Science Journal of Chemistry
2023; 11(3): 78-87
http://www.sciencepublishinggroup.com/j/sjc
doi: 10.11648/j.sjc.20231103.12
ISSN: 2330-0981 (Print); ISSN: 2330-099X (Online)

Study of the Quantitative Structure Activity Relationship (QSAR) of a Series of Molecules Derived from Thioureas with Anticancer Activities in the Liver

Doumbia Siriki[1], Dembele Georges Stephane[1, 2], Tuo Nanou Tieba[1, *], Konate Bibata[1, 2], Kodjo Charles[1], Ziao Nahosse[1, 2]

[1]Laboratory of Thermodynamics and Physico-Chemistry of the Environment, UFR SFA, Université Nangui Abrogoua, Abidjan, Ivory Coast
[2]Ivorian Research Group in Disease Modeling (GIR2M), UFR-SFA, Université Nangui Abrogoua, Abidjan, Ivory Coast

Email address:
nanoutuo07@gmail.com (Tuo Nanou Tieba)
[*]Corresponding author

To cite this article:
Doumbia Siriki, Dembele Georges Stephane, Tuo Nanou Tieba, Konate Bibata, Kodjo Charles, Ziao Nahosse. Study of the Quantitative Structure Activity Relationship (QSAR) of a Series of Molecules Derived from Thioureas with Anticancer Activities in the Liver. *Science Journal of Chemistry*. Vol. 11, No. 3, 2023, pp. 78-87. doi: 10.11648/j.sjc.20231103.12

Received: May 9, 2023; **Accepted**: May 27, 2023; **Published**: June 9, 2023

Abstract: The study of the quantitative structure activity relationship (QSAR) of liver cancer was carried out using a series of twenty-five (25) molecules derived from thioureas. The molecular descriptors were obtained after optimization of all these molecules at the B3LYP/6-31+ G (d, p) computational level. The multiple linear regression (MLR) method was used to carry out this study. The use of this method has thus made it possible to obtain a model from the molecular descriptors that are the lipophilicity LogP, the bond lengths d(C=N2) and d(N2-Cphen1), the vibration frequency υ (C =O) and the number of atoms. The results of the statistical indicators obtained from the model (R2=0.906; RMCE=0.198; F= 21.170), allow us to say that this model is acceptable, robust and has good predictive power. Also, the vibration frequency of the carbon-oxygen double bond (C=O), the length of the C-N2 bond and the lipophilicity (LogP) were found to be the priority descriptors in the prediction of the anticancer activity of the liver. Moreover, all the criteria of Tropsha et al. were verified by our model. Moreover, the analysis of the domain of applicability of this model shows that a prediction of the anticancer activity of new derivatives of thiourea is acceptable when its leverage value is less than 1.06, otherwise the anticancer activity of the liver of this compound could not be reliably predicted.

Keywords: QSAR, RML, Thiourea Derivatives, Lipophilia (LogP), Area of Applicability

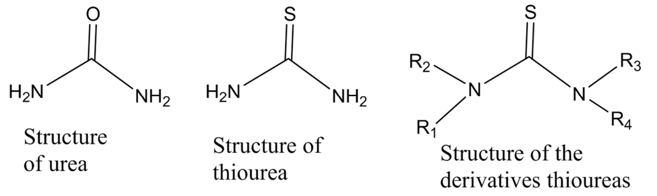
1. Introduction

Liver cancer is the third leading cause of death among cancers worldwide [1]. Africa and Asia remain the two continents with the highest incidence of liver cancer [2]. Treatments include surgery, radiotherapy and chemotherapy, among others. Current treatment modalities including chemotherapy are often accompanied by high toxicity, including normal or healthy cells, and resistance to drugs, even if these are combined. [3-7]. The treatment of cancer is indeed difficult and remains a huge challenge because of a lack of specificity of treatment on malignant cells, without healthy or normal cells being also affected [5]. However, chemotherapy is the most common method and remains less expensive than surgery and radiotherapy. Tumors in the metastatic phase can be easily treated with chemotherapy [8]. It is in this context that there is a constant demand for new anti-cancer treatments and agents [9].

The research, design and development of drugs remains a tedious process strewn with pitfalls and which requires a colossal investment of time and money [10]. The discovery of an actual drug molecule requires about 15 years of research [11]. It is in view of all this that new lines of research based on predictive methods of the activities and properties of molecules have emerged, in particular the QSAR (Quantitative Structure Activity Relationship) methods. These predictive methods have greatly reduced biological tests and facilitated the design of new therapeutic compounds [12]. Nowadays, the development and constant performance of increasingly reliable computer tools has enabled a boom in the use of molecular modeling in drug research and design. The use of alternative methods to experimentation, based on mathematical models, including the Quantitative Structure Activity Relationship or RQSA method, has thus become of great interest because of its many advantages [13].

Thioureas remain one of the most active families of anticancer molecules and constitute excellent building blocks or "scaffolds" in the discovery of new cancer candidates.

They have a variety of therapeutic applications, particularly in cancer treatment [14, 15]. Thioureas are a class of organic compounds with the general formula (R1R2N) (R3R4N) C=S. Their structures are similar to those of ureas except that the sulfur atom is replaced by that of oxygen [16, 14] (Figure 1).



Structure of urea

Structure of thiourea

Structure of the derivatives thioureas

Figure 1. Structures of urea, thiourea and thiourea derivatives.

