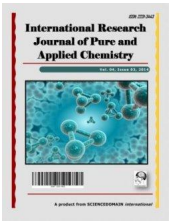


International Research Journal of Pure & Applied Chemistry

16(2): 1-13, 2018; Article no.IRJPAC.40895
ISSN: 2231-3443, NLM ID: 101647669

Prediction of the Inhibitory Concentration of Hydroxamic Acids by DFT-QSAR Models on Histone Deacetylase 1

Doh Soro[1], Lynda Ekou[1], Mamadou Guy-Richard Koné[1*], Tchirioua Ekou[1], Sopi Thomas Affi[1], Lamoussa Ouattara[1] and Nahossé Ziao[1]

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

Authors' contributions

This work was carried out in collaboration between all authors. Author DS designed the study, performed the statistical analysis, wrote the protocol and first draft of the manuscript. Authors LE, TE, MGRK and NZ managed the analyses of the study. Authors TA and LO managed the literature searches. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/IRJPAC/2018/40895
Editor(s):
(1) Farzaneh Mohamadpour, Department of Organic Chemistry, University of Sistan and Baluchestan, Iran.
Reviewers:
(1) Nobuaki Tanaka, Shinshu University, Japan.
(2) B. C. Revanasiddappa, Nitte-Deemed to be University, India.
Complete Peer review History: http://www.sciencedomain.org/review-history/24085

Original Research Article

Received 27th January 2018
Accepted 2nd April 2018
Published 10th April 2018

ABSTRACT

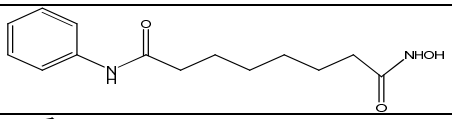
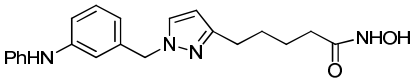
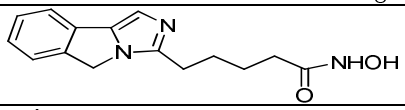
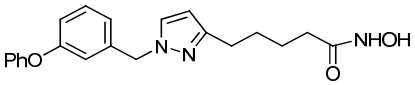
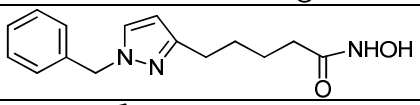
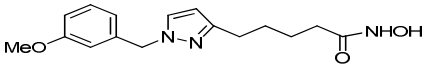
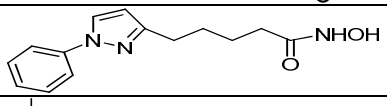
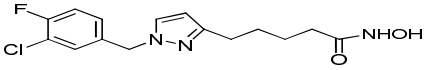
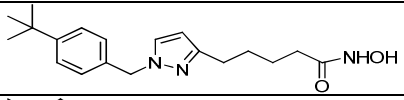
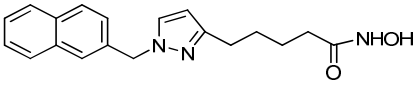
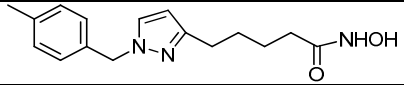
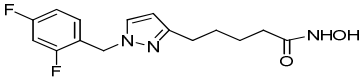
In order to study the relationship between inhibitory concentration and the molecular structures of hydroxamic acids, a Quantitative Structure Activity Relationship (QSAR) study is applied to a set of 31 histone deacetylase inhibitors (HDACi). This study is performed by using the Principal Component Analysis (PCA) method, the Ascendant Hierarchical Classification (AHC), the Linear Multiple Regression Method (RML) and the nonlinear regression (RMNL). Multivariate statistical analysis allowed to obtain two quantitative models (RML model and RMNL model) by the means of the quantum descriptors those are the dipole moment (μ), the bond length d(C=O) and the valence angles α°(O=C-N) and α°(H-N-O). The RMNL model gives statistically significant results and shows a good predictability R^2 = 0.967, S = 0.379 and F = 557.031. The valence angle α°(O=C-N) is the priority descriptor in the prediction of the inhibitory concentration of the studied hydroxamic acids. The obtained results show that geometric descriptors could be useful for predicting the inhibitory concentration of histone deacetylase inhibitors.

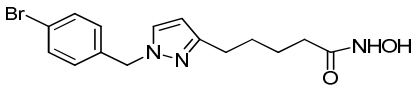
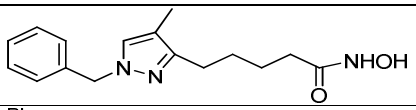
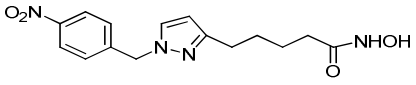
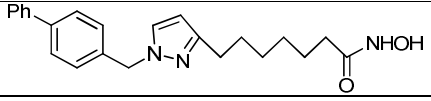
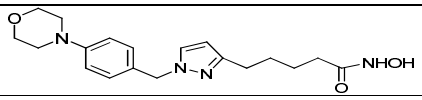
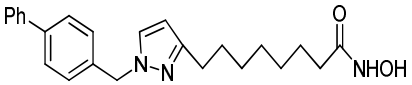
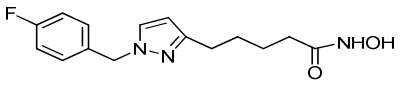
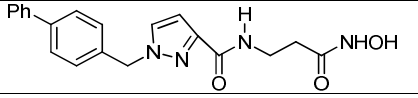
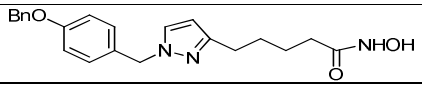
**Corresponding author: E-mail: guyrichardkone@gmail.com, guyrichardkone@gamail.com;*

Keywords: Histone Deacetylases; QSAR; hydroxamic acid; DFT method.

