J. Mater. Environ. Sci. 7 (7) (2016) 2252-2258
ISSN : 2028-2508
CODEN: JMESC

DFT-based QSAR study of N-hydroxyfurylacrylamide and its derivatives, principal inhibitors of histone deacetylase

F. Z. Tassine*, M. Elhallaoui

LIMME, Department of chemistry Faculty of Sciences Dhar El Mahraz, University Sidi Mohammed Ben Abdellah, B.P. 1796, Atlas. FES, Morocco.

Received 14 Jan 2016, Revised 02 Apr 2016, Accepted 11 Apr 2016
**Corresponding author. E-mail: tassinefatima@hotmail.fr F. Z. Tassine; Phone: +212 649053650*

Abstract

A data set of 38 N-hydroxyfurylacrylamide derivatives, principal inhibitors of histone deacetylase (HDAC) is used for quantitative structure-activity relationship (QSAR) study. The physicochemical properties of the tested compounds have been described by a total of six descriptors comprising four quantum chemical descriptors and two molecular descriptors. Molecules geometries optimization is performed using firstly DFT method with B3LYP/6-31G (d) level, to generate descriptors based on electronic properties, and secondly molecular mechanics method with MM+ force field to generate molecular descriptors. QSAR models were constructed using Multiple Linear Regression method (MLR) and Artificial Neural Network (ANN) techniques. Descriptors based on structural and electronic properties were shown to be important in classifying the compounds. Good correlations are shown both for MLR and ANN towards HDAC inhibitory activities (0.80 and 0.98, respectively). So, ANN model established with ANN techniques considering the relevant descriptors selected by MLR, is statistically significant and show very good stability towards data variation with leave-one-out (LOO) test, the most convenient procedure of cross validation method (r_{LOO} =0.77).

Keywords: N-hydroxyfurylacrylamide, Histone deacetylases (HDACs), DFT, 3D-QSAR, MLR, ANN.

1. Introduction

Epigenetics refers to the heritable changes in gene expression or phenotype that are stable between cell division, and sometimes between generation, but do not involve changes in the underlying DNA sequence of the organism [1,2]. Epigenetic alteration including DNA methylation, Histone modifications, nucleosome positioning, and small noncoding (mi RNA, si RNA), play an important role in cancer progression [3-5]. The histone proteins are widely described as essential players in the control of expression via modification though chemical such as acetylation, phosphorylation, and methylation. The identification of cancer-related epigenetic changes that can be genome-wide, or more restricted and involving altered expression or activity of defend epigenetic regulatory protein; provide a strong rationale for the use of epigenetic-based therapies such as HDACs. Most studies to date have focused on the aberrant recruitment of HDACs to promoters though their physical association with oncogenic, or overexpression of repressive transcription factors that physically interact with HDACs. Histone deacetylases (HDACs) catalyse the removal of acetyl groups from lysine residues in histone amino termini, leading to chromatin condensation and transcriptional repression [6,7].

Quantitative structure-activity relationship (QSAR) method is generally defined as an application of mathematical and statistical methods to establish empirical relationship (QSAR models) of the form Pi= K (D1, D2,…, Dn), where Pi are biological activities (or other properties of interest) of molecule, D1, D2,…., Dn representing structural properties (molecular descriptors) of compound, which are calculated (or, sometimes,

experimentally measured), and K characterizes an empirically established transformation that captures the mathematics relationship between the molecular descriptors and their bioactivities.
In this study, Multiple Linear Regression (MLR) analysis and Artificial Neural Network (ANN) calculations are applied to a series of 38 N- hydroxyfurylacrylamide derivatives, in order to set up a 3D-QSAR model able to predict inhibitory histone deacetylases.

2. Materials and methods

2.1. Data set

The data set constituted of 38 N-hydroxyfurylacrylamide derivatives and their HDACs inhibitory activities values as presented in table 1, is obtained from the work of Taotao Feng and Hai Wang [8]. The pharmacological activity IC_{50} has been expressed in [μM]. In this work the activity HDACs is expressed in the logarithmic form ($logIC_{50}$).
SAHA (suberoylanilide hydroxamic acid), a potent HDACs inhibitor that has been registered as antitumor drug, was used as the positive control of HDAC-8 assay.