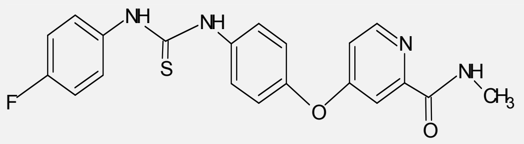
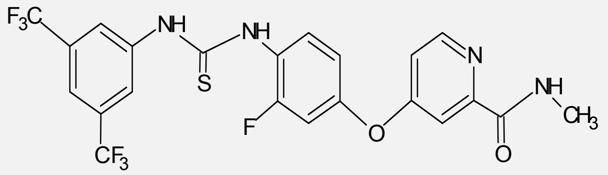
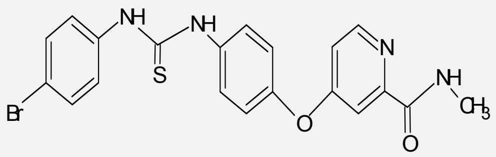
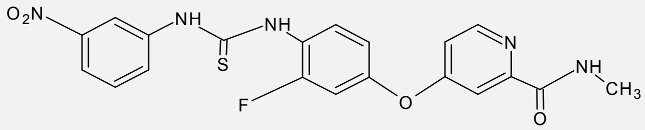
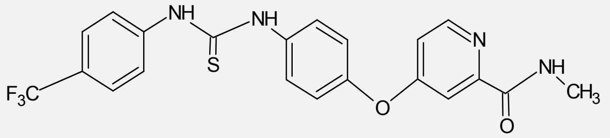
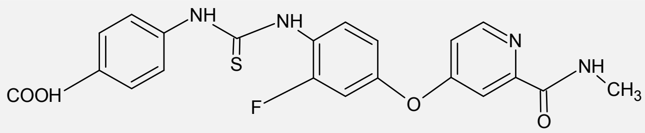
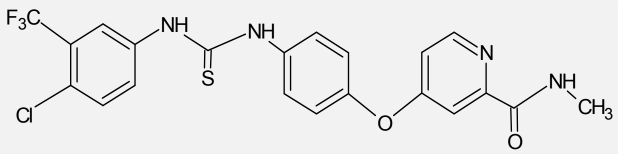
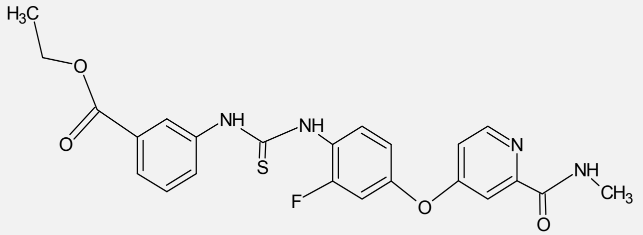
The properties of ureas and thioureas differ significantly due to the difference in electronegativities between oxygen and sulfur [16]. Thioureas and their derivatives are widely used in the medical field as drugs and have shown numerous biological activities. One of the most important applications of thiourea derivatives is their anticancer activity. Many thioureas are used in many cancer therapies and many of them are in the clinical trial phase. Thiourea derivatives are among the most active anticancer drugs that remain effective in cytotoxicity [15, 17]. In the present work, a QSAR study was conducted on a series of thiourea derivatives active as potential anticancer agents using DFT (Density Functional Theory).

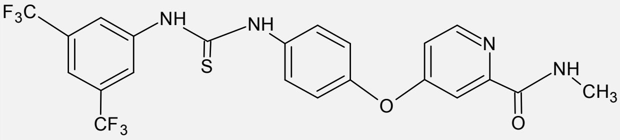
2. Materials and Methods

2.1. Materials

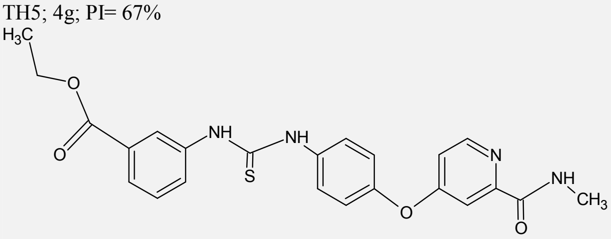
W. Bai *et al.* [18] carried out in vitro tests of a series of molecules of the thiourea family and their derivatives on cancerous cells, among others HepG2 (human hepatoma cell line) in which the percentages of inhibition at 10 μM of anticancer activities, in particular antiproliferative and antitumor. The twenty-five (25) molecules derived from Thioureas the subject of our QSAR study. They were coded from TH1 to TH 25. Their designation codes, numbers and PI inhibition percentages are listed in Table 1.

***Table 1.** 2D Structures, Designation Codes, Numbers and PI Inhibition Percentages Twenty-Five (25) Molecules of Proposed Thiourea Derivatives.*

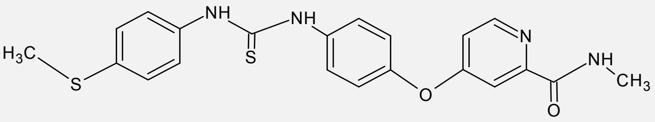
 TH1; 4b; PI=15.6%	 TH 14; 5g; PI=67.8%
 TH2; 4d; PI=23.2%	 TH 15 5h; PI= 23.2%
 TH3; 4e; PI=27.9%	 TH 16 5i; PI=7.8%
 TH4; 4f; PI= 68.8%	



