

# DFT Study, Linear and Nonlinear Multiple Regression in the Prediction of HDAC7 Inhibitory Activities on a Series of Hydroxamic Acids

Doh Soro<sup>1</sup>, Lynda Ekou<sup>1</sup>, Bafétigué Ouattara<sup>2</sup>, Mamadou Guy-Richard Kone<sup>1\*</sup>, Tchirioua Ekou<sup>1</sup>, Nahossé Ziao<sup>1</sup>

<sup>1</sup>Laboratoire de Thermodynamique et de Physico-Chimie du Milieu, UFR SFA, Université Nangui Abrogoua, Abidjan, Côte-d'Ivoire

<sup>2</sup>Laboratoire de Physique Fondamentale et Appliquée, Université Nangui Abrogoua, Côte d'Ivoire

Email: \*guyricharkone@gmail.com

**How to cite this paper:** Soro, D., Ekou, L., Ouattara, B., Kone, M.G.-R., Ekou, T. and Ziao, N. (2019) DFT Study, Linear and Nonlinear Multiple Regression in the Prediction of HDAC7 Inhibitory Activities on a Series of Hydroxamic Acids. *Computational Molecular Bioscience*, 9, 63-80.

<https://doi.org/10.4236/cmb.2019.93006>

**Received:** June 8, 2019

**Accepted:** August 29, 2019

**Published:** September 2, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

In this work, we conducted a QSAR study on 18 molecules using descriptors from the Density Functional Theory (DFT) in order to predict the inhibitory activity of hydroxamic acids on histone deacetylase 7. This study is performed using the principal component analysis (PCA) method, the Ascendant Hierarchical Classification (AHC), the linear multiple regression method (LMR) and the nonlinear multiple regression (NLMR). DFT calculations were performed to obtain information on the structure and information on the properties on a series of hydroxamic acids compounds studied. Multivariate statistical analysis yielded two quantitative models (model MLR and model MNLR) with the quantum descriptors: electronic affinity (AE), vibration frequency of the OH bond ( $\nu(\text{OH})$ ) and that of the NH bond ( $\nu(\text{NH})$ ). The LMR model gives statistically significant results and shows a good predictability  $R^2 = 0.9659$ ,  $S = 0.488$ ,  $F = 85$  and  $p\text{-value} < 0.0001$ . Electronic affinity is the priority descriptor in predicting the activity of HDAC7 inhibitors in this study. The results obtained suggest that the descriptors derived from the DFT could be useful to predict the activity of histone deacetylase 7 inhibitors. These models were evaluated according to the criteria of Tropsha *et al.*

## Keywords

Hydroxamic Acids, Histones Deacetylases, DFT, Histones Deacetylases Inhibitors, QSAR

## 1. Introduction

Histone deacetylases (HDAC) have become essential transcriptional corepressors in a variety of physiological systems. To date, 18 human HDACs have been identified and grouped into four classes. HDAC Class I (HDAC1, 2, 3 and 8), HDAC Class II (HDAC4, 5, 6, 7, 9 and 10), HDAC Class III, also called sirtuins (SIRT1, 2, 3, 4, 5, 6 and 7) and HDAC Class IV (HDAC11). Class II HDACs are subdivided into classes IIa (HDAC4, 5, 7, 9) and class IIb (HDAC6 and 10) [1]. HDACs have also been identified as one of the major players in tumorigenesis and the inhibition of HDAC function has been shown to be an effective strategy in the treatment of cancer [2]. They are ubiquitously expressed and are widely implicated in many chronic diseases such as cancer and inflammation [3]. Class II is closely related to HDAC1 class I. HDAC class IIa has large N-terminal extensions with conserved binding sites for transcription factor 2 (MEF2) and chaperone 14-3-3 protein that make the signal sensitive to HDACs [4]. Regulated phosphorylation of HDAC class IIa provides a mechanism for binding extracellular signals to transcription and plays a key role in many tissues during development and disease. In contrast to other HDACs, class IIa HDACs show relatively restricted expression patterns. HDAC5 and HDAC9 are highly enriched in muscles, heart and brain [5]. HDAC4 is highly expressed in the brain and growth plate of the skeleton [6], and HDAC7 is enriched in endothelial cells and thymocytes [7]. The class IIa HDACs repress transcription has not been fully elucidated. Highly purified recombinant class IIa HDACs possess only minimal catalytic activity, and the activity of class IIa HDACs purified from mammalian cells has been shown to be due to contaminating class I HDACs [8]. Studies in the cell culture by silencing or over-expressing HDAC7 have shown that HDAC7 is involved in the regulation of cell proliferation, apoptosis, differentiation and migration. Some reports have investigated the role of HDAC7 in cancers, and a high level of HDAC7 protein has been observed in nine different human pancreatic cancer cases [9]. HDAC7 is also overexpressed in breast cancer stem cells and is required to maintain cells [10]. HDAC7 belongs to Class IIa HDACs and has conflicting functions in different types of cancer [11]. Some reports indicate an ambivalent role of HDAC7 in cancers: for example, HDAC7 has oncogenic functions in children with acute lymphoblastic leukemia and pancreatic cancer [12] but acts as a tumor suppressor in acute lymphoblastic leukemia B and Burkitt's lymphoma [13]. In addition, the high expression of HDAC7 was related to a prognosis in lung cancer [14]. As part of the prediction of biological activities of HDAC7, certain descriptors from quantum chemistry were used during our work. Quantitative Structure Activity Relationship (QSAR) is one of the best methods used to design new therapeutic agents [8] [9] [10]. 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 and expensive and the cost of drug production by pharmaceutical compa-

nies [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 goal is to conduct a descriptive and predictive study of the anticancer activity of a series of eighteen (18) compounds of HDAC7 inhibitors. By implementing quantum chemistry methods, this work aims at modeling the anticancer activities observed, the molecular descriptors being calculated solely from the chemical structure of the compounds, and subsequently predicting the inhibitory concentration of analogous molecules. In the specific case of the QSAR study, twelve (12) histone deacetylase inhibitors (HDACi) were used for the test set and six (6) others from the same series were used for the test. external validation (Table 1). Yao *et al.* [15] in their work on the synthesis of hydroxamic acids have determined the experimental inhibitory activities of these inhibitors used in our work. of 38 molecules synthesized, only 19 provided satisfactory activity on HDAC7 classified in the following table.