1. INTRODUCTION

Histone Deacetylases (HDACs) have been discovered as a class of enzymes that regulate the elimination of acetyl groups from lysine residues of histones. They play an important regulatory role in epigenetics [1]. HDACs have been identified as one of the main agents in tumorigenesis and the inhibition of HDAC function has been shown to be an effective strategy in the treatment of cancer [2]. The well-known HDACs are eighteen (18) molecules and are grouped into four classes. They are ubiquitously expressed and are widely implicated in many chronic diseases such as cancer and inflammation [3]. Class I contains HDACs 1, 2, 3 and 8. They have a high homology (likeness) with the RPD3 transcriptional regulator and are a subunit of multiprotein nuclear complexes. The members of Class I which are most physically similar to each other are HDAC1 and HDAC2 because of HDAC1 shares nearly 85% sequence homology with HDAC2, followed by HDAC3 which shares about 57% and HDAC8 (nearly 38%). HDAC1 and 2 are associated in most corepressor complexes as heterodimers or homodimers [4]. Moreover HDAC1, 2 and 3 are highly expressed in renal cancer [5] and prostate one [6]. Class II, i.e. HDAC4, 5, 6, 7, 9, and 10 are closely related to HDAC1. The different types of cancer due to HDAC1 are related to peripheral tissues [7]. In our study, the IC_{50} inhibitory activity on HDAC1 was used. Quantitative Structure Activity Relationship (QSAR) is one of the best methods used to design new therapeutic agents [8-10]. It permits to correlate quantitatively with a mathematical model the structure of the compounds with their biological activities. It is increasingly used to reduce the excessive number of experiments, sometimes long and expensive and of course the cost of drug's production for pharmaceutical companies [11, 12]. This QSAR approach has its origins in the studies carried out by Hansch [13] and by Free and Wilson [14]. In this work, the aim is to conduct a descriptive and predictive study of the anticancer activity of a series of thirty one (31) HDACi by implementing the methods of quantum chemistry in order to model the observed anticancer activities. The molecular descriptors have been calculated by using only the chemical structure of the compounds. Those descriptors help us to predict the inhibitory concentration of similar molecules. In the specific case of the QSAR study, twenty (21) histone deacetylase inhibitors (HDACi) were used for the test set and ten (10) others from the same series were used for the test external validation (Fig. 1). These molecules were synthesized by Yao et al. [15].

Training set			
Code		**Code**	
1		12	
2		13	
3		14	
4		15	
5		16	
6		17	

7		18	
8		19	
9		20	
10		21	
11			

Validation set

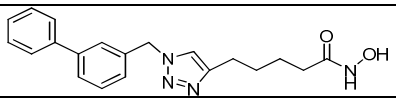
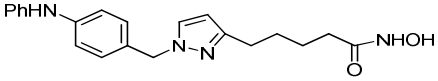
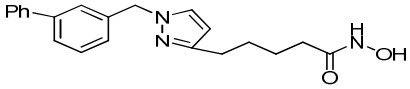
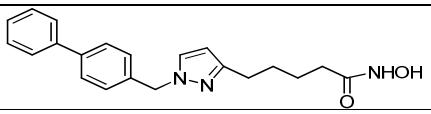
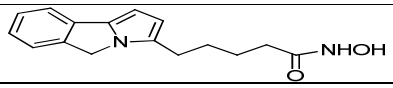
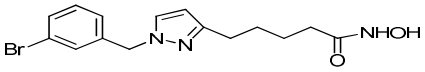
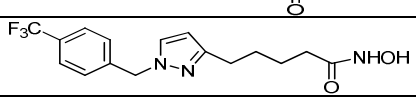
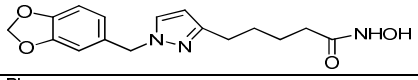
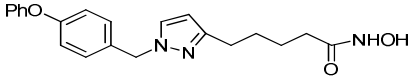
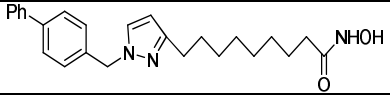
Codes		Codes	
22		27	
23		28	
24		29	
25		30	
26		31	

Fig. 1. Molecular structures of training and validation sets of histone deacetylase inhibitors used for QSAR models

2. MATERIALS AND METHODS

2.1 Data Sources

In order to establish a descriptive and predictive theory of the anticancer activity of HDACs I, the methods of Theoretical Chemistry are used at the B3LYP/6-311G (d, p) level. Gradient-corrected functionalities and hybrid functionals such as B3LYP give better energies and are in good agreement with high-level ab initio methods [16,17]. Gaussian 09 [18] was used in order to evaluate the quantitative structure-activity relationship between the inhibitory activity of HDACi and quantum descriptors. The split-valence and triple-zeta bases (6-311G (d,p)) which is sufficiently extended and the fact of taking into account the polarization functions is important because it takes into account the pairs of electron of heteroatoms those are not engaged in a link. The modeling was done using the multilinear regression method implemented in Excel spreadsheets [19] and XLSTAT version 2014 [20].

2.2 Molecular Descriptors

Some physico-chemical descriptors have been used for the development of QSAR models. In particular, the descriptors related to the geometry that is to say the structure and the dipole moment (μ_D). These descriptors are all determined from the optimized molecules. The used geometric descriptors are the bond length d (C = O) in angstrom (Å) and the valence angles α°(O = C-N) and α°(H-N-O) in degrees (Fig. 2). It should be noted that these geometric descriptors were all measured on the common core of studied HDACi. Several studies have demonstrated that geometric descriptors provide better models as well as global reactivity descriptors [21,22,23]. The dipole moment (μ_D)

indicates the stability of a molecule in water, in particular an aqueous solution. Thus, with a strong dipole moment we will get low solubility in organic solvents and high solubility in water [24, 25].

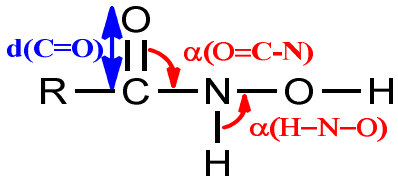


Fig. 2. Geometric descriptors of Class I Histone Deacetylase Inhibitors

2.3 Statistical Analysis

2.3.1 Data analysis

The structures of 31 HDACi were studied by using both statistical methods based on Principal Component Analysis (PCA) [26,27] and the XLSTAT version 2014 software [20] to determine the descriptors that are straightly linked to the inhibitory activity. PCA is a useful statistical technique for summarizing all the information contained in the structure of the different compounds. It is also very important because it allows us both to understand the distribution of compounds and to select descriptors that are directly related to biological activity [28]. It is a method which is essentially descriptive. This method aims to present the maximum information of the physicochemical descriptors graphically. The Ascendant Hierarchical Classification (AHC) aims to partition a set of molecules into homogeneous classes [29]. It organizes molecules by grouping them hierarchically on a dendrogram according to a number of variables and modalities. It gathers them by aggregating those are most similar by using the measure of distance between molecules in order to form classes. This classification is made by taking into account the biological activity of molecules and calculated descriptors. AHC has established a typology of molecules as a function of the dipole moment (μ_D), the bond length d(C=O), and the α°(O=C-N) and α°(H-N-O) valence angles.