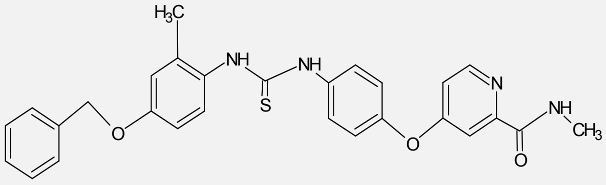
TH5; 4g; PI= 67%



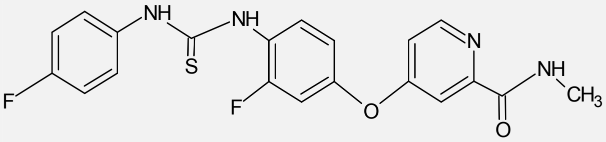
TH6; 4j; PI=21.2%



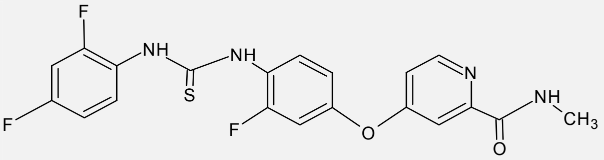
TH7; 4l; PI=16.6%



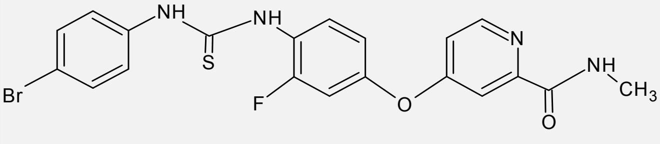
TH 8; 4o; PI=7.6%



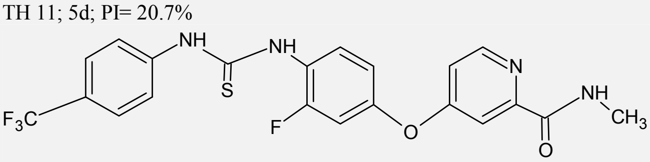
TH 9 5b; PI= 7.3%



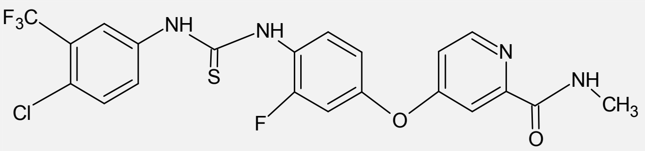
TH 10 5c; PI=7%



TH 11; 5d; PI= 20.7%

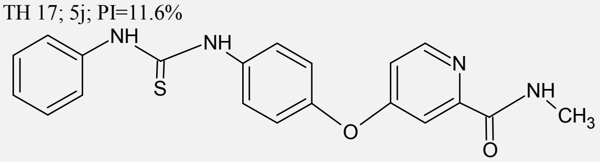


TH 12; 5e; PI=23.5%

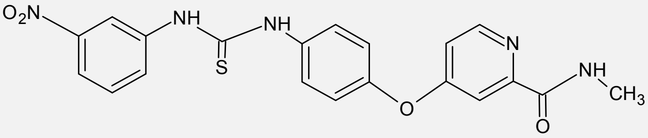


TH 13; 5f; PI=60%

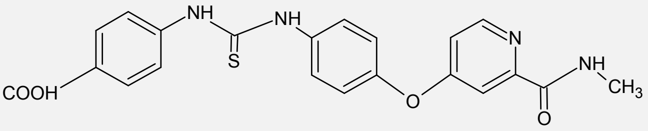
TH 17; 5j; PI=11.6%



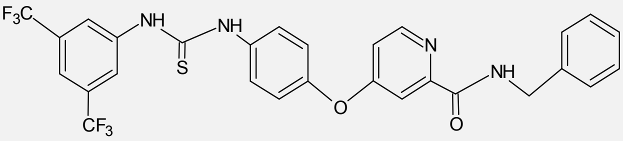
TH 18; 4a; PI=15.6%



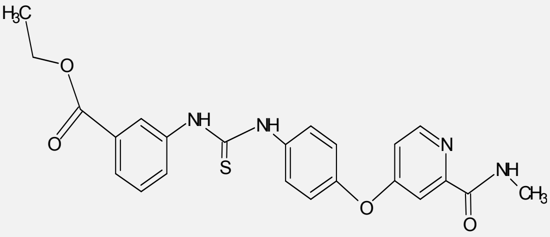
TH 19; 4h; PI=29.5%



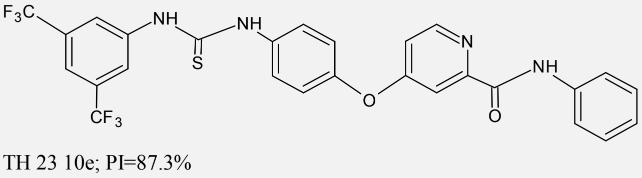
TH 20; 4i; PI= 6.9%



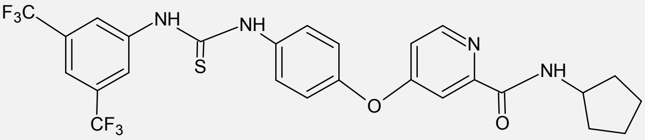
TH 21; 10d; PI=65.8%



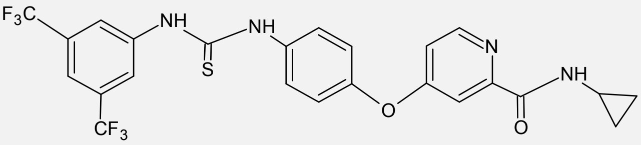
TH 22 4j; PI=21.2%



TH 23 10e; PI=87.3%



TH 24; 10a; PI=57.9%



TH 25; 10b; PI= 49.4%

2.2. Methods

2.2.1. Level of Computational Theory

In QSAR studies, DFT methods are generally able to satisfactorily generate a variety of molecular properties because they have a better predictive capacity by reducing computation times and design costs for new drugs [19-22]. The prediction of anticancer activity of thiourea derivatives quantum chemical calculations were performed using Gaussian 09 software [23] with its GaussView 05 graphical interface. In this work All molecules were optimized and the theory level B3LYP/6-31+G (d, p) was used to determine the molecular descriptors. The modeling was developed using the statistical method of linear multiple regression (RML) which is implemented in *Excel* [24] and *XLSTAT* [25] spreadsheets.

Biological data are usually expressed in logarithm to obtain better mathematical values when the structures are biologically active [26, 27].

Anti-cancer activity activity will be expressed logAAC as a function of the PI inhibition potential defined by equation (1) [28-30]:

$$\text{logAAC} = \log\frac{PI}{(100-PI)} - \log\frac{D}{M} \quad (1)$$

PI: Percentage inhibition

D: Molar volume concentration or absorbed dose

M: Molecular molar mass

2.2.2. Molecular Descriptors Used

Five (05) theoretical molecular descriptors were determined for the development of our QSAR model. These are the LogP lipophilicity, the C=N2 and N2-Cphen1 bond lengths, the vibrational frequency of the υ (C=O) bond, and the number of fluorine atoms F in the molecule.

A lipophilic substance is a substance that "likes oil". Lipophilia is a physico-chemical parameter which therefore expresses the affinity of a molecule for a lipid environment [31]: oil, cell membrane, lipid solvent. It is commonly measured by the distribution of the molecule, neutral, soluble, between water and another immiscible solvent: generally n-octanol (or octan-1-ol) [32-35]. The lipophilicity is evaluated in a practical way from the decimal logarithm value of the partition coefficient logP, and which is equal to the logarithm of the ratio of the concentrations of the substance studied in octanol and in water logP = log (Coct/Ceau). Thus: If logP > 0; then P>1, the molecule is lipophilic. It is soluble in the lipid phase: it is then non-polar. If logP < 0 then P<1, the molecule is hydrophilic, it is soluble in water: it is then polar. Molecular lipophilia occupies a prominent place in the study of chemical substances with medicinal properties [36]. It is the most important physicochemical descriptor for the biological activity of a drug appearing in 70% of QSAR methods [33]. Lipophilia is involved in determining the pharmacokinetic and pharmacodynamic properties as well as the bioavailability and permeability of drugs [34, 36]. According to Gnewuch et *al* [37], logP gives a measure of the infiltration of a drug through cell membranes and its subsequent migration into the cell nucleus. For these authors; If logP < 0 the molecule is too hydrophilic, it has poor lipid, and therefore membrane, permeability. If 0 < logP < 3 the molecule has biological activity, good permeability and solubility. If logP > 0 the molecule has high solubility and good penetration into cell membranes but low aqueous solubility and cannot be transported by blood plasma. In this work, the Chemsketck software [38] allowed us to determine the logP values.