2.2. Calculation of molecular descriptors

3-D modeling and calculations were performed using the Gaussian 03 quantum chemistry package [9] for the whole of molecules. To save computational time, initial geometry optimizations were carried out with molecular mechanics (MM) method using the MM+ force field. The lowest energy conformations of the molecules obtained by the MM method were further optimized by the DFT method [10], by employing Becke's three-parameter hybrid functional (B3LYP) [11], with a 6-31G basis set. In the other hand, Advanced Chemistry Development's ChemDraw Ultra 8.0 is used to calculate Melting Point (MP) and Molar Refractivity (MR). The totality of descriptors used in this work is listed in table 2.

2.3. Multiple Linear Regression (MLR)

Multiple linear regression (MLR) is one of the most frequently applied methods in generating QSAR models [12]. The ability of MLR to treat intercorrelated variables and missing data, as well as the fact that can consider only one dependent variable in each model can be overcome though partial least squares which is also widely used in QSAR studies. So For the purpose of this work, the data set is submitted to MLR Analysis, using the software SYSTAT, version 12, so that, the proposed MLR model served primarily to select the descriptors used as the input parameters for a back propagation network (NN), and in the other hand it could be used to predict inhibitory activities $logIC_{50}$.

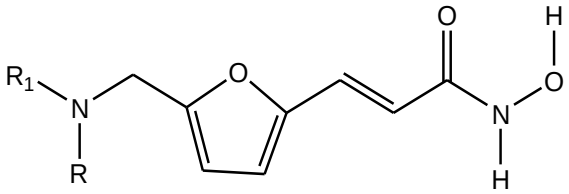
2.4 Artificial Neural Network (ANN)

Neural networks are artificial systems; they use a large number of interrelated data-processing neurons to emulate the function of brain. Although there are a number of different ANN models in use today, the most frequently used type of ANN in QSAR, and the one employed in our research, is the three-layered back propagation neural network in the back propagation strategy, the neurons are arranged in an input layer, a hidden layer, and an output layer. Each neuron in any layer is fully connected with the neurons of another layer, and there are no connections between neurons in the same layer. The network received a set of inputs as the training set. After training, a nonlinear transfer function was applied to each node in the hidden layer. The goal of training the network is to optimize the weights between the layers so as to minimize the output errors [13, 14]. All calculations of ANN are done on Matlab 7 using our program written in C language. The final ANN architecture is developed in part of neural networks.

2.5. Cross Validation

Leave-one-out cross-validation (LOO-CV) was used as a data sampling approach to separate the data set as training and testing sets [15]. For each data set, an input-output model is developed. The model is evaluated by measuring its accuracy in predicting the responses of the remaining data (the ones that have not been used in the development of the model).

Table 1: Studied compounds and their observed Inhibitory activities against HDACs $logIC_{50}$ (Obs), and calculated $logIC_{50}$ with MLR; ANN and CV methods.