2.3.2 Multiple Linear and Nonlinear Regressions (RML and RMNL)

The Multiple Linear Regression (RML) statistical technique 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. It also allowed to select the descriptors used as input parameters in multiple nonlinear regression (RMNL). RMNL analysis is a technique that improves the structure-activity relationship in order to evaluate quantitatively the biological activity. This technique is the well-known tool for studying multidimensional data which takes into account several parameters. The background of this technique is a preprogram XLSTAT functions which is given below (1):

$$y = a + (bx_1 + cx_2 + dx_3 + ex_4) + (fx_{12} + gx_{22} + hx_{32} + ix_{42}) \quad (1)$$

Where a, b, c, d... represent the parameters and x1, x2, x3, x4... represent the variables.

The (RML) and the (RMNL) were generated using the XLSTAT software version 2014 [20] to predict the inhibitory activity IC_{50} HDAC1. The equations of the different models were specified by the coefficient of determination (R^2), the mean squared error (S), the Fischer test (F) and the cross correlation coefficient (Q^2_{CV}) [30,31].

2.4 Estimation of the Predictive Capacity of a QSAR Model

Studied Histone deacetylase I inhibitors have various inhibitory concentrations ranging from 0.033 to 9.58 µM. This range of concentrations makes it possible to define a quantitative relationship between the cancer activity and the theoretical descriptors of these molecules. The quality of a model is determined by based on various statistical analysis criteria including the coefficient of determination R^2, the standard deviation S, the correlation coefficients of the cross validation Q^2_{CV} and Fischer F. R^2, S and F are related to the adjustment of calculated and experimental values. They describe the predictive ability within the limits of the model, and make it possible to estimate the accuracy of the calculated values on the test set [32,33]. The cross validation coefficient Q^2_{CV} gives information on the predictive power of the model. This predictive power is called "internal" because it is calculated from the structures used to build this model. The correlation coefficient R^2 gives an evaluation of the dispersion of the theoretical values around the experimental ones. The quality of the modeling is better when the dots are closed to the adjustment line [34]. The

adjustment of points to this line can be evaluated by the coefficient of determination (2).

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

$y_{i,exp}$: The experimental value of inhibitory activity

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

$\bar{y}_{i,exp}$: The average value of the experimental values of the inhibitory activity.

More the value of R^2 will be closer to 1 more the theoretical and experimental values will be correlated. Moreover, the variance σ^2 was determined by the following relation:

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

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

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

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

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

The coefficient of determination of the cross validation Q^2_{CV} permits to evaluate the accuracy of the prediction on the test set. It has been calculated by using the following relation:

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

2.5 Criterion for Acceptance of a QSAR Model

According to Eriksson et al. [35,36], the performance of a mathematical model is characterized by a value of $\boldsymbol{Q^2_{cv}}$. A model is said to be satisfactory when $\boldsymbol{Q^2_{cv} > 0.5}$ and to be excellent when $\boldsymbol{Q^2_{cv} > 0.9}$. According to them, a given test set will be qualified as efficient model if the acceptance criterion $\boldsymbol{R^2 - Q^2_{cv} < 0.3}$ is respected.

According to Tropsha et al. [37,38,39], concerning the external validation set, the predictive power of a model can be obtained from the five follow criteria.

1) $\boldsymbol{R^2_{Test} > 0.7}$,
2) $\boldsymbol{Q^2_{CvTest} > 0.6}$,
3) $\boldsymbol{\left|R^2_{Test} - R^2_0\right| \leq 0.3}$,
4) $\frac{|R^2_{Test} - R^2_0|}{R^2_{Test}} < 0.1$ and $0,8 \leq k \leq 1.15$,
5) $\frac{|R^2_{Test} - R'^2_0|}{R^2_{Test}} < 0.1$ and $0.8 \leq k' \leq 1.15$

Otherwise, Roy and Roy [40] have further refined the predictive ability of a QSAR model. They have developed quantities such as $\boldsymbol{r^2_m}$ and $\boldsymbol{\Delta r^2_m}$, called metric values. $\boldsymbol{r^2_m}$ determines the proximity between the observed activity and the predicted one. The metric values $\boldsymbol{r^2_m}$ and $\boldsymbol{\Delta r^2_m}$ are calculated from the observed and predicted activities. Currently, these two different parameters (variants) $\boldsymbol{r^2_m}$ and $\boldsymbol{\Delta r^2_m}$ can be calculated for the internal validation or for external validation. According to these authors a QSAR model is acceptable, when both criteria are met.

$$\overline{r^2_m} = \frac{(r^2_m + r'^2_m)}{2} > 0.5$$

$$\boldsymbol{\Delta r^2_m = \left|r^2_m - r'^2_m\right| < 0.2}$$

With $\boldsymbol{r^2_m = r^2 * \left(1 - \sqrt{(r^2 - r^2_0)}\right)}$ and $\boldsymbol{r'^2_m = r^2 * \left(1 - \sqrt{(r^2 - r'^2_0)}\right)}$

3. RESULTS AND DISCUSSION

3.1 Descriptors and Experimental Inhibitory Activities

All the descriptors values for the twenty one (21) molecules of class 1 of histone deacetylase inhibitors of the test set and the ten (10) other molecules of the validation set are given in Table 1.

3.2 Principal Component Analysis (PCA) and Ascendant Hierarchical Classification (AHC)

All the four descriptors of the 31 inhibitors are submited to PCA. The two main axes are enough to describe the information provided by the data matrix. Indeed, the variance percentages are

55.57% and 22.93% for the F1 and F2 axes, respectively. The total information's estimation is 78.50%. (PCA) [29]. PCA was conducted to identify the relationship between the different descriptors. Bold values are different from 0 to a significance level of p = 0.05. The correlations between the four descriptors are presented in Table 2 as a correlation matrix and in Fig. 3 where these descriptors are represented in a correlation circle. The Pearson correlation coefficients are summarized in Table 2. The resulting matrix provides information on the negative or positive correlation between the variables.