Bond lengths are geometric descriptors that can be determined from the relative positions of a molecule in space and require knowledge of the 3D structure of the molecule. This can be obtained experimentally but most often by empirical or ab initio molecular modeling, thus requiring a certain calculation time [39]. The bond lengths d(C-N2) and d(N2-Cphen1) and the vibration frequency of the carbon-oxygen bond υ (C=O) used in our work were obtained using Gaussian 09 software [23] with its GaussView 05 graphical interface and are illustrated in Figure 2 below.

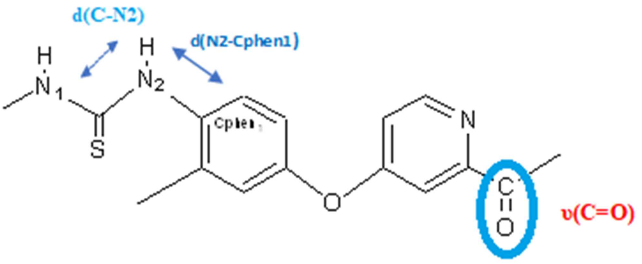


Figure 2. *Bond lengths d(C-N2) d(N2-Cphen1) and bond vibration frequency υ (C=O).*

The number of fluorine F atoms in the molecule is a constitutional molecular descriptor. The number of atoms in a molecular system remains the simplest descriptor to determine or represent because it does not take into account any geometric or electronic consideration.

2.2.3. Estimating the Predictive Ability of a QSAR Model

The quality of a QSAR model, once developed: is linked to its reliability, its robustness and its predictive nature, must be established using statistical analysis criteria including the coefficient of determination R^2, the standard deviation (S) or the Root Mean Square Error (RMCE), the cross-validation and Fischer correlation coefficients F, R^2, S and F relate to the adjustment 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 [27, 40]. The cross-validation coefficient provides information on the predictive power of the model; which 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. Thus, the quality of the modeling is better when the points are close to the fitted line [41]. The adjustment of the points to this line can be evaluated by the coefficient of determination. These different parameters are calculated according to the following expressions:

The Coefficient of determination R^2

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

With:

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

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

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

The Root Mean Square RMCE Error is another statistical indicator used. It is used to assess the reliability and accuracy of a model:

$$\text{RMCE} = \sqrt{\frac{\sum\left(y_{i,exp} - y_{i,théo}\right)^2}{n-k-1}} \tag{3}$$

The closer the value of R² is to 1, the more the theoretical and experimental values are correlated.

The Fisher F test is also used to measure the level of statistical significance of the model, ie the quality of the choice of descriptors constituting the model.

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

The coefficient of determination of the cross-validation makes it possible to evaluate the accuracy of the prediction on the learning set. It is calculated using the following relationship:

$$Q_{cv}^2 = \frac{\sum\left(y_{i,théo} - \bar{y}_{i,exp}\right)^2 - \sum\left(y_{i,théo} - y_{i,exp}\right)^2}{\sum\left(y_{i,théo} - \bar{y}_{i,exp}\right)^2} \tag{5}$$

2.2.4. Model Acceptance Criteria

According to Eriksson et *al.* [42], The performance of a mathematical model is characterized by a value of Q_{cv}^2 ; $Q_{cv}^2 > 0.5$ for a satisfactory model, and $Q_{cv}^2 > 0.9$ when for the excellent model. 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 respected.

According to Tropsha et *al.* [43-46], 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) $Q_{Cv\,Test}^2 > 0.6$,

3) $|R_{Test}^2 - R_0^2| \leq 0.3$,

4) $\frac{|\mathrm{R}_{\mathrm{Test}}^2 - \mathrm{R}_0^2|}{\mathrm{R}_{\mathrm{Test}}^2} < 0.1$ et $0.85 \leq k \leq 1.15$,

5) $\frac{|R_{Test}^2 - R'^2_0|}{R_{Test}^2} < 0.1$ et $0.85 \leq k' \leq 1.15$.

2.2.5. Domain of Applicability (DA)

The domain of applicability (DA) of a QSAR model is the physico-chemical, structural or biological space, in which the model equation can be applied in order to be able to make predictions for new compounds [47]. It is the region of the chemical space which includes the compounds of the training set and the similar or analogous compounds, which are close in this same space [48]. It therefore appears necessary and even mandatory to determine the domain of applicability (DA) of a QSAR model. It is a method based on the variation of the standardized residuals of the dependent variables with the distance between the values of the descriptors and their, called levers. These levers constitute the elements of a matrix H called hat matrix, which is a projection of the experimental values of the explained variable in the space of values of the predicted variable as follows

$$Y_{préd} = HY_{expé} \tag{6}$$

is given by expression (7):

$$H = X(X^tX)^{-1}X^t \tag{7}$$

The value of its standardized residual of Any compound that falls within the scope of the model included in the interval $[-3\sigma\,;+3\sigma\,]$., where σ is the standard deviation.

The critical value of the lever (h*) is fixed at:

$$h^* = \frac{3\,(k+1)}{n} \tag{8}$$

Where n is the number of test compounds used; k is the number of model descriptors.

3. Results and Discussion

Our QSAR study was conducted on a series of twenty-five (25) molecules derived from Thioureas active as potential anticancer agents using DFT by the B3LYP method in 6-31+G (d, p). Their HepG2 liver anticancer activities are given by the decimal logarithm Log AAC HepG2 calculated from their Inhibition percentages at 10µM. These thiourea derivatives were grouped into two groups, seventeen (17) were used for the learning game and eight (08) for the validation game. The modeling of the anticancer activity was carried out from five descriptors which are the lipophilic LogP, the bond lengths d(C=N2) and d(N2-Cphen1), the vibration frequency υ (C=O) and the number of fluorine F atoms in the molecule. The values of the descriptors as well as those of the experimental biological activities of the compounds are recorded in Table 2.