## 2. Materials and Methods

### 2.1. Theoretical Calculations

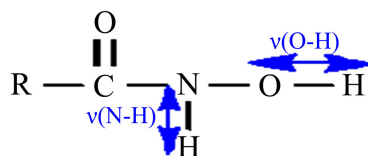
In order to establish a descriptive and predictive theory of the anticancer activity HDAC7 of hydroxamic acids, the methods of Theoretical Chemistry are used at the B3LYP/6-311G (d, p) level. The gradient-corrected functionalities and the hybrid functionals such as B3LYP give better energies and agree with high-level ab initio methods [16] [17]. In this work, to evaluate the quantitative structure-activity relationship between the anticancer activity of HDAC7 and the descriptors, the Gaussian 09 [18] quantum chemistry software was used. The base 6-311G (d, p) being sufficiently extended, considering the polarization functions are important for considering the free doublets of the hetero atoms in order to obtain satisfactory results. The modeling was done using the multilinear regression method implemented in Excel spreadsheets [19] and XLSTAT version 2014 [20].

### 2.2. Chemical Descriptors

Some physico-chemical descriptors have been used for the development of QSAR models. In particular, the electronic affinity (AE) and the vibration descriptors that are the vibration frequency  $\nu$ (O-H) and the vibration frequency  $\nu$ (N-H). These two vibration descriptors are shown in Figure 1.

It should be noted that the descriptors related to the molecular frontier orbitals have been calculated as part of the Koopmans approximation [21]. The LUMO energy characterizes the sensitivity of the molecule to a nucleophilic attack. The electronic affinity AE is the descriptor that translates the ability of a molecule to capture an electron. This descriptor is obtained from the relation (1) below:

$$AE = -E_{\text{LUMO}} \quad (1)$$



**Figure 1.** Vibration descriptors of the hydroxamic acids used  $\nu(O-H)$  and  $\nu(N-H)$ .

**Table 1.** Molecular structures of test sets and validation of hydroxamic acids used for QSAR models.

Training set		Validation set	
Code	Structure	Code	Structure
1		11	
2		12	
5		14	
7		15	
9		17	
10		18	
19			
3		8	
4		13	
6		16	

Several studies have shown that geometric descriptors provide better models as well as global responsivity descriptors [22] [23] [24].

## 2.3. Statistical Analyzes

### 2.3.1. Data Analysis

The structures of 18 hydroxamic acid compounds were studied by statistical

methods based on Principal Component Analysis (PCA) [25] [26] [27] using the software XLSTAT version 2014 [20] to determine the descriptors that are related direct with anticancer activity. PCA is a useful statistical technique for summarizing all the information encoded in the structures of the compounds. It is also very useful for understanding the distribution of compounds and for selecting descriptors that are directly related to biological activity [28]. It is an essentially descriptive statistical method that aims to present, in graphical form, the maximum information of physicochemical descriptors. The Ascendant Hierarchical Classification (AHC) aims to partition a set of molecules into homogeneous classes [29]. It organizes molecules, defined by several variables and modalities, by grouping them hierarchically on a dendrogram. It aggregates those that are most like each other by using dissimilarity or distance measurements between molecules to form classes. It is made from the data of molecules and descriptors. AHC has established a typology of molecules based on electronic affinity (AE) and vibration frequencies  $\nu(\text{O-H})$  and  $\nu(\text{N-H})$ .

### 2.3.2. Multiple Linear and Nonlinear Regressions (MLR and NMR)

The Multiple Linear Regression (MLR) 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 nonlinear multiple regression (NMR).

Nonlinear multiple regression (NMR) analysis is a technique that improves the structure-activity relationship to quantitatively evaluate biological activity. It considers several parameters. It is the most common tool for studying multidimensional data. It is based on preprogrammed XLSTAT functions as follows:

$$y = a + (bx_1 + cx_2 + dx_3 + ex_4) + (fx_{12} + gx_{22} + hx_{32} + ix_{42}) \quad (2)$$

where  $a, b, c, d, \dots$ : are the parameters and  $x_1, x_2, x_3, x_4, \dots$ : are the variables.

The (MLR) and the (NRM) were generated using the XLSTAT software version 2014 [20] to predict the anticancer activity IC<sub>50</sub> HDAC7. The equations of the different models were evaluated by the coefficient of determination ( $R^2$ ), the mean squared error ( $S$ ), the Fischer test ( $F$ ) and the cross-correlation coefficient ( $Q_{cv}^2$ ) [30] [31].

The potential of the inhibitory concentration is calculated according to the following expression:

$$\text{pIC}_{50} = -\log(\text{IC}_{50} * 10^{-6}) \quad (3)$$

### 2.4. Estimation of the Predictive Capacity of a Model

Histone deacetylase 7 has various inhibitory concentrations ranging from 0.311 to 38.9  $\mu\text{M}$ . 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 based on various statistic-

al analysis criteria including the coefficient of determination  $R^2$ , the standard deviation  $S$ , the correlation coefficients of the cross validation  $Q_{CV}^2$  and Fischer  $F$ .  $R^2$ ,  $S$  and  $F$  relate 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 values calculated on the test set [32] [33]. As for the cross-validation coefficient  $Q_{CV}^2$ , it 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 values. The quality of the modeling is better when the points are close to the adjustment line [34]. The adjustment of points to this line can be evaluated by the coefficient of determination.

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

where:

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

$y_{i,theo}$  : Theoretical value of anticancer activity and,

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

The more the value of  $R^2$  will be close to 1, the more the theoretical and experimental values are correlated.

Moreover, the variance  $\sigma^2$  is determined by the relation 4:

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

where  $k$  is the number of independent variables (descriptors),  $n$  is the number of molecules in the test or learning set, and  $n - k - 1$  is the degree of freedom.