Table 1. Quantum descriptors and Inhibition Concentrations (IC_{50}) of training and validation sets

	μ_D	d(C=O)	α°(O=C-N)	α°(H-N-O)	IC_{50}
			Training set		
1	5.782	1.22611	119.550	111.081	0.131
2	6.157	1.22554	119.335	110.828	1.160
3	5.178	1.22598	119.371	110.942	0.32
4	4.807	1.22594	119.367	110.938	0.313
5	5.231	1.22611	119.344	110.957	1.310
6	5.158	1.22606	119.360	110.961	0.323
7	5.537	1.22590	119.369	110.930	0.218
8	7.251	1.22573	119.385	110.904	0.242
9	5.347	1.22616	119.335	110.947	0.376
10	5.472	1.22592	119.378	110.943	0.293
11	6.224	1.22614	119.321	110.943	0.067
12	4.448	1.22595	119.342	110.906	0.145
13	5.667	1.22606	119.329	110.922	0.719
14	5.259	1.22610	119.354	110.971	0.342
15	4.718	1.22591	119.385	110.926	0.116
16	5.161	1.22605	119.349	110.952	0.043
17	4.528	1.22595	119.343	110.886	0.539
18	5.109	1.22606	119.343	110.916	0.646
19	4.474	1.22603	119.353	110.958	0.233
20	5.353	1.22599	119.353	110.989	2.660
21	3.190	1.22583	118.935	110.771	9.58
			validation set		
22	6.777	1.22604	119.310	110.902	0.112
23	5.192	1.22602	119.351	110.939	0.064
24	3.823	1.22554	119.387	110.851	1.760
25	6.000	1.22585	119.376	110.919	0.197
26	6.222	1.22605	119.343	110.935	0.086
27	5.777	1.22610	119.331	110.946	0.035
28	5.151	1.22603	119.342	110.923	0.033
29	3.831	1.22591	119.376	110.938	0.068
30	5.265	1.22606	119.348	110.949	0.195
31	4.592	1.22606	119.362	111.029	2.490

μ_D in Debye (D), d (C = O) in angstrom (Å), α°(O = C-N) and α°(H-N-O) in degrees, IC_{50} (μM)

Table 2. Correlation matrix (Pearson (n)) between the different descriptors

Variables	IC_{50}	μ_D	d(C=O)	α°(O=C-N)	α°(H-N-O)
IC_{50}	1				
μ_D	-0.4937	1			
d(C=O)	-0.2513	0.1250	1		
α°(O=C-N)	-0.8235	0.3729	0.1130	1	
α°(H-N-O)	-0.4795	0.2225	0.6502	0.6939	1

Bold values are different from 0 to a significant level for p <0.05.
Very significant for p <0.01. Very significant for p <0.001.

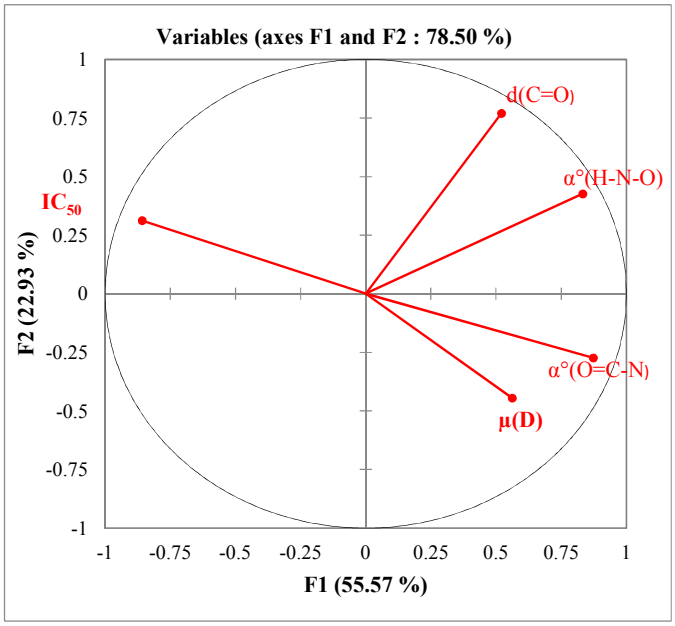


Fig. 3. Circle of correlation

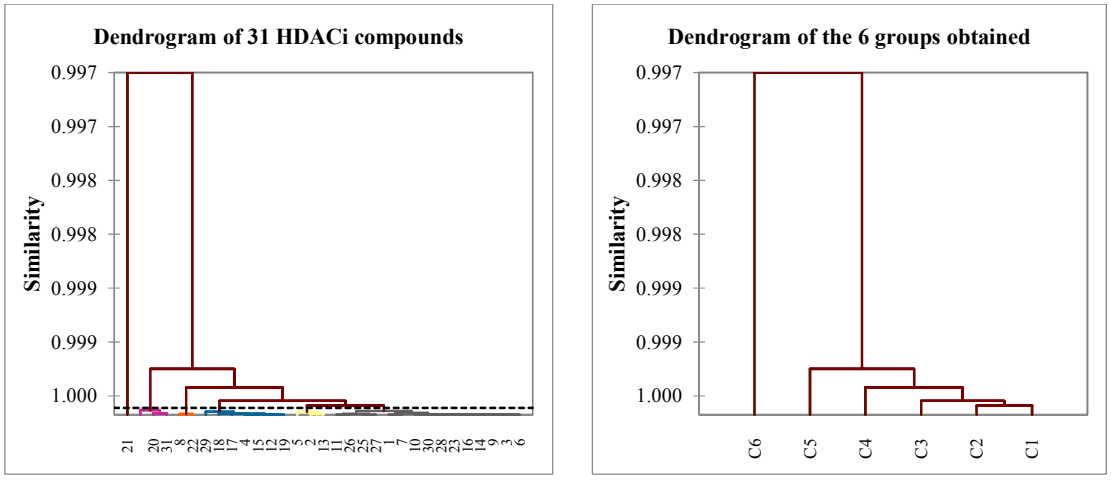


Fig. 4. Dendrograms of class I histone deacetylase inhibitors

The dipole moment, the valence angles (α°(O=C-N) and α°(H-N-O)) are negatively correlated with the IC_{50} (respectively with -0.4937, -0.8235, -0.4795 and $p < 0.05$) at a significant level.