Table 2. *Experimental physicochemical and log AAC HepG2 descriptors of the training and validation sets.*

Molecules	Log P	d(C=N2) (Å)	d(N2-Cphen1) (Å)	υ (C=O) (cm^{-1})	F	logAAC HepG2
	Learning Set					
TH1	3.580	1.374	1.417	1734.860	1.000	6.881
TH2	4.250	1.371	1.422	1730.660	0.000	7.140
TH3	4.350	1.372	1.418	1733.790	4.000	7.255
TH4	4.900	1.376	1.422	1742.020	3.000	8.025
TH5	5.270	1.375	1.423	1741.780	6.000	8.019

Molecules	Log P	d(C=N2) (Å)	d(N2-Cphen1) (Å)	υ (C=O) (cm^{-1})	F	logAAC HepG2
TH6	3.580	1.374	1.417	1735.500	0.000	7.084
TH7	3.860	1.376	1.417	1732.220	0.000	6.927
TH8	5.520	1.377	1.423	1728.770	0.000	6.613
TH9	3.740	1.379	1.420	1742.290	2.000	6.514
TH10	3.900	1.380	1.421	1742.140	3.000	6.513
TH11	4.410	1.380	1.419	1742.450	1.000	7.094
TH12	4.500	1.380	1.421	1742.650	4.000	7.154
TH13	5.060	1.380	1.410	1742.320	4.000	7.874
TH14	5.420	1.376	1.421	1742.290	7.000	8.050
TH15	3.470	1.377	1.423	1742.410	1.000	7.125
TH16	3.410	1.379	1.420	1742.140	1.000	6.571
TH17	3.740	1.380	1.420	1742.190	1.000	6.789
	Validation Set					
TH18	3.420	1.375	1.418	1735.490	0.000	6.845
TH19	3.340	1.376	1.422	1741.980	0.000	7.248
TH20	2.980	1.378	1.421	1741.820	0.000	6.496
TH21	7.000	1.376	1.423	1734.140	6.000	8.055
TH22	3.580	1.381	1.419	1742.430	1.000	6.170
TH23	6.930	1.375	1.423	1740.360	6.000	8.598
TH24	6.400	1.375	1.424	1734.000	6.000	7.893
TH25	5.560	1.375	1.423	1743.210	6.000	7.722

3.1. Interdependence of Descriptors

The different descriptors used for the model must be independent of each other. This interdependence of these descriptors is measured using the partial correlation coefficients aij, The values of the partial correlation coefficients aij in Table 3.

Table 3. *Correlation matrix between the different physico-chemical descriptors.*

Variables	Log P	d(C=N2)	d(N2-Cphen1)	υ (C=O)	F
Log P	1.000	-0.037	0.125	-0.045	0.589
d(C=N2)	-0.037	1.000	-0.124	0.664	0.059
d(N2-Cphen1)	0.125	-0.124	1.000	0.037	0.033
υ (C=O)	-0.045	0.664	0.037	1.000	0.504
F	0.589	0.059	0.033	0.504	1.000

The partial correlation coefficient aij contained in Table 3 between the pairs of descriptors is less than 0.7 (aij < 0.70). This result reflects the independence of the descriptors used to develop the model.

3.2. Multiple Linear Regression (MLR)

Equation (8) represents the determined QSAR model equation.

$$logAAC\ (HepG2) = 114.86329 + 0.68393 * Log\ P - 173.87212 * d(C - N2) - 59.89 * d(N2 - Cphen1) + 0.12305 * \upsilon\ (C = O) - 0.08679 * F \tag{9}$$

The negative signs of the coefficients of the C-N2 and N2-Cphen1 bond lengths and of the number of fluorine atoms F reflect the fact that the anticancer activity is improved for small values of these different molecular descriptors. On the other hand, this anticancer activity is improved for lipophilicity (LogP) and the vibration frequency of the C=O bond because of the positive sign of their respective coefficients. The statistical indicators of the model are given in Table 4.

Table 4. *RML model statistical analysis report.*

Number of observations N	17
Coefficient of determination R^2	0.9059
Standard deviation RMCE	0.198
Fisher test F	21.170
Coefficient of determination of the cross-validation Q_{CV}^2	0.8851
R^2-Q_{CV}^2	0.021
Confidence Level α	> 95%

Table 5. *Tropsha criteria checks of the RML model external validation set.*

Statistical Parameters	Tropsha criteria checks [46-49]	
R^2	> 0.7	0.9059
Q_{CV}^2	> 0.6	0,8851
$\|R^2 - R_0^2\|$	≤ 0.3	0.0001
$\frac{\|R^2 - R_0^2\|}{R^2}$	< 0.1	0.0001
k	$0.85 \leq k \leq 1.15$	1,0162
$\frac{\|R^2 - R_0'^2\|}{R^2}$	< 0.1	0.0001
k'	$0.85 \leq k' \leq 1.15$	0.9831

The value of the coefficient of determination which is 0.9059 indicates that the estimated values of logAAC HepG2 contain practically 90.1% of the experimental values. The standard deviation (RMCE = 0.198) expresses the small variation of the predicted values with respect to the experimental mean. The value of the Fisher test which 21.170 is relatively high compared to the critical value Fcr = 4.74

[49]. This shows that the error made is less than what the model explains [49]. For this model, the cross-validation correlation coefficient is equal to = 0.8851 and =0.9059-0.8851= 0.021 < 0. 3.. These different values reflect a satisfactory model in accordance with the acceptance criterion according to Eriksson et *al.* [42], the verifications of the Tropsha criteria are recorded respectively in Table 5.

The values in Table 4 show that all the Tropsha criteria are met. All these statistical indicators clearly show that the model developed explains the anticancer activity of thiourea derivatives in a statistically significant and satisfactory manner. These different results are confirmed by the regression graph of the RML model showing the theoretical anticancer activity as a function of the experimental activity shown in Figure 3.