The standard deviation or standard deviation  $S$  is another statistical indicator used. It allows to evaluate the reliability and the precision of a model:

$$s = \sqrt{\frac{\sum (y_{i,exp} - y_{i,theo})^2}{n - k - 1}} \quad (6)$$

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

$$F = \frac{\sum (y_{i,theo} - y_{i,exp})^2}{\sum (y_{i,exp} - y_{i,theo})^2} * \frac{n - k - 1}{k} \quad (7)$$

The coefficient of determination of the cross-validation  $Q_{CV}^2$  makes it possible to evaluate the accuracy of the prediction on the test set. It is calculated using the following relation:

$$Q_{cv}^2 = \frac{\sum (y_{i,theo} - \bar{y}_{i,exp})^2 - \sum (y_{i,theo} - y_{i,exp})^2}{\sum (y_{i,theo} - \bar{y}_{i,exp})^2} \quad (8)$$

## 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  $Q_{cv}^2 > 0.5$  for a satisfactory model when for the excellent model  $Q_{cv}^2 > 0.9$ . According to them, given a test game, a model will be efficient if the acceptance criterion  $R^2 - Q_{cv}^2 < 0.3$  is respected.

According to Tropsha *et al.* [37] [38] [39], for the external validation set, the predictive power of a model can be obtained from five criteria. These criteria are:

- 1)  $R_{Test}^2 > 0.7$ , 2)  $Q_{Cv\,Test}^2 > 0.6$ , 3)  $|R_{Test}^2 - R_0^2| \leq 0.3$ ,
- 4)  $\frac{|R_{Test}^2 - R_0^2|}{R_{Test}^2} < 0.1$  and  $0.85 \leq k \leq 1.15$ , 5)  $\frac{|R_{Test}^2 - R_0'^2|}{R_{Test}^2} < 0.1$  and  $0.85 \leq k' \leq 1.15$

## 2.6. Applicability Domain

The applicability domain principle helps modelers to specify the scope of proposed models, thereby defining the model's limitations with respect to its structural domain and chemical space. If an external compound exceeds the defined scope of a model, it is outside the applicability domain of that model and cannot be associated with reliable prediction. There are several methods for determining the applicability domain of a QSAR model, among which we find the lever method that is used the most. If a compound has a residual and a lever that exceeds the threshold  $h^* = 3p/n$  (where  $p$  is the number of descriptors plus 1 and  $n$  the number of observations), this compound is considered outside the field of applicability of the elaborate model. The field of applicability will be discussed using the Williams diagram which represents the standardized prediction residuals as a function of the values of the hi levers [40]. For each compound  $i$  in the original space of the independent variables ( $X_i$ ), the value of  $h_i$  is calculated by the following relation [41]:

$$h_i = X_i (X^T X)^{-1} X_i^T \quad (9)$$

where  $i = (1, \dots, n)$

With:  $X_i$  is the line vector of the descriptors of the compound  $i$ ,  $X(n * k - 1)$  is the matrix of the model deduced from the values of the descriptors of the training set; the index T designates the transposed matrix of the matrix. The critical value of the lever ( $h^*$ ) is set [42] to:

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

With  $n$ , the number of test compounds used;  $k$  is the number of the descriptors of the model.

If  $h_i < h^*$ , the probability of agreement between the measured and predicted values of the compound "I" is as high as that of the compounds in the database. Compounds with  $h_i > h^*$  reinforce the model when they belong to the training set but will otherwise have dubious predicted values without being necessarily

aberrant, the residues being low [43].

### 3. Results and Discussion

The set of descriptor values of the thirteen (13) hydroxamic acid molecules of the test set and the six (6) other molecules of the validation set are presented in **Table 2**.

#### 3.1. Principal Component Analysis (PCA) and Ascendant Hierarchical Classification (AHC)

All three descriptors for the 19 hydroxamic acid compounds are subjected to the PCA analysis. The two main axes are enough to describe the information provided by the data matrix. The correlations between the three descriptors are presented in **Table 3** according to a correlation matrix and in **Figure 2** where these descriptors are represented in a correlation circle.

**Table 2.** Experimental quantum and potential descriptors of test and validation sets.

Compounds	AE	$\nu(\text{O-H})$	$\nu(\text{N-H})$	pIC <sub>50</sub>
<b>Training set</b>				
1	1.305	3562.210	3631.280	5.535
2	0.511	3560.910	3630.910	6.069
3	0.768	3561.380	3630.410	6.202
4	0.630	3560.950	3629.760	6.018
5	0.502	3561.920	3631.460	6.507
6	1.123	3562.720	3631.190	6.114
7	0.470	3561.710	3631.710	6.444
8	1.022	3562.170	3632.030	5.836
9	0.585	3562.160	3632.450	6.276
10	0.848	3559.450	3619.450	6.312
11	1.138	3560.340	3630.790	5.348
12	1.122	3560.730	3632.510	5.098
13	1.157	3560.480	3632.420	5.115
<b>Validation set</b>				
14	0.586	3561.610	3630.480	6.450
15	1.082	3557.640	3631.620	4.410
16	1.074	3560.81	3632.76	5.100
17	1.409	3561.980	3631.790	5.252
18	1.397	3560.790	3631.690	4.897
19	1.424	3561.520	3631.710	5.357

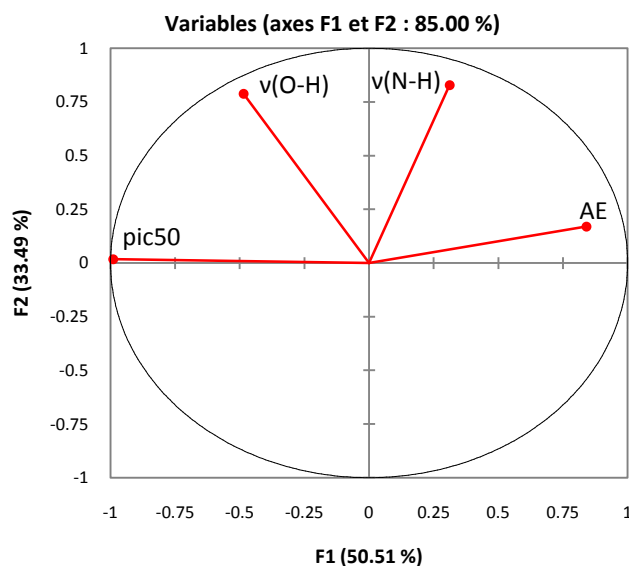
AE in electron Volt (eV),  $\nu(\text{O-H})$  in  $\text{Cm}^{-1}$  and  $\nu(\text{N-H})$  in  $\text{Cm}^{-1}$ , IC<sub>50</sub> ( $\mu\text{M}$ ).