The correlation circle was made to detect the connection between the different descriptors. The analysis of the principal components from the correlation circle (Fig. 3) revealed that the F1 axis (55.57% of the variance) seems to represent the valence angles α°(O=C-N) and α°(H-N-O), and the F2 axis (22.93% of the variance) seems to represent the bond d(C=O).

The AHC in Fig. 4 illustrates the distribution of inhibitors in six (6) classes according to their affinity. The six (6) classes are: C1 (1; 3; 6; 7; 9; 10; 11; 14; 16; 23; 25; 26; 27; 28; 30); C2 (2; 5; 13), C3 (4; 12; 15; 17; 18; 19; 29), C4 (8; 22), C5 (20; 24; 31) and C6 (21).

3.3 Multiple Linear Regression (RML)

The equation of the QSAR model with statistical data is presented below. Fig. 5 illustrates the correlation between the experimental and theoretical IC_{50} of the test set (blue dots) and the validation set (red dots). This obtained model

relates the inhibitory activities and the theoretical descriptors of histone deacetylase inhibitors. The negative or positive sign of the coefficient of a model's descriptor reflects the effect of proportionality between the evolution of the biological activity and this parameter. The negative sign indicates that when the value of the descriptor increases, the biological activity decreases. The positive sign reflects the opposite effect. The obtained equation is shown below:

IC_{50}^{theo} = 9213 – 0.19339*μ_D - 7645*d(C=O) - 28.98873*α°(O=C-N) + 32.64810*α°(H-N-O) N=21 $R^2 = 0.913$ $Q_{CV}^2 = 0.913$S= 0.599F= 198.197

The negative signs of the dipole moment (μ), the bond d(C=O) and valence angle α°(O=C-N) coefficients indicate that inhibitory activity will be improved for low values of these quoted descriptors. And the positive sign of α°(H-N-O) also indicates that the inhibitory activity will be improved for high values of this valence angle. The significance of a model is expressed by the Fischer coefficient F = 198.197: the correlation coefficient of the cross validation Q_{CV}^2 = 0.913. This model is acceptable with $R^2 - Q_{CV}^2 = 0.913 - 0.913 = 0.000 < 0.3$. The regression line between experimental and theoretical inhibitory activity of the training set and validation set is shown in Fig. 5.

The low value of the standard error S = 0.599 attests the good similarity between the predicted and experimental values (Fig. 6). This curve reflects a similar evolution of these data by the multilinear model of the inhibitory activity.

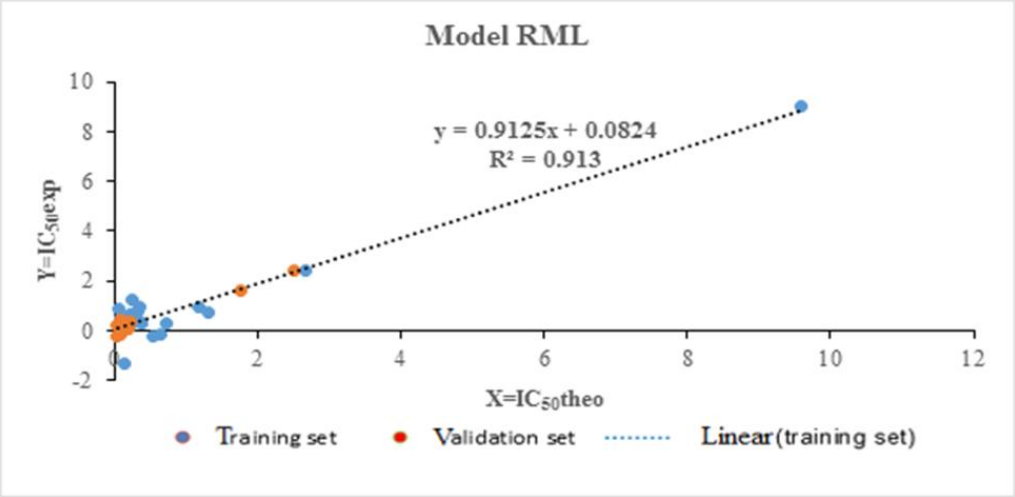


Fig. 5. The regression line of the RML model

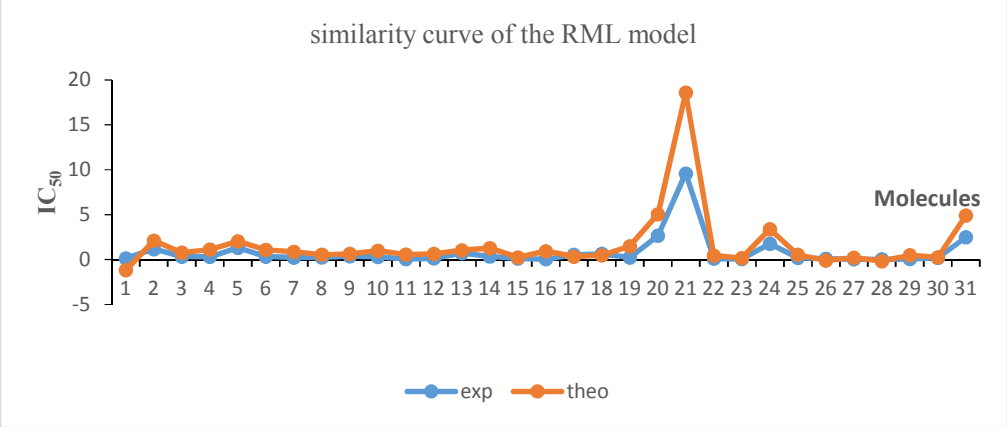


Fig. 6. Similarity curve of the experimental and predicted values of the RML model

Verification of Tropsha Criteria

R^2_{Test}= 0.944 > 0.7 Q^2_{CvTest} = 0.998 > 0.6
$\left|R^2_{Test} - R^2_0\right| = \mathbf{0.0543} \leq \mathbf{0.3}$

$\frac{|R^2_{Test} - R^2_0|}{R^2_{Test}} = \mathbf{0.0575} < \mathbf{0.1}$ and $\mathbf{.85} \leq \mathbf{k} = \mathbf{0.959} \leq \mathbf{1.15}$;

$\frac{|R^2_{Test} - R'^2_0|}{R^2_{Test}} = \mathbf{0.0593} < \mathbf{0.1}$ and $\mathbf{0.85} \leq \mathbf{k'} = \mathbf{1.00} \leq \mathbf{1.15}$

All values meet the Tropsha criteria, so the model is acceptable for predicting inhibitory activity.