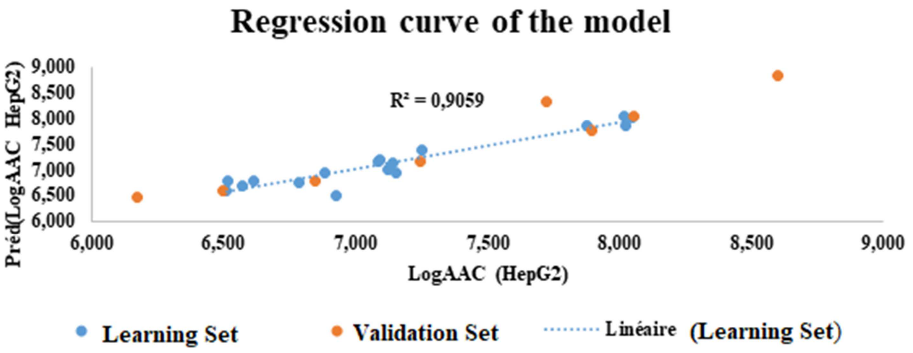


Figure 3. *The regression line of the RML model.*

The regression curve of the RML model, after analysis, shows that all the points are around the regression curve of the RML model, confirmed by a low value of the difference RMCE = 0.198 between the values of Log AAC (HepG2) experimental and that of theoretical Log AAC (HepG2). This good similarity between these values is illustrated by the similarity curve of the model in Figure 4.

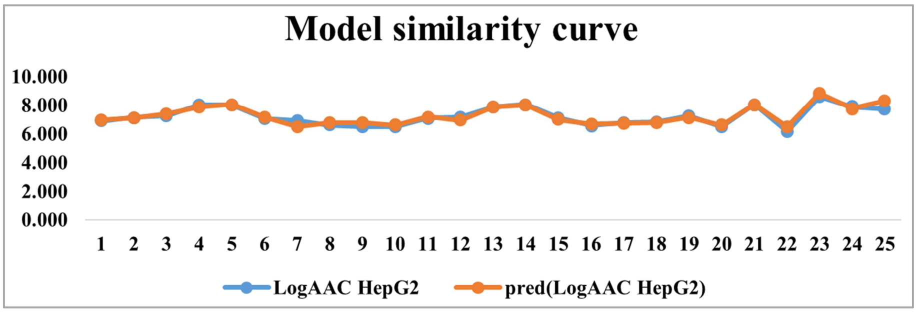


Figure 4. *The regression line of the RML model.*

In view of all these observations, our model presents a good reliability with a satisfactory predictive power on all the compounds of the learning set. Our model is acceptable and suitable for predicting the anticancer activity of our series of thiourea derivatives.

3.3. Contribution of Model Descriptors

The predictive power of this model depends on five theoretical descriptors. However, these different descriptors do not have the same weight in this activity. It is important to determine the contribution of each of these theoretical descriptors to the anticancer activity of molecules that may belong to the chemical space of application. The determination of the contribution makes it possible to define an order of priority of the descriptors thus facilitating the arbitration at the level of the possible choice of the parameters likely to be modified to achieve an optimal activity. The values obtained are represented through the diagram of figure 5.

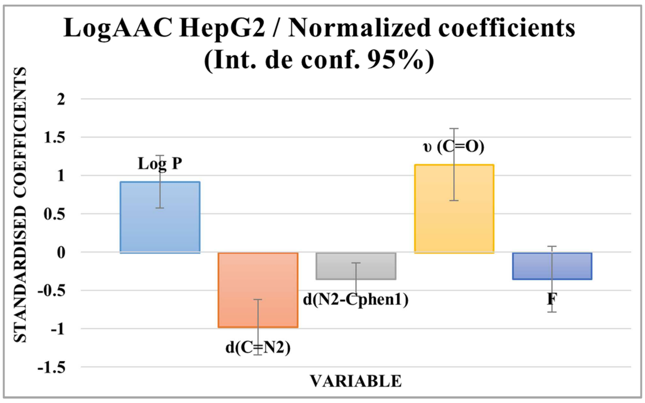


Figure 5. *Normalized coefficients of Anticancer Activity Model descriptors.*

According to this graph, the importance of the weight of the descriptors involved in our model decreases in absolute value in the following order υ (C=O) >d(C-N2) >logP> F > d

(N2 –Cphen1). The contribution analysis shows that the vibration frequency υ (C=O) is the significant descriptor in the prediction of the anticancer activity of the liver of our model. The weights of the two other descriptors, namely the length of the C-N2 bond remain just as preponderant. A high lipophilicity value also improves the activity according to. However, high logP values are a sign of high lipid solubility and good penetration into cell membranes, but this implies low aqueous solubility. This can impair the transport of the drug by the blood plasma. According to their negative signs of the coefficient, the other parameters a decrease in the lengths of C-N2 and N2 –Cphen1 bonds and the number of fluorine F atoms in the molecule increases the anticancer activity. The bond length between two atoms, which is inversely proportional to the bond energy, depends, in all rigor, of the molecule in which these are found. It depends, in the case of organic compounds, on various factors such as the hybridization of the orbitals and the electronic and steric nature of the substituents. In practice, reducing the number of fluorine atoms in the molecule seems to be the easiest operation to perform among these three molecular descriptors in order to improve the anticancer activity of thiourea derivatives according to the model developed.

3.4. Model Applicability Domain (DA)

The domain of applicability (DA) of our model was obtained using the method of levers. This method consists of analyzing through a diagram called the Williams diagram. In this diagram is represented the variation of the standardized prediction residuals according to the values of the levers *hi* of each of the compounds. These compounds having its activity predicted by the model. Outliers are those whose values differ significantly from the patterns and trends of other values within the series. As for the influential compounds, their values of the variation of the standardized prediction residuals are outside the limit set by the critical value h* symbolized by the line parallel to the axis of the variation of the standardized prediction residuals. Any compound that falls within the scope of the model must have the value of its standardized residual included in the interval [-3σ ; $+3\sigma$].

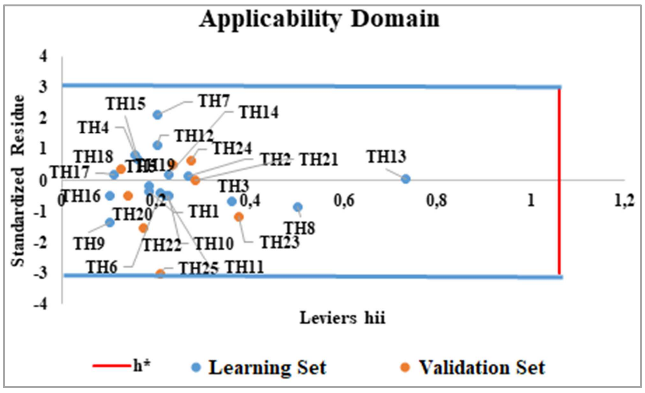


Figure 6. Williams diagram of standardized residues according to the levers of the compounds used.