**Table 3.** Correlation matrix (Pearson (n)) between the different descriptors.

Variables	AE	$\nu(\text{O-H})$	$\nu(\text{N-H})$	pIC <sub>50</sub>
AE	1			
$\nu(\text{O-H})$	0.0917	1		
$\nu(\text{N-H})$	-0.1845	-0.6225	1	
pIC <sub>50</sub>	<b>-0.9264</b>	-0.1242	-0.0209	1

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

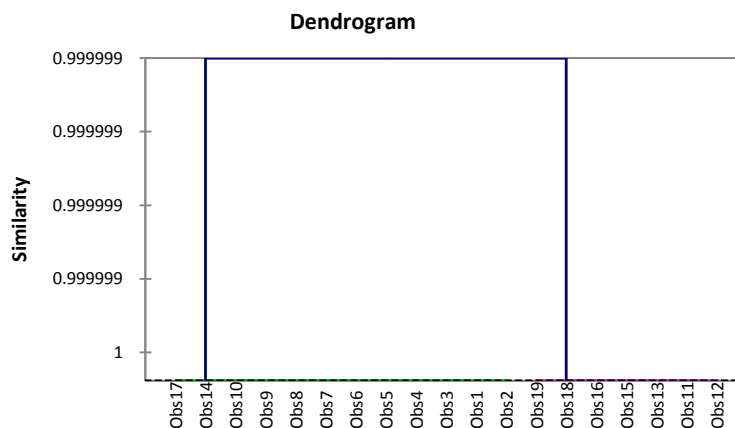
**Figure 2.** Correlation Circle descriptors and the explained variable.

The two main axes are enough to characterize the different descriptors. In fact, the variance percentages are 50.51% and 33.49% for the F1 and F2 axes, respectively. The total information is estimated at 85%. Principal Component Analysis (PCA) [29] was conducted to identify the link between the different descriptors. Bold values are different from 0 at a significance level of  $p = 0.05$ .

The matrix obtained provides information on the negative or positive correlation between the variables. The Pearson correlation coefficients are summarized in **Table 3**. The resulting matrix provides information on the negative or positive correlation between the variables.

The correlation circle was made to detect the connection between the different descriptors. The analysis of the principal components from the correlation circle (**Figure 2**) revealed that the F1 axis (50.51% of the variance) seems to represent the vibration frequencies  $\nu(\text{OH})$  and  $\nu(\text{NH})$ , and the axis F2 (33.49% of the variance) seems to represent the electronic affinity AE.

The AHC of **Figure 3** distributes the iHDAC7 in two classes according to the affinity of one compound to another. The two major classes consist of compounds as follows: **C**<sub>1</sub> (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 14, 17) and **C**<sub>2</sub> (11, 12, 13, 15, 16, 18, 19).



**Figure 3.** Dendrogram of the partition hierarchy of the 19 hydroxamic acids for 3 homogeneous classes.

### 3.2. Prediction of HDAC7 Anti-Cancer Activity from MLR and NMR Models

The equations of the QSAR models obtained for cancer activity from MLR and NMR as well as the statistical indicators are given in **Table 4**. It should be emphasized that these models were established using the same test and validation **Table 3**.

The equations of the different models are obtained by using three descriptors (AE,  $\nu(\text{O-H})$  and  $\nu(\text{N-H})$ ) determined from the optimized molecules. It is important to note that the negative or positive sign of a model descriptor coefficient reflects the proportionality effect between the evolution of the biological activity and this parameter of the MLR model equation. Thus, the negative sign indicates that when the value of the descriptor is high, the biological activity decreases while the positive sign reflects the opposite effect. For the MLR model, the negative signs of the coefficients of the three descriptors (AE,  $\nu(\text{O-H})$  and  $\nu(\text{N-H})$ ) indicate that the HDAC7 activity will be improved for low values of these descriptors. The study of the significance of these different models is led by the evaluation of the statistical indicators and by the acceptance criteria of Erickson *et al.* and Tropsha *et al.* The values of the statistical indicators determined for each model are reported in **Table 5**. The values of the statistical indicators listed in this table reflect a good correlation of the inhibitory activity on HDAC7 with the different descriptors.

In these respective models (MLR and NMR), 96.56% and 97.62% of the descriptors (AE,  $\nu(\text{O-H})$  and  $\nu(\text{N-H})$ ) are considered with the standard deviation  $S$  of prediction from 0.488 to 0.663. The significance of these models is given by the Fischer  $F$  test, 85.00 to 172.79 respectively for the MLR and NMR. The correlation coefficient of the cross-validation  $Q_{CV}^2$  is 0.918 to 0.938 respectively for the RMNL and RML models. These values reflect excellent models according to Erikson *et al.* [35] [36]. These models are also acceptable because they agree with the acceptance criteria of these authors:  $R^2 - Q_{CV}^2 < 0.3$

**Table 4.** Most Significant QSAR Models for Modeling HDAC7 Activity from MLR and NMR.

Regression equations	
MRL	$pIC_{50}^{pred} = -902.89167 - 1.16813 * AE + 0.35349 * \nu(O-H) - 0.09615 * \nu(N-H)$
NMR	$pIC_{50}^{pred} = 186539 - 1.27488 * AE - 121.60150 * \nu(O-H) + 16.28351 * \nu(N-H) + 0.03043 * AE^2 + 0.01712 * \nu(O-H)^2 - 0.00226 * \nu(N-H)^2$