3.4 Nonlinear Multiple Regression (RMNL)

The statistical nonlinear regression method was used to improve the predicted inhibitory activity quantitatively. It has taken into account the four chosen descriptors (μ_D, d(C=O), α°(O=C-N), α°(H-N-O)). This method is the most common tool for studying multidimensional data. This statistical method is applied to the data in Table 1 containing 31 molecules associated with the four descriptors. The resulting equation is:

IC^{theo}_{50} = -1783E^{+4} + 1.29040*μ_D + 3224E^{+4}*d(C=O) + 5141*α°(O=C-N) - 40320*α°(H-N-O) - 0.09718*μ_D^2 -1315E^{+4}*d(C=O)^2 – 21.64688*α°(O=C-N)^2 + 181.80443*α°(H-N-O)^2

N=21 $R^2 = \mathbf{0.967}$ $Q^2_{CV} = \mathbf{0.967}$
S= 0.379 F= 557.031

The significance of the model is expressed by the Fischer coefficient F = 557.031 and the correlation coefficient of the cross validation Q^2_{CV} = 0.967. This model is acceptable with $R^2 - Q^2_{CV} = \mathbf{0.967} - \mathbf{0.967} = \mathbf{0.000} < \mathbf{0.3}$. The regression line between the experimental and theoretical anticancer activities of the training set (blue dots) and the test set (red dots) is shown in Fig. 7.

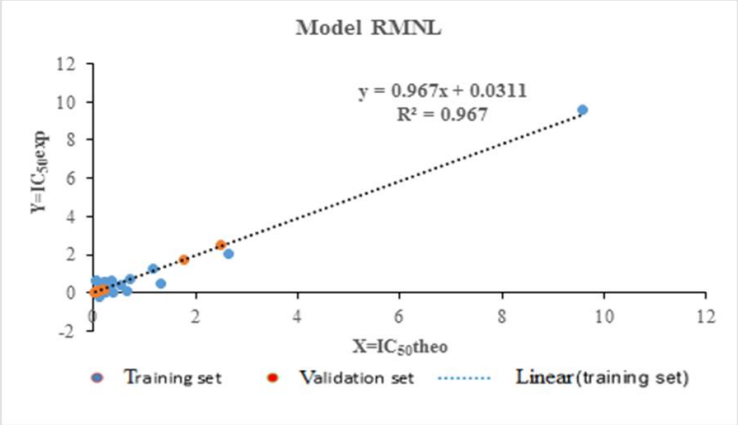


Fig. 7. The regression line of the RMNL model

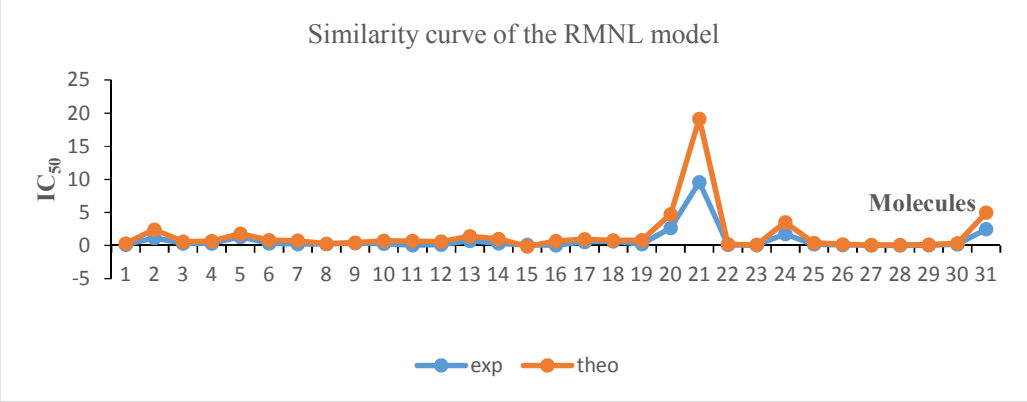


Fig. 8. Similarity curve of the experimental and predicted values of the RMNL model

The very low value of the standard error, S = 0.379 also certify the good similarity between the predicted and experimental values (Fig. 8). This curve reflects a very good analogical evolution of the experimental values predicted by the RMNL model of the inhibitory activity, despite some recorded discrepancies.

Verification of Roy's Criteria

$$\overline{r_m^2} = \frac{(r_m^2 + r'^2_m)}{2} = \mathbf{0.943} > 0.5; \qquad \Delta r_m^2 = |r_m^2 - r'^2_m| = \mathbf{0.000} < 0.2$$

$$\textbf{With } r_m^2 = \mathbf{0.971} \textbf{ and } r'^2_m = \mathbf{0.971}$$

All values meet the Tropsha criteria, so the model is acceptable for predicting the inhibitory activity of hydroxamic acid compounds.

Among the two models, the model obtained by the RMNL statistical method has a much better predictive capability than the RML approach.

However, these models are based on four theoretical descriptors, it is useful to determine the contribution of each of them in predicting the inhibitory activity of the test compounds. Indeed, the knowledge of this contribution allows to establish the order of priority of the various descriptors. It also permit to define the choice of the parameters which must be optimized in order to get an improved activity.

3.5 Analysis of the Contribution of Descriptors

The contribution of the four descriptors of this model in the prediction of the inhibitory activity of hydroxamic acid compounds was determined from the XLSTAT software version 2014 [20]. The different contributions are gathered in Fig. 9.

The valence angle α°(O=C-N) displays a large proportion followed by the valence angle α°(H-N-O) and the bond length d(C=O). And finally the dipole moment (μ) displays the smallest proportion. It should be noted that geometric descriptors globally bring a fairly important contribution in predicting the inhibitory activity of hydroxamic acid compounds.