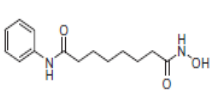
N°	R	R_1	$logIC_{50}$(Obs)	$logIC_{50}$(MLR)	$logIC_{50}$(ANN)	$logIC_{50}$(CV)
1	Ethyl	Benzyl	-0.8860	-0.1825	-0.8809	-0.9095
2	Propyl	Benzyl	-0.7695	-0.0678	-0.7859	-0.6402
3	Iso-propyl	Benzyl	0.0718	0.0874	0.1322	0.0740
4	Butyl	Benzyl	-0.2676	-0.0310	-0.2768	-0.0572
5	Iso-butyl	Benzyl	0.5010	-0.1317	0.4846	0.2720
6	Hexyl	Benzyl	0.4232	0.1199	0.4205	0.3682
7	Octyl	Benzyl	0.4065	0.3725	0.3474	0.3042
8	2-Hydroxy-ethyl	Benzyl	-0.7695	-0.6325	-0.7701	-0.9800
9	2-Phenoxyethyl	Benzyl	0.4149	0.0267	0.3300	0.3041
10	2-(p-Tolyloxy)ethyl	Benzyl	0.2718	0.1517	0.3367	0.2970
11	3-Phenoxypropyl	Benzyl	0.1702	0.3098	0.2879	0.2861
12	3-(p-Tolyloxy)propyl	Benzyl	0.4048	0.2871	0.2895	0.3162
13	4-Phenoxybutyl	Benzyl	0.3263	0.2049	0.2943	0.3222
14	4-(p-Tolyloxy)butyl	Benzyl	0.1172	0.3991	0.2729	0.1403
15	Ethyl	4-Phenylbenzyl	0.0961	-0.0226	0.2219	0.1777
16	Iso-propyl	4-Phenylbenzyl	0.2068	0.2498	0.3771	0.3114
17	Amyl	4-Phenylbenzyl	0.3802	0.2930	0.3013	0.3350
18	Octyl	4-Phenylbenzyl	0.3360	0.5147	0.2590	0.2893
19	Benzyl	4-Phenylbenzyl	0.4149	0.2145	0.2949	0.3051
20	2-Hydroxy-ethyl	4-Phenylbenzyl	-1	-1.0002	-1.0077	0.2210
21	2-Phenoxyethyl	4-Phenylbenzyl	0.0827	0.2423	0.2663	0.2458
22	2-(p-Tolyloxy)ethyl	4-Phenylbenzyl	0.2329	0.4529	0.2577	0.2632
23	3-Phenoxypropyl	4-Phenylbenzyl	0.3424	0.3160	0.2570	0.2824
24	3-(p-Tolyloxy)propyl	4-Phenylbenzyl	0.2201	0.3249	0.2595	0.2630
25	4-Phenoxybutyl	4-Phenylbenzyl	0.3222	0.4704	0.2529	0.3157
26	4-(p-Tolyloxy)butyl	4-Phenylbenzyl	0.3010	0.5406	0.2520	0.2916
27	Ethyl	Homopiperonyl	0.3636	-0.0076	0.2039	0.0488
28	Propyl	Homopiperonyl	0.2013	0.0423	0.3060	0.2991
29	Butyl	Homopiperonyl	0.4471	0.3148	0.3834	0.3234
30	Iso-butyl	Homopiperonyl	0.4471	0.2303	0.4004	0.4507
31	Amyl	Homopiperonyl	0.5428	0.1933	0.3314	0.3232
32	Benzyl	Homopiperonyl	0.3579	0.0359	0.3574	0.3161
33	2-Hydroxyethyl	Homopiperonyl	-0.3979	-0.2642	-0.3759	-0.2107
34	2-Phenoxyethyl	Homopiperonyl	0.3729	0.2820	0.2806	0.2778
35	2-(p-Tolyloxy)ethyl	Homopiperonyl	-0.0757	0.1499	0.2891	0.2992
36	3-Phenoxypropyl	Homopiperonyl	0.3617	0.3560	0.2583	0.2832
37	4-Phenoxybutyl	Homopiperonyl	0.1959	0.2717	0.2607	0.2319
SAHA			-1	-0.9459	-1.0001	-0.7832

Table 2: List of employed descriptors.

Descriptors calculated using Gaussian 03W	❖ *Hardness (η)* $\eta = (E_{LUMO}-E_{HOMO})/2$ ❖ *Electron Affinity (EA)* $EA= -E_{HUMO}$ ❖ *Ionization Potential (IP)* $IP= -E_{LUMO}$ ❖ *Total Energy (TE)*
Descriptors calculated using ChemDraw Ultra, 8.0	❖ *Melting point(MP)* ❖ *Molar Refractivity (MR)*

3. Results and Discussion

3.1. Multiple Linear Regression

The MLR analysis, made with six descriptors, led to the derivation of one model with five descriptors formulated by the following equation:

$$\log IC_{50}=7.142-0.004\ TE + 678.075\ \eta+326.119\ EA-366.53\ IP-0.006\ MP \quad (1)$$
$$n = 38 \qquad r = 0.80 \qquad s = 0.288$$

Where n is the number of compounds, r is the correlation coefficient, and s is the standard deviation. The selected descriptors, their coefficients and t-values are shown in table 3.

Table 3: Results of multiple linear regression analysis.