**Table 5.** Statistical indicators of multilinear regression.

Statistical indicators of multilinear regression	MLR	NMR
Number of compounds ( <i>n</i> )	19	19
Coefficient of determination $R^2$	0.9659	0.9762
Standard deviation $S$	0.488	0.663
Fisher's test $F$	85.00	172.79
Coefficient of correlation of cross validation $Q_{cv}^2$	0.95	0.9759
$R^2 - Q_{cv}^2$	0.0159	0.000
Activity field $IC_{50}^{exp}$ ( $\mu M$ )	0.311 à 38.9	
Trust level $\alpha$	95 %	

The regression line between the experimental and theoretical nematocidal potentials of the test set and the validation set is illustrated in **Figure 4**. This figure illustrates the correlation between the experimental and theoretical IC50 inhibition concentrations of the test set (blue dots) and the validation game (red dots). These models obtained relate the HDAC7 activity and the theoretical descriptors of hydroxamic acids.

The low values of the standard error of 0.448 to 0.663 respectively of the NMR and RML models attest to the good similarity between the predicted and experimental values of the HDAC7 activity despite some differences recorded (**Figure 5**).

#### Verification of Tropsha Criteria

The statistical indicators of the five (5) Tropsha criteria of these two models (MLR and NMR) of the validation set are given in **Table 6**.

$$R_{Test}^2 > 0.7, Q_{Cv\ Test}^2 > 0.6, |R_{Test}^2 - R_0^2| \leq 0.3$$

$$\frac{|R_{Test}^2 - R_0^2|}{R_{Test}^2} < 0.1 \text{ and } 0.85 \leq k \leq 1.15; \frac{|R_{Test}^2 - R_0'^2|}{R_{Test}^2} < 0.1 \text{ and } 0.85 \leq k' \leq 1.15$$

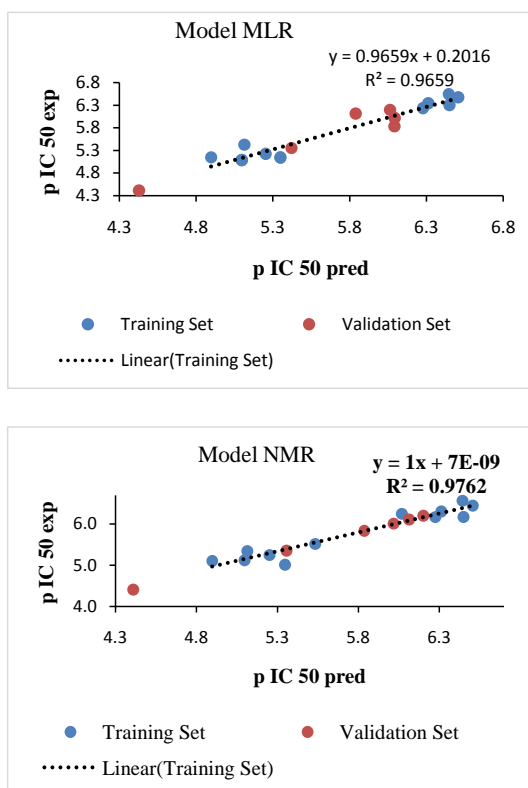
All values meet the Tropsha criteria, so these models are acceptable for predicting HDAC7 anticancer activity.

However, these two models being a function of three theoretical descriptors, it is essential to determine the contribution of each in the prediction of the anticancer activity HDAC7 of the hydroxamic acids studied. Indeed, the knowledge of this contribution makes it possible to establish the order of priority of the

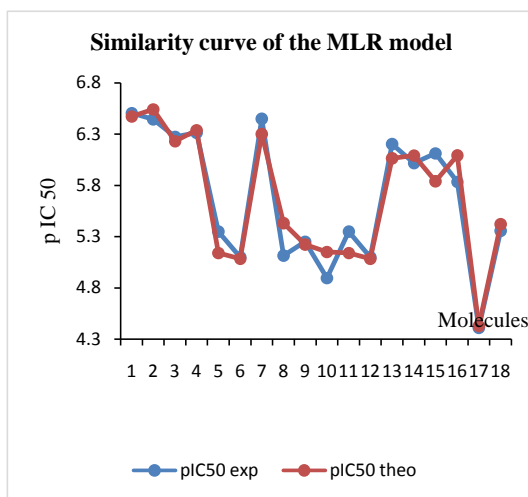
various descriptors and to define the choice of the parameters to be optimized for the realization of a better activity HDAC7.

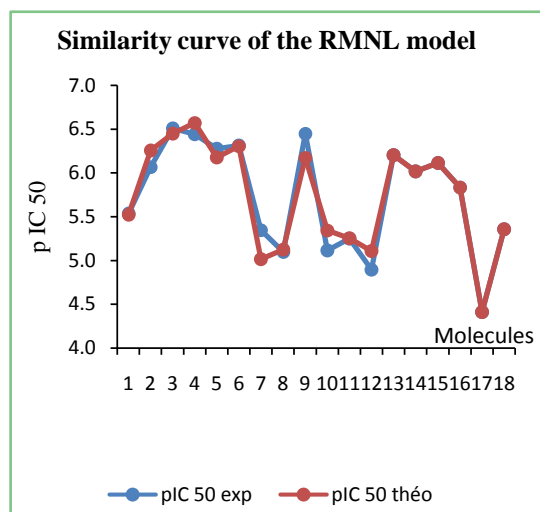
**Table 6.** Tropsha criteria for different models.

HDAC7	$R_{Test}^2$	$Q_{CV\ Test}^2$	$R_{Test}^2 - R_0^2$	$\frac{ R_{Test}^2 - R_0^2 }{R_{Test}^2}$	$\frac{ R_{Test}^2 - R_0^2 }{R_{Test}^2}$	$k$	$k'$
MLR	0.9809	0.9504	0.0301	0.000	0.000	1.001	0.998
NMR	0.9807	0.9511	0.000	0.000	0.001	0.9974	1.002



**Figure 4.** Regression lines of the different models (MLR and NMR).





**Figure 5.** Similarity curve of the experimental and predicted values of the RML and NMR models.

### 3.3. Analysis of the Contribution of Descriptors

The contribution of the four descriptors of this model in the prediction of the anticancer activity HDAC7 of the hydroxamic acids was determined from the software XLSTAT version 2014 [20]. The different contributions are illustrated in **Figure 6**.