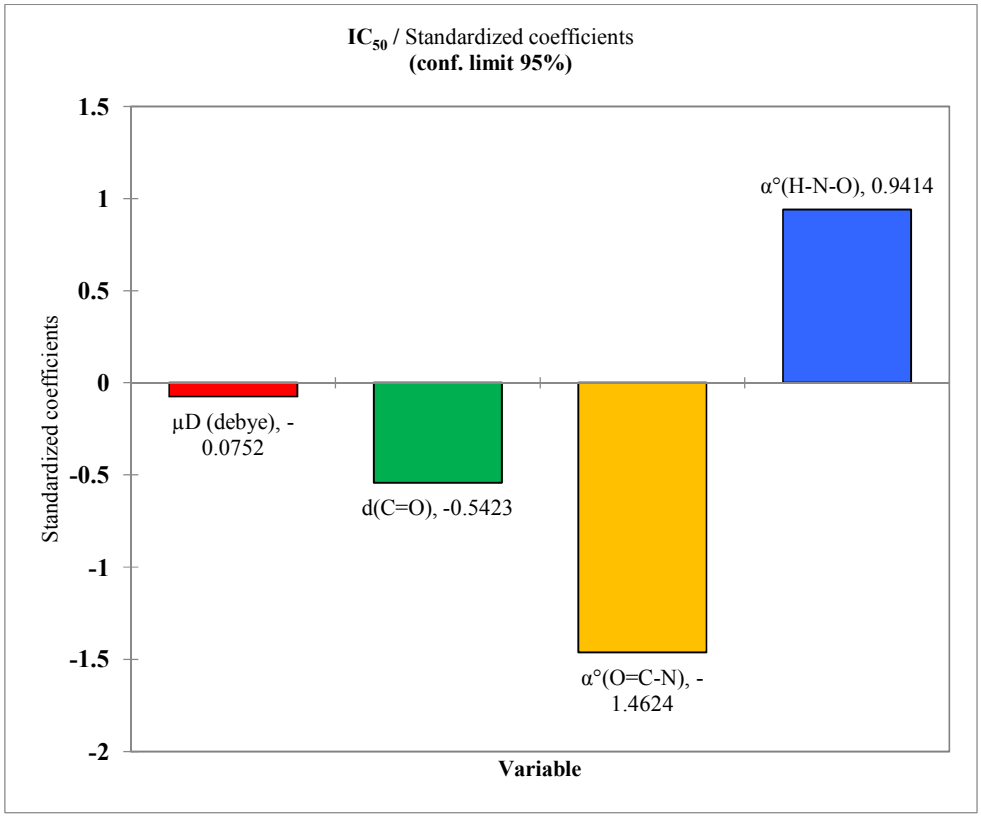


Fig. 9. Contribution of descriptors in models

4. CONCLUSION

This study revealed quantitative relationships between the inhibitory activity (IC_{50}) and the physicochemical descriptors of HDACi. From the chemical point of view, these obtained descriptors depend on mainly the geometric structure of the hydroxamic acids. The descriptors like the dipole moment (μ_D), the bond length d(C=O) and the α°(O=C-N) and α°(H-N-O) valence angles explain and predict the inhibitory concentration of HDACi. Statistical methods such as Principal Component Analysis (PCA), Ascendant Hierarchical Classification (AHC), Multilinear and Nonlinear Regression were used. The study of the robustness of the two built models (RML and RMNL) has a good stability and an excellent power of prediction. In addition, the RMNL model (R^2 = 0.967, S = 0.379, F = 557.031) is better than RML one and is an effective tool for predicting the inhibitory activity of the best analogs of studied HDACi. Moreover, the study of the contribution of the descriptors showed that the valence angle α°(O=C-N) is the first descriptor in terms of priority in the prediction of the inhibitory concentration of studied hydroxamic acids. The negative sign of the coefficient of valence angle α°(O=C-N) in the equation of the RML model indicates that the low values of the valence angle α°(O=C-N) could improve the inhibitory concentration of HDACi.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Arrowsmith C, Bountra C, Fish P, Lee K, Schapira M. Epigenetic protein families: A new frontier for drug discovery. Nat. Rev. Drug. Discov. 2012;11:384-400.
2. Marks P, Rifkind R, Richon V, Breslow R, Miller T, Kelly W. Histone deacetylases and cancer: causes and therapies. Nat. Rev. Cancer. 2001;1:194-202.
3. Villagra A, Sotomayor E, Seto E. Histone deacetylases and the immunological network: implications in cancer and inflammation. Oncogene. 2010;9:157-173.
4. Jurkin J, Zupkovitz G, Lagger S, Grausenburger R, Hagelkruys A, Kenner L. Distinct and redundant functions of histone deacetylases HDAC1 and HDAC2 in proliferation and tumorigenesis. Celle Cycle. 2011;10:406-412.
5. Fritzsche FR, Weichert W, Röske A, Gekeler V, Beckers T, Stephan C, Jung K, Scholman K, Denkert C, Dietel M, Kristiansen G. Class I histone deacetylases 1, 2 and 3 are highly expressed in renal cell cancer. BMC Cancer; 2008.
6. Weichert W, Röske A, Gekeler V, Beckers T, Stephan C, Jung K, Fritzsche F, Niesporek S, Denkert C, Dietel M, Kristiansen G. Histone deacetylases 1, 2 and 3 are highly expressed in prostate cancer and HDAC2 expression is associated with shorter PSA relapse time after radical prostatectomy. Br. J. Cancer. 2008;98:604-610.
7. Zupkovitz G, Tischler J, Posch M, Sadzak I, Ramsauer K, Egger GGRSN, Chiocca S, Decker T, Seiser C. Negative and positive regulation of gene expression by mouse histone deacetylase 1. Mol Cell Biol. 2006;26:7913-7928,
8. Chhabria M, B Mahajan, Brahmkshatriya P. QSAR study of a series of Acyl Coenzyme A (CoA): Cholesterol Acyl-transferase inhibitors using genetic function approximation. Medicinal Chemistry Research. 2011;20:1573-1580.
9. Buha V, Rana D, Chhabria M, Chikhalia K, Mahajan B, Brahmkshatriya P, Shah N, Synthesis, biological evaluation and QSAR study of a series of substituted Quinazolines as antimicrobial agents. Medicinal Chemistry Research. 2013;22:4096-4109.
10. A Tropsha. Best practices for QSAR model development, validation, and exploitation. Molecular Informatics. 2010;29:476-488.
11. Oprea TI. Chemoinformatics in drug discovery. Ed. Wiley-VCH Verlag. Allemagne; 2005.
12. Rekka EA, Kourounakis PN. Chemistry and molecular aspects of drug design and action. Ed. Taylor & Francis Group, LLC. Etats Unies; 2008.
13. Hansch C, Fujita T. ρ–σ–π, analysis: Method for correlation of biological activity and chemical structure. J. Am. Chem. Soc. 1964;86:1616-1626.
14. Free SM, Wilson JW. A mathematical contribution to structure-activity studies. J. Med. Chem. 1964;7:395-399.
15. Yao Y, Liao C, Li Z, Wang Z, Sun Q, Liu C, Yang Y, Tu Z, Jiang S. Design, synthesis, and biological evaluation of 1, 3-