Descriptor	Coefficient	Standard error	t-value
TE	-0.004	0.001	-3.837
η	678.075	256.418	2.644
EA	326.119	129.683	2.515
IP	-366.53	127.942	-2.865
MP	-0.006	0.002	-2.887

The correlation of the observed activities with the MLR calculated ones is illustrated in figure 1. A good correlation was obtained with MLR method (r_{MLR}= 0.80) so the predictive power of the model is very significant. The quantique descriptors having the highest contribution, Hardness (η), Electron Affinity (EA), and Ionization Potential (IP), seem to be the most important parameters in the establishment of HDAC inhibitors based hydroxyfurylacryl-N-amide.

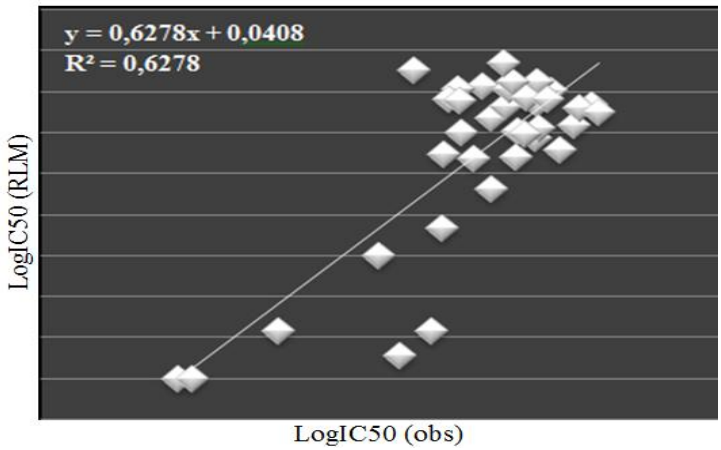


Figure 1: Correlation of observed activities and MLR predicted ones

3.2. Artificial Neural networks

Artificial Neural Networks (ANN) is used to generate predictive models of quantitative structure–activity relationships (QSAR) between a set of molecular descriptors obtained from the MLR analysis and observed activities. One major problem in neural network is how to determine the number of nodes in the hidden layer. Though there is no rigorous rule to rely on [16], a practical way is to use a ratio, ρ, to determine the number of hidden units [17]. ρ is defined as following:

ρ = (Number of data points in the training set / Sum of the number of connections in the ANN).

The reasonable value of ρ should be scaled between 1.8 and 2.3. If ρ < 1.8, the network simply memorizes the data, whereas if ρ > 2.3 the network is not able to generalize [18]. In our study, the five selected variables were used as the input and the HDAC inhibitory activity was used as the output, the optimum number 3 was determined after a training process, so the final architecture of the ANN used in this study is (5-3-1) (Figure 2) .

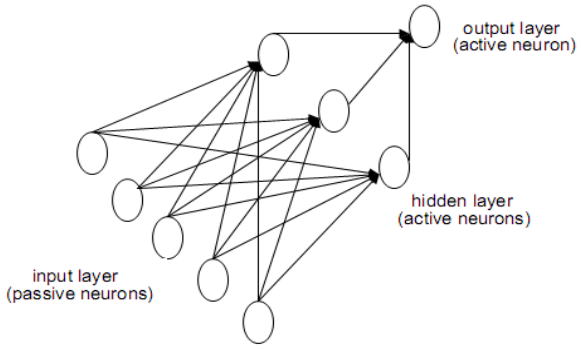


Figure 2: Schematic representation of the network architecture (5-3-1) used in this work

The correlation of the observed activities with the ANN calculated ones, is illustrated in figure 3.

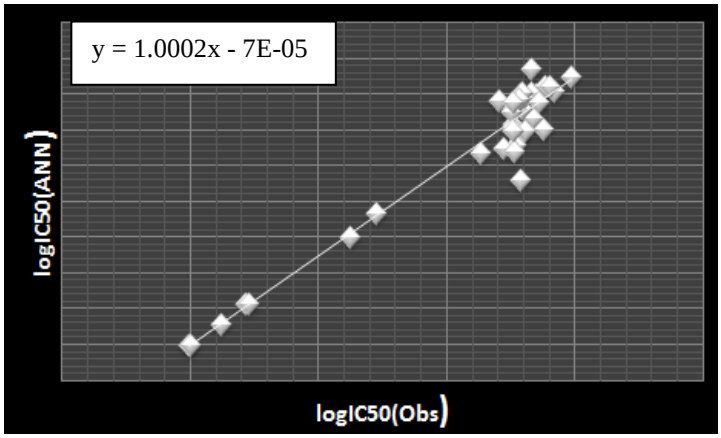


Figure 3: Correlation of observed activities and ANN predicted ones

The correlation coefficient value of Inhibitory activities against HDACs, r_{ANN} = 0.98, shows that the selected descriptors by MLR are pertinent and that the ANN model proposed to predict activity is relevant.