In this study, the electronic affinity (AE) has a nearly identical weight as the vibration frequencies  $\nu(\text{O-H})$  and  $\nu(\text{N-H})$ . The absence of one of these descriptors in the model could destabilize this one. It should be noted that these quantum descriptors in a global way make a rather important contribution in the prediction of the anticancer activity of HDAC7.

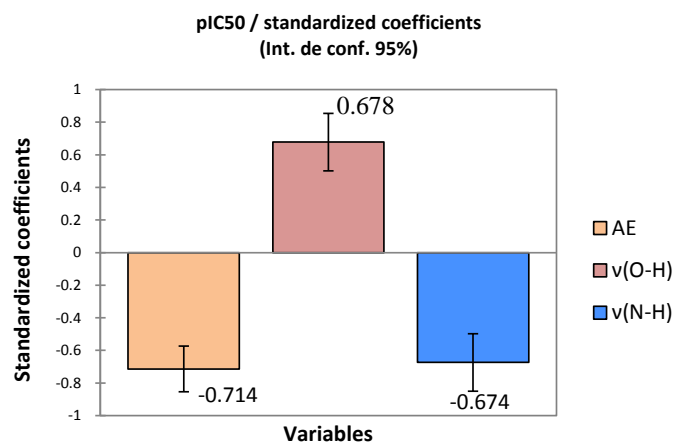
### 3.4. Applicability Domain Analysis

The values of leverage and standardized residuals of the observables of the model used to develop the applicability domain of this model are shown in **Table 7**.

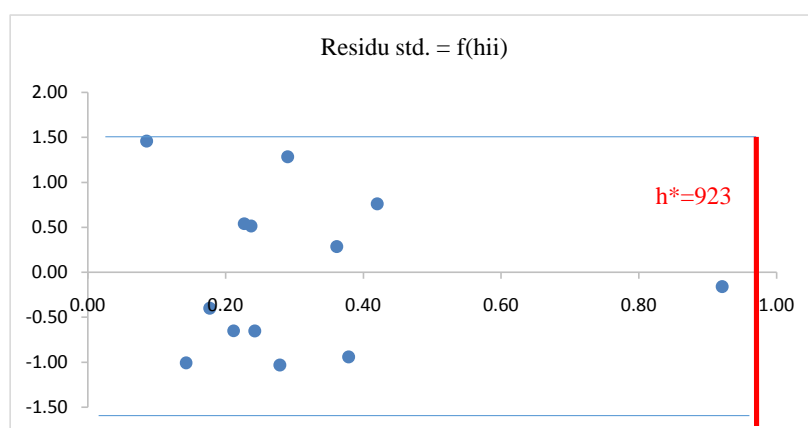
Analysis of the data in **Table 7** shows that all other observations have their standardized residuals between  $-1.5$  and  $1.5$ . The lifts obtained are all below the critical value  $h^* = 0.923$ . This fact is elucidated by the Williams diagram (**Figure 7**). The results of the external validation and the domain of applicability show that the established model can be used reliably for the prediction of the inhibitory concentration of future hydroxamic acids.

## 4. Conclusion

In this work, the anticancer activity HDAC7 of nineteen (19) hydroxamic acid compounds was correlated with the theoretical descriptors calculated by the DFT methods. The descriptors electronic affinity (AE), vibration frequencies  $\nu(\text{N-H})$  and  $\nu(\text{O-H})$  can explain and predict the anticancer activity HDAC7. Statistical methods such as Principal Component Analysis (PCA), Ascending Hierarchical Classification (AHC), Multilinear and Nonlinear Regression were used.



**Figure 6.** Contribution of descriptors in models.



**Figure 7.** Williams diagram of the MLR model.

**Table 7.** Values of the levers and residuals of observations of the model.

Compounds	pic50-exp	Pred (pic50)	Residu	Residu std.	hii
1	5.535	5.633	-0.098	-0.942	0.38
2	6.069	6.136	-0.068	-0.653	0.24
3	6.202	6.050	0.152	1.460	0.09
4	6.018	6.123	-0.105	-1.008	0.14
5	6.507	6.451	0.056	0.540	0.23
6	6.114	6.034	0.079	0.760	0.42
7	6.444	6.390	0.053	0.514	0.24
8	5.836	5.877	-0.042	-0.399	0.18
9	6.276	6.343	-0.068	-0.651	0.21
10	6.312	6.329	-0.017	-0.159	0.92
11	5.348	5.214	0.134	1.285	0.29
12	5.098	5.205	-0.107	-1.032	0.28
13	5.115	5.085	0.030	0.285	0.36

Two QSAR models (MLR, NMR) showed that the descriptors used would predict, at an acceptable level of confidence, the inhibitory activity of hydroxamic acids. The use of two different methods in this work was to show on the one hand that from these descriptors, we can predict in a different way the inhibitory activity of hydroxamic acids on HDAC7 and on the other hand, the relevance has these descriptors. However, the MLR model ( $R^2 = 0.9659$ ,  $S = 0.488$ ,  $F = 85$  and **p-value** < **0.0001**) is an effective tool for predicting HDAC7 anticancer activity. Moreover, the study of the contribution of the descriptors showed that these descriptors are almost equivalent in the prediction of the inhibitory activity of the HDAC7 studied. For this model, the future compounds must have their standardized residue between  $-1.5$  and  $+1.5$  with a threshold lever  $h^* = 0.923$ . A study of the applicability domain of these models is envisaged. From the field of applicability and quantum descriptors elaborated in this work, we plan to propose new molecules with improved activities.

### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

### References

- [1] Yoon, S. and Eom, G.H. (2016) HDAC and HDAC Inhibitor: From Cancer to Cardiovascular Diseases. *Chonnam Medical Journal*, **52**, 1-12. <https://doi.org/10.4068/cmj.2016.52.1.1>
- [2] Witt, O., Deubzer, H.E., Milde, T. and Oehme, I. (2009) HDAC Family: What Are the Cancer Relevant Targets? *Cancer Letters*, **277**, 8-21. <https://doi.org/10.1016/j.canlet.2008.08.016>
- [3] Micelli, C. and Rastelli, G. (2015) Histone Deacetylases: Structural Determinants of Inhibitor Selectivity. *Drug Discovery Today*, **20**, 718-735. <https://doi.org/10.1016/j.drudis.2015.01.007>
- [4] Haberland, M., Montgomery, R.L. and Olson, E.N. (2009) The Many Roles of Histone Deacetylases in Development and Physiology: Implications for Disease and Therapy. *Nature Reviews Genetics*, **10**, 32-42. <https://doi.org/10.1038/nrg2485>
- [5] Zhang, C.L., McKinsey, T.A., Chang, S., Antos, C.L., Hill, J.A. and Olson, E.N. (2002) Class II Histone Deacetylases Act as Signal-Responsive Repressors of Cardiac Hypertrophy. *Cell*, **110**, 479-488. [https://doi.org/10.1016/S0092-8674\(02\)00861-9](https://doi.org/10.1016/S0092-8674(02)00861-9)
- [6] Vega, R.B., *et al.* (2004) Histone Deacetylase 4 Controls Chondrocyte Hypertrophy during Skeletogenesis. *Cell*, **119**, 555-566. <https://doi.org/10.1016/j.cell.2004.10.024>
- [7] Chang, S., Young, B.D., Li, S., Qi, X., Richardson, J.A. and Olson, E.N. (2006) Histone Deacetylase 7 Maintains Vascular Integrity by Repressing Matrix Metalloproteinase 10. *Cell*, **126**, 321-334. <https://doi.org/10.1016/j.cell.2006.05.040>
- [8] Jones, P., *et al.* (2008) Probing the Elusive Catalytic Activity of Vertebrate Class IIa Histone Deacetylases. *Bioorganic & Medicinal Chemistry Letters*, **18**, 1814-1819. <https://doi.org/10.1016/j.bmcl.2008.02.025>
- [9] Ouaiissi, M., *et al.* (2008) High Histone Deacetylase 7 (HDAC7) Expression Is Significantly Associated with Adenocarcinomas of the Pancreas. *Annals of Surgical Oncology*, **15**, 2318-2328. <https://doi.org/10.1245/s10434-008-9940-z>

- [10] Witt, A.E., *et al.* (2017) Identification of a Cancer Stem Cell-Specific Function for the Histone Deacetylases, HDAC1 and HDAC7, in Breast and Ovarian Cancer. *Oncogene*, **36**, 1707-1720. <https://doi.org/10.1038/onc.2016.337>
- [11] Parra, M. (2015) Class IIa HDACs—New Insights into Their Functions in Physiology and Pathology. *The FEBS Journal*, **282**, 1736-1744. <https://doi.org/10.1111/febs.13061>
- [12] Ouaiissi, M., Giger, U., Sielezneff, I., Pirrò, N., Sastre, B. and Ouaiissi, A. (2011) Rationale for Possible Targeting of Histone Deacetylase Signaling in Cancer Diseases with a Special Reference to Pancreatic Cancer. *Journal of Biomedicine and Biotechnology*, **2011**, Article ID 315939. <https://doi.org/10.1155/2011/315939>
- [13] Barneda-Zahonero, B., *et al.*, (2015) The Transcriptional Repressor HDAC7 Promotes Apoptosis and c-Myc Downregulation in Particular Types of Leukemia and Lymphoma. *Cell Death & Disease*, **6**, 16-35. <https://doi.org/10.1038/cddis.2014.594>
- [14] Lei, Y., *et al.* (2017) Hdac7 Promotes Lung Tumorigenesis by Inhibiting Stat3 Activation. *Molecular Cancer*, **16**, 170. <https://doi.org/10.1186/s12943-017-0736-2>
- [15] Yao, Y., Liao, C.Z., Li, Z., Wang, Z., *et al.* (2014) 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*, **86**, 639-652. <https://doi.org/10.1016/j.ejmech.2014.09.024>
- [16] Kapp, J., Remko, M. and Schleyer, P.V.R. (1996) H<sub>2</sub>XO and (CH<sub>3</sub>)<sub>2</sub>XO 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*, **118**, 5745-5751. <https://doi.org/10.1021/ja953846p>
- [17] Johnson, B.G., Gill, P.M. and Pople, J.A. (1993) The Performance of a Family of Density Functional Methods. *The Journal of Chemical Physics*, **98**, 5612-5626. <https://doi.org/10.1063/1.464906>
- [18] Frisch, M., Trucks, G., Schlegel, H. and Scuseria, G. (2009) Gaussian 09, Revision C.01. Gaussian Inc., Wallingford.
- [19] Microsoft Excel (2013) (15.0.4420.1017) MSO (15.0.4420.1017) 64 Bits Partie de Microsoft Office Professionnel Plus.
- [20] XLSTAT (2014) Version 2014.5.03 Copyright Addinsoft 1995-2014 XLSTAT and Addinsoft Are Registered Trademarks of Addinsoft. <https://www.xlstat.com>
- [21] Koopmans, T. (1934) Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den einzelnen Elektronen eines Atoms. *Physica*, **1**, 104-113. [https://doi.org/10.1016/S0031-8914\(34\)90011-2](https://doi.org/10.1016/S0031-8914(34)90011-2)
- [22] Balaban, A.T. (1997) From Chemical Topology to Three-Dimensional Geometry. Plenum Press, New York, 1-24.
- [23] Randić, M. and Razinger, M. (1995) Molecular Topographic Indices. *The Journal for Chemical Information and Computer Scientists*, **35**, 140-147. <https://doi.org/10.1021/ci00023a021>
- [24] Rastija, V. and Medic-Š., M. (2009) QSAR Modeling of Anthocyanins, Anthocyanidins and Catechins as Inhibitors of Lipid Peroxidation Using Three-Dimensional Descriptors. *Medicinal Chemistry Research*, **18**, 579-588. <https://doi.org/10.1007/s00044-008-9151-y>
- [25] Larif, M., Adad, A., Hmamouchi, R., Taghki, A.I., Soulaymani, A., Elmidaoui, A., Bouachrine, M. and Lakhlifi, T. (2013) Biological Activities of Triazine Derivatives. Combining DFT and QSAR Results. *Arabian Journal of Chemistry*.