Figure 6 presents the Williams diagram which represents the variation of the prediction residuals according to the levers of the compounds where the points in blue represent the samples of the learning set and in brown the samples of the validation set.

An analysis of this diagram shows that all the samples of the training games all have levers below the critical value. In this series, there are no aberrant or influential compounds. Thus, all the compounds that fall within the scope of the model all have the value of their standardized residue included in the interval of standardized residue limits −3 and +3 of their standard deviation. All compounds therefore have their levers below the threshold lever.

4. Conclusion

Our work consisted in establishing a relationship between the anticancer activity of the liver and their physicochemical properties characterized by molecular descriptors. Five molecular descriptors, namely log P lipophilicity, C-N2 and N2-Cphen1 bond lengths, C=O vibrational frequency and the number of fluorine atoms in the different compounds allowed us to explain the anticancer activity of the liver. Multiple linear regression (MLR) was used as a method to establish our mathematical model and provided the values of the statistical indicators of the model (R^2=0.9059; RMCE=0.198; F=21.170). These values clearly show that our model is acceptable, robust and has a good predictive power of the anticancer activity of the liver studied. The analysis of the contributions made it possible to show that the vibration frequency of the carbon-oxygen bond (C=O), the length of the C-N2 bond and the lipophilicity (LogP) appear as the significant descriptors in the prediction of the anticancer activity of the liver of our model. An increase in the vibrational frequency of the C=O bond and reasonable lipophilicity, with a decrease, among others, in the number of fluorine atoms in the molecule improve the anticancer activity of the liver according to our model. The analysis of the domain of applicability of this model shows that a prediction of the anticancer activity of new thiourea derivatives is acceptable when its leverage value is less than 1.06, otherwise the anticancer activity of the liver of this compound does not could be reliably predicted. Any proposal for molecules aimed at improving the anticancer activity of these thiourea derivatives should take this into account.

References

[1] V. A. Arzumanian, O. I. Kiseleva and E. V. Poveranna, The Curious Case of The HepG2 cell line: 40 years of Expertise; *Int. J. Mol Sci.*, 2021. Dec 4, 22 (23): 13135.

[2] J. Thusyan, N. S. Wickramatne, K. H., I. Thabrew, S. R. Samarakoon, Cytotoxicity against Human Hepatocellular Carcinoma (HepG2) cells and Antioxydant Activity of Selected Endemic or Medicinal Plants in Sri Lanka, *Adv. Pharmacol. Pharm. Sci.*, 2022, Mar 30, 2022, 6407688974.

[3] K. D. Miller, L. Nogueira, A. B. Mariotto, J. H. Rowland, K. R. Yabroff, C. M. Alfano, A. Jemal, J. L. Kramer, Cancer treatment and survivorship statitics, *C. A. Cancer J. Clin*., 2019, 69 (5), 363-385.

[4] A. Mahapatra, T. Prasad and T. Sharma, Pyrimidine: a review on anticancer activity with key emphasis on SAR, *Fut. J. of Pharm. Sci*., 2021, 7, 123.

[5] N. M. Ahmed, M. M. Youns, M. K. Soltan and A. M. Said, Design, Synthesis, N. M. Ahmed, M. M. Youns, M. K. Soltan and A. M. Said, Design, Synthesis, Molecular Modeling and Antitumor Evaluation of Novel Indolyl-Pyrimidine Derivatives with EGFR Inhibitor Activity. *Molecule*s, 2020, 26, 1838.

[6] H. Zhuang, W. Jiang, W. Cheng et al., Down-regulation of HSP27 sensitizes TRAIL-resistant tumor cell of TRAIL-induce apoptosis, *Lung Cancer*, 2010, 68, 27-38.

[7] L. C. Hamming, B., J. Slotmaman, H. M. W. Verheul, Angiogenesis, 2017, 20, 217-232.

[8] S. Saeed,; N. Rashid,;, P. G Jones.; M. Ali,; R. Hussain,; *Eur. J. of Med; Chem*.. 2010, 45, 1323-31.

[9] S. Fortin, Molecular modeling, chemical synthesis, evaluation of antiproliferative activity and determination of mechanism of action of novel hybrid arylchloroethylurea and 2-imidazolidone derivatives, Université de Laval Québec, 2010 p 4.

[10] B. Werth, The billion dollard molecule, one company's quest for the perfect drug, 1995, 1.

[11] M. R. Yadav New drug discovery: Where are you heating to ?, *J. Adv. Pharm. Tech. Res.* 2020, 4 (1) 2.

[12] R. Hmamouchi, M. Bouachrine, T. Lakhlifi, Pratique de la Relation Quantitative Structure Activité/Propriété (RQSA / RQSP), *Rev. Interdisc.*, 2016, 1 (1).

[13] M. Ghamali, S. Chiita, M. Bouachrini, T. Lakhlifi, General methodology of a RQSA/RQSP study, *Rev. Interdisc.*, 2016, 1 (1).

[14] V. Kumar, S. Chimni, Recent Developments on Thiourea Based Anticancer Chemotherapeutics, *Anticancer Agents Med*. Chem., 2015, 15, 163-175.

[15] A. Shakeel, A. A. Altaf, A. M. Qureshi, A. Badshah, Thiourea Derivatives in Drug Design and Medicinal Chemistry: A. Short Review, *J. of Drug Med. Chem*., 2016, 2, 10-20.

[16] N. A. Meanwell, Symposis of some Recent Tactical Application of Bioisosteres in Drug Design, *J. Mod. Chem*., 2011, 54, 2539-2591.

[17] S. Saeed, N. Rachid: M. Ali, R. Hussain, P. G. Jones, *Eur. J. of Chem*., 2010, 1 (3), 221-227.

[18] W. Bai, J. Ji, Q. Huang, Synthesis and evaluation of new thiourea derivatives as antitumor and antiangiogenic agents, *Tetrahed. Lett.*, 2020, 61, 15236,

[19] P. K. Chattaraj, A. Cedillo et R. G. Parr, *J. Phys. Chem.,* 1991, 103, 7645.

[20] P. Ayers et M. Levy, «Density Functional Approach to the Frontier-Electron Theory of Chemical Reactivity,» *Theorical Chemistry Accounts-springer,* 2000, 103 (13-4), 353-360.