3.3. Cross-Validation

The cross validation analysis was performed using leave one out (LOO) method in which one compound is removed from the data set and the activity is correlated using the rest of the data set. The cross-validated R^2 in each case was found to be very close to the value of R^2 for the entire data set and hence these models can be termed as statistically significant. The correlation of the observed activities with the CV calculated ones is illustrated in figure 4.

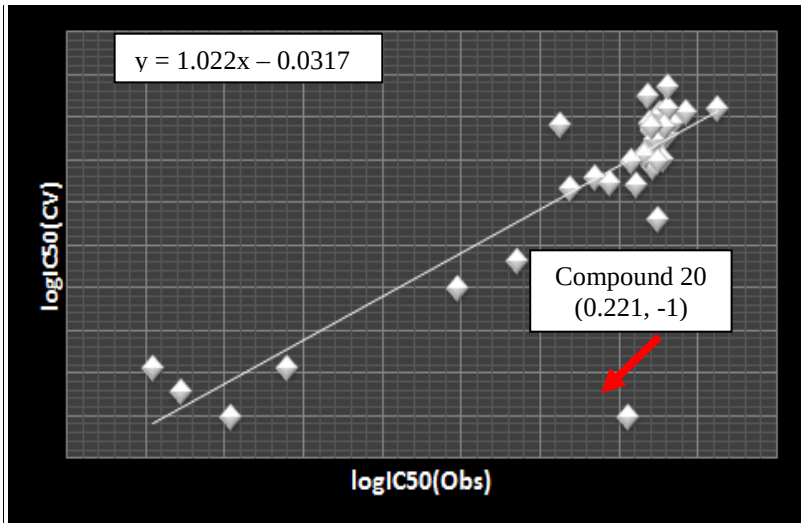


Figure 4: Correlation of observed activities and LOO predicted ones for 38 compounds (r_{CV}= 0.77)

The correlation illustration (figure 4) of predicted activities shows good predictability generally of all compounds except for compound 20 which causes decrease of the correlation coefficient (0.80). Despite that the elimination of this compound of the cross-validation prediction process (Figure 5) increases the correlation coefficient (0.90) we are not able to claim that it is an outlier point, because of the fact that the prediction of this compound's activity is good in the case of both MLR and ANN methods.

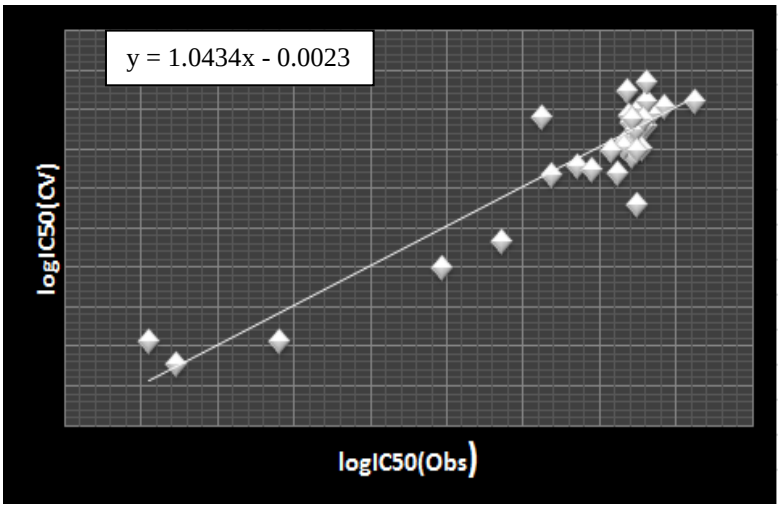


Figure 5: Correlation of observed activities and LOO predicted ones, without molecule 20 (r_{CV}= 0.90)