- [26] Traoré, Y., Koné, M.G.-R., Ouattara, O. and Ziao, N. (2018) Qsar Approach to Estimating the Analgesic Activity of a Serie of Tri-Substituted Pyrimidine Derivatives. *Journal of Computational Chemistry & Molecular Modelling*, **2**, 1-14.
- [27] Kpidi, Y.H., Yapo, O.B., Koné, M.G.-R., Gadj, G.A., Gnagne, A.E.J.E.Y., N'dri, J. and Ziao, S.N. (2018) Monitoring and Modeling of Chlorophyll-a Dynamics in a Eutrophic Lake: M'koa Lake (Jacqueville, Ivory Coast). *American Journal of Environmental Protection*, **6**, 1-9. <https://doi.org/10.12691/env-6-1-1>
- [28] Hmamouchi, R., Taghki, A.I., Larif, M., Adad, A., Abdellaoui, A., Bouachrine, M. and Lakhli, T. (2013) The Inhibitory Activity of Aldose Reductase of Flavonoid Compounds: Combining DFT and QSAR Calculations. *Journal of Chemical and Pharmaceutical Research*, **5**, 198-209.
- [29] Escofier, B. and Pagès, J. (2008) Analyses factorielles simples et multiples: Objectifs, méthodes et interprétation. Dunod, Paris.
- [30] Rücker, C., Rücker, G. and Meringer, M. (2007)  $\gamma$ -Randomization and Its Variants in QSPR/QSAR. *Journal of Chemical Information and Modeling*, **47**, 2345-2357. <https://doi.org/10.1021/ci700157b>
- [31] N'guessan, K.N., Koné, M.G.-R., Bamba, K., Ouattara, W.P. and Ziao, N. (2017) Quantitative Structure Anti-Cancer Activity Relationship (QSAR) of a Series of Ruthenium Complex Azopyridine by the Density Functional Theory (DFT) Method. *Computational Molecular Bioscience*, **7**, 19-31. <https://doi.org/10.4236/cmb.2017.72002>
- [32] Snedecor, G.W. and Cochran, W.G. (1967) *Methods, Statistical*. Oxford and IBH, New Delhi, 381.
- [33] Kangah, N.J.-B., Koné, M.G.-R., Kodjo, C.G., N'guessan, B.R., Kablan, A.L.C., Yéo, S.A. and Ziao, N. (2017) 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*, **6**, 38-43.
- [34] Esposito, E.X., Hopfinger, A.J. and Madura, J.D. (2004) Methods for Applying the Quantitative Structure-Activity Relationship Paradigm. *Methods in Molecular Biology*, **275**, 131-213. <https://doi.org/10.1385/1-59259-802-1:131>
- [35] Eriksson, L., Jaworska, J., Worth, A., Cronin, M.D., Dowell, R.M.M. and Gramatica, P. (2003) Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs. *Environmental Health Perspectives*, **111**, 1361-1375. <https://doi.org/10.1289/ehp.5758>
- [36] N'Dri, J.S., Koné, M.G.-R., Kodjo, C.G., Affi, S.T., Kablan, A.L.C., Ouattara, O., Soro, D. and Ziao, N. (2017) Relation Quantitative Structure Activité (QSAR) d'une série d'azetidines dérivés de Dapsone par la méthode de Théorie de la fonctionnelle de la densité(DFT). *International Journal of Applied Sciences*, **8**, 55-62. <https://doi.org/10.21013/jas.v8.n2.p2>
- [37] Golbraikh, A. and Tropsha, A. (2002) Beware of QSAR. *Journal of Molecular Graphics and Modelling*, **20**, 269-276. [https://doi.org/10.1016/S1093-3263\(01\)00123-1](https://doi.org/10.1016/S1093-3263(01)00123-1)
- [38] Ouattara, O., Koné, M.G.-R., Affi, T.S., Bamba, K., Traore, Y. and Ziao, N. (2018) Contribution to the Molecular Lipophilicity Scale by Qspr Models of Lipophilicity Prediction. *Journal of Engineering Research and Application*, **8**, 55-61.
- [39] Ouattara, O., Affi, T.S., Koné, M.G.-R., Bamba, K. and Ziao, N. (2017) Can Empirical Descriptors Reliably Predict Molecular Lipophilicity? A QSPR Study Investiga-

- tion. *International Journal of Engineering Research and Applications*, **7**, 50-56. <https://doi.org/10.9790/9622-0705015056>
- [40] Sahigara, F., *et al.* (2012) Comparison of Different Approaches to Define the Applicability Domain of QSAR Models. *Molecules*, **17**, 4791-4810. <https://doi.org/10.3390/molecules17054791>
- [41] Tropsha, A., Gramatica, P. and Gombar, V.K. (2003) The Importance of Being Earnest, Validation Is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR & Combinatorial Science*, **22**, 69-77. <https://doi.org/10.1002/qsar.200390007>
- [42] Netzeva, T.I., *et al.* (2005) Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships. The Report and Recommendations of ECVAM Workshop 52. *Alternatives to Laboratory Animals*, **33**, 155-173. <https://doi.org/10.1177/026119290503300209>
- [43] Dearden, J.C. (2016) The History and Development of Quantitative Structure-Activity Relationships (QSARs). *International Journal of Quantitative Structure-Property Relations*, **1**, 44. <https://doi.org/10.4018/IJQSPR.2016010101>