disubstituted pyrazole derivatives as new class I and IIb histone deacetylase inhibitors. European Journal of Medicinal Chemistry. 2014;86:639-652.
16. Kapp J, Remko M, Schleyer PVR. 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. 1996;118:5745-5751.
17. Johnson BG, Gill PM, Pople JA. The performance of a family of density functional methods. The Journal of Chemical Physics. 1993;98:5612-5626.
18. Frisch M, Trucks G, Schlegel H, Scuseria G. Gaussian 09, Revision C.01, Gaussian, Inc., Wallingford CT; 2009.
19. Microsoft ® Excel ® 2013 (15.0.4420.1017) MSO (15.0.4420.1017) 64 Bits Partie de Microsoft Office Professionnel Plus; 2013.
20. XLSTAT Version 2014.5.03 Copyright Addinsoft 1995-2014 XLSTAT and Addinsoft are Registered Trademarks of Addinsoft; 2014.
Available:https://www.xlstat.com
21. Balaban AT. From chemical topology to three-dimensional geometry. Plenum Press, New York (NY). 1997;1-24.
22. Randic M, Razinger M. Molecular topographic indices. J. Chem. Inf. Comput. Sci. 1995;35:140-147.
23. Rastija V, Medic-Š M. QSAR modeling of anthocyanins, anthocyanidins and catechins as inhibitors of lipid peroxidation using three-dimensional descriptors. Medicinal Chemistry Research. 2009;18:579-588.
24. Rivelino R, Canuto S. Conformational stability of furfural in aqueous solution: The role of hydrogen bonding. Brazilian Journal of Physics. 2004;34(11):84-89.
25. Wong MW, Frisch MJ, Wiberg KB. Solvent effects: The mediation of electrostatic effects by solvents. J. Am. Chem. Soc. 1991;113(113):4776-4782.
26. Larif M, Adad A, Hmamouchi R, Taghki AI, Soulaymani AB, Elmidaoui A, Bouachrine M, Lakhlifi T. Biological activities of triazine derivatives. Combining DFT and QSAR results. Arab. J. Chem., in press; 2013.
27. Kpidi YH, Yapo OB, Koné MGR, GA Gadji, Gnagne AEJEY, N'dri JS, Ziao N. Monitoring and modeling of chlorophyll-a dynamics in a Eutrophic Lake: M'koa Lake (Jacqueville, Ivory Coast) American Journal of Environmental Protection. 2018;6(11):1-9.
28. Hmamouchi R, Taghki AI, Larif M, Adad A, Abdellaoui A, Bouachrine M, Lakhlifi T. The inhibitory activity of aldose reductase of flavonoid compounds: Combining DFT and QSAR calculations. J. Chem Pharm. Research. 2013;5(19):198-209.
29. Escofier B, Pagès J. Analyses factorielles simples et multiples: Objectifs, méthodes et interprétation., Paris: Dunod ; 2008.
30. Rücker C, Rücker G, Meringer M. y-Randomization and its variants in QSPR/QSAR. J. Chem. Inf. Model. 2007;47:2345-2357.
31. N'guessan KN, Koné MGR, Bamba K, Ouattara WP, Ziao N. Quantitative structure anti-cancer activity relationship (QSAR) of a series of ruthenium complex azopyridine by the density functional theory (DFT) method. Computational Molecular Bioscience. 2017;7:19-31.
32. Snedecor GW, Cochran WG. Methods, Statistical. Oxford and IBH: New Delhi, India. 1967;381.
33. Kangah NJB, Koné MGR, Kodjo CG, N'guessan BR, Kablan ALC, Yéo SA, Ziao N. Antibacterial activity of Schiff bases derived from Ortho Diaminocyclohexane, Meta-Phenylenediamine and 1,6-Diaminohexane: Qsar study with quantum descriptors. International Journal of Pharmaceutical Science Invention. 2017;6(13):38-43.
34. EX Esposito, Hopfinger AJ, Madura JD. Methods for applying the quantitative structure-activity relationship paradigm. Methods in Molecular Biology. 2004; 275:131-213.
35. Eriksson L, Jaworska J, Worth A, Cronin MD, Dowell RMM, Gramatica P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. Environmental Health Perspectives. 2003;111(110):1361-1375.
36. N'Dri JS, Koné MGR, Kodjo CG, Affi ST, Kablan ALC, Ouattara O, Soro D, Ziao N. Relation quantitative structure activite(QSAR) d'une série d'azetidinones dérivés de Dapsone par la méthode de Théorie de la fonctionnelle de la densité(DFT). IRA-International Journal of Applied Sciences. 2017;8(12):55-62.
37. Golbraikh A, Tropsha A. Beware of qsar. J. Mol. Graph. Model. 2002;20:269-276.

38. Tropsha A, Gramatica P, Gombar VK. The importance of being earnest, validation is the absolute essential for successful application and interpretation of QSPR models. QSAR Comb. Sci. 2003;22; 69-77,
39. Ouattara O, Affi TS, Koné MGR, Bamba K, Ziao N. Can empirical descriptors reliably predict molecular lipophilicity? A QSPR study investigation. Int. Journal of Engineering Research and Application. 2017;7(15):50-56.
40. Roy PP, Roy K. On some aspects of variable selection for partial least squares regression models. QSAR Comb Sci. 2008;27:302-313.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history/24085