Conclusion

The MLR and ANN methods have been applied to derive 3D-QSAR model for N-hydroxyfurylacrylamides HDAC inhibitors. The model obtained using these methods showed high correlative and predictive abilities. Our results could be summarized as follow:

- The correlation coefficient obtained with MLR method (r_{MLR}= 0.80) is very significant.
- The artificial neural network (ANN) techniques considering the relevant descriptors obtained from the MLR, showed good agreement between observed values and predicted ones (r_{NN} = 0.98).
- These results show that quantum chemistry theory (DFT precisely) associated with QSAR methods are effective tools to predict activity of inhibitory histone deacetylase.
- The QSAR model proposed in this work is expected to be a useful tool in the conception of novel active molecules.

ISSN : 2028-2508
CODEN: JMESC

References

1. Arrowsmith C. H., Bountra C., Fish P.V., Lee K., Schapira M., *Rev. Drug Disc*.11 (2012) 384.
2. Dhanak D., *Chem. Lett*. 3 (2012 521.
3. M. Brait., *Toxicol Mech Methods*. 21 (2011) 275.
4. Dawson M. A., Kouzarides T., *Cell*. 150 (2012) 12.
5. Roth S. Y., Denu J. M., Allis C. D., *Annu. Rev Biochem*. 70 (2001) 81–120.
6. Thiagalingam S., Cheng. K. H., Mineva N., Thiagalingam A., *Ann N Y Acad. Sci*. 983 (2003) 84-100.
7. Tropsha A., *Mol. Inf*. 29 (2010) 476.
8. Feng T., Wang H., Su H., Lu H., *Bioorganic and Medicinal Chemistry*. 21 (2013) 5339–5354.
9. Frisch M.J., Trucks.G.W., Schlegel H.B., Scuseria G.E., Robb M.A., Cheeseman J.R., Montgomery J.J. A. , Vreven T., Kudin K.N., Burant J.C., Millam J.M., Iyengar S.S., Tomasi J., Barone V., Mennucci B., Cossi M., Scalmani G., Rega N., Petersson G.A., Nakatsuji H., Hada M., Ehara M., Toyota K., Fukuda R., Hasegawa J., Ishida M., Nakajima T., Honda Y., Kitao O., Nakai H., Klene M., Li X., Knox J.E., Hratchian H.P., Cross J.B., Bakken V., Adamo C., Jaramillo J., Gomperts R., Stratmann R.E., Yazyev O., Austin A.J., Cammi R., Pomelli C., Ochterski J.W., Ayala P.Y., Morokuma K., Voth G.A., Salvador P., Dannenberg J.J., Zakrzewski V.G., S., Dapprich A.D. Daniel., Strain M.C., Farkas O., Malick D.K, Rabuck A.D., Raghavachari K., Foresman J.B., Ortiz J.V., Cui Q., Baboul A.G., Clifford S., Cioslowski J., Stefanov B.B., Liu G., Liashenko A., Piskorz P., Komaromi I., Martin R.L., Fox D.J., Keith T., Al-Laham M.A., Peng C.Y., Nanayakkara A., Challacombe M., Gill P.M.W., Johnson B., Chen W., Wong M.W., Gonzalez C., Pople J.A., *Revision C.02, Gaussian, Inc., Wallingford CT*. 2004.
10. Parr R., Yang W., *Oxford University Press. New York.* 1989.
11. Becke A.D., *J. of Chemical Physics*. 98 (1993) 5648.
12. Neter J., Wasserman W., Kutner M., *third ed. Irwin Inc*. 81 (1995) 40-48.
13. Douali L., Villemin D., Cherqaoui D., *J. Chem. Inf. Comput. Sci*. 43 (2003) 1200–1207.
14. Andrea T.A., Kalayeh H., *J. Med. Chem.* 34 (1991) 2824–2836.
15. Worachartcheewan A., Nantasenamat C., al., *J. homepage*. 109 (2011) 207–216.
16. Golbraikh A., Tropsha A., *J. Mol. Graph. Model*. 20 (2002) 269–276.
17. Andrea T.A., Kalayeh H., *J. Med. Chem.* 34 (1991) 2824–2836.
18. So S., Richards G., *J. Med. Chem.* 35 (1992) 3201-3207.