3,971
M8	488,390	0,197	10,152	3,907
M10	344,080	0,067	29,638	4,027
M14	280,800	0,094	21,305	4,013
M16	401,650	0,109	18,330	3,845
M18	356,120	0,170	11,761	4,060

Table 5: Values of the bivariate linear correlation coefficients of the descriptors

	VM	ΔE	S
VM	**1**		
ΔE	0,506	**1**	
S	-0,257	-0,923	**1**

Citation: Koffi Alexis Respect Kouassi, *et al.*, Predictive study of human colon (CaCo2) and pulmonary tumor lines (NCI-h727) of a series of bis- (5-arylidene-rhodanine-3-yl) diamine. Australian Journal of Basic and Applied Sciences, 13(6): 106-114. DOI: 10.22587/ajbas.2019.13.6.12

The partial correlation coefficients aij between the descriptors (**Table 3** and **5**) are less than 0.70 (aij < 0.70). These values demonstrate the independence of the descriptors (*VM*, *ΔE* and *S*) used to develop the models taken in pairs.

3.2. Structure-activity quantitative relationship models

These models were developed using different test and validation sets shown in **Tables 2** and **4**. In these models, the negative or positive sign of the model descriptor coefficient reflects the proportionality effect between the evolution of the biological activity of interest and this parameter of the regression equation. Thus, the negative sign indicates that when the descriptor value is high, biological activity decreases while the positive sign reflects the opposite effect. The model equations obtained using the theoretical descriptors for optimized molecules and statistical indicators are presented in **Table 6**, and have been developed using different test and validation sets in **Tables 2** and **4**. In these models, the negative or positive sign of the model descriptor coefficient reflects the proportionality effect between the evolution of the biological activity of interest and this parameter of the regression equation. Thus, the negative sign indicates that when the descriptor value is high, biological activity decreases while the positive sign reflects the opposite effect. The model equations obtained using the theoretical descriptors for optimized molecules and statistical indicators are presented in **Table 6.**

Table 6: The most significant QSAR models for modeling antiproliferative activities on the Caco2 and NCI-h727 cell lines

Cell line	Regression equation	R^2	S	Q^2_{CV}	F	R^2- Q^2_{CV}
CaCo2	$pIC_{50}^{préd}$ = 3,8528-1,4659.10^{-03}***VM** + 3,0674***ΔE**+1,5662.10^{-02}***S**	0.909	0.033	0.909	26.790	0.000
NCI-h727	$pIC_{50}^{préd}$ = 4,1836-1,2644.10^{-03}***VM**+1,4460***ΔE**+5,6051.10^{-03}***S**	0.915	0.025	0.915	28.690	0.000

The negative sign of the molecular volume coefficient (MV) indicates that antiproliferative activity will be improved for low values of this descriptor. In addition, the positive signs of the coefficients of ΔE and S indicate that antiproliferative activity will be improved for high values of these descriptors. The significance of these models is given by the Fisher coefficient F estimated at 26.790 for CaCo2 and 28.690 for NCI-h727. The correlation coefficient of the cross-validation $Q^2_{CV} > 0.6$ for both models. These models are acceptable because R^2cv - $Q^2_{CV} = 0.000 < 0.3$. The different regression lines between the experimental and theoretical antiproliferative activities of the test set (blue dots) and the validation set (red dots) for each cell line are shown in **Figure 1**.