[21] C. Hansch, P. G. Sammes et J. B. Taylor, «in: Comprehensive Medicinal Chemistry,» *Computers and the medicinal chemist,* 1990, 4, 33-58.

[22] R. Franke, «Theoretical Drug Design Methods,» *Elsevier,* 1984.

[23] M. J. Frisch, G. W. Trucks, H. B. Schlegel, et al. Gaussian 09, Inc., Wallingford CT, 2009.

[24] Microsoft Excel, «(15.0.4420.1017) MSO (15.0.4420.1017) 64 Bits,» Microsoft Office Professionnel, 2016.

[25] XLSTAT, « XLSTAT and Addinsoft are Registered Trademarks of Addinsoft,» Copyright Addinsoft, 2014.

[26] Tammo, Theoretical Analysis of Molecular Membrane Organization, B. Raton, Éd., Florida: CRC, 1995.

[27] G. W. Snedecor et W. G. Cochran, Methods Statistical, India: Oxford and IBH: New Delhi, 1967, p. 381.

[28] A. Nayyar, A. Malde, R. Jain, and E. Coutinho, 3D-QSAR study of ring-substituted quinoline class of anti-tuberculosis agents, *Bio. & Med. Chem*., 2006, 14, 847-856.

[29] A. Manvar, A. Malde, J. Verma, V. Virsodia, A. Mishra, K. Upadhyay, H. Acharya, E. Coutinho, A. Shah, Synthesis, anti-tubercular activity and 3D-QSAR study of coumarin-4-acetic acid benzylidene hydrazides, Eur. J. of Med. Chem., 2008, 43, 2395-2403.

[30] M. Song, M. Clark, Development and Evaluation of an in Silico Model for hERG Binding, *J. Chem. Inf. Model.* 2006, 46, 392-400.

[31] E. Rutkowska, K. Pajak and K. Jozwiak, *Lipophilicity - Methods of Determination and its Role in Medicinal Chemistry, Act. Pol. Pharm. - Drug Res*., 2013, 70 (1), 3-18,

[32] A. Cozma, V. Zaharia, A. Ignat, S. Gocan et N. Grinberg, Prediction of the Lipophilicity of Nine New Synthesized Selenazoly and Three Aroyl–Hydrazinoselenazoles Derivatives by Reversed-Phase High Performance Thin-Layer Chromatography, J. of Chrom. Sci., 2012; 50, 157– 161.

[33] R. Mannhold, G. I. Poda, C. Ostermann, I. V. Tetko, Calculation of Molecular Lipophilicity: State-of-the-Art and Comparison of LogP Methods on More Than 96,000 Compounds, *J. of Pharm. Sci*., 2009, 98 (3), 861-893.

[34] M. A. Bakht, M. F. Alajmi, P. Alam, A. Alam, P. Alam, T. M. Aljarba, Theoretical and experimental study on lipophilicity and wound healing activity of ginger compounds, *Asian Pac. J. of Trop. Biomed*. 2014; 4 (4): 329-333.

[35] J. Kujawski, H. Popielarska, A. Myka, B. Drabińska, M. K. Bernard, The logP Parameter as a Molecular Descriptor in the Computer-aided Drug Design - an Overview, *Comp. Meth. In Sci. And Techn*. 2012, 18 (2), 81-88.

[36] J. Dearden and A. Worth, In Silico Prediction of Physicochemical Properties, JRC Sci. and techn. Rep., 2015, 19-23.

[37] Gnewuch CT, Sosnovsky G (2002). Critical appraisals of approaches for predictive designs in anticancer drugs. Cell Mol Life Sci. 59: 959-1023.

[38] acdlabs, *Advanced Chemistry Development / Chemskecht,* 1994-2010.

[39] Guillaume Fayet, Development of a QSPR model for predicting the explosive properties of nitroaromatic compounds, Doctoral thesis, Université Pierre and Marie Curie. 30 Mars 2010, p 63.

[40] N. J.-B. Kangah, M. G.-R. Koné, C. G. Kodjo, B. R. N'guessan, S. A. Kablan, Yéo et N. Ziao, «Antibacterial Activity of Schiff Bases Derived from Ortho Diaminocyclohexane, Meta-Phenylenediamine and 1,6-Diaminohexane: Qsar Study with Quantum Descriptors,» *Int. J. of Pharm. Sci. Inv.,* 2017. 6 (113), 38-43.

[41] E. X. Esposito, A. J. Hopfinger et J. D. Madura, «Methods for Applying the Quantitative Structure-Activity Relationship Paradigm» *Met. in Mol. Bio.* 2004*,* vol. 275, pp. 131-213.

[42] L. Eriksson, J. Jaworska, A. Worth, M. D. Cronin, R. M. Mc Dowell et P. Gramatica, «Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs,» *Environmental Health Perspectives,* 2003*,* 111,(110), 1361-1375.

[43] K. Roy et al. A Primer on QSAR/QSPR Modeling, Chapter 2 Statistical Methods in QSAR/QSPR, *Springer. Briefs in Mol. Sci*., 2015, pp 37-59.

[44] R. Veerasamy, H. Rajak, A. Jain, S. Sivadasan, C. P. Varghese et R. K. Agrawal, Validation of QSAR Models - Strategies and Importance International *J; of Drug Des; and Disc;*, 2011, 2 (3), 511-519.

[45] M. Shahlaei. Descriptor Selection Methods in Quantitative Structure−Activity Relationship Studies: A Review Study, *Chem. Rev., Am. Chem. Soc. Public*., 2013.

[46] T. M. Martin, P. Harten, D. M. Young, E. N. Muratov, A. Golbraikh, H. Zhu and A. Tropsha. Does Rational Selection of Training and Test Sets Improve the Outcome of QSAR Modeling? J. Chem. Inf. Model. 2012, 52, 2570-2578,

[47] N. N.-Jeliazkova and J. Jaworska. *An Approach to Determining Applicability Domains for QSAR Group Contribution Models: An Analysis of SRC KOWWIN*, *ATLA*, 2005, 33, 461–470.

[48] F. Sahigara, K. Mansouri, D. Ballabio, A. Mauri, V. Consonni and R. Todeschini, Comparison of Different Approaches to Define the Applicability Domain of QSAR Models, *Mol.* 2012, 17, 4791-4810.

[49] A. Fortuné, «Molecular Modeling Techniques applied to the Study and Optimization of Immunogenic Molecules and Chemoresistance Modulators. Drugs,» 2006.