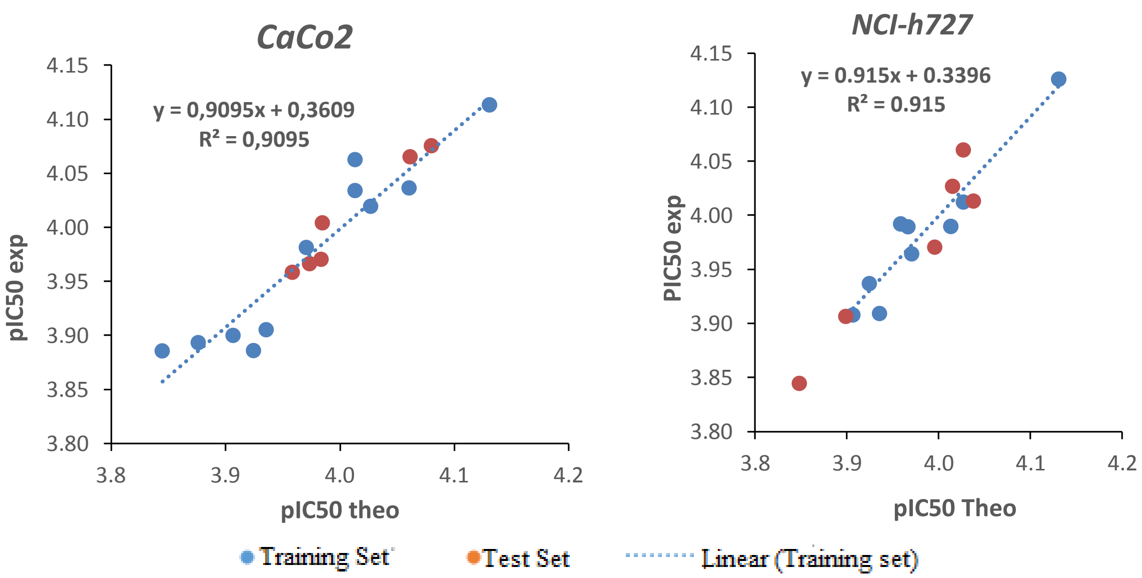


Figure 1: Linear regression of the different models.

Verification of Tropsha criteria for both cell line models

Cell lines	R^2_{Test}	$Q^2_{Cv\ Test}$	$\lvert R^2_{Test}-R^2_0\rvert$	$\left\lvert\frac{R^2_{Test}-R^2_0}{R^2_{Test}}\right\rvert$	$\left\lvert\frac{R^2_{Test}-R'^2_0}{R^2_{Test}}\right\rvert$	k	k'
CaCo2	0.953	0.953	0.000	0.000	0.000	1.000	1.000
NCI-h727	0.922	0.922	0.078	0.084	0.084	1.000	1.000

The five Tropsha criteria of these two models are verified for validation sets. This means that these models can predict and explain the antiproliferative activity of bis-(5-arylidene-rhodanine-3-yl) diamines. As these two models are based on three descriptors each, we will determine the contributions of each descriptor to the prediction of the antiproliferative activity of cell lines.

3.3. Analysis of the contribution of descriptors

The contribution of the three descriptors of each model to the prediction of the antiproliferative activity of bis-(5-arylidene rhodanine-3-yl) diamines was determined from XLSTAT software version 2016 (Johnson *et al.*, 1993). The different contributions are illustrated in **Figure 2**.

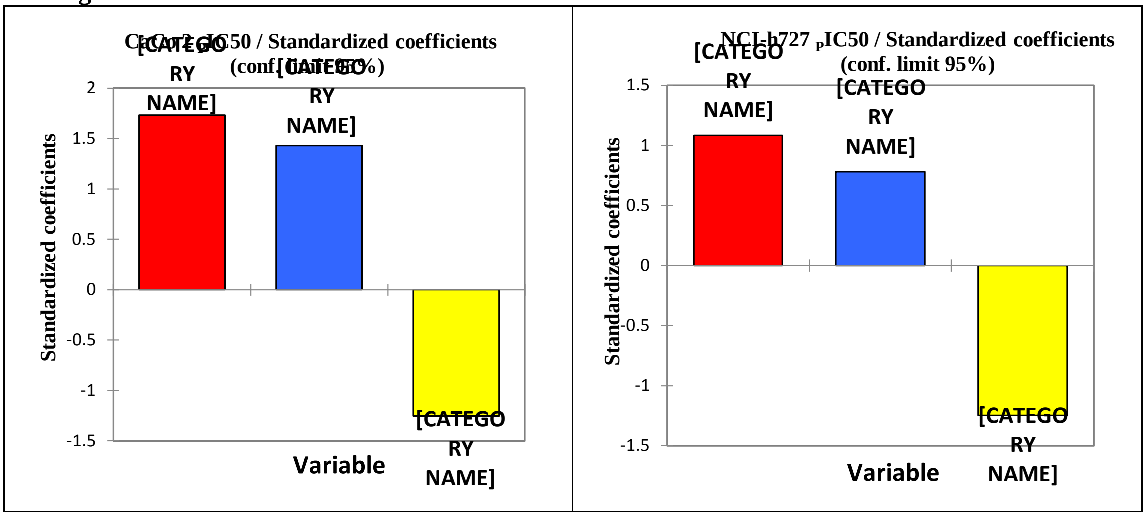


Figure 2: Contribution of the different descriptors in the different models.

The predictive power of the different descriptors with their respective standardized coefficients is classified according to the following sequences:

CaCo2: ΔE > S > VM
NCI-H727: VM > ΔE > S

According to the order of the different contributions of the descriptors, for human colon tumor line (CaCo2), the energy gap (ΔE) is the priority descriptor. For human pulmonary tumor line (NCI-h727), the molecular volume (VM) is the priority descriptor. Thus, the energy gap (ΔE) and molecular volume (VM) are the priority descriptors in predicting the anti-proliferative activity of the bis-(5-arylidene rhodanine-3-yl) diamines studied. The low standard error values of the two cell lines Caco2 and NCI-h727 being 0.033 and 0.025 respectively attest to the good similarity between the predicted and experimental values (**Figure 3**).

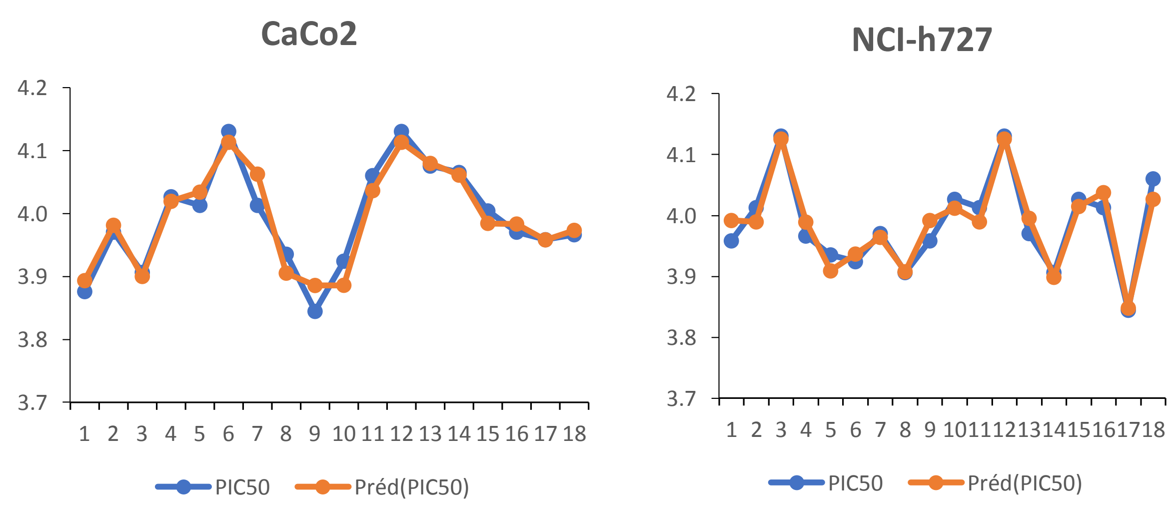


Figure 3: Similarity curves of experimental and predicted model values

These curves reflect a very good analogical evolution of the experimental values and predicted by these models of the anti-proliferative activity of the bis-(5-arylidene rhodanine-3-yl) diamines studied despite some recorded differences.

4. CONCLUSION

This study demonstrated a relationship between the antiproliferative activity pIC50 (M) and the calculated theoretical physico-chemical descriptors. The strong correlations between the calculated values and the experimental cytotoxicity of the two cancer cell strains CaCo2 and NCI-h727 made it possible to identify the energy gap (ΔE), the global softness (S) and the molecular volume (VM) as the descriptors that best influence the cytotoxicity of the compounds studied on the CaCo2 and NCI-h727 lines. The models (MLR) obtained for each cancer cell studied indicate good robustness ($Q^2 > 0.9$): which reflects good stability and excellent predictive power. The QSAR models obtained can predict the activity of new molecules based on the structure of compounds studied and identify descriptors that improve antiproliferative activity to guide the design of new molecules that are more active against cancer cells. For human colon tumor line (CaCo2), the energy gap (ΔE) is the priority descriptor while for the human pulmonary tumor line (NCI-h727), the molecular volume (VM) is the priority descriptor.

REFERENCES

Agrawal, V.K., Sohgaura, R. and Khadikar, P.V. (2002), “QSAR studies on biological activity of piritrexim analogues against pc DHFR”, *Bioorganic & medicinal chemistry*, Vol. 10 No. 9, pp. 2919–2926.

Buha, V.M., Rana, D.N., Chhabria, M.T., Chikhalia, K.H., Mahajan, B.M., Brahmkshatriya, P.S. and Shah, N.K. (2013), “Synthesis, biological evaluation and QSAR study of a series of substituted quinazolines as antimicrobial agents”, *Medicinal Chemistry Research*, Vol. 22 No. 9, pp. 4096–4109.

Chaltterjee, S., Hadi, A.S. and Price, B. (2011), “Regression Analysis by Examples Wiley VCH, New York (2000)”, *Original scientific paper*.

Chattaraj, P.K., Cedillo, A. and Parr, R.G. (1995), “Variational method for determining the Fukui function and chemical hardness of an electronic system”, *The Journal of chemical physics*, Vol. 103 No. 17, pp. 7645–7646.

(2004), *Chemoinformatics*, Springer.

Chhabria, M.T., Mahajan, B.M. and Brahmkshatriya, P.S. (2011), “QSAR study of a series of acyl coenzyme A (CoA). Cholesterol acyltransferase inhibitors using genetic function approximation”, *Medicinal Chemistry Research*, Vol. 20 No. 9, pp. 1573–1580.

Connolly, M.L. (1985), “Computation of molecular volume”, *Journal of the American Chemical Society*, Vol. 107 No. 5, pp. 1118–1124.

COULIBALY, W., Paquin, L., Bénié, A., Bekro, Y.-A., Durieux, E., Meijer, L., Le Guével, R., Corlu, A. and Bazureau, J.-P. (2012), “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*, Vol. 80 No. 4, pp. 825–836.

Csizmadia, I. and Enriz, R.D. (2000), “The role of computational medicinal chemistry in the drug discovery process”, *Journal of Molecular Structure: THEOCHEM*, Vol. 504.

Dennington, R., Keith, T. and Millam, J. (2009), “GaussView, version 5”, *Semichem Inc.: Shawnee Mission, KS*.

Eriksson, L., Jaworska, J., Worth, A.P., Cronin, M.T.D., McDowell, R.M. and Gramatica, P. (2003), “Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs”, *Environmental health perspectives*, Vol. 111 No. 10, pp. 1361–1375.

Esposito, E.X., Hopfinger, A.J. and Madura, J.D. (2004), “Methods for applying the quantitative structure-activity relationship paradigm”, in *Chemoinformatics*, Springer, pp. 131–213.

Free, S.M. and Wilson, J.W. (1964), “A mathematical contribution to structure-activity studies”, *Journal of Medicinal Chemistry*, Vol. 7 No. 4, pp. 395–399.

Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, JR, Scalmani, G., Barone, V., Mennucci, B. and Petersson, G.A. (2013), “Gaussian 09, Revision D. 01, 2013, Gaussian”, *Inc., Wallingford CT*.

Ghanbari, Z., Housaindokht, M.R., Izadyar, M., Bozorgmehr, M.R., Eshtiagh-Hosseini, H., Bahrami, A.R., Matin, M.M. and Khoshkholgh, M.J. (2014), “Structure-activity relationship for Fe (III)-salen-like complexes as potent anticancer agents”, *The Scientific World Journal*, Vol. 2014.

Golbraikh, A. and Tropsha, A. (2002), “Beware of q2!”, *Journal of Molecular Graphics and Modelling*, Vol. 20 No. 4, pp. 269–276.

Hansch, C. and Fujita, T. (1964), “p-σ-π Analysis. A method for the correlation of biological activity and chemical structure”, *Journal of the American Chemical Society*, Vol. 86 No. 8, pp. 1616–1626.

Johnson, B.G., Gill, P.M.W. and Pople, J.A. (1993), “The performance of a family of density functional methods”, *The Journal of chemical physics*, Vol. 98 No. 7, pp. 5612–5626.

Kamchonwongpaisan, S., Quarrell, R., Charoensetakul, N., Ponsinet, R., Vilaivan, T., Vanichtanankul, J., Tarnchompoo, B., Sirawaraporn, W., Lowe, G. and Yuthavong, Y. (2004), “Inhibitors of multiple mutants of Plasmodium f alciparum dihydrofolate reductase and their antimalarial activities”, *Journal of Medicinal Chemistry*, Vol. 47 No. 3, pp. 673–680.

Kapp, J., Remko, M. and Schleyer, P.v.R. (1996), “H2XO and (CH3) 2XO Compounds (X= C, Si, Ge, Sn, Pb). Double Bonds vs Carbene-Like Structures Can the Metal Compounds Exist at All?”, *Journal of the American Chemical Society*, Vol. 118 No. 24, pp. 5745–5751.

Koopmans, T. (1934), “Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den einzelnen Elektronen eines Atoms”, *Physica*, Vol. 1 No. 1-6, pp. 104–113.

Mattioni, B.E. and Jurs, P.C. (2003), “Prediction of dihydrofolate reductase inhibition and selectivity using computational neural networks and linear discriminant analysis”, *Journal of Molecular Graphics and Modelling*, Vol. 21 No. 5, pp. 391–419.

"molinspiration cheminformatics software - Recherche Google", available at: https://www.google.com/search?q=molinspiration+cheminformatics+software&oq=Molinspiration+Software+cheminformatics&aqs=chrome.1.69i57j0.1940j0j7&sourceid=chrome&ie=UTF-8 (accessed 30 June 2019).

Oprea, T.I., Mannhold, R., Kubinyi, H. and Folkers, G. (2005), *Chemoinformatics in drug discovery*, Wiley-VCH.

OUATTARA, O., Thomas Sopi, A., KONE, M.G.-R., BAMBA, K. and Nahossé, Z. (2017), "Can Empirical Descriptors Reliably Predict Molecular Lipophilicity? A QSPR Study Investigation", *Int. Journal of Engineering Research and Application*, Vol. 7 No. 15, pp. 50–56.

Parr, R.G., Donnelly, R.A., Levy, M. and Palke, W.E. (1978), "Electronegativity. The density functional viewpoint", *The Journal of chemical physics*, Vol. 68 No. 8, pp. 3801–3807.

Rekka, E.A. and Kourounakis, P.N. (2008), *Chemistry and molecular aspects of drug design and action*, CRC Press.

Rücker, C., Rücker, G. and Meringer, M. (2007a), "y-Randomization and its variants in QSPR/QSAR", *Journal of chemical information and modeling*, Vol. 47 No. 6, pp. 2345–2357.

Rücker, C., Rücker, G. and Meringer, M. (2007b), "y-Randomization and its variants in QSPR/QSAR", *Journal of chemical information and modeling*, Vol. 47 No. 6, pp. 2345–2357.

SIMON, D. (1998), "Enseignement des modèles de liaison chimique", *Didaskalia (Paris)*.

Sohn, E.J., Li, H., Reidy, K., Beers, L.F., Christensen, B.L. and Lee, S.B. (2010), "EWS/FLI1 oncogene activates caspase 3 transcription and triggers apoptosis in vivo", *Cancer research*, Vol. 70 No. 3, pp. 1154–1163.

Thomas, A.S., Claude, K.A.L. and Adama, O.Z., "Quantitative Structure antifungal Activity Relationship (QSAR) study of a series of Schiff bases derivatives from4-aminobenzenesulphonamide by DFT method".

Tropsha, A. (2010), "Best practices for QSAR model development, validation, and exploitation", *Molecular informatics*, Vol. 29 No. 6-7, pp. 476–488.

Victora, C.G., Requejo, J.H., Barros, A.J.D., Berman, P., Bhutta, Z., Boerma, T., Chopra, M., Francisco, A. de, Daelmans, B. and Hazel, E. (2016), "Countdown to 2015. A decade of tracking progress for maternal, newborn, and child survival", *The Lancet*, Vol. 387 No. 10032, pp. 2049–2059.

Zanuncio, J.C., Mourão, S.A., Martínez, L.C., Wilcken, C.F., Ramalho, F.S., Plata-Rueda, A., Soares, M.A. and Serrão, J.E. (2016), "Toxic effects of the neem oil (Azadirachta indica) formulation on the stink bug predator, Podisus nigrispinus (Heteroptera. Pentatomidae)", *Scientific reports*, Vol. 6, p. 